# Refractive Multi-view Stereo

Matthew Cassidy, Jean Mélou, Yvain Quéau, François Lauze, Jean-Denis
Durou

# Refractive Multi-view Stereo

Matthew Cassidy
INP-ENSEEIHT
Toulouse, France
matthew.cassidy@etu.enseeiht.fr

Jean Mélou
IRIT, UMR CNRS 5505
Toulouse, France
jean.melou@toulouse-inp.fr

Yvain Quéau
GREYC, UMR CNRS 6072
Caen, France
yvain.queau@ensicaen.fr

François Lauze
DIKU
Copenhagen, Denmark
francois@di.ku.dk.

Jean-Denis Durou
IRIT, UMR CNRS 5505
Toulouse, France
durou@irit.fr

## Abstract

*In this article we show how to extend the multi-view stereo technique when the object to be reconstructed is inside a transparent – but refractive – material, which causes distortions in the images. We provide a theoretical formulation of the problem accounting for a general, non-planar shape of the refractive interface, then a discrete solving method, which are validated by tests on synthetic and real data.* **Keywords:** *3D-reconstruction, Multi-view Stereo, Refraction.*

## 1. Introduction

Many Natural History Museums have collections of fragile specimens embedded in a transparent medium, for instance a prehistoric insect trapped in a block of amber, or a small animal conserved in a solution of formaldehyde or alcohol in a glass jar (see Fig. 1). They constitute an invaluable source of information for evolutionary scientists.
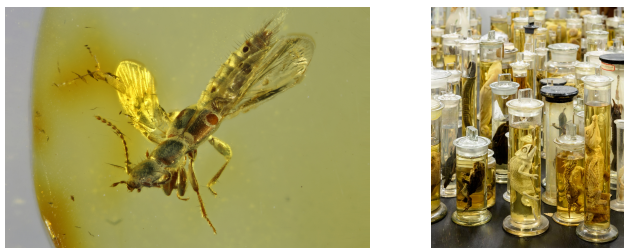


Figure 1: Prehistoric beetle trapped in amber (seen under a microscope), and reptiles specimens in jars, from the Natural History Museum in Copenhagen.

Exploitation of these resources necessitates "see through" 3D-estimation techniques. Computerised tomography (CT) provides a 3D-reconstruction, however it requires special CT scanners that can handle objects of various sizes. It is expensive and time consuming. This may be feasible for a limited amount of samples but out of reach for large collections (thousands of amber pieces, or linear kilometers of specimens in jars). Photogrammetric 3D-scanning is a reasonable alternative, which however requires overcoming several difficulties.

First of all, if the *refractive* medium deflects the trajectory of light rays that pass through it, the shape of the *interface*, i.e. the separation surface between air and this medium, has a strong influence on the appearance of the object to be reconstructed. In the example on the left in Fig. 1, it is not forbidden to cut the amber block to give it a particular shape, as long as this does not alter the insect. The shape of the refractive interface will thus be assumed to be known. However, unlike most existing techniques, we do not restrict our attention to the planar case.

Refraction can cause other difficulties. Because it is wavelength dependent, it causes *light dispersion*. In addition, an inhomogeneous transparent medium may have spatially varying refractive index, resulting in light rays that are not straight. Taking a varying refractive index into account does not present an intractable difficulty, but estimating this index at each point in the medium is in itself a difficult problem. We will therefore assume that the refractive medium is *homogeneous* and that the refractive index varies little with wavelength.

**Contributions** – We introduce a general 3D-reconstruction method for objects trapped inside a homogeneous refractive medium with arbitrary, known interface, from a set of calibrated multi-view stereo images. The paper is organized as follows. In Section 2, we review the main studies taking into account refraction. We show in Section 3 how to extend the multi-view stereo technique in the presence of an interface. The main difficulty of this extension, which is detailed in Section 4, consists in predicting the projection in an image of a 3D point located in the refractive medium. A first set of tests is conducted in Section 5 to validate this method. In Section 6, we show how to extend it to any shape of the interface. Section 7 allows us to validate it on real data. Finally, we conclude in Section 8 by mentioning some possible extensions.

## 2. State of the Art

Light rays from a point source naturally tend to diverge. By refraction, a lens allows to deviate their trajectories to make them converge in an *image point*. Refraction is therefore the key to optical instruments, with the caveat that manufacturing a photographic lens assembly requires very precise alignment of the lenses that make it up, in order to limit undesired effects, called *aberrations*. However, if the scene itself contains transparent, i.e. refractive, objects, this modifies the appearance of opaque objects, which appear distorted.

**Correcting refraction in the images** – A number of papers have addressed the correction of these distortions when the transparent object, such as a window pane, is attached to the front of the camera, in which case the distortions can be calibrated, and standard 3D-reconstruction pipelines can then be employed. For example, while it is relevant to use photogrammetry to study convection in a water-filled tank, Maas shows in [28] how to take into account refraction through both sides of the glass to improve measurement accuracy. In [27], Luczynski et al. show how to correct the images acquired by a pair of calibrated underwater cameras, protected by a glass plate, to restore epipolar geometry. The same idea of pre-correcting the images has also been explored in [1, 34], as well as in [2] where correction is carried out by neural networks. It has also been shown in [20, 37] that light field cameras can provide useful clues for correcting refraction in the images. Another example where the consideration of refraction is required is the 3D-reconstruction of the seabed from aerial images [7, 30].

**Active refraction techniques** – On the other hand, several studies, which can be grouped under the term *active refraction*, take advantage of refraction to allow 3D-reconstruction from a single view, thanks to image duplication, using either a bi-prism [25, 35], or a glass plate rotating around the optical axis [12, 17]. A problem involving refraction, but with a different goal than ours, is that of estimating the surface of a transparent object. The solution proposed by Morris and Kutulakos consists in mapping points of interest seen through transparency [29]. Another approach is that of Ben-Ezra and Nayar [6], who fit the parameters of a surface model in order to reproduce the distorted image of an object of known geometry. The 3D-reconstruction of transparent objects has also been revisited recently under the neural network perspective [26].

**Refractive structure-from-motion** – Chari and Sturm showed in [11] how to extend epipolar geometry in the case where the camera is separated from the 3D-scene by a planar interface, using a $12 \times 12$ fundamental matrix. The explicit consideration of refraction in the estimation of pose by *structure-from-motion* was also studied by Jordt et al. [21, 22], Kang et al. [23], Qiao et al. [31], Suresh et al. [32], and most recently by Chadebecq et al. [9]. In most of these papers, validation is carried out on underwater images – a recent survey of underwater 3D-imaging techniques can be found in [8].

**Refractive multi-view stereo** – In this paper, we rather assume that the camera poses have been pre-calibrated, e.g. by structure-from-motion, and we focus on 3D-reconstruction by multi-view stereo (MVS). Extensions of the PMVS algorithm (*patch-based multi-view stereo*) by Furukawa and Ponce [16] have been proposed by Kang et al. [24] and by Chang and Chen [10], but the extension process is not really detailed. Agrawal et al. have also studied MVS in a refractive medium in [3], but their work remains limited to a (multi-layer) planar interface. To the best of our knowledge, the only work considering a non-planar interface is the recent work of Yoon et al. [36]. However, their objective is different, because the curved refractive medium is placed between the observer and the object, while in our case the object lies inside the refractive medium.

Overall, there is a lack of multi-view stereo method adapted to the case of an object placed in a refractive medium with non-planar shape. The goal of our work is to fill this gap.

## 3. From Multi-view Stereo to Refractive Multi-view Stereo

Multi-view stereo (MVS) provides a dense 3D-reconstruction by maximizing the photometric consistency across a set of images. An overview of MVS-related approaches is provided in [15]. Knowing $t + 1$ images of the same 3D-scene and the corresponding $t + 1$ camera poses, we choose the first pose as reference. For a 3D-point $\mathbf{P}$ visible in all the images, we note $\mathbf{p} = \pi(\mathbf{P})$ the projection of $\mathbf{P}$ into the reference image and $\mathbf{p}_j = \pi_j(\mathbf{P})$, $j \in \{1, \dots, t\}$, its projections into the $t$ other images, called *control images*. The Lambertian hypothesis is written:

$$I_j \circ \underbrace{\pi_j \circ \pi^{-1}(\mathbf{p})}_{\mathbf{p}_j} = I(\mathbf{p}), \quad j \in \{1, \dots, t\} \qquad (1)$$

with $I_j$ and $I$ denoting the grayscale functions of the $j$-th control image and of the reference image.

The MVS method consists in searching the 3D-point $\mathbf{P} = \pi^{-1}(\mathbf{p})$ conjugate to pixel $\mathbf{p}$ that "best" satisfies the system of Eqs. (1). This can be done by solving this system in the least-squares sense:

$$\min_{\mathbf{P} = \pi^{-1}(\mathbf{p})} \sum_{j=1}^{t} [I_j \circ \pi_j(\mathbf{P}) - I(\mathbf{p})]^2 \qquad (2)$$

In practice, this comparison is performed between the neighborhoods of two pixels, using a robust estimator (see the overview presented in [15, chapter 2]).

**MVS in a homogeneous medium** – When the medium is homogeneous, it is easy to parameterize the back-projection $\pi^{-1}$ from the reference camera to the surface by using perspective projection. Denoting by $\mathbf{P} = [x, y, z]^\top$ the 3D-point conjugate to pixel $\mathbf{p}$, and by $\mathbf{K}$ the intrinsics camera matrix:

$$\pi^{-1}(\mathbf{p}) = z\,\mathbf{K}^{-1} \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} \qquad (3)$$

The projection from the surface into the $j$-th control image is also given by central projection, along with a rigid change of coordinates system. Denoting by $(\mathbf{R}_j, \mathbf{t}_j)$ the known pose of the $j$-th camera (rotation and translation, respectively), and defining $\rho([a, b, c]^\top) = [a/c, b/c]^\top$:

$$\pi_j(\mathbf{P}) = \rho(\mathbf{K}\,[\mathbf{R}_j\mathbf{P} + \mathbf{t}_j]) \qquad (4)$$

Inserting (3) and (4) into (2), the 3D-reconstruction problem is easily reparameterized in terms of the sole depth value $z$. The function to be minimized may however be nonlinear, non-differentiable and/or non-convex, which makes optimization potentially difficult. For this reason, minimization is usually performed by an exhaustive search (*brute-force*) in a list of predefined depth values. This apparently simplistic strategy proves to be very efficient for 3D-reconstruction of scenes with sufficiently textured surfaces [18]. As depicted in Fig. 2, the scenario is similar, but clearly more complex, in the presence of a refractive medium.
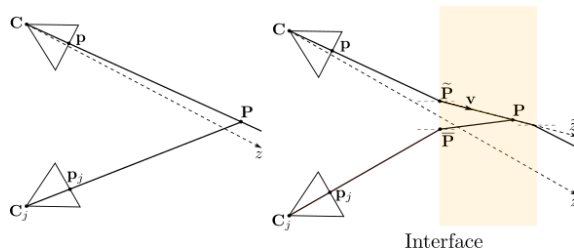


Figure 2: Left: MVS in a homogeneous medium. Right: MVS in the presence of a refractive medium. In this case, the projection of a 3D-point $\mathbf{P}$ into the images is clearly more complex, due to refraction.

**MVS in the presence of a refractive medium** – The presence of a medium whose refraction index is strictly more than 1, i.e. of an *interface* between the object to be reconstructed and the camera, has the effect of distorted images. Among other consequences, the epipolar geometry of a pair of images is lost. Instead of corresponding to an epipolar straight line in the second image, a point in the first image now corresponds to a curve, the shape of which obviously depends on the position of the interface, its geometry, and the refractive index involved.

Chari and Sturm showed in [11] how to generalize the matrix formalism of the epipolar geometry with a $12 \times 12$ fundamental matrix, but epipolar geometry is primarily used for the estimation of the pose by structure-from-motion. Since this problem has already been studied in several articles [9, 21], we focus on the extension of MVS, which amounts to solving Problem (2) in each point $\mathbf{p}$ of the reference image. Some adaptations are necessary, in comparison with the case of a homogeneous medium:

- To calculate $\mathbf{P} = \pi^{-1}(\mathbf{p})$ in the presence of an interface, we need to draw the ray from $\mathbf{C}$ through $\mathbf{p}$. While this ray is straight in the absence of an interface, it now follows a broken line (see on the right in Fig. 2). The back-projection $\pi^{-1}$ is thus not as simple as in (3):

$$\pi^{-1}(\mathbf{p}) = \widetilde{\mathbf{P}} + \tilde{z}\,\mathbf{v} \qquad (5)$$

where we denote by $\widetilde{\mathbf{P}}$ the *point of refraction*, by $\mathbf{v}$ the *direction of refraction*, and by $\tilde{z} \geq 0$ the distance travelled along that direction. Finding the point of refraction $\widetilde{\mathbf{P}}$ amounts to finding the intersection between the viewing direction and the interface, while the direction of refraction $\mathbf{v}$ is given by the Descartes-Snell's laws (see Section 4). The first of these tasks can be tedious, depending on the parameterization of the shape of the interface.

- The calculation of $\mathbf{p}_j = \pi_j(\mathbf{P})$ is much more difficult if there is an interface, in comparison with (4). Indeed, function $\pi_j$ is no longer a central projection into the $j$-th control image. We will see in Section 4 that this calculation requires solving a *shortest optical path* problem.

- As shown in Fig. 3, in the case of a non-planar interface, a 3D-point $\mathbf{P}$ can project in several points of the $j$-th control image, which correspond to as many *local minima* of the optical path between $\mathbf{P}$ and $\mathbf{C}_j$, according to Fermat's principle. However, this does not complicate the problem, since each of these projections can equivalently be used in Problem (2).
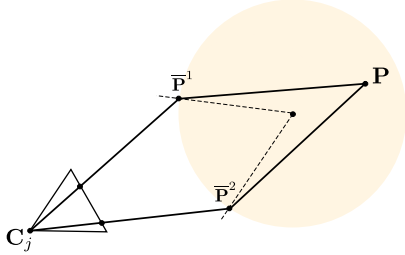


Figure 3: Due to the non-planarity of the interface, a 3D-point $\mathbf{P}$ may project in several points of the $j$-th control image.

In comparison with the classical MVS problem, the main difficulty of refractive MVS is therefore to compute $\mathbf{p}_j = \pi_j(\mathbf{P})$, which amounts to finding the point of refraction $\overline{\mathbf{P}}$ defined on the right in Fig. 2.

## 4. Projections into the Control Images

Descartes-Snell's first law of refraction tells us that the refracted ray lies in the *plane of incidence*, defined by the incident ray and the normal to the point of refraction $\overline{\mathbf{P}}$. For this reason, one can represent the refraction phenomenon within a plane (see Fig. 4).

Denoting by $i_1$ the angle between the incident light ray and the normal to the interface, in a medium of refractive index $n_1$, and by $i_2$ the angle between the refracted ray and the same normal, in a medium of
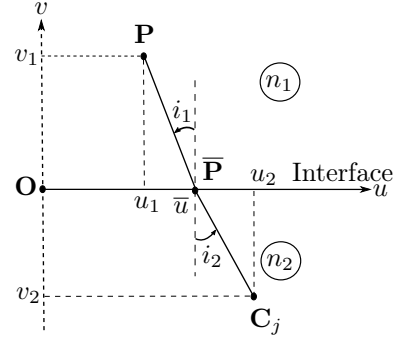


Figure 4: Descartes-Snell's second law of refraction.

refractive index $n_2$, Descartes-Snell's second law of refraction tells us that:

$$n_1 \sin i_1 = n_2 \sin i_2 \tag{6}$$

In the case of a planar interface, let us square both members of Eq.(6), using the notations of Fig. 4:

$$n_1^2 \frac{(u_1 - \overline{u})^2}{(u_1 - \overline{u})^2 + v_1^2} = n_2^2 \frac{(u_2 - \overline{u})^2}{(u_2 - \overline{u})^2 + v_2^2} \tag{7}$$

To determine the point of refraction $\overline{\mathbf{P}}$, we thus are led to solve a polynomial equation of degree 4 in $\overline{u}$:

$$a_4 \overline{u}^4 + a_3 \overline{u}^3 + a_2 \overline{u}^2 + a_1 \overline{u} + a_0 = 0 \tag{8}$$

whose coefficients are expressed in function of $u_1$, $v_1$, $u_2$, $v_2$, and $\alpha = n_2/n_1$ (see [3] for more details).

Since the optical path has a unique minimum in this case, Eq.(8) has a unique real solution. Therefore, the calculation of $\mathbf{p}_j = \pi_j(\mathbf{P})$ can be done in two steps. First, to find the shortest optical path between $\mathbf{P}$ and $\mathbf{C}_j$, an equation such as (8) must be solved to determine the point of refraction $\overline{\mathbf{P}}$. Then, $\overline{\mathbf{P}}$ must be projected into the $j$-th control image, according to (4).

One may wonder whether the transformation $\pi_j$ keeps the alignment of points i.e., if the image by $\pi_j$ of a light ray after refraction, which is a straight line, is still a straight line in the $j$-th control image, in which case we could simply look along this straight line to find the point conjugate to a point $\mathbf{p}$ of the reference image. This does not hold true in the general case, and even in the simplest case of a planar interface, it has been shown in [11] that at a point $\mathbf{p}$ of the reference image corresponds, in the second image, a curve whose equation is deduced from a $12 \times 12$ matrix.

However, this equation is not very useful to solve Problem (2). Indeed, we already mentioned in Section 3 that it is recommended to carry out the resolution of this problem by an exhaustive search in a list of 3D-points $\mathbf{P}^1$, $\mathbf{P}^2$, $\mathbf{P}^3$, etc., located for instance on parallel slices of the refractive medium (see Fig. 5).
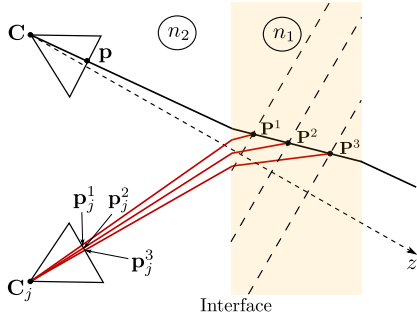
Figure 5: We must find the projection $\mathbf{p}_j^k$ into the $j$-th control image of each of the points $\mathbf{P}^k$, $k \in \{1, 2, 3, \dots\}$, located on parallel slices of the refractive medium. The point $\mathbf{P}^k$ selected is the one whose image $\mathbf{p}_j^k$ is most similar to $\mathbf{p}$.

# 5. Validation on Synthetic Images

A plausible strategy for solving the refractive MVS (RMVS) problem is therefore to proceed exactly as for the classical MVS problem, except that, for each pixel $\mathbf{p}$ of the reference image, and for each of the 3D-points $\mathbf{P}^1$, $\mathbf{P}^2$, $\mathbf{P}^3$, etc., Problem (2) now requires solving $t$ polynomial equations of degree 4 such as (8), in addition to computing $t$ projections such as (4).

Images of the 3D-model *Stanford Bunny* cast in a box-shaped block of glass are simulated using the ray tracing features of the free software *Blender* (see Fig. 6): a wooden texture is applied on the surface of the bunny to facilitate the matching; a realistic refractive index of 1.5 is chosen for the glass; the scene is placed in front of a green background, to facilitate the background/foreground segmentation.



Figure 6: Simulated image of the *Stanford Bunny* embedded in a box-shaped block of glass.

Let us evaluate the method described in Sections 3 and 4. Fig. 7 shows two views of the colored 3D-point cloud obtained from 24 images such as that in Fig. 6, by merging four clouds. Each cloud is estimated from six images: five images taken through the same vertical face of the box are complemented by one image taken from the side.
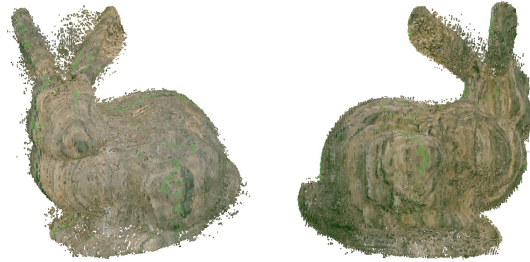


Figure 7: RMVS: Two views of the 3D-point cloud obtained from 24 images such as that in Fig. 6. The space between the two ears of the rabbit is dusty, because this part is only visible in a few images.

The RMSE values corresponding to the four 3D-point clouds which are merged to form the cloud in Fig. 7, are reported in Table 1. These very low RMSE values, compared to the scale of the reference 3D-model (in the order of 20 $mm$), and to the distance of the 3D-model from the camera (between 90 $mm$ and 110 $mm$), validate the relevance of the RMVS method.

| Face | Front | Back | Left | Right |
|---|---|---|---|---|
| RMSE ($mm$) | 0.28 | 0.29 | 0.34 | 0.38 |

Table 1: RMSE of each 3D-point cloud obtained from six images: five images taken through the same face, complemented by one image taken from the side.

Let us remark that, in the same series of simulated images as that of Fig. 6, under certain angles, the image of the bunny is duplicated (see Fig. 8). This makes it possible to implement single-view stereoscopy, in a similar way as in [25, 35].



Figure 8: Under certain angles, the image of the bunny in a box-shaped block of glass is duplicated.

In addition to being very resource-intensive, our solving strategy presents another serious limitation: it is difficult to generalize to an interface with a more complex shape. These two problems can be overcome by opting for a fully-discrete solving strategy, which consists of discretizing the interface as well.

# 6. Discretizing the Interface

Huygens-Fresnel's principle allows us to predict the wave surfaces of a luminous flux, which are orthogonal to the light rays. A discrete version of the wave surfaces calculation is provided by Dijkstra's algorithm [13], that finds the shortest path between two vertices in a graph. To calculate the path of the light rays, we could therefore cut out the full scene into voxels, and consider them as the vertices of an indirected graph. However, the problem is greatly simplified under the assumption that the refractive medium is homogeneous.

Indeed, if the refractive index is the same at any point in the block, the light propagates inside the medium along straight lines (as in the air). The path of a light ray between a point $\mathbf{P}^k$ in the block and the optical center $\mathbf{C}_j$ is thus a broken line, with just one break at the interface (see the red lines in Fig. 5). As already stated at the end of Section 3, finding the shortest path between $\mathbf{P}^k$ and $\mathbf{C}_j$ thus reduces to the search for the point of refraction $\overline{\mathbf{P}}$ (see Fig. 4). For a planar interface, we know from Section 4 that this amounts to solving a degree 4 equation, but for more complex shapes, the analytical solving may be intractable.

To find $\overline{\mathbf{P}}$, a simpler solution consists in discretizing the interface, and calculating the optical path through each point of the discretized interface. Problem (2) can then be solved as follows: for each pixel $\mathbf{p}$ of the reference image, and for each tested point $\mathbf{P}^k$, one just has to find the shortest path between $\mathbf{P}^k$ and the optical center $\mathbf{C}_j$ (see Fig. 9). To define the discretization scale, a compromise has to be found between accuracy and a reasonable computing time. In this respect, we only test the points on the interface that simultaneously face the camera, and project inside the silhouette of the object in each of the control images.
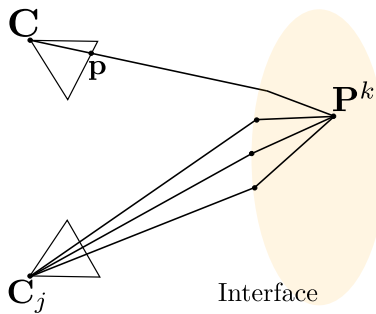


Figure 9: For each pixel $\mathbf{p}$ of the reference image, and each tested point $\mathbf{P}^k$, the optical path between $\mathbf{P}^k$ and $\mathbf{C}_j$ is calculated through each point on the discretized interface. The one which corresponds to the minimum optical path is chosen as the point of refraction $\overline{\mathbf{P}}$.

To validate this fully-discrete solving strategy, we use the 3D-model of an insect called *graphosoma*. Fig. 10 shows two simulated images of this insect trapped in a spherical block of glass, seen from two viewpoints, as well as the same views of this insect outside the glass. In addition to the obvious magnification effect due to the 3D-shape of the block of glass, which is equivalent to a convex lens, strong deformations appear, particularly near the limbs.



Figure 10: Top row: Simulated images of a graphosoma trapped in a spherical block of glass, seen from two viewpoints. Bottom row: Images from the same viewpoints, and at the same scale, but outside the block of glass. Besides the magnification effect due to the block convexity, deformations clearly appear. Source of the 3D-model: Digital Archive of Natural History [5].

Fig. 11 shows the result by our fully-discrete RMVS solving method, obtained by merging four colored 3D-point clouds: each cloud is obtained from six images (one reference image plus five control images). Not only the insect's legs and antennas from these four clouds do coincide, but fine details are reconstructed. At this point, let us insist that we merge the 3D-point clouds without any post-processing.



Figure 11: Fully-discrete RMVS: Two views of the 3D-reconstruction provided by our method, using 24 images such as those of the top row in Fig. 10.

The result in Fig. 11 validates our fully-discrete RMVS solving strategy, at least qualitatively. To confirm this assertion, Table 2 lists the RMSE values of the four 3D-point clouds, to be compared to the scale of the reference 3D-model (around 20 $mm$), and to the distance of the 3D-model from the camera (between 90 $mm$ and 110 $mm$). Without taking refraction into account, the legs and antennas are duplicated (the absence of post-processing accentuates the duplication effect), as shows the result on the left in Fig. 12, which is confirmed by the RMSE values listed in Table 2.

| RMVS ($mm$) | 0.22 | 0.23 | 0.18 | 0.16 |
|---|---|---|---|---|
| MVS ($mm$) | 0.51 | 0.83 | 0.72 | 0.31 |

Table 2: RMSE of each of the 3D-point clouds which, by merging, provide the 3D-reconstruction in Fig. 11 (RMVS) and that on the left in Fig. 12 (simple MVS).

As the result on the left in Fig. 12 is obtained by (wrongly) setting the index of the refractive medium to 1, this shows how crucial a precise knowledge of the refractive index is. An additional result provided by AliceVision, which is a classical SfM/MVS free software [4], however not intended to consider refraction, is shown on the right in Fig. 12. Unsurprisingly, this result is far from being satisfactory, even qualitatively.
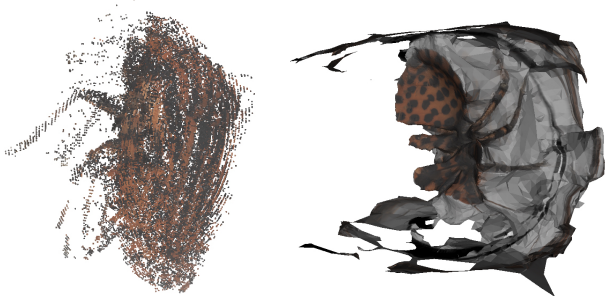


Figure 12: Left: Ghost legs and antennas appear when an erroneous value of the refractive index (here equal to 1) is used. Right: Misunderstanding of the graphosoma 3D-shape by the AliceVision free software [4].

The image of a straight rod passing through a refractive ellipsoid (see Fig. 13) shows to which extent such an interface can distort the image. Since the projections of the 3D-points inside the ellipsoid are very difficult to predict, our fully-discrete RMVS solving method could be less accurate in such a case.

Indeed, Fig. 14 shows how distorted the bunny's images are, when the glass block is ellipsoidal. Even if the 3D-reconstruction provided by our fully-discrete RMVS solving method, using six images, is faithful to the original, it is much less accurate than that of Fig. 7,
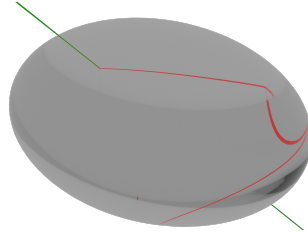


Figure 13: Simulated image of a straight rod passing through a refractive ellipsoid. The projections of the points inside the medium are highlighted in red.

because for some rays, the angle $i_2$ of the Descartes-Snell's law (6) is very close to $\pi/2$ (grazing angle). Since the derivative of the arcsin function tends towards infinity in 1, this causes strong inaccuracies in the calculation of the angle $i_1$ in Eq.(6).
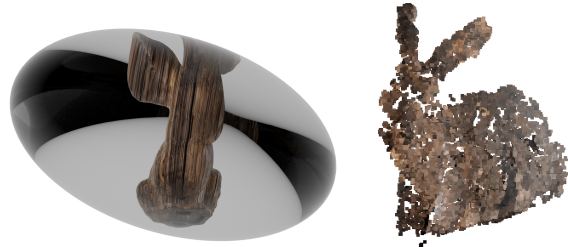


Figure 14: Left: Simulated image of the *Stanford Bunny* inside an ellipsoidal block of glass. Right: Due to the huge distortions of the images, the 3D-reconstruction by RMVS, using six images, is far less accurate than that of Fig. 7. However, the general shape of the reconstructed 3D-point cloud is still much more significant of a bunny than the image on the left.

## 7. Validation on Real Images

Now, let us test this solving method on real images. As sets of images from Natural History Museums have not yet been produced, we use a present time insect from a private collection. It is a grasshopper cast in a box-shaped block of resin. The photographs were taken under poorly controlled operating conditions, placing the scene in front of a blue background to facilitate segmentation (see Fig. 15).

The main difficulty in implementing our 3D-reconstruction method is to estimate the camera poses. We could have used a method of estimation of the pose by *refractive structure-from-motion* [9], but we also need to know the position of the faces of the resin block. So we simply glued a small grid on each face, which allowed us to obtain estimates that were consistent with each other.

Figure 15: Two real images of a grasshopper cast in a box-shaped block of resin. The small grid glued on each side allows us to estimate the camera pose and the position of the planar interface.

Fig. 16 shows two views of the 3D-reconstruction of the grasshopper obtained from five images taken through the same face, plus one image taken from the side (we set the resin refraction index to the plausible value $n_1 = 1.6$). Although imperfect, this result is encouraging. In particular, the reconstructed legs and antennas seem realistic, which is probably an important criterion in determining whether or not we meet the needs of entomologists.



Figure 16: 3D-reconstruction of the grasshoper by RMVS: Two views of the 3D-point cloud obtained from five images through the same face, plus one side shot.

## 8. Conclusion and Perspectives

In this paper, we have shown how to adapt the multiview stereo technique to the case where the object of interest is trapped in a refractive medium. Knowing that refraction distorts the images, it is necessary to take it into account to model the path of the light rays. More specifically, we have proposed a fully-discrete solving method. Our first results on real data are encouraging,

although there are still many obstacles to overcome before this preliminary study leads to an operational method that can be used by entomologists.

Among the tasks to be carried out as a priority, it is necessary to improve the estimation of the camera poses, using *refractive structure-from-motion* methods, which are based on the matching between points of interest seen by transparency, provided that the epipolar geometry is adapted [11].

Another improvement of our method will consist in estimating the block 3D-shape, resorting for instance to *shape-from-silhouettes* [19]. By taking care to place the block in front of a colored background, as we already did (see Fig. 15), this technique allows one to estimate the volume of any convex object, provided a sufficient number of images are used.

As already noticed, another essential characteristic of the refractive medium, whose precise estimation would improve the results on real data, is its index, which was arbitrarily set at 1.6 for the test in Fig. 16. Recall that the problem would be much more complex if the refractive index of the material were not homogeneous.

The efficiency of our RMVS solving method could also be improved. To give an order of magnitude of the complexity of this resolution strategy, each 3D-point cloud giving, by merging, the result on the left of Fig. 7, is obtained with $t = 5$ control images. The reference image has about $1.5 \times 10^5$ pixels. Since the block of glass is discretized into 200 slices, the complete resolution of Problem (2) requires solving about $10^8$ equations of degree 4, which takes about one hour of computing time on a standard PC, using either an analytical or a bisection method of solving.

Finally, in the longer term, we plan to adapt to refraction the photometric 3D-reconstruction methods that are *shape-from-shading* [14] and photometric stereo [33]. To our knowledge, these problems have been very rarely addressed in the refractive case. In addition to the need to model the phenomenon of light attenuation inside the refractive medium, according to *Beer-Lambert's law*, it will also be necessary to take color into account, i.e. to consider that this attenuation depends on the wavelength. And while we have deliberately ignored light scattering, this phenomenon will also have to be taken into account in the context of photometric 3D-reconstruction.

## References

[1] P. Agrafiotis, K. Karantzalos, A. Georgopoulos, and D. Skarlatos. Correcting image refraction: Towards accurate aerial image-based bathymetry mapping in shallow waters. *Remote Sensing*, 12(2):322, 2020. 2

[2] P. Agrafiotis, D. Skarlatos, A. Georgopoulos, and K. Karantzalos. DepthLearn: learning to correct the refraction on point clouds derived from aerial imagery for accurate dense shallow water bathymetry based on SVMs-fusion with LiDAR point clouds. *Remote Sensing*, 11(19):2225, 2019. 2

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

[4] AliceVision. https://alicevision.org/. 7

[5] Digital Archive of Natural History. https://sketchfab.com/disc3d. 6

[6] M. Ben-Ezra and S. K. Nayar. What does motion reveal about transparency? In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pages 1025–1032, 2003. 2

[7] B. Cao, R. Deng, and S. Zhu. Universal algorithm for water depth refraction correction in through-water stereo remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 91:102108, 2020. 2

[8] M. Castillón, A. Palomer, J. Forest, and P. Ridao. State of the Art of Underwater Active Optical 3D Scanners. *Sensors*, 19(23):5161, 2019. 2

[9] F. Chadebecq, F. Vasconcelos, R. Lacher, E. Maneas, A. Desjardins, S. Ourselin, T. Vercauteren, and D. Stoyanov. Refractive Two-View Reconstruction for Underwater 3D Vision. *International Journal of Computer Vision*, 128:1101–1117, 2020. 2, 3, 7

[10] Y. J. Chang and T. Chen. Multi-View 3D Reconstruction for Scenes under the Refractive Plane with Known Vertical Direction. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 351–358, 2011. 2

[11] V. Chari and P. Sturm. Multiple-View Geometry of the Refractive Plane. In *Proceedings of the British Machine Vision Conference*, pages 1–11, 2009. 2, 3, 4, 8

[12] Z. Chen, K.-Y. K. Wong, Y. Matsushita, and X. Zhu. Depth from refraction using a transparent medium with unknown pose and refractive index. *International Journal of Computer Vision*, 102(1–3):3–17, 2013. 2

[13] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959. 6

[14] J.-D. Durou, M. Falcone, and M. Sagona. Numerical Methods for Shape-from-shading: A New Survey with Benchmarks. *Computer Vision and Image Understanding*, 109(1):22–43, 2008. 8

[15] Y. Furukawa and C. Hernàndez. Multi-View Stereo: A Tutorial. *Foundations and Trends in Computer Graphics and Vision*, 9(1–2):1–148, 2015. 3

[16] Y. Furukawa and J. Ponce. Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010. 2

[17] C. Gao and N. Ahuja. A Refractive Camera for Acquiring Stereo and Super-resolution Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2316–2323, 2006. 2

[18] M. Goesele, B. Curless, and S. M. Seitz. Multi-View Stereo Revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2402–2409, 2006. 3

[19] C. Hernández and F. Schmitt. Silhouette and stereo fusion for 3D object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004. 8

[20] K. Ichimaru and H. Kawasaki. Underwater Stereo Using Refraction-Free Image Synthesized From Light Field Camera. In *Proceedings of the IEEE International Conference on Image Processing*, pages 1039–1043, 2019. 2

[21] A. Jordt, K. Köser, and R. Koch. Refractive 3D reconstruction on underwater images. *Methods in Oceanography*, 15–16:90–113, 2016. 2, 3

[22] A. Jordt-Sedlazeck, D. Jung, and R. Koch. Refractive Plane Sweep for Underwater Images. In *Proceedings of the German Conference on Pattern Recognition*, pages 333–342, 2013. 2

[23] L. Kang, L. Wu, Y. Wei, S. Lao, and Y.-H. Yang. Two-view underwater 3D reconstruction for cameras with unknown poses under flat refractive interfaces. *Pattern Recognition*, 69:251–269, 2017. 2

[24] L. Kang, L. Wu, and Y. H. Yang. Two-View Underwater Structure and Motion for Cameras under Flat Refractive Interfaces. In *Proceedings of the European Conference on Computer Vision*, volume 7575 of *Lecture Notes in Computer Science*, pages 303–316, 2012. 2

[25] D. H. Lee, I.-S. Kweon, and R. Cipolla. A biprism-stereo camera system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, page 87, 1999. 2, 5

[26] Z. Li, Y.-Y. Yeh, and M. Chandraker. Through the Looking Glass: Neural 3D Reconstruction of Transparent Shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1262–1271, 2020. 2

[27] T. Luczynski, M. Pfingsthorn, and A. Birk. Image rectification with the pinax camera model in underwater stereo systems with verged cameras. In *OCEANS 2017*, pages 1–7, 2017. 2

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

[29] N. J. W. Morris and K. N. Kutulakos. Dynamic Refraction Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1518–1531, 2011. 2

[30] T. Murase, M. Tanaka, T. Tani, Y. Miyashita, N. Ohkawa, S. Ishiguro, Y. Suzuki, H. Kayanne, and H. Yamano. A Photogrammetric Correction Procedure for Light Refraction Effects at a Two-Medium Boundary. *Photogrammetric Engineering and Remote Sensing*, 9(8):1129–1136, 2008. 2

[31] X. Qiao, Y. Ji, A. Yamashita, and H. Asama. Structure from Motion of Underwater Scenes Considering Image Degradation and Refraction. *IFAC-PapersOnLine*, 52(22):78–82, 2019. 2

[32] S. Suresh, E. Westman, and M. Kaess. Through-water stereo slam with refraction correction for AUV localization. *IEEE Robotics and Automation Letters*, 4(2):692–699, 2019. 2

[33] R. J. Woodham. Photometric Method For Determining Surface Orientation From Multiple Images. *Optical Engineering*, 19(1):134–144, 1980. 8

[34] X. Wu and X. Tang. Accurate binocular stereo underwater measurement method. *International Journal of Advanced Robotic Systems*, 16(5):1729881419864468, 2019. 2

[35] A. Yamashita, Y. Shirane, and T. Kaneko. Monocular Underwater Stereo – 3D Measurement Using Difference of Appearance Depending on Optical Paths. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3652–3657, 2010. 2, 5

[36] S. Yoon, T. Choi, and S. Sull. Depth estimation from stereo cameras through a curved transparent medium. *Pattern Recognition Letters*, 129:101–107, 2020. 2

[37] C. Zhang, X. Zhang, D. Tu, and P. Jin. On-site calibration of underwater stereo vision based on light field. *Optics and Lasers in Engineering*, 121:252–260, 2019. 2