REFRACTIVE STEREO RAY TRACING FOR RECONSTRUCTING UNDERWATER STRUCTURES

Scott Sorensen, Abhishek Kolagunda, Philip Saponaro, Chandra Kambhamettu

Video/Image Modeling and Synthesis (VIMS) Lab
Computer and Information Sciences
University of Delaware, Newark DE, USA

ABSTRACT

Underwater objects behind a refractive surface pose problems for traditional 3D reconstruction techniques. Scenes where underwater objects are visible from the surface are commonplace, however the refraction of light causes 3D points in these scenes to project non-linearly. Refractive Stereo Ray Tracing allows for accurate reconstruction by modeling the refraction of light. Our approach uses techniques from ray tracing to compute the 3D position of points behind a refractive surface. This technique aims to reconstruct underwater structures in situations where access to the water is dangerous or cost prohibitive. Experimental results in real and synthetic scenes show this technique effectively handles refraction.

***Index Terms*—** Underwater, Stereo Vision, Refraction,

1. INTRODUCTION

Typical stereo techniques are not well suited to deal with specular refractive surface. These surfaces can lead to falsely matched correspondences and incorrectly reconstructed objects. Many common materials refract light, water being the most common, so dealing with this problem is important for accurate scene reconstruction. The bending of light as it crosses a boundary between materials, or an interface, is known as refraction and is governed by the Snell's law [1]. Snell's law dictates that the trajectory of light crossing an interface is governed by the initial direction, the interface surface and the indices of refraction, a physical property of the mediums. Ray tracing models the trajectory of light in reverse from the camera into the scene and is used widely in graphical applications. Ray tracing is well suited for complex refraction through multiple interfaces, however this requires complete knowledge of the 3D scene. Stereo reconstruction aims to accurately model a 3D scene with no prior model of the scene itself. Refraction in stereo vision techniques and the various limitations have been studied, and we give a brief background in section 2. Refractive Stereo Ray Tracing (RSRT) models the refracting surface of water as a plane. This approximation is simple, but justified in many applications, as we explain in section 3.3.

We would like to thank NSF CDI Type I grant 1124664 for funding.

2. RELATED WORKS

Stereo vision is a major area of study. An excellent overview of standard stereo approaches can be found in [2], and we direct the reader to [3] for a comparison of many modern techniques. Many techniques have been developed for specialized problems [4][5][6]. These techniques assume a linear mapping of scene points onto the image plane, however under refraction this assumption does not hold. Reconstructing specular surfaces is an ongoing research area in computer vision with much work aiming to reconstruct the refracting surface[7][8][9]. [10],[11], and [12] studied refractive planes. [13] explored triangulation under refraction. Unlike these works we do not assume ray correspondences, we do not impose any temporal constraints, nor is our approach constrained to specific poses or material interfaces.

A surface stereo system could potentially be deployed instead of an underwater system. Underwater systems have been an active topic [14] [15] [16] [17]. ROV's and divers have been used to monitor animals and inspect structures. These techniques require a physical presence underwater, which can be costly and dangerous. Our approach could allow similar work to be carried out from the surface.

3. METHODS

3.1. Ray Tracing

Ray tracing is the rendering process of projecting rays through each pixel into a 3D scene to compute intensity. For a given pixel $p_i = [x_i, y_i]$, the equation for a ray V_i is given by

$$V_i = C_0 + t \cdot \frac{\beta}{norm(\beta)} \tag{1}$$

where C_0 is the camera center, and

$$\beta = R' \cdot A^{-1} \cdot [x_i, y_i, 1] \tag{2}$$

where A is the camera matrix,and R is the camera rotation.

These rays are intersected with surfaces in the scene. For our approach the relevant equation is ray-plane intersection. A plane is as defined

$$P \cdot n + d = 0 \tag{3}$$

where P is a point in the plane, n is the plane normal and d is some constant. To solve for the intersection we substitute P with V_i from equation 1, and solve for t. Plugging t back into equation 1 yields intersection point I_i.

Refraction is governed by Snell's law, which is formulated in 3D as

$$V_{refract} = r \cdot l + (r \cdot c - \sqrt{1 - r^2 \cdot (1 - c^2)}) \cdot n \tag{4}$$

where n is the interface normal, l is the light vector, r is the ratio of the index of refraction's (IOR) of the two materials n_1/n_2 and $c = -n \cdot l$. For our application the refracted ray thus has an origin of I_i, and a direction of $V_{refract}$.

3.2. Physical Properties of Water

Water has a number of physical properties being utilized in this work, but here we focus on just a few of them. The IOR of fresh water is 1.333, which differs significantly from air with an IOR of 1. Water exerts a force on objects which displace it, called buoyancy. Buoyancy causes less dense objects to float, and they come to rest at the interface[1]. We model water as a refracting plane. This accurately models scenes with mostly still water where wind creates capillary waves (small, irregular naturally occuring wind generated waves (different from larger gravity waves)). These waves have wavelengths of no more than 1.74cm and a maximum wave height (amplitude) of 0.243cm [18]. While this water is not completely planar, at the sampling size of a large scene of say 100 m^2 a network of maximum amplitude capilary waves would mean a depth variation of only 0.1% of the scene width.

Considering capilary waves with maximum amplitude of 0.243 cm and wavelength of 1.74 cm as idealized sine waves in 2 dimensions, they can be expressed as

$$sin(2x/1.7\pi) * (0.243/2) \tag{5}$$

If we differentiate this function to compute the tangent and compute an orthogonal vector, we find that capillary waves have a maximum surface perturbation angle of 24.1801°.

3.3. Plane Extraction

We extract the plane by leveraging buoyancy. In our controlled experiments we add small strips of colored paper to the surface of the water. Contrasting color can aid in segmentation, and SIFT matching is used to reconstruct just these objects on the surface. To extract plane parameters we perform a principle component analysis (PCA) of the reconstructed SIFT matches of floating objects described above. We then compute the centroid, and define our plane as the computed normal and centroid for a plane origin as defined in equation 3.

3.4. Stereo Matching

Stereo matching is an active area of research in computer vision, and feature based techniques are quite common as are disparity based approaches. Feature points are used in numerous applications to find correspondences and among these SIFT matching [19] is one of the most common and best performing. In rectified stereo images disparity estimation techniques are typically used to find dense correspondences. Under refraction however, the surface normal and camera position will affect rectification. Therefore it is necessary to calculate new rectification parameters for each image pair in which scene has changed. We crop our stereo pairs around the relevant objects, recompute rectification parameters[2] and dense correspondences are calculated using the semi global block matching technique[20]. SIFT and disparity based correspondences are used for experiments in section 4.

3.5. Refraction Based Reconstruction

To reconstruct points behind a refractive plane, we employ ray tracing techniques. Correspondences in stereo images can be thought of as 2 rays from the camera centers into the 3D scene. These rays intersect the plane and are refracted according to equation 4.

We then compute the closest intersection of these rays using a least squared error by looking at the squared error function for a parametrically defined ray. For line i, the squared error function is

$$D^2(t) = (x - x_i - a_i * t)^2 + (y - y_i - b_i * t)^2 + (z - z_i - c_i * t) \tag{6}$$

where our point is defined as $[x, y, z]$, and our ray is defined as initial point $[x_i, y_i, z_i]$ and unit direction vector $[a_i, b_i, c_i]$.

$$l_i = [x_i, y_i, z_i] + t_i * [a_i, b_i, c_i] \tag{7}$$

To minimize the error we take the derivative of the function to find the minima at 0. This allows us to solve a system of 6 equations with 5 unknowns $[x, y, z, t_1, t_2]$. Solving this system gives the intersection, $[x, y, z]$, and by using the calculated t_1 value we can determine triangulation error. The point on ray 1 closest to the $[x, y, z]$ is $p_1 = [x_1, y_1, z_1] + t_1 * [a_1, b_1, c_1]$. Triangulation error is then the Euclidian distance $E_t = dist(p_1, [x, y, z])$ and is very useful for classifying points as inliers or outliers. We discard points where $E_t \geq \mu$ for our choice of threshold μ.

4. EXPERIMENTS

4.1. Synthetic Experiments

In this subsection we will test the basic reconstruction technique with synthetic scenes as well as quantify sources of error. We test various sources of error in our approach using rendered scenes for which we have ground truth. The scene consists of a textured object, either a cube or sphere, and a refractive plane. The objects have been rendered with a highly textured surface to facilitate dense SIFT matching. We measure RMS from the surface to the reconstruction normalized to the radius of the sphere or half cube length. We will focus on 3 sources of error specific to our problem, namely errors introduced in estimating the plane normal, errors in estimating the plane origin, and errors in estimating the refracting material. We synthetically vary our estimates for these parameters and observe errors in the resulting reconstruction. We conduct 8 experiments with each shape varying IOR for the refractive plane. Results are discussed in section 5.1.

To quantify the effect of an inaccurate estimate of the surface normal we randomly perturb the refractive plane normal by varying amounts. To do this we take our ground truth normal vector and perturb it as follows. The set of all vectors with θ angle to unit vector v_1 form a circle of the unit sphere. The parametric equation for a circle is

$$p(t) = r \cdot cos(t)u + r \cdot sin(t) \cdot u \times n + c \tag{8}$$

where r is the circle radius, n is the unit normal to the circle, c is the center, and u is a unit vector orthogonal to the normal. For our applications n is the original normal, $r = sin(\theta)$, $c = n \cdot cos(\theta)$. To generate a random point on the circle we find the plane on which the circle lies, defined by n and c. we generate a random linear basis in the plane for the intersection with the circle. We then randomly select a point on this circle, and take this to be the new estimate for the plane normal. We perturb the estimate normal from $0°$ to $30°$ in steps of $0.05°$. Since this is a stochastic process we repeat the experiment 100 times for each increment of perturbation, and report the mean. Additionally we experiment with artificially altering the origin of the plane, by moving the origin along the normal and measure error. For each synthetic scene we vary the position of the estimated plane centroid from [0,0,1.5] to [0,0,9] in increments of 0.01 and report the resulting RMS. Finally to explore how misestimation of the index of refraction affects reconstruction we vary the estimated IOR from 0.5 to 1.5 in increments of 0.001 and report the resulting RMS.

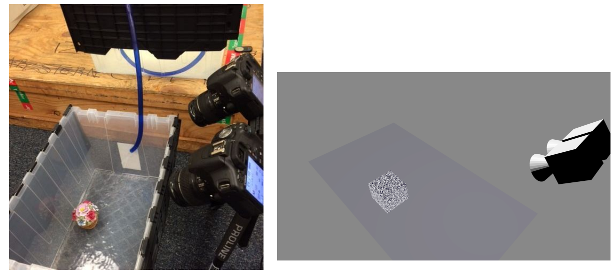
4.2. Controlled Experiments

To demonstrate that this approach works in a real world scene we have conducted a series of experiments with real objects. We have constructed a calibrated stereo system which captures images of objects in a bin that is subsequently filled with water as illustrated in Fig. 1a. We place objects in a stable position in the bin, capture stereo pairs as ground truth, and siphon water from an upper reservoir so as to not disturb the object. We then add small pieces of colored paper that float to the water for use in extracting the refractive plane parameters and stereo pairs are captured again. We mask the region with colored paper and reconstruct SIFT points in this region, erroneous points are manually removed, and PCA is applied. Quantitative results are obtained by measuring point cloud to point cloud distance using the Cloud Compare utility [21].

5. ANALYSIS OF RESULTS

5.1. Synthetic Results

Results for synthetic experiments are shown in Fig. 2. The color key is in Fig 2d.



(a) The controlled setup with water being siphoned from upper reservoir to the imaging vessel.

(b) An illustration of the synthetic stereo setup.

Fig. 1: The experimental setup.

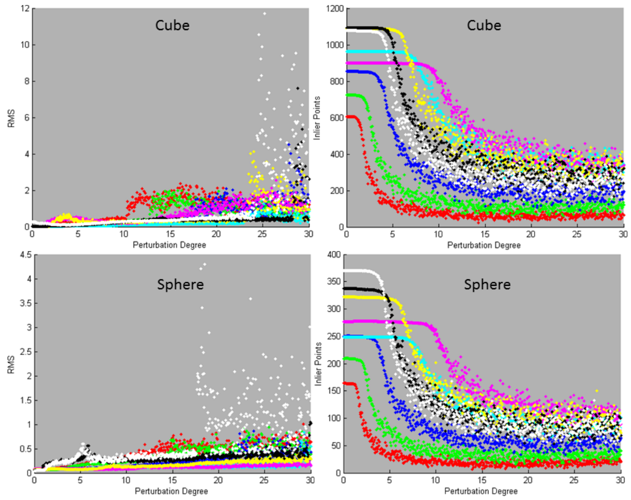
From the results in Fig. 2 we see that with increased perturbation, we do not see a large initial increase in RMS nor do we see a large decline in the number of points classified as inliers. This suggests that for small perturbance in the normal there is not much effect on the reconstructed surface, and triangulation error increases a small amount. With a perturbation of more than $5°$ we see a rapid decline in the number of points that are classified as inliers, coupled with a slow but erratic increase in RMS. This suggests that while we do see an increase in error, many points are correctly discarded by thresholding on triangulation error. At more extreme angles however, RMS reaches the highest of any of our experiments (Note: each graph has a different scaling).

Results for varying the plane origin are shown in Fig. 2b. In our scene the plane is at $z = 4$ and we see that the error reaches a minimum at this point. These results show that the plane position, and therefore intersection point has a less significant effect on the total error than the IOR and surface normal, however it is also more difficult to classify points as outliers.

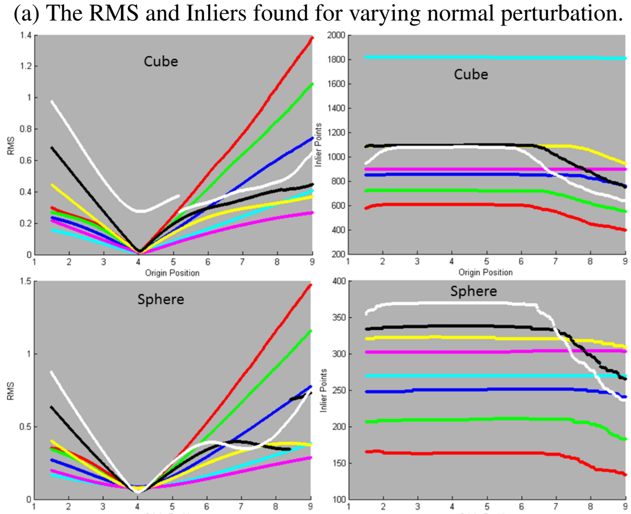
Results for varying the IOR of the refracting plane are shown in Fig. 2c. These results indicate a good estimate of IOR is important, but in practice this is easily done for fresh water, as even many inclusions do not drastically affect IOR[22]. Our results show within a small neighborhood IOR minimally affects error.

5.2. Controlled Experiment Results

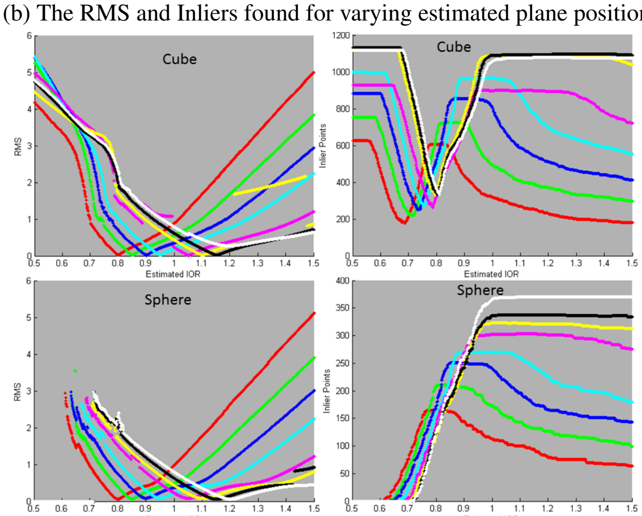
In this section we compare the reconstruction of our real world objects with and without refraction. In Fig. 3a and 4a we see qualitative results of reconstructing the model brain and flower pot respectively. The models, reconstructed with and without refraction, are presented occupying the same coordinate space. It worth noting that in both sets of reconstructions slightly different portions of the objects are reconstructed. This is because with a refractive surface the cameras see a different view of the object, and in the case of our experimental setup they see a more direct view of the top



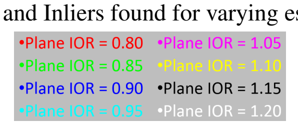
(a) The RMS and Inliers found for varying normal perturbation.



(b) The RMS and Inliers found for varying estimated plane position



(c) The RMS and Inliers found for varying estimated IOR.

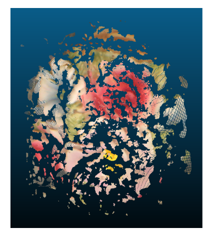


(d) the color key for all synthetically rendered scenes.

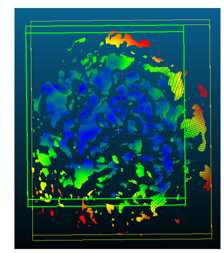
Fig. 2: Results for synthetic scenes.

of the object.

For quantitative results we show the distance map from the refracted reconstruction to the ground truth model in Fig. 3b and 4b. The bounding boxes for both point clouds are shown with the reference (non refracted) box in green and the refracted reconstruction bounding box in yellow. For the flower pot we achieve a mean distance of 5.007mm with a standard deviation of 4.712mm. For the model brain we obtain a mean distance of 8.536mm, and a standard deviation of 7.584mm.

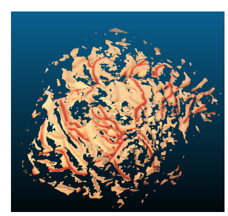


(a) The ground truth and refracted models

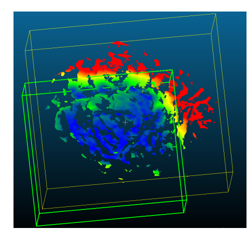


(b) The cloud-cloud distance and bounding boxes for the refracted model

Fig. 3: Results for the reconstructed flower pot.



(a) the ground truth and refracted models



(b) The cloud-cloud distance and bounding boxes for refracted model

Fig. 4: Results for the reconstructed brain model.

6. CONCLUSION

In this work we have presented a scheme for reconstructing underwater objects from surface based stereo systems by Refractive Stereo Ray Tracing. These scenes pose problems to typical techniques. We model water as a refracting plane and have shown this holds for many scenes. We have presented methods to extract plane parameters and reconstruct underwater objects. We have quantified potential error using rendered scenes with refraction. Our results show that correctly estimating the refractive surface is important but that it is also possible to eliminate erroneous points. We have further tested this technique with real scenes and demonstrated that plane parameters can be extracted by leveraging buoyancy. This work aims to allow surface based reconstruction of submerged structures, and does not require a physical presence underwater which can be costly and dangerous.

7. REFERENCES

[1] Hugh Young, *Sears and Zemansky's university physics : technology update*, Pearson Education, San Francisco Toronto, 2014.

[2] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, NY, USA, 2 edition, 2003.

[3] Jannik Fritsch, Tobias Kuehnl, and Andreas Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.

[4] M.V. Rohith, S. Sorensen, S. Rhein, and C. Kambhamettu, "Shape from stereo and shading by gradient constrained interpolation," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, Sept 2013, pp. 2232–2236.

[5] Sing Bing Kang, R. Szeliski, and Jinxiang Chai, "Handling occlusions in dense multi-view stereo," in *Computer Vision and Pattern Recognition, 2001.*, 2001, vol. 1, pp. I–103–I–110 vol.1.

[6] Peter J. McNerney, J. Konrad, and Margrit Betke, "Block-based map disparity estimation under alpha-channel constraints.," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 17, no. 6, pp. 785–789, 2007.

[7] Nigel J. W. Morris and Kiriakos N. Kutulakos, "Dynamic refraction stereo.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1518–1531, 2011.

[8] Guillaume Gomit, Ludovic Chatellier, Damien Calluaud, and Laurent David, "Free surface measurement by stereo-refraction," *Experiments in Fluids*, vol. 54, no. 6, 2013.

[9] Jinwei Ye, Yu Ji, Feng Li, and Jingyi Yu, "Angular domain reconstruction of dynamic 3d fluid surfaces.," in *CVPR*, 2012, pp. 310–317.

[10] Visesh Chari and Peter F. Sturm, "Multi-view geometry of the refractive plane.," in *BMVC*. 2009, British Machine Vision Association.

[11] Yao-Jen Chang and Tsuhan Chen, "Multi-view 3d reconstruction for scenes under the refractive plane with known vertical direction," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, Nov 2011, pp. 351–358.

[12] Zhihu Chen, Kwan-Yee Kenneth Wong, Yasuyuki Matsushita, and Xiaolong Zhu, "Depth from refraction using a transparent medium with unknown pose and refractive index.," *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 3–17, 2013.

[13] M. Alterman, Y.Y. Schechner, and Y. Swirski, "Triangulation in random refractive distortions," in *IEEE International Conference on Computational Photography (ICCP), 2013*, April 2013, pp. 1–10.

[14] Luis B. Gutierrez, Carlos A. Zuluaga, Juan A. Ramirez, Rafael E. Vasquez, Diego A. Florez, Elkin A. Taborda, and Raul A. Valencia, "Development of an underwater remotely operated vehicle (ROV) for surveillance and inspection of port facilities," in *Volume 11: New Developments in Simulation Methods and Software for Engineering Applications: Safety Engineering, Risk Analysis and Reliability Methods: Transportation Systems*. 2010, ASME.

[15] Shahriar Negahdaripour and Pezhman Firoozfam, "An rov stereovision system for ship hull inspection," 2005.

[16] Yohay Swirski and Yoav Y. Schechner, "Caustereo: structure from underwater flickering illumination," 2012, vol. 8480, pp. 84800D–84800D–8.

[17] Julia Santana-Garcon, Matias Braccini, Tim J. Langlois, Stephen J. Newman, Rory B. McAuley, and Euan S. Harvey, "Calibration of pelagic stereo-bruvs and scientific longline surveys for sampling sharks," *Methods in Ecology and Evolution*, vol. 5, no. 8, pp. 824–833, 2014.

[18] Alan P. Trujillo Harold V. Thurman, *Essentials of Oceanography*, Prentice Hall, seventh edition, 2001.

[19] D. Lowe, "Distinctive image features from scale-invariant keypoints," in *International Journal of Computer Vision*, 2003, vol. 20.

[20] Heiko Hirschmller, "Accurate and efficient stereo processing by semi-global matching and mutual information.," in *CVPR (2)*. 2005, pp. 807–814, IEEE Computer Society.

[21] Telecom ParisTech EDF R&D, "Cloudcompare (version 2.5.5.2)[gpl software]," 2014, Retrieved from http://www.cloudcompare.org/.

[22] CRC Handbook, *CRC Handbook of Chemistry and Physics, 88th Edition*, CRC Press, 88th edition, 2007.