2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Daejeon Convention Center
October 9-14, 2016, Daejeon, Korea

Towards Real-time Underwater 3D Reconstruction with Plenoptic Cameras

Katherine A. Skinner[1] and Matthew Johnson-Roberson[2]

Abstract— **Achieving real-time perception is critical to developing a fully autonomous system that can sense, navigate, and interact with its environment. Perception tasks such as online 3D reconstruction and mapping have been intensely studied for terrestrial robotics applications. However, characteristics of the underwater domain such as light attenuation and light scattering violate the brightness constancy constraint, which is an underlying assumption in methods developed for land-based applications. Furthermore, the complex nature of light propagation underwater limits or even prevents subsea use of real-time depth sensors used in state-of-the-art terrestrial mapping techniques. There have been recent advances in the development of plenoptic (also known as light field) cameras, which use an array of micro lenses capturing both intensity and ray direction to enable color and depth measurement from a single passive sensor. This paper presents an end-to-end system to harness these cameras to produce real-time 3D reconstructions underwater. Our system builds upon the state-of-the-art in online terrestrial 3D reconstruction, transferring these approaches to the underwater domain by gathering real-time color and depth (RGB-D) data underwater using a plenoptic camera, and performing dense 3D reconstruction while compensating for attenuation effects of the underwater environment simultaneously, using a graphics processing unit (GPU) to achieve real-time performance. Results are presented for data gathered in a water tank and the proposed technique is validated quantitatively through comparison with a ground truth 3D model gathered in air to demonstrate that the proposed approach can generate accurate 3D models of objects underwater in real-time.**

Index Terms— **plenoptic cameras, light field cameras, underwater 3D reconstruction, real-time 3D reconstruction, computer vision**

I. Introduction

The aqueous medium presents many challenges for the operation of subsea robotic systems. Sensing capabilities are limited by the attenuation of electromagnetic signals, which hinders the application of current state-of-the-art terrestrial perception techniques in achieving similar success in underwater environments. Real-time depth sensing has been an incredible boon to land-based mobile robotics [1], with advances in sensor modalities such as stereo cameras, light detection and ranging (LIDAR), and more recently color and depth (RGB-D) sensors [2]. Unfortunately, the latter two – time-of-flight range sensing and pattern projection structured light approaches – are still quite limited in their success underwater [3]. Conversely, stereo cameras have met with great success underwater [4], as passive optical sensing can be achieved even at depth using artificial light at close range. This range limitation is complicated by the fact that typical submersible platforms lack the instant control authority to follow an undulating benthos precisely, leading to rapidly varying ranges between camera and target. This creates a challenge for traditional optics as the depth of field (DoF) is insufficient to maintain focus across the typical range of platform altitudes. Additionally, the physical stereo camera setup is often quite cumbersome as the underwater housing for the cameras becomes heavier and more expensive as the inner diameter increases to accommodate two cameras with even a modest baseline of separation. In response to these issues, plenoptic or light field cameras have been proposed as an alternative to traditional stereo cameras for underwater applications [5]. These cameras increase the DoF enabling focus across a much greater range of altitudes from the target and they allow for depth recovery with a single imaging sensor and single optics resulting in a compact form factor, reducing the challenge, complication and weight of designing a larger depth-tolerant pressure housing. This paper discusses the use of these cameras in a real-time 3D reconstruction system for underwater applications, an enabling technology for many tasks carried out by autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs), including, but not limited to, navigation, obstacle avoidance, mapping, and grasping and manipulation.

This paper includes the following contributions: (i) a novel end-to-end system that uses plenoptic cameras to generate a real-time 3D reconstruction of a scene; (ii) a novel integration of an underwater propagation model to account for attenuation effects of the water column in online 3D model building; and (iii) quantitative results that compare these models to terrestrially scanned ground truth to validate the accuracy of the proposed approach. The paper is organized as follows: §II will present background and prior work; §III will describe the proposed water column compensated 3D reconstruction approach; §IV will discuss the experimental design for ground truth and present results of the validation; §V will discuss these results; and finally, §VI will present our conclusions and future work.

II. Background

A. Real-time Dense 3D Reconstruction

Real-time dense 3D reconstruction is a critical perception task for fully autonomous robotic systems. Recently developed RGB-D simultaneous localization and mapping (SLAM) systems have demonstrated the ability to perform

[1]K. Skinner is with the Robotics Program, University of Michigan, Ann Arbor, MI 48109, USA kskin@umich.edu

[2] M. Johnson-Roberson is with the Department of Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, MI 48109, USA mattjr@umich.edu

this task with a high degree of success for land-based applications [6] [7]. The reconstruction framework in this paper is adopted from fusion-based methods, specifically Elastic-Fusion [8]. Fusion-based methods perform online alignment of overlapping color images and depth maps to maintain and update a single 3D model, while tracking an estimated pose to enable global loop closures. Other fusion-based methods include KinectFusion [9] and Kintinuous, a spatially extended KinectFusion [10]. Each of these methods exploits RGB-D sensors capable of returning high-resolution color images with corresponding depth or range measurements to sense the surrounding environment.

B. Underwater 3D Reconstruction

Unfortunately, the aqueous medium presents unique challenges to transferring these terrestrial approaches to underwater environments. Complex effects on light propagation limit the effectiveness of active RGB-D sensors subsea and lead to a violation of the brightness constancy constraint (BCC) [11], each of which are commonly used to achieve real-time mapping in terrestrial applications. There have been recent advances in methods that use passive optical sensors for underwater applications. Several offline dense 3D reconstruction techniques have been developed and tested underwater with stereo camera data [4] [12]. Jordt-Sedlazeck et al. developed an underwater light propagation model to account for the effects of refraction through a flat port camera housing in both camera calibration and structure-from-motion [13] [14]. Additionally, Bryson et al. developed methods to perform color correction of images captured in underwater environments for color consistent reconstructions [15]. While these methods provide a basis for overcoming several fundamental limitations of underwater vision, they are all offline techniques. The goal of this work is to advance towards the online dense 3D reconstruction required for many autonomous navigation and intervention tasks.

C. Plenoptic Cameras

Plenoptic, or light field, cameras have recently gained interest as increased computing power and improved image resolution have enabled the development and release of consumer-grade instruments from Raytrix [16] and Lytro [17]. Instead of a traditional single lens, light field cameras have an array of micro lenses behind the main lens that captures the position and direction of each light ray to the imaged scene, making it possible to refocus the scene, or to obtain different viewpoints from a single image after the image is captured. From this information, it is possible to extract depth information as well as a high-resolution image from a single passive optical sensor [18]. There have been several methods recently developed to achieve real-time tracking and visual odometry with plenoptic cameras for use in real-time robotics applications [19] [20] [21]. Prior work on 3D reconstruction using light field cameras has focused on static single scene reconstruction [22]. Tao et al. [23] exploited both the defocus and correspondence cues to create high quality depth maps of single frame images, which could then be interpolated to form a 3D surface reconstruction. Our focus for this work is on performing the reconstruction of an online video feed rather than a single static scene in order to obtain a complete 3D model of a large object spanning multiple frames. Furthermore, our proposed method is developed specifically for the underwater domain.

III. TECHNICAL APPROACH

Figure 1 shows a flowchart of our processing pipeline, broken into three main parts: processing of data from the light field camera, incorporation of an underwater light propagation model for this data, and the real-time 3D reconstruction procedure.

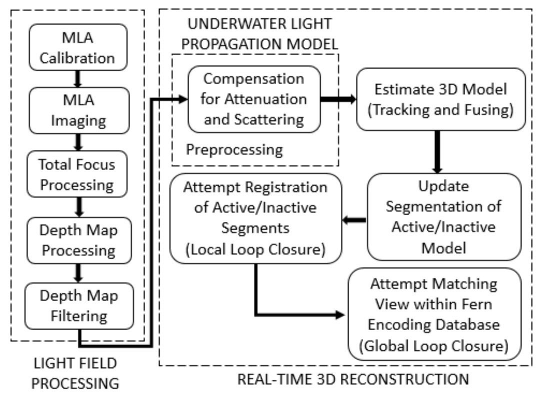


Fig. 1. Flowchart of overall pipeline for the methods presented in this paper. The key points are organized into three main sections: light field processing, compensation for underwater lighting effects, and real-time 3D reconstruction.

A. Light Field Processing

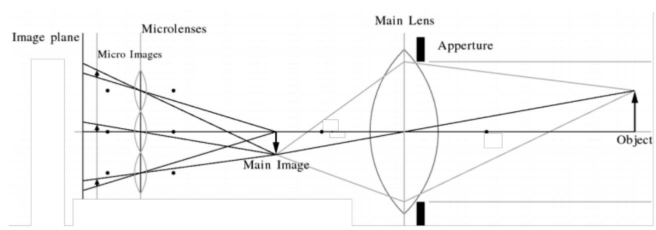


Fig. 2. (Reproduced from [16]) Camera geometry of a Raytrix light field camera showing the placement of the micro lens array between the main lens and the image plane. This allows the ray intensity and direction to be stored to provide a color image and depth map from a single camera.

Figure 2 shows the basic geometry of a light field camera, where the main difference from traditional cameras is the micro lens array (MLA) placed between the main lens and the image plane. The following outlines the basic pipeline for light field processing shown in Fig. 1. For further details of development and implementation of these steps, the reader is referred to [16], as these details are out of the scope of this work.

1) Calibration: First an MLA calibration is required to compute intrinsic camera parameters, including parameters of the micro lens structure for the light field camera. The raw uncalibrated image returned by the light field camera displays the raw intensity values read by the imaging sensor during image capture. This may be either monochrome or, for our use, color. As shown in Fig. 3a, the individual micro lenses can still be distinguished in the raw image.

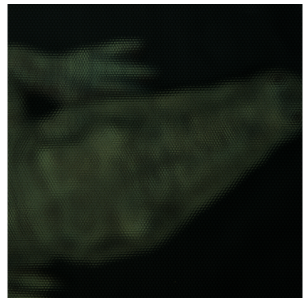
To remove the segmentation of micro lenses from the raw image, a gray image is taken during the MLA calibration with a white filter for continuous illumination to highlight only the edges and vignetting of each micro lens. This provides a template for eliminating the micro lens vignetting from the raw image to produce a processed image. Finally, a metric calibration must also be performed to provide conversion from pixel space units to metric units.

The result of the calibration provides intrinsic parameters, which include micro lens diameter and offset, x- and y-direction field-of-view, depth-of-field, working distance (effective focus distance within which an object could move and remain in focus), distortion parameters, as well as lateral and depth resolution.

2) MLA imaging: Each of these micro lenses captures a small fraction of the total view of the imaging plane. These micro images are synthesized to form the processed image at a plane of focus (Fig. 3b). To do this, the depth (or virtual depth – the distance from the object to the *micro* lens) must be assumed for each pixel. For each point (x, y) in object space, a set of micro lenses is compiled that captures that point in their image plane. Then the point's projection through each of these micro lenses onto the full image plane is determined. The final color value for the pixel in the full image plane is taken as an average of pixel intensities from the corresponding micro images.

3) Total focus processing: With a plenoptic camera, DoF increases and working distance – or the range of depths a scene can remain in focus – is larger than a traditional camera. Additionally, plenoptic cameras can be designed as multi-focus cameras, where there are different types of micro lenses with each type set to a different focus throughout the MLA. Each group of similar micro lenses is considered as a subarray. An image with "total focus," or one in which a large DOF is maintained can be synthesized using multiple subarrays of different focal lengths to compute an image in focus over a large working distance. Figure 3c shows a total focus image. Note the clarity versus the single plane of focus in Fig. 3b.

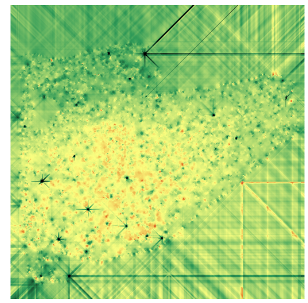
4) Depth map processing: Plenoptic cameras are designed such that each micro image shares correspondences with its neighbor, so that when the MLA is calibrated to determine geometric parameters between micro lenses, these correspondence cues can be used for determining depth via triangulation. Note that the best triangulation performance is achieved in regions of high contrast. Figure 3d shows a sample depth map from our light field camera. Note the noise that appears in the low contrast region. The following section describes an approach to ameliorate this issue.



(a) Raw MLA image initially returned by the light field camera prior to calibration. The edges of each micro lens are visible due to vignetting.

(b) Refocused image after MLA calibration and refocusing for single focal length.

(c) Image with total focus showing the large DoF achievable. Note that the scene is in focus through the full depth of the object.



(d) Depth map after MLA and metric calibration with colors ranging from yellow to green as depth from the camera increases.

Fig. 3. Set of processing steps used to go from raw MLA image to focused images and depth maps. This data serves as the input for our 3D reconstruction pipeline. The use of the plenoptic camera affords us RGB-D images at high framerate in a domain where other techniques like pattern projection and time-of-flight are difficult or impossible to employ. Note the high noise in the depth image (d). Such images necessitate the use of a 3D reconstruction pipeline where noise can be suppressed through multiple observations.

5) Depth map filtering: To decrease noise in the final depth map, we filtered the depth with a combination of mean, median, and bilateral filtering. Additionally, we implemented a maximum depth filter which eliminates noise or erroneous depth measurements beyond the plane of interest by eliminating these points from consideration in the reconstruction process. Robustness to remaining noise is afforded by the subsequent processing pipeline to generate a full 3D model from many noisy observations.

B. Real-time Dense 3D Reconstruction

With depth maps and images calculated, the next step in the proposed pipeline is to perform reconstruction to produce a dense 3D model. Our architecture is built around the fusion-based reconstruction framework developed by Whelan, Leutenegger, Moreno, *et al.* [8], with modifications for the underwater domain. The following subsections summarize the relevant sections of the fusion-based reconstruction pipeline and detail our novel integration of a real-time water column compensation approach. As with the light field processing, the entire framework relies on the graphics processing unit (GPU) to facilitate real-time performance.

1) Fusion-based reconstruction framework: The global 3D model follows the surfel representation proposed by Keller, Lefloch, Lambers, *et al.* [7], where a surfel $\mathcal{M}^{\mathcal{S}}$ contains position, color, radius, normal, weight, timestamp of initialization, and timestamp of last update. The set of all surfels $\mathcal{M}$ is segmented into active and inactive subsets. The active subset consists of points that have been seen within a set time threshold δ_t, where the inactive subset consists of points that have not been seen for greater than δ_t. This bounds the computation complexity and memory footprint of the matching process and enables real-time performance.

In real-time, each incoming RGB-D frame is added to the active map and compared to the inactive model that, based on pose estimates, lies in the same spatial region. Tracking of the estimated pose is done by solving for the motion parameters (position and orientation) that minimize the following cost function:

$$E_{track} = E_{icp} + w_{rgb} E_{rgb} \tag{1}$$

where w_{rgb} is a tuning parameter, E_{icp} is the point-to-plane error function for geometric pose estimation, and E_{rgb} is the photometric error function based on intensity change from pixel to pixel.

If the active model can be matched to an inactive portion, a local loop closure is made so the inactive map segment is made active again, aligned with the matching active portion through iterative closest point alignment (ICP) [24] and fused to the global 3D model. The RGB values are fused using a moving average of colors.

2) Underwater light propagation model: In the 3D reconstruction framework described above, the term E_{rgb} represents the photometric consistency of an observation to the model. It is a energy term that expresses the BCC assumption underlying many terrestrial vision algorithms. As discussed previously, underwater optical imaging is a challenging problem because of light attenuation, light scattering, moving light sources, and other violations of that assumption.

We propose a modified photometric constraint E_{rgb} that harnesses the significant work that has gone into understanding, modeling, and compensating for these range-dependent effects. By accounting for water column effects directly in the fusion step of the reconstruction framework, we can correct the RGB values of input images based on input depth estimation, and enforce a photometric constraint that is more consistent with reality simultaneously to structure computation. As such, this proposed method seeks to improve both the geometric accuracy as well as photometric quality of the resulting 3D model.

The physics of light in water are now well modeled. One of the major effects on received light in the aqueous environment is absorption. As photons travel through water both water molecules and particulates in the water absorb the light's energy. The degree to which the light is absorbed is dependent on the wavelength of the light (color) and the properties of the water it is traveling through. For example, seawater, freshwater, coastal, and deep ocean waters exhibit different rates of attenuation.

Another important water column effect is scattering, which can occur upon a photon's collision with a particle in its path. The result is a new photon traveling in a random direction. Some proportion of this scattered light ends up being received at the camera and is referred to as backscatter. This work is focused on a real-time system for underwater 3D reconstruction and as such a full discussion of the underwater light propagation is beyond the scope of this paper. For a discussion of a more detailed offline color correction system see Bryson, Johnson-Roberson, Pizarro, *et al.* [15]. Here we employ a modified Jaffe-McGlamery model [25] [26], which when implemented on the GPU can achieve real-time performance in our online 3D reconstruction system.

The core water column effects in the Jaffe-McGlamery model are range and wavelength dependent. Lambert's law states that the absorption of photons is exponential following this form:

$$L(d, \lambda) = Re^{-a(\lambda)d} \tag{2}$$

where λ is wavelength of the light, R is the irradiance before traveling through the water column, and d is the distance traveled.

In addition, to properly account for scattering effects we compute a scattering coefficient $b(\lambda)$ that depends on the volume scattering function $\beta(\psi, \lambda)$ [27], which unlike absorption is parametrized over the scattering angle ψ and therefore must be integrated over all directions:

$$b(\lambda) = 2\pi \int_0^{\pi} \beta(\psi, \lambda) \sin\psi \, d\psi \tag{3}$$

Scattering is modeled in a similar form to absorption.

$$L(d, \lambda) = Re^{-b(\lambda)d} \tag{4}$$

where again λ is wavelength of the light, R is the irradiance before traveling through the water column, and d is the distance traveled. This enables us to summarize absorption and scattering into one single attenuation coefficient:

$$\eta(\lambda) = a(\lambda) + b(\lambda) \tag{5}$$

$$L(\lambda, d) = Re^{-\eta(\lambda)d} \tag{6}$$

where R is the initial irradiance and $L(\lambda, d)$ is final irradiance after traveling distance d. Values for the attenuation coefficient $\eta(\lambda)$ can be estimated using previous approaches [15] and the proposed technique assumes that $\eta(\lambda)$ has been estimated a priori. In implementation we discretize $\eta(\lambda)$ to the three major wavelengths captured by our camera – red, green, and blue light – and use the known pure water attenuation coefficients shown in Table I.

In the GPU both color and range information are available and so the entire water column compensation can be implemented as a pre-processing fragment shader in the frame integration step of the 3D reconstruction pipeline (see Fig. 1). Note that the water column effects *within* a single frame in the MLA processing steps are currently being ignored for implementation simplicity and in practice the intra-frame variations in depth are much less significant than the inter-frame camera movement in effecting lighting path lengths. However, such a correction could be easily added to the single frame depth estimation step and is planned for future work.

TABLE I

PURE WATER ATTENUATION COEFFICIENTS ACCOUNTING FOR BOTH ABSORPTION AND SCATTERING.

λ in $[nm]$	Attenuation η in $[m^{-1}]$
440 (blue)	0.015
510 (green)	0.036
650 (red)	0.350

IV. EXPERIMENTAL RESULTS

We tested the above approach in a controlled lab setting (see Fig. 4) with a Raytrix R5 light field camera (see Fig. 5a and Table II for technical specifications) in a custom flat port underwater housing (Fig. 5b). For our optical setup we mounted a 25 mm lens, focused at infinity, on the Raytrix R5. This focus was chosen so the total target would be in focus from a survey depth of 1 m above the object. The achievable lateral resolution of this setup is 0.88 mm, and the achievable depth resolution is 7.38 mm with an approximate x- and y-plane field-of-view of 1.33 m, and depth-of-field of 2.64 m.

TABLE II

TECHNICAL SPECIFICATIONS FOR THE RAYTRIX R5 LIGHT FIELD CAMERA.

Specification	Value
Resolution	2048x2048
Pixel Size	5.5 micrometer square pixels
Max Frames Per Second	56
Dimensions	52x52x37mm
Focal Length	25mm

The test procedure was as follows: the object was placed at the bottom of a large water tank, the camera was calibrated in its housing in water, and then the camera was moved over the object at a distance of approximately 1 meter from the object in multiple controlled passes while the data was processed on two networked and GPU equipped Intel i7 3.0Ghz Desktop PCs with 16GB of RAM. With our system consisting of an NVIDIA GTX 980, we achieved processing rates around 25-30Hz for images at an output resolution of 1024$\times$1024. This frame rate is sufficient for typical 3D reconstruction tasks we envision on an AUV or ROV. Note that the calibration provides intrinsic and extrinsic camera

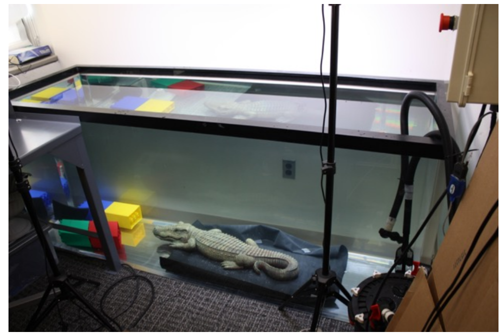


Fig. 4. Laboratory water tank setup and target object for reconstruction. Lighting and water clarity can be tightly controlled enabling us to test the limits of the proposed approach.

(a) Raytrix R5 light field camera (b) Flat port housing

Fig. 5. Underwater imaging setup for light field experimentation, with (a) the Raytrix R5 camera contained in (b) a custom underwater housing. Calibration through the flat viewport proved reliable in laboratory experimentation.

parameters including distortion effects, which are accounted for during light field processing to obtain online depth maps and total focus images for input to our proposed 3D reconstruction methods. We used the open-source implementation of ElasticFusion as a base for our reconstruction methods [8], and modified this existing implementation to integrate our underwater light propagation model and account for water column effects. Since we have RGB and depth information directly from the plenoptic camera, this model consists of a single constant-time operation, and thus this integration does not significantly impact the run-time of the proposed pipeline.

A. Validation Procedure

To demonstrate that the proposed method has validity we gathered a 3D model in real-time underwater and compared the accuracy and fidelity of that model to ground truth gathered in air. To do this we generated a ground truth dataset by scanning the object with a Microsoft Kinect structured light sensor. Here we present several metrics to demonstrate the accuracy of the model both structurally and photometrically with respect to the terrestrial ground truth.

To produce a quantitative comparison we use the well established point-to-surface distance metric and calculate it over the experimentally gathered mesh compared with the ground truth mesh. The point-to-surface distance $d(p, S)$ for a point p and a surface S is given by [28]:

$$d(p, S) = \min_{p \in S} d(p, p') \tag{7}$$

As a point-to-point distance measure $d(p, p')$, we use Euclidean distance in $\mathbb{R}^3$. Then to determine the distance between two surfaces S_1, S_2 we define a one-sided distance $E(S_1, S_2)$ as follows:

$$E(S_1, S_2) = \max_{p \in S_1} d(p, S_2) \tag{8}$$

Since this distance is not necessarily symmetric, we then take the Hausdorff distance, so the final result is the maximum of $E(S_1, S_2)$ and $E(S_2, S_1)$. Typically this measure is used to express the structural or geometric differences between two aligned models. However, this same approach can be expanded to measure attribute distance (e.g. color, normal, etc.) by modifying the function $d(p, p')$ to express things like photometric distance [28]. Here we will present results that show both geometric and photometric consistency with the ground truth using this formulation.

B. Detailed Results

After a complete model was gathered with the proposed system it was stored and then aligned to the ground truth model through ICP. The distance metric described above was then used to measure the geometric accuracy and photometric consistency with and without the proposed water column attenuation compensation. Results are shown in Tables III and IV for geometric and photometric comparisons, respectively.

TABLE III

HAUSDORFF GEOMETRIC DISTANCE BETWEEN RECONSTRUCTED MODEL AND GROUND TRUTH.

	Uncorrected (m)	Proposed method (m)
Max	0.117685	0.100227
Mean	0.026717	0.023818
RMS	0.037340	0.033677

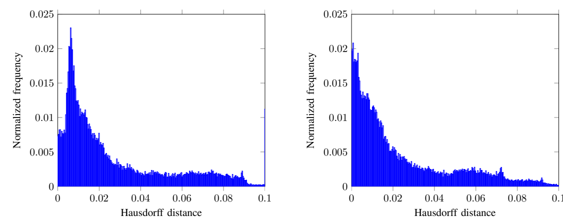
TABLE IV

HAUSDORFF PHOTOMETRIC DISTANCE BETWEEN RECONSTRUCTED MODEL AND GROUND TRUTH.

	Uncorrected (pixel intensity)	Proposed method (pixel intensity)
Max	1.56031	1.55126
Mean	0.977898	0.952538
RMS	1.02169	0.997918

The uncorrected model is obtained by performing real-time dense 3D reconstruction with the light field camera in water, with no compensation for water column effects, whereas the proposed method performs the reconstruction while simultaneously accounting for these effects using the light propagation model explained above. The maximum, mean, and root mean square geometric Hausdorff distances are presented in meters from the ground truth model obtained in air with a structured light sensor. The photometric Hausdorff distances are presented as pixel intensity units across all three wavelengths (red, green, and blue). The resulting 3D model obtained with the proposed method shows improvements in both geometric and photometric accuracy compared to the model obtained without water column compensation. For the purposes of this comparison, we used the same light field data as input to the uncorrected method and our proposed method to produce the corrected and uncorrected results shown here. Note the decrease in geometric error facilitated by improved photometric matching in the fusion process.

In addition, the histograms of these distances appear in Fig. 6, demonstrating the distribution of normalized frequency, or count, of individual point-to-plane distances of vertices across the entire model. Note that the error between the resulting 3D model and ground truth is consistently lower with our proposed method for compensation of water column effects compared to the model produced with no correction.



(a) No correction (b) Water column corrected

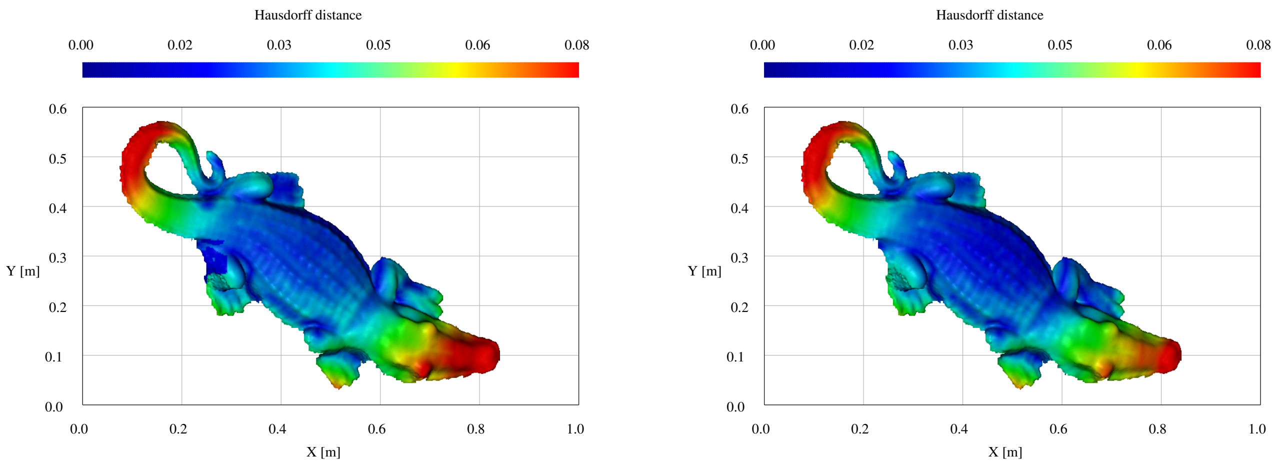
Fig. 6. Plot of Hausdorff geometric distance error (m) between the resulting 3D models and ground truth, with and without the light propagation model. Note the decrease in error in (b) where there is an increase in occurrence of low distance faces indicating the model more closely matches reality.

In addition to the overall distance metrics, we present the spatial layout of error in Fig. 7a and Fig. 7b for the uncorrected and attenuation corrected modeling procedure, respectively. Note the head of the model shows a decrease in error, or distance from the ground truth, particularly around the snout, in the water column corrected approach. This is attributable to the improved tracking and the decrease in falsely induced curvature as the camera moved along the flat model. Finally, Fig. 8 shows the full reconstructed 3D model with and without inclusion of the light propagation compensation for qualitative comparison. Note the restoration of red light in the visual texture of Fig. 8b. The power of the proposed approach is that not only does it improve tracking and modeling, but the resulting models are more visually appealing and more similar to those gathered in air.

V. DISCUSSION

Plenoptic cameras are still a nascent field but these results suggest that they offer a promising alternative to other sensors for underwater perception. As they improve, so too will their applicability and our understanding of how they can be used in the robotics domain.

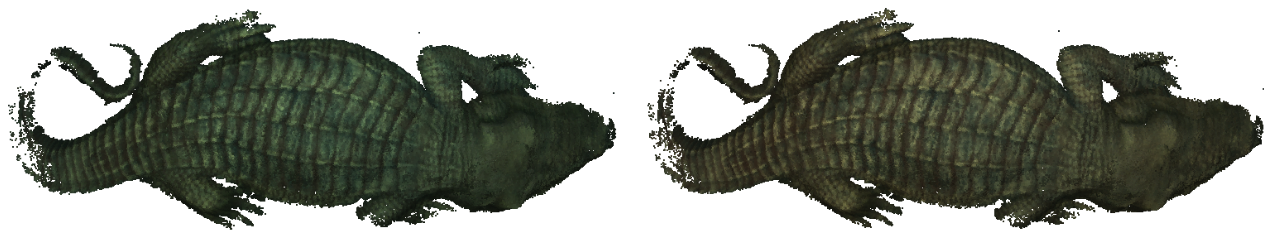
We believe that real-time underwater 3D reconstruction will be an important technology for pushing grasping and



(a) Results without incorporation of light propagation model

(b) Results with incorporation of light propagation model

Fig. 7. Spatial layout of Hausdorff distance between uncorrected and water column corrected model and ground truth. The color bar indicates the degree to which the gathered model is consistent with the ground truth. The values are projected onto the ground truth model for visualization. Note the decrease in error around the snout of the model. This highlight the higher robustness to drift in the lighting corrected fusion as this was the last area to be scanned in this run.



(a) Results without incorporation of light propagation model

(b) Results with incorporation of light propagation model

Fig. 8. Textured 3D models gathered in water in real-time using the proposed approach. The resolution of the models was high enough to detect the ridges on the back of the target bolstering support for this approach as a pathway to real-time grasping and manipulation underwater. Note in (a) the absence of and in (b) the compensation for the red spectrum light typically lost in the water.

manipulation forward in the ocean domain. Underwater manipulation has many applications including scientific data gathering (e.g. collecting samples, cores, gas capture, manipulative experiments), construction, infrastructure inspection, ordinance disposal, and equipment retrieval. Currently ROVs perform these tasks using a human operator in-the-loop. This proposed approach is a necessary first step for automating the perception tasks associated with underwater manipulation. Just as RGB-D sensors have helped push personal robotics forward in the indoor environment, the continued push to improve both the resolution and fidelity of underwater reconstruction may have a similar effect on AUVs and their capabilities.

Still, there are limitations to our proposed approach, which lead to motivations for future work. The light propagation model presented here was chosen for its relative ease of implementation and integration into the reconstruction framework. Additionally, it runs in real-time as the correction is computed through a single operation to account for both attenuation and scattering effects. However, it is important to note that this model does not account for several other water column effects or for refraction through the flat viewport housing. These unmodeled effects must be realized when considering the limitations of the resulting 3D model and future work will focus on mitigating these effects.

Next, as we rely on external software for light field processing to provide online images and depth maps from the plenoptic camera, we neglect the influence of the water column on depth map estimation and only correct for these effects during tracking and fusing. Future work will integrate water column compensation in the light field processing step to improve results.

Lastly, our experiments presented here are controlled in-lab experiments with pure water only. It is important to extend this testing to a variety of water conditions and environments in order to ensure development of a robust approach that can be transferred to implementation on ROVs and AUVs in the field.

VI. Conclusions & Future Work

Overall this work presents a fully implemented end-to-end system that uses a plenoptic camera to perform 3D reconstruction in an underwater environment. This is done by leveraging the RGB-D data gathered online with a plenoptic camera for input into an existing state-of-the-art real-time 3D reconstruction framework, which we have adapted to simultaneously account for effects of underwater light propagation during model reconstruction. These contributions provide a basis for employing the current state-of-the-art in terrestrial vision and 3D reconstruction to the underwater domain, and results presented with data gathered in an in-lab water tank show promise for achieving accurate 3D models of submerged objects with our proposed pipeline. Future work will focus on further integrating compensation for water column effects directly into the light field processing step to mitigate these effects during depth map estimation. We will also seek to extend the current model of underwater light propagation to account for the effects of refraction while maintaining real-time operation. Lastly, we plan to test these developed methods in more complex and uncontrolled environments to further validate the use of plenoptic cameras for real-time underwater perception tasks. We ultimately hope to achieve online dense 3D reconstruction on a robot in situ, which is key for enabling full autonomy for subsea robotic systems.

ACKNOWLEDGMENTS

The authors would like to thank Eduardo Iscar for assistance with data gathering and the University of Michigan Water Center - Graham Sustainability Institute for Equipment support. This material is based in part upon work supported by the National Science Foundation under Award Number: 1452793 and by work supported by National Oceanic and Atmospheric Administration: Ocean Exploration and Research Award NA14OAR0110265.

REFERENCES

[1] DARPA. (2007). DARPA urban challenge, [Online]. Available: `http://archive.darpa.mil/grandchallenge/`.

[2] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10, 2012.

[3] J. S. Jaffe, "Development of a laser line scan lidar imaging system for auv use," UC San Diego, Tech. Rep., 2011.

[4] M. Johnson-Roberson, O. Pizarro, S. B. Williams, and I. Mahon, "Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys," *J. Field Robot.*, vol. 27, no. 1, pp. 21–51, 2010.

[5] D. Dansereau, "Plenoptic signal processing for robust vision in field robotics," Dissertation, Australian Centre for Field Robotics, School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, 2014.

[6] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "Dtam: Dense tracking and mapping in real-time," in *Proc. of IEEE Int. Conf. on Comp. Vision*, 2011, pp. 2320–2327.

[7] M. Keller, D. Lefloch, M. Lambers, S. Izadi, T. Weyrich, and A. Kolb, "Real-time 3d reconstruction in dynamic scenes using point-based fusion," in *Proc. of IEEE Int. Conf. on 3D Vision*, 2013, pp. 1–8.

[8] T. Whelan, S. Leutenegger, R. S. Moreno, B. Glocker, and A. Davison, "ElasticFusion: Dense SLAM without a pose graph," in *Robotics: Science and Systems*, Rome, Italy, 2015.

[9] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking," in *Proc. of IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2011, pp. 127–136.

[10] T. Whelan, M. Kaess, M. Fallon, H. Johannsson, J. Leonard, and J. McDonald, "Kintinuous: Spatially extended KinectFusion," in *Robotics: Science and Systems Workshop on RGB-D: Advanced Reasoning with Depth Cameras*, Sydney, Australia, 2012.

[11] P. H. K. Berthold and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17.1-3, pp. 185–203, 1980.

[12] "Out-of-core efficient blending for underwater georeferenced textured 3d maps," in *Proc. of IEEE Int. Conf. on Computing for Geospatial Research and Application*, 2013, pp. 8–15.

[13] A. Jordt-Sedlazeck and R. Koch, "Refractive calibration of underwater cameras," in *Proc. of European Conference on Computer Vision*, vol. 7576, 2012, pp. 846–859.

[14] A. Jordt-Sedlazeck and R. Koch, "Refractive structure-from-motion on underwater images," in *Proc. of IEEE Int. Conf. on Computer Vision*, 2013, pp. 57–64.

[15] M. Bryson, M. Johnson-Roberson, O. Pizarro, and S. B. Williams, "True color correction of autonomous underwater vehicle imagery," *J. Field Robot.*, pp. 1–22, 2015.

[16] C. Perwass and L. Wietzke, "Single lens 3d-camera with extended depth-of-field," in *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, vol. 8291, 2012, p. 829 108.

[17] Lytro. (2016). Lytro, [Online]. Available: `https://www.lytro.com/`.

[18] E. Adelson and J. Wang, "Single lens stereo with a plenoptic camera," in *Proc. of IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, 1992, pp. 99–106.

[19] F. Dong, S.-H. Ieng, X. Savatier, R. Etienne-Cummings, and R. Benosman, "Plenoptic cameras in real-time robotics.," *Int. J. Robotic Research*, vol. 32, no. 2, pp. 206–217, 2013.

[20] D. G. Dansereau, I. Mahon, O. Pizarro, and S. B. Williams, "Plenoptic flow: Closed-form visual odometry for light field cameras," in *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2011, pp. 4455–4462.

[21] N. Zeller, F. Quint, and U. Stilla, "Narrow field-of-view visual odometry based on a focused plenoptic camera," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3/W4, pp. 285–292, 2015.

[22] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, "Scene reconstruction from high spatio-angular resolution light fields," *ACM Trans. on Graphics (Proc. of ACM SIGGRAPH)*, vol. 32, no. 4, 73:1–73:12, 2013.

[23] M. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Proc. of the IEEE Int. Conf. on Computer Vision*, 2013, pp. 673–680.

[24] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Computer Vision*, vol. 13, no. 2, pp. 119–152, 1994.

[25] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE Journal of Oceanic Engineering*, vol. 15, no. 2, pp. 101–111, 1990.

[26] B. L. McGlamery, "Computer analysis and simulation of underwater camera system performance," UC San Diego, Tech. Rep., 1975.

[27] C. Gabriel, M. A. Khalighi, S. Bourennane, P. Leon, and V. Rigaud, "Channel modeling for underwater optical communication," in *IEEE GLOBECOM Workshops*, 2011, pp. 833–837.

[28] M. Roy, S. Foufou, and F. Truchetet, "Mesh comparison using attribute deviation metric," *Int. J. Image and Graphics*, vol. 4, no. 1, pp. 127–140, 2004.