2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Underwater Stereo SLAM with Refraction Correction

James Servos*, Michael Smart*, and Steven L. Waslander†
University of Waterloo, Waterloo, ON, Canada, N2L 3G1

Abstract— **This work presents a method for underwater stereo localization and mapping for detailed inspection tasks. The method generates dense, geometrically accurate reconstructions of underwater environments by compensating for image distortions due to refraction. A refractive model of the camera and enclosure is calculated offline using calibration images and produces non-linear epipolar curves for use in stereo matching. An efficient block matching algorithm traverses the precalculated epipolar curves to find pixel correspondences and depths are calculated using pixel ray tracing. Finally the depth maps are used to perform dense simultaneous localization and mapping to generate a 3D model of the environment. The localization and mapping algorithm incorporates refraction corrected ray tracing to improve map quality. The method is shown to improve overall depth map quality over existing methods and to generate high quality 3-D reconstructions.**

I. Introduction

Underwater simultaneous localization and mapping (SLAM) has a wide breadth of applications, including coral reef inspection and monitoring [1] [2], inspection of archaeological sites [3], geological surveying [4], and profiling of northern icebergs [5]. Operating underwater presents many challenges such as unpredictable motion due to surf and currents, particulate matter and sediment floating in the water, and diverse and often unstructured environments. Also, many common sensors do not function underwater including GPS and LIDAR sensors. This work focuses on the offline generation of detailed underwater 3D reconstructions, which can be challenging to perform underwater.

Many researchers have developed submersible vehicles which are capable of SLAM operations in various forms. Zhou and Bachmayer [5] and Clark et al. [3] have successfully used sonar sensors for the mapping of underwater structures. Sonar sensors however are limited in that they can only map macro-features due to their limited resolution.

Camera based methods are commonly used to get more detailed reconstructions but present additional challenges including poor lighting conditions, washed out colors, and distortion due to refraction. One common method of generating maps from camera images is image mosaicking [6]. The mosaicking method by Bagheri et al. can produce detailed two dimensional maps of the sea floor but is not capable of reconstructing 3-D features. Mosaicking, however, has been combined with sonar sensing to generate 3-D top down topological maps by Elibol et al. [7].

* M.A.Sc. Candidate, Mechanical and Mechatronics Engineering, University of Waterloo; jdservos@uwaterloo.ca, msmart@uwaterloo.ca

† Assistant Professor, Mechanical and Mechatronics Engineering, University of Waterloo; stevenw@uwaterloo.ca

Recent developments in stereo SLAM have also been implemented underwater in order to create better reconstructions. These techniques use stereo cameras to generate depth maps which can be used to create detailed reconstructions of the environments. Warren et al. [1] were able to generate high quality maps of a coral reef by using visual odometry for pose estimation and overlaying consecutive stereo frames to create a 3-D mesh. Alternatively, in [2], a feature based approach is used to to track features over time and and align successive frames.

One of the major complications in underwater stereo SLAM is caused by the refractive interface between the air in the enclosure and the exterior water. It is insufficient to directly apply terrestrial techniques to the underwater mapping problem. When not taken into account, refraction can cause major distortions to the individual camera images as well as to the calculated depths produced from the stereo correspondences. These errors accumulate over time and result in larger errors in the overall map. Hogue et al. [2] used an additional inertial measurement unit (IMU) to compensate for the drift but this could not remove it completely.

Several researchers have presented methods to directly account for refraction in images. Mechanical methods such as using calibrated view-ports that physically compensate for the difference of refractive indices have been proven effective [8], but are costly and reliant on precise alignment. Radial correction is commonly used for refraction compensation by including it in the lens distortion parameters [9]. This method takes advantage of the fact that for any given depth, refraction exactly matches a radial distortion provided the refractive plane is perpendicular to the optical axis. However, this equivalent radial distortion varies with depth, creating a situation where the single view perspective model fails when scenes move away from the calibrated depth range [8]. For underwater stereo, radial correction has been shown to lead to inconsistent results [10], as it becomes difficult to establish correct correspondences. This occurs because the traditional epipolar assumption is no longer valid: the epipolar rays through space no longer appear as lines in the second camera, but rather as curves that are dependent on the parameters of the refractive interface [10][11].

Recently new methods for performing dense SLAM have emerged, particularly for real time applications. RGB-D SLAM [12] uses stereo camera data and leverages scale invariant features to perform frame to frame matching and optimizes over the global network. The maps are generated by projecting the data from all frames into 3-D space. This method is also susceptible to degradation due to lighting and contrast effects underwater. In contrast to the above methods, KinectFusion [13] does not rely on RGB information to construct dense 3-D models but instead uses only depth map information from the Microsoft Kinect sensor. KinectFusion aligns frames to a global map representation using an iterative closest point optimization and refines the map by averaging new information over time. This gives KinectFusion the ability to generate a detailed map even from noisy input data and to be unaffected by changes in lighting conditions.

The contribution of this work is the definition of a new stereo SLAM method which correctly accounts for refraction at both the stereo matching and 3D reconstruction stages of processing to generate dense geometrically accurate maps of underwater environments. The method is evaluated for both stereo depth map and overall map reconstruction quality in an underwater scene and is compared against the radial correction method.

II. Refractive Projection

The method used for correcting refractive effects in underwater stereo imaging follows the approach first proposed by [14] and used in [10]. First, it is important to define the methods by which pixels are transformed into and out of 3D space from a given camera image. Given a known physical configuration and depth measurement, it is possible to perform refraction-corrected projection to obtain the correct location of a 3D point observed beyond a refractive surface and to project 3D points back into the camera frame. These projections are necessary in order to generate the epipolar curves needed for stereo matching, calculate accurate depth maps, and to generate accurate 3D map reconstructions.

In this work it is assumed that the refraction interface is planar and occurs as a single refraction between the air and water. In reality the enclosure panel lies between the air and water, and refraction occurs at both air/enclosure and enclosure/water interfaces. However, the error induced by the panel is bounded to be no more than the thickness of the enclosure panel and therefore it is assumed that the enclosure is ideal, having no thickness and causing no refraction [10].

Refraction corrected projection is performed by casting rays through the refractive interface where the refractive interface P in front of the a camera is defined by the plane unit normal $N_P \in \mathbb{R}^3$ and the distance to the plane $d_P \in \mathbb{R}$. A projective ray, $r \in \mathbb{R}^3$, in air for a given pixel, $u \in \mathbb{R}^2$, is calculated by applying the inverse of the air-calibrated intrinsic camera matrix $K \in \mathbb{R}^{3\times 3}$, such that $r = K^{-1}\dot{u}$. The ray r is then scaled to intersect the refractive interface at the point p_0, given by

$$p_0 = \frac{d_P}{-r^T(N_P)} r \tag{1}$$

The direction of the ray is then refracted at the interface according to Snell's Law:

$$n_i \sin\theta_i = n_r \sin\theta_r \tag{2}$$

where θ_i is the incident angle between the ray and the surface normal, θ_r is the angle of the refracted ray, and n_i and n_r are the refractive indices of the incident medium and the refractive medium respectively. Computing the unit vector $\hat{r}$ of r allows us to apply Snell's Law and yields the refracted ray direction $r' \in \mathbb{R}^3$ [14]:

$$r' = \frac{n_i}{n_r}\hat{r} - \left(-\frac{n_i}{n_r}\cos\theta_i + \cos\theta_r\right) N_P \tag{3}$$

The direction r' and the point p_0 define the ray projected from pixel u through the refractive plane. The function $\Psi : \mathbb{R}^2 \to \mathbb{R}^3$ maps the given a pixel u to the refracted ray r', and $\gamma : \mathbb{R}^2 \to \mathbb{R}^3$ to the ray offset p_0 in the frame of the camera. Given d_P and N_P for a refractive plane, and a depth measurement at pixel u, the combination of $\Psi(u)$, and $\gamma(u)$ can be used to calculate the pixel location in 3D space.

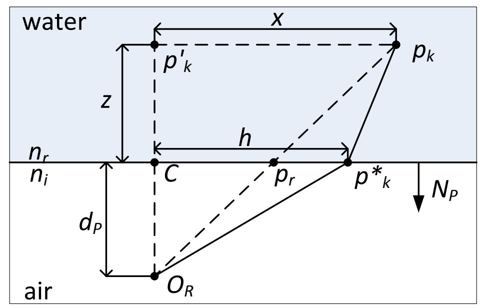


Fig. 1. Reprojection of point p_k across a refractive interface. The optical axis of the camera need not align with the plane's normal

The reprojection of a point p_k into the camera as in Figure 1 is found by the following process. First, p_k is projected onto the refractive interface's normal to obtain p'_k. Then, using the points p_k and p'_k we can define $z = \|p'_k - d\|$ and $x = \|p'_k - p_k\|$ where z represents the depth of the point p_k beyond the refractive plane and x represents the distance from the point p_k to the vector N_P through O_R. The values of x and z can be used along with the refractive indices n_i and n_r, to formulate the quartic equation for h,

$$\left[\left(\frac{n_r}{n_i}\right)^2 \left(d_P^2 + h^2\right) - h^2\right](x-h)^2 - h^2 z^2 = 0 \tag{4}$$

where Equation (4) is derived by applying Fermat's principle of least time as in [10]. In this application, only one of the four roots of Equation (4) is physically valid, and it will always lie in the interval $h \in [0, x]$ and be the only root in that interval. The valid root of (4) corresponds to the point where the refracted ray emanating from the source intersects the refractive plane. The point, C, which defines the location of intersection of the vector N_R from the camera origin with the refractive interface is found by

$$C = -d_P N_P \tag{5}$$

and the point p_r where the ray from the camera origin to the point p_k would pass through the interface if no refraction occurred is found by

$$p_r = \frac{d_P}{-p_k^T(N_P)} p_k \tag{6}$$

The point, p_k^*, where the refracted ray from the object to the camera origin intersects the interface is finally given by

$$p_k^* = h\frac{p_r - C}{\|p_r - C\|} + C \tag{7}$$

The point p_k^* can now be projected into the image without concern for the refractive plane by using the camera's intrinsic camera matrix, K and the function $\pi = (x/z, y/z)$. The reprojected pixel, u_k, is then

$$u_k = \pi(K p_k^*) \tag{8}$$

The function $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$, $u_k = \Phi(p_k)$, is defined using the reprojection equations and maps a a 3D point to a pixel.

III. Epipolar Curve Generation and Triangulation

A standard stereo camera algorithm attempts to match pixels in the reference image with those in the target image by searching linear epipolar curves through the target image. However, when refraction is introduced to the stereo camera the epipolar curves become non-linear and are dependent on the refractive properties of the system. The stereo setup is assumed to be a binocular baseline stereo setup with a known transformation, T_c, linking the left and right camera frames and corresponding rotation, $R_c \in \mathbb{SO}3$.

The epipolar curve for a pixel is obtained by casting a ray through the pixel, u, from the left reference image and then finding the set of the ray's reprojections, $Q_u = \{q_u^1, q_u^2, ...q_u^m\}$, in the right camera image. The set Q_u defines the epipolar curve of pixel u and contains correspondence points in the second image. The ray points used to generate Q_u are obtained by sampling points along the refracted ray, converting them to the right camera's local frame using T_c, and finding the reprojections of the points, p_k, in the right camera image such that $u_k = \Phi(p_k)$.

We exhaustively sample the curve to obtain all of the relevant elements $q_u^i \in Q_u$ for a specified range of depths and store the values in a lookup table for depth map calculation.

Once the set of pixel correspondences are found, the depth measurements can be calculated based on the corresponding pixel locations, u_L and u_R. The depth measurements are computed by triangulating the intersection of the refracted rays projected from each of the pixels, using the interface parameters N_R, d_R and N_L, d_L for the right and left camera respectively. The rays from the left and right pixels, r'_L and r'_R are computed such that $r'_L = R_c \Psi(u_L)$ and $r'_R = \Psi(u_R)$ with the corresponding offests. Ideally the rays would intersect exactly at one point, but in reality this does not occur. The points, p'_L, p'_R, along each ray that form the shortest line segment connecting the two rays can be found and the midpoint can then be taken as the estimate of the 3D triangulated point $\tilde{p} \in \mathbb{R}^3$.

IV. Refraction Calibration

It is assumed that the physical position and orientation of the camera system relative to the refractive plane, i.e. relative to the camera enclosure, is constant and therefore aquatic calibration need only be performed once as long as the camera is not moved relative to the enclosure. A second calibration routine is performed to identify the parameters of the refraction model once the stereo camera has been calibrated in air. Using the air-calibrated system, images of a known checker-board are captured underwater and used to establish the corners as a set of known feature correspondences, A non-linear minimization can then be performed to determine the refractive parameters $\phi = \{n_i, n_r, d_L, d_R, N_L, N_R\}$. The error function is chosen as the sum of the squared distances between the corresponding points p'_L, p'_R, determined using triangulation as seen in Section III

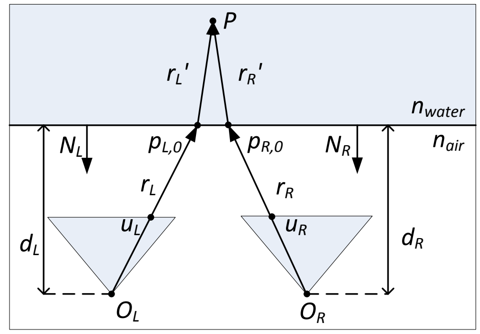


Fig. 2. Refraction corrected triangulation of 3-D point P. The ray directions from the left and right cameras, r_L and r_R respectively, are refracted at points $p_{L,0}$ and $p_{R,0}$ on the interface to the new directions r'_L and r'_R.

It is important to ensure that the calibration images include the checker-board at multiple depths and cases where the corresponding corners appear near the edges of the image as these are the areas where refraction will be most prominent.

V. Stereo Block Matching

To perform stereo matching between the cameras, a sum of absolute differences (SAD) block matching algorithm [15] is used. Performing this 1-D optimization along the epipolar curve for each pixel in the reference image generates a stereo disparity map. The depths for each of the corresponding pixel pairs can then be determined by using pixel ray tracing through the refractive interface to find the point of intersection between the left and right image ray.

The non-refractive block matching approach can be performed relatively quickly by leveraging the linear search paths and can be easily parallelized. The search for correspondence is simply the minimum of the precalculated SAD sum for each pixel. The use of non-linear epipolar curves maintains the inherent possibility for parallelized implementations but is slowed down considerably due to the non-linear search paths not having a simple mapping to and from the pixel difference space. The optimization of the non-linear block matching is not investigated here but remains a challenge for future work.

VI. Dense Localization and Mapping

The localization and mapping method used is adapted from KinectFusion[13]. The KinectFusion algorithm has many beneficial characteristics which make it specifically suitable for underwater inspection tasks and can be modified for use with stereo camera and to specifically account for refraction. The algorithm generates detailed maps while allowing for denoising, and is largely lighting invariant.

The map is represented by a global 3-D voxel grid which contains discretized truncated signed distance function

(TSDF) values [16]. The algorithm proceeds by iterating through four main stages:

1) **Measurement Pre-processing**: New measurements are used to generate dense vertex and normal maps.
2) **Pose Estimation**: The new position of the camera is estimated using an iterative closest point approach.
3) **Map Update**: The map TSDF values are updated with the current measurement information
4) **Surface Prediction**: A predicted surface is generated by projecting the map into the current camera estimate.

One of the vital aspects of the algorithm is the ability to project image pixels into and back from 3-D space. Under normal operation in open air this is a trivial task which is accomplished using the camera's intrinsic parameter matrix. However, when the camera is submerged underwater this task is no longer trivial and requires taking into account the effect of the refractive interface.

The measurement pre-processing stage incorporates refraction corrections in order to correctly project depth image pixels into 3-D space. A new measurement consists of a depth map, D_t, where each image pixel, u, is a raw depth measurement, $D_t(u) \in \mathbb{R}$. A pixel in the vertex map can then be defined from the corresponding pixel location and the depth value such that the vertex point in the 3-D camera frame, $p_c \in \mathbb{R}^3$ is defined as

$$p_c = D_t(u)\Psi(u) + \gamma(u) \tag{9}$$

for all u, in the depth image. This generates the corrected vertex map, $V_{c,t}$, in the camera frame, c such that $V_{c,t}(u) \in \mathbb{R}^3$. The corresponding normal map, $N_{c,t}$ can be computed by calculating the cross product of a vertex point and its neighboring pixels.

The pose estimation process is performed by employing point to plane iterative closest point (ICP) optimization [17] and the fast projective data association algorithm for correspondences. We denote the camera pose in the global frame, g, at time t by the transformation matrix, $T_{g,t}$. Vertex correspondences are computed by calculating the predicted pixel location

$$\hat{u} = \Phi(\hat{T}^k_{g,t-1} V_{c,t}) \tag{10}$$

where $\hat{T}^k_{g,t}$ is the current transform estimate at ICP iteration k. The measured and predicted vertex maps must contain a valid value at u, and correspondences are thresholded based on surface normals and point to plane distances. Invalid correspondences are not included in the map update stage to remove outliers from distorting the map. In this manner the alignment of the measured surface ($V_{c,t}$,$N_{c,t}$) and the predicted surface ($\hat{V}_{g,t}$,$\hat{N}_{g,t}$) is performed and generates the new camera pose $T_{g,t}$.

The map update stage takes the aligned measurement information and fuses it into the global map using the TSDF values. The TSDF is implemented as a discrete voxel grid where each cell has both the TSDF value, $F \in \mathbb{R}$, and a confidence weighting $W \in \mathbb{R}_+$. The new TSDF value of a cell located at point, $p_g \in \mathbb{R}^3$, is calculated by projecting p_g to a specific camera pixel, $\hat{u}$, and calculating the difference between the measured depth, $D_t(u)$, and the distance to p_g. The depth difference $\delta \in \mathbb{R}$ is

$$\delta = ||T^{-1}_{g,t} p_g||_2 - D_t(\Phi(T^{-1}_{g,t} p_g)) \tag{11}$$

The difference value, δ, is then truncated and normalized to get the measured value of F_m. The new value of the TSDF, F_t is the weighted average between the measured, F_m, and current value, F_{t-1}.

The surface prediction stage generates a predicted vertex map, $\hat{V}_{g,t+1}$, and surface normal map, $\hat{N}_{g,t+1}$, for use in the pose estimation stage of the next iteration of the algorithm. The predicted maps are generated by ray tracing from the current estimate of the camera position into the map. The camera rays used for ray tracing are calculated using the refraction equations seen previously. By accounting for refraction in the surface prediction stage the predicted surface more accurately represents what the raw camera measurements produce. As such, the pose estimation stage will provide more accurate matching results which will improve the resulting map.

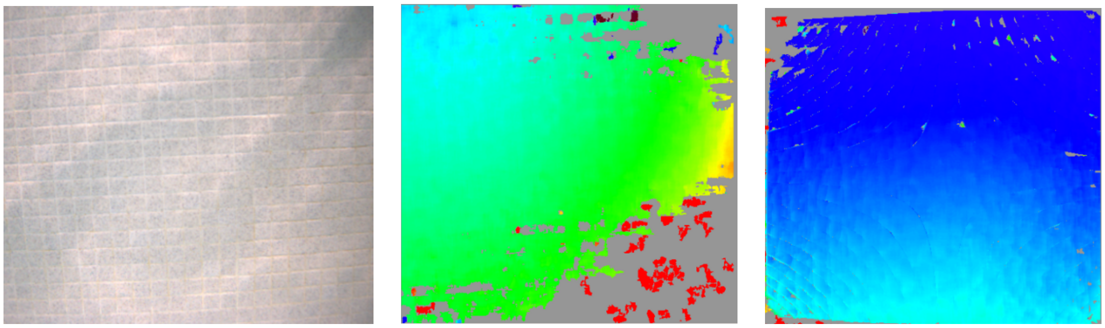
The Kinect fusion algorithm leverages GPU parallelization to vastly increase performance and allow for high frame rates and real time operation. Our modified approach still maintains the same ability to be parallelized but increases computation by a non-trivial constant factor due to the increased number of operations needed to account for refraction. However, since the bottle neck for this type of GPU programming has generally been memory and not computation the modified algorithm is still able to be run at near real time speeds on current graphics processors.

VII. Experimental Results

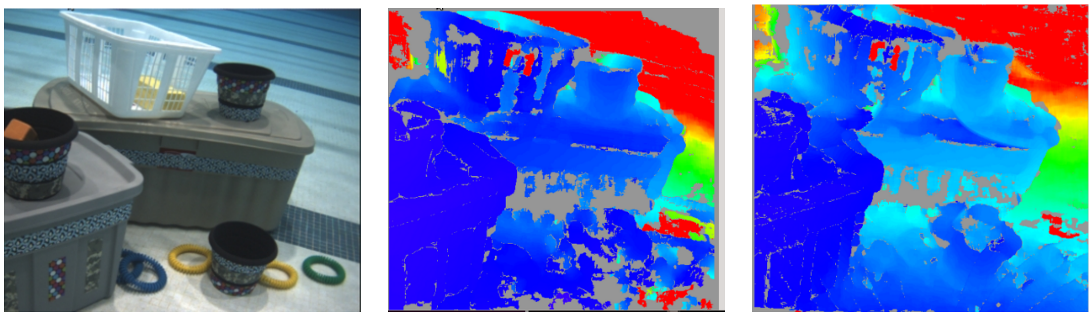
The proposed method was evaluated using a stereo camera setup which included a Point Grey Bumblebee2 stereo camera mounted in a custom waterproof enclosure. The data was collected at the University of Waterloo pool with the addition of artificial submerged objects of various sizes and shapes to generate static environment. The camera was moved manually by an operator through the environment for the various test cases. Two main results of system are presented. First, the improvement in the raw stereo depth map generation is demonstrated and evaluated based on pixel correspondences and observed curvature on a flat surface. Second, the results of the overall mapping are presented and demonstrate the improvements to the global localization and mapping accuracy over time. The results in both cases are compared against radial correction calibrated at a depth of 2.75m.

A. *Stereo Depth Image Correction*

The quality of the final map is directly correlated to the quality of the depth maps that can be produced by the stereo camera setup. In the case where the refraction distortion is not corrected, the depth maps can be degraded. In Figure 3 two representative depth maps are presented. The top row of images presents depth maps with and without refractive correction for a flat wall parallel to the image plane. The bottom row presents depth maps for a varied scene containing



(a) Raw image of flat textured surface 0.7m away from the camera (b) Radial correction on a flat textured surface 0.7m away from the camera (c) Full refraction correction on a flat textured surface 0.7m away from the camera



(d) Raw image of example scene (e) Radial correction on an example scene (f) Full correction on example scene

Fig. 3. Comparison of standard radial correction versus our full refraction correction technique on depth map quality and density. False coloring is applied to the images based on depth values. Grey represents a point which failed to be corresponded.

multiple non-planar objects such as pots and containers. The main characteristics of interest when evaluating the depth maps are that of erroneous curvature of flat surfaces as well as correspondence quality of the stereo matching.

Figure 3 shows that in the case of the flat surface there are still significant errors present when using radial correction causing a false impression of curvature on the surface. The refraction corrected image does not show the false curvature and correctly calculates a constant gradient across the whole image. The image shows a slight vertical gradient because the test rig was not aligned perfectly parallel to the surface. The images also show that the number of incorrect correspondences is significantly higher in the radial correction case, particularly in the corners of the images. The image corners are the most affected by the effects of refraction and as such the pixel correspondences fail most often in these regions. The depths of the example scene are better in the case of the refractive correction. Similar to the flat surface case, areas in the corners of the image tend to have significant degradation in the radial correction case as do areas which vary significantly in depth. Overall the refraction corrected depth map of the example scene shows smoother, more dense, and more accurate measurements.

The curvature of the flat wall in each of the images can be calculated quantitatively using principle curvature estimation. Principle curvature is calculated at a point by performing principle component analysis on the surface normals in a neighborhood in the tangent plane of the given point.

Correspondence can also be quantitatively accessed by taking the percent of the image which has a valid correspondence. In this case we will assume that any correspondence which was not filtered is valid and will be included in the total. Table I presents the curvature and correspondence results, which demonstrate that the refraction corrected depth maps are notably improved over the radial correction case.

TABLE I

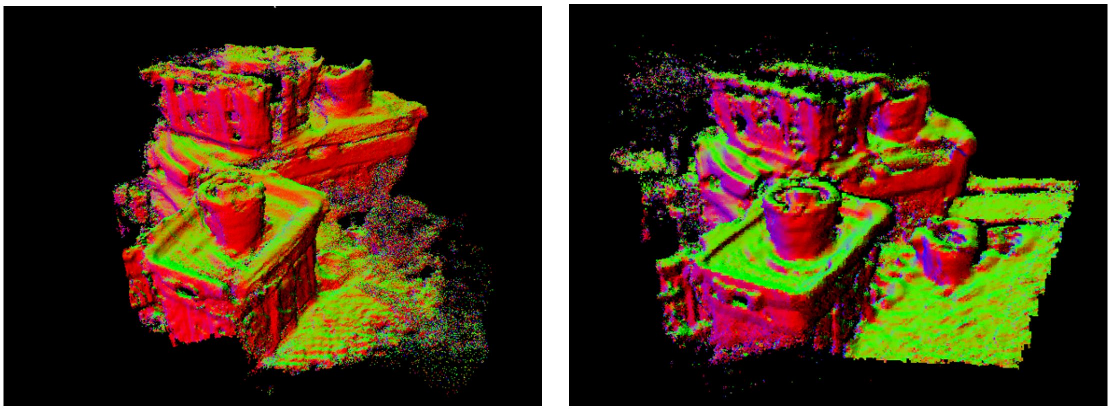
SUMMARY CURVATURE AND CORRESPONDENCE PERCENTAGES FOR THE RADIAL AND FULL CORRECTION CASES ON THE FLAT SURFACE.

	No Correction	Radial Correction	Refractive Correction
Principal Curvature	0.0540	0.0114	0.0045
% Correspondence	65.3	75.6	91.2

B. Map Reconstruction Quality

To evaluate the overall map reconstruction quality the presented method was performed on an underwater dataset in which an artificial scene of interest was constructed out of plastic containers of various sizes and dive weights. The main characteristics of interest when evaluating the map are that of erroneous curvature of flat surfaces as well as general geometric consistency in size and shape of objects. Figure 4 shows the comparison of the map generated with radial correction versus the map generated using our full refraction correction method.

The refraction corrected map has smoother features and flat surfaces show less curvature. This is due to the fact that consecutive frames have much better global consistency and thus align more accurately to the global map. It should also be noted that the radially corrected version has significantly worse convergence properties since the ICP has



(a) With radial correction

(b) With refraction correction

Fig. 4. Comparison of SLAM performance in mapping an underwater scene using both a)radial correction and b) full refraction correction

worse correspondence information to use when aligning the frames. The global consistency and ICP alignment of the map can be demonstrated quantitatively by comparing the average error in the ICP minimization and the percentage of valid correspondences. This information shows how closely the depth image matches to the global map. If a given depth image matches the global map with minimal error and has a high number of correct correspondences the map can be considered to be an accurate representation of the measured scene. Table II summarizes the results of the ICP alignment.

TABLE II

SUMMARY OF ICP ERROR AND CORRESPONDENCE REJECTION.

	No Correction	Radial Correction	Refraction Correction
Mean ICP error	0.6912	0.3988	0.2644
% Correspondence	93.77	97.25	97.73

The ICP matching results on the refraction corrected map are significantly better, showing a 33.7% reduction in error over radial correction, indicating that the generated reconstruction is a better match to the actual measured scene.

VIII. CONCLUSION

This work demonstrates that stereo SLAM results for underwater applications can be improved by accounting for refraction in the stereo matching and SLAM algorithms. A method for refraction compensated SLAM is presented and shown to improve both depth map quality from stereo camera sensors as well as the overall resulting 3-D reconstruction. Future work includes optimization of implementation for real time performance as well as outdoor field trials under differing environmental conditions.

REFERENCES

[1] M. Warren, P. Corke, O. Pizzaro, S. Williams, and B. Upcroft, "Visual sea-floor mapping from low overlap imagery using bi-objective bundle adjustment and constrained motion," in *Australasian Conference on Robotics and Automation*, Brisbane, Australia, December 2012, pp. 3–5.

[2] A. Hogue, A. German, and M. Jenkin, "Underwater environment reconstruction using stereo and inertial data," in *International Conference on Systems, Man and Cybernetics (ISIC)*, Montreal, Canada, 2007, pp. 2372–2377.

[3] C. Clark, C. Olstad, K. Buhagiar, and T. Gambin, "Archaeology via underwater robots: Mapping and localization within maltese cistern systems," in *International Conference on Control, Automation, Robotics and Vision (ICARCV)*, Hanoi, Vietnam, December 2008, pp. 662–667.

[4] D. R. Yoerger, M. B. Albert, B. W. Barrie, H. Singh, and R. Bachmayer, "Surveying a subsea lava flow using the autonomous benthic explorer (ABE)," *International Journal of Systems Science*, vol. 29, no. 10, pp. 1031–1044, 1998.

[5] M. Zhou and R. Bachmayer, "Working towards single beam acoustic iceberg profiling using active roll control on a slocum glider," in *IEEE Symposium on Underwater Technology (UT)*, Tokyo, Japan, 2011, pp. 1–7.

[6] H. Bagheri, A. Vardy, and R. Bachmayer, "Strategies for filtering incorrect matches in seabed mosaicking," in *OCEANS 2011*, Santander, Spain, September, pp. 1–5.

[7] A. Elibol, N. Gracias, and R. Garcia, "Augmented state–extended kalman filter combined framework for topology estimation in large-area underwater mapping," *Journal of Field Robotics*, vol. 27, no. 5, pp. 656–674, 2010.

[8] T. Treibitz, Y. Schechner, C. Kunz, and H. Singh, "Flat refractive geometry," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 51–65, January 2008.

[9] Y. Kwon, "A camera calibration algorithm for the underwater motion analysis," in *International Symposium on Biomechanics in Sports (ISBS)*, vol. 1, no. 1, Perth, Australia, 1999, pp. 257–260.

[10] J. Gedge, M. Gong, and Y.-H. Yang, "Refractive epipolar geometry for underwater stereo matching," in *Canadian Conference on Computer and Robot Vision (CRV)*, St. Johns, Canada, May 2011, pp. 146–152.

[11] A. Yamashita, R. Kawanishi, T. Koketsu, T. Kaneko, and H. Asama, "Underwater sensing with omni-directional stereo camera," in *International Conference on Computer Vision Workshops (ICCV Workshops)*, November 2011, pp. 304–311.

[12] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard, "An evaluation of the RGB-D SLAM system," in *International Conference on Robotics and Automation (ICRA)*, St. Paul, USA, 2012, pp. 1691–1696.

[13] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *International Symposium on Mixed and Augmented Reality (ISMAR)*, Basel, Switzerland, 2011, pp. 127–136.

[14] R. Li, H. Li, W. Zou, R. Smith, and T. Curran, "Quantitative photogrammetric analysis of digital underwater video imagery," *Journal of Oceanic Engineering*, vol. 22, no. 2, pp. 364–375, April 1997.

[15] M. J. Atallah, "Faster image template matching in the sum of the absolute value of differences measure," *Image Processing, IEEE Transactions on*, vol. 10, no. 4, pp. 659–663, April 2001.

[16] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Conference on Computer graphics and Interactive Techniques*, 1996, pp. 303–312.

[17] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," in *International Conference on Robotics and Automation (ICRA)*, vol. 3, Sacramento, USA, April 1991, pp. 2724–2729.