

Low-Constraint Ray Optimization for Photogrammetric Shallow Water Bathymetry

Christoph Liebender, Pascal Meyer, Michael Bleier, Andreas Nüchter

Informatics XVII - Robotics, Julius Maximilian University of Würzburg, Germany
{christoph.liebender,michael.bleier,andreas.nuechter}@uni-wuerzburg.de
pascal.meyer1@stud-mail.uni-wuerzburg.de

Keywords: Raytracing, Snell's law, Optimization, Multi-View Stereo, Watersurface estimation.

Abstract

This paper presents a raytracing-based approach for estimating the depth of shallow water bodies. Based on a calibrated, multi-view camera setup above the water surface, we find correspondences that lie below the surface in pairs of two views. We trace rays from each camera to each triangulated correspondence, as if no interface is present in the scene. Using Snell's law, we then compute refracted rays to optimize an objective function with a theoretical optimum as the Hesse normal form of the water surface. The challenge of our work is to keep the number of constraints in our mathematical model as low as possible. For a practical evaluation, we record images with two different setups: a laboratory setting with a box filled with water at a depth of 14.5 cm and a water roll-off container at 2.2 m, and use both fiducial markers as well as computed features as correspondences. Our results show that with a well-setup optimizer, our method converges towards the true water plane and is usable to estimate the depth of our setups.

1. Introduction

PHOTOGRAMMETRY through a water surface is a challenging problem. Here, calibrated cameras are placed above the water surface, with the object to be measured being located below. For above water stereo vision as one form of above water photogrammetry, the epipolar constraint enforced by known extrinsics is used for 3D reconstruction. It states that any single point in one camera's view maps to a line in another camera's view. However, this does not hold as soon as an interface between two (or multiple) media is placed in the light's path; refraction according to Snell's law makes the initial constraint unapplicable. To this end, different methods for 3D reconstruction are required.

It is possible to consider naive triangulations of found correspondences, as if no air-water interface is in the scene. For instance, placing fiducial markers below the water surface, for them to be detected in each camera's image is an option. Using these unique correspondences, triangulation from stereo-geometry gives a 3D point. However, this introduces errors in the computed points w.r.t. the true 3D points, depending on the marker-interface and interface-camera distances as well as their orientation. Once a flat calibration pattern is placed below the water surface this effect is especially noticeable: in that case, its triangulated points appear curved.

This issue motivates modeling and estimating the air-water interface, as errors are correctable once the pose and shape of the air-water interface is known. Typically, many constraints are imposed on the modeling problem, even more so when the water surface is distorted by waves as the latter quickly becomes an ill-posed problem. Current research does not make it clear whether or not a generalized, low-constraint approach to the mathematical modeling problem of multi-view, multi-media photogrammetry of unknown underwater scenes exists that successfully estimates water surfaces and therefore would be usable to correct refraction errors in 3D measurements.

Our approach explores the implications of reducing the number of constraints on the modeling problem, using an objective

function with a novel criterion. The main contributions of this paper are: 1. an objective function with a novel refractive-plane and backrefractive based disparity criterion. 2. Its evaluation using four calibrated cameras applied in both: 3. a laboratory setting of a filled water tank, with both fiducials and detected features, and 4. an outdoor setting with a filled roll-off container (van der Lucht et al., 2019) with CCTag fiducials (Calvet et al., 2016). We additionally observe how CCTag fiducials are robust against distortions caused by waves.

2. Related Work

Previous work on defining refractive objective functions imposes different types of constraints on the modeling problem. Maas (1995) already showed how strict geometric modeling of a windowed underwater camera enclosure w.r.t. the camera itself and the scene makes it possible to correct 3D points. He requires the camera to be at a fixed, known distance from a glass window of known thickness that separates air from water of known depth. In another work of Ferreira et al. (2005), the authors show how highly constrained measurement setups allow to treat multi-media underwater reconstruction as a conventional stereo problem that utilizes the epipolar constraint. For this, they require the camera orientations to be in a certain range, in order to keep the incident angle of light small.

Morris and Kutulakos (2005) base their model on the fact that normals at a media intersection need to be consistent across two views with each different rays from that intersection – this normal consistency constraint has since seen multiple applications (Qian et al., 2018, 2017; Ding et al., 2011). To model an arbitrarily shaped refractive surface, Morris and Kutulakos position two cameras in a stereo setup with known position w.r.t. a reference pattern. They show that in their case, two views are sufficient to reconstruct the shape of the surface, and coin the term 'refractive disparity', defined as the distance between points on that reference pattern that are projected onto the image, to provide a more robust approach for comparing refractive normals. We also define a disparity between points that are the

intersections of rays and planes. However, we do not strictly require any known reference plane on the bottom of the underwater scene, only correct correspondences are a requirement. We compute planes from the cross-product of each uncorrected, direct ray with its refracted ray at the intersection with an interface estimate.

More recently, multi-media photogrammetry has seen an increase of interest in the research community. Zhang et al. (2014) use distortion and defocus of a video sequence to recover the immersed surface, starting from a reference image that features a flat water surface. Qian et al. (2018) deploy a multi-camera system with nine adjacent views from a near-nadir perspective. By utilizing single-shot correspondences of optical flow, they numerically optimize an objective function to recover both the water surface and shape of a visually rich underwater scene. Instead of requiring nine views, Xiong and Heidrich (2021) recover both water and underwater surfaces by taking a video sequence of a static underwater scene with a single static camera. They then treat the different media intersections of the same subsurface point at different points in time as distinct views of the subsurface point. Most recently, deep learning (Thapa et al., 2020) and neural radiance fields (Wang et al., 2023; Zhan et al., 2023) have also been applied to water and underwater surface reconstruction. There is also focus on the analysis of water effects to reduce the complexity down to a planar surface (Mulrow et al., 2024; Sardemann et al., 2024), while Rofallski et al. (2024) present an application of flat water plane estimation and refraction correction with waterlogged wood conservation. Our approach is more closely related to photogrammetry and bathymetry, since initially, our goal is to estimate the water surface, and in case of a wavy surface, its average surface. As a result, we receive the (average) water height computed from corrections based on naively triangulated points.

3. Methodology

We now introduce our raytracing model and error criterion, shown in Figure 1. Afterwards, we present our objective function that is then applied to pair-of-two combinations from multiple, calibrated cameras.

3.1 Raytracing Model

Starting with sparse correspondences of underwater points that are visible in at least two views, we triangulate a set of warped points $\{\tilde{X}^i\}$ from a combination of two views. Since the views' intrinsics and extrinsics are calibrated, tracing rays v_1^i from each camera's center O_1 to each warped point \tilde{X}^i is possible. With an initial guess of the water surface S^* , we calculate the intersection $I_1^i = v_1^i \wedge S^*$. Given the surface normal n of S^* at I_1^i , we get a refracted ray r_1^i from v_1^i (both without sub- and superscripts) with Snell's law :

$$r = \rho v - n \left[\rho n \cdot v + \sqrt{1 - \rho^2 (1 - (n \cdot v)^2)} \right], \quad (1)$$

with fixed $\rho = 1.33$ as the refractive index. This refracted ray does not intercept \tilde{X}^i – with the exception of $n \parallel v_1^i$. Both $v_1^i \times r_1^i$ and I_1^i now span the *plane of refraction* P_1^i . This is repeated for a second camera O_2 , such that we get r_2^i and P_2^i . For the true water surface S , $r_{1,2}^i$ need to intersect in the true point X^i . For any other estimate S^* , this needn't be the case.

3.1.1 Refractive-Plane Residual Given these two refracted rays $r_{1,2}^i$, one approach is to compute the shortest distance between them and define this as an error for optimization. However, this results in a similar problem as it was previously described by van der Lucht et al. (2018). There, the authors intersect camera and laser rays that are refracted at a known flat water surface. They note that calculations solely on rays do not provide a robust solution, as measurements as well as calibration parameters are prone to errors that get amplified when projected onto rays. Instead, they propose to intersect rays and refractive planes. We now apply this principle to our problem of intersecting two camera rays and define our *refractive-plane* residual as:

$$e^i = \underbrace{r_1^i \wedge P_2^i}_{J_{1 \rightarrow 2}^i} - \underbrace{r_2^i \wedge P_1^i}_{J_{2 \rightarrow 1}^i}, \quad (2)$$

where $\{J^i\}$ define the intersection of the refracted ray from one camera with the plane of refraction of another camera's ray. Thus, for $S^* \approx S$, we get $J_{1 \rightarrow 2}^i \approx J_{2 \rightarrow 1}^i \approx X^i$, and explicitly $e_i \approx 0$.

3.1.2 Backrefractive Residual Optimizing e_i for $\{\tilde{X}^i\}$ will unsuitably lead to the trivial solution where $J_{1 \rightarrow 2}^i$ and $J_{2 \rightarrow 1}^i$ meet in \tilde{X}^i . To this end, we need to define a second category of residuals that counterweights this trivial solution. We choose to trace $J_{1 \rightarrow 2}^i$ back through the opposite camera's intersection I_2^i , calling it $b_{2 \rightarrow 1}^i$. Refracting this ray at I_2^i gives us a ray $\tilde{b}_{1 \rightarrow 2}^i$ that intersects the origin of camera O_2 for $S^* \approx S$. This is again repeated for $J_{2 \rightarrow 1}^i$, enabling us to define the *backrefractive* residual:

$$e_2^i = O_2 - I_2^i - \tilde{b}_{1 \rightarrow 2}^i \left[(O_2 - I_2^i) \cdot \tilde{b}_{1 \rightarrow 2}^i \right] \quad (3)$$

as the orthogonal distance e_2^i from $\tilde{b}_{1 \rightarrow 2}^i$ to O_2 . (e_1^i analogous.)

3.2 Objective Function

We implement our optimization routine using the Ceres framework of Agarwal et al. (2023). Thus, the objective function for a single stereo pair of our problem is defined as:

$$f(\{\tilde{X}\}) = \sum_i \left[\lambda_{\text{rp}} \left(\|e^i\|^2 \right) + \sum_{c \in \{1,2\}} \lambda_{\text{br}} \left(\|e_c^i\|^2 \right) \right] \quad (4)$$

with $\lambda_{\text{rp,br}}$ as refractive-plane and backrefractive loss functions. Since we use more than two cameras in our experimental setups, we optimize $\binom{4}{2}$ camera combinations as one bundle. In practice, we set:

$$\lambda_{\text{rp}}(x) = x \quad (5)$$

$$\lambda_{\text{br}}(x) = 2(\sqrt{1+x} - 1) \quad (6)$$

To both trivial and pseudo-huber losses respectively. λ_{br} is chosen for optimizer stability since otherwise, backrefractive errors disproportionately increase with larger interface-camera distances. Ceres further defines a scale factor that can be used to change the robustification behaviour. Our evaluation shows that different scale factors α_{br} are required for distinct measurement setups and scenes.

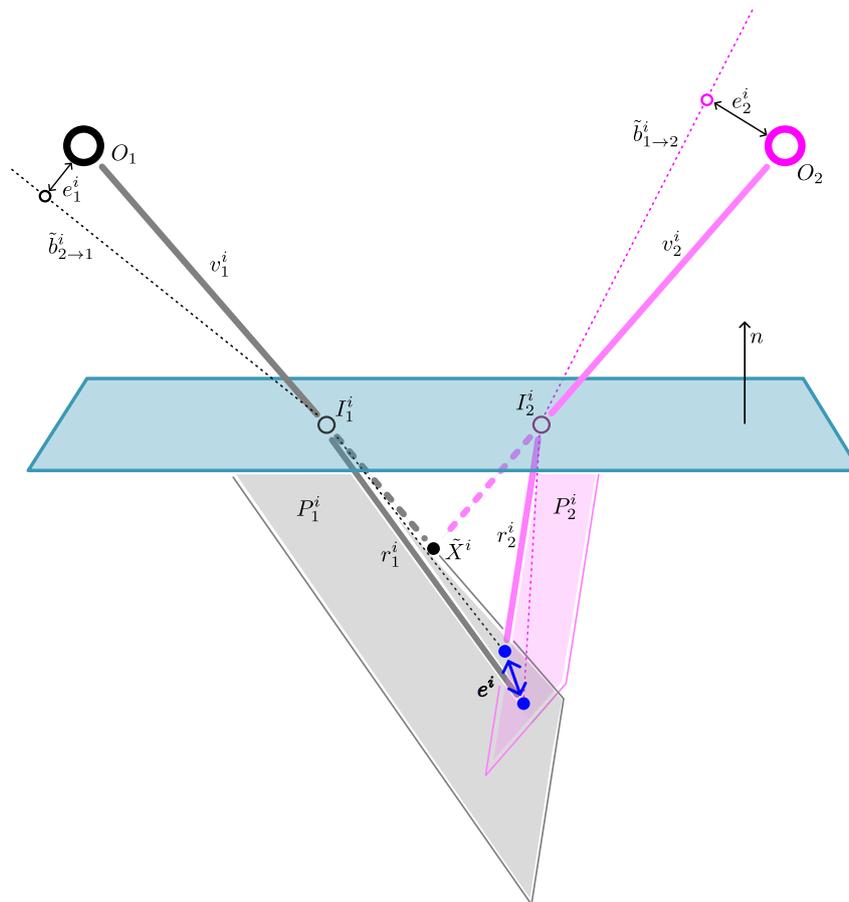


Figure 1. Geometrical refraction model for a single, naive subsurface triangulation. Initial rays $v_{1,2}^i$ originating from cameras $O_{1,2}$ meet at the naive triangulation \tilde{X}^i . An estimated water plane refracts $v_{1,2}^i$ at intersections $I_{1,2}^i$. The plane spanned by v_1^i and r_1^i is intersected by r_2^i (vice versa). The distance between these intersections is our refractive-plane disparity. Finally, tracing these intersections back through the opposite interface intersections should intersect the opposite camera. The distance between these back-refractions and the respective camera origin coins the second backrefractive disparity.

4. Evaluation

Our evaluation consists of two measurement setups: 1. a laboratory setting with an euro container filled with 14.5 cm of a strictly flat water surface, and 2. a 40 000 l Roll-on/Roll-of container (van der Lucht et al., 2019) with both a nearly flat water surface of approximately 2.2 m and a wavy water surface, introduced by a wave generator.

4.1 Laboratory Setting

Figure 2 shows our laboratory setup. To obtain both intrinsic and extrinsic parameters per camera, we use the method of Zhang (2000) and 150 images of a 29×29 ChArUco pattern with a square length of 2.5 cm. We set one camera as the reference camera and transform each other camera’s observations into this reference coordinate frame.

4.1.1 Apriltags First, we run the detector from the apriltag library (Olson, 2011) with an immersed planar pattern to obtain four accurate, corner-refined correspondences per marker. Since our optimization routine is an iterative approach, we need to set an initial guess for the Hesse normal form of the interface. In this case, we compute it from a PCA-decomposition of the naive triangulations, of which the normal vector with smallest eigenvalue is a suitable candidate for a plane normal. Then, we

choose the mean of the naive triangulations plus a scaled normal vector as the support point for the initial guess. Finally, we scale this point such that the initial guess is at the midpoint between naive triangulations and our cameras. This is applicable because the distance from the cameras to the true water surface is larger than the surface-marker distance.

4.1.2 Features Placing fiducial markers in the submerged scene may not always be feasible. If the visual structure of the scene allows for it, detected feature keypoints may be used instead. However, this requires that the computed feature points are consistent between views. For an initial evaluation of the robustness of our approach with less robust correspondences, we use SIFT features (Lowe, 2004) detected in the images of Figure 2.

4.2 Water Container

Figure 3 depicts our camera setup mounted on a crane above the water container. The height of the camera setup above the water surface is approximately 60 cm. Because we lack large enough calibration patterns to facilitate the intrinsic calibration at the measurement distance of almost 3 m, we take the intrinsic calibration from the laboratory setup and only re-calibrate the extrinsics. This is required as we change the physical arrangement of the cameras into a square.

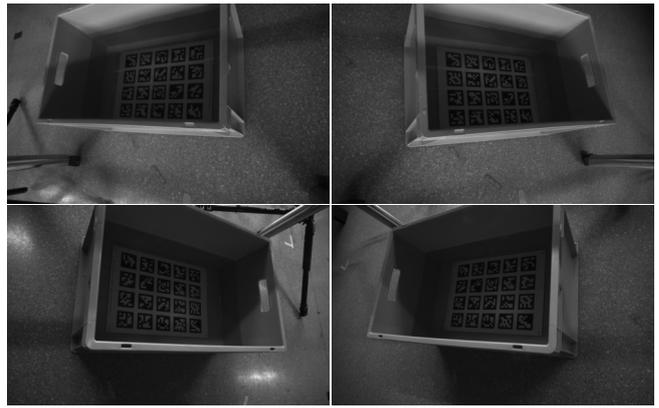
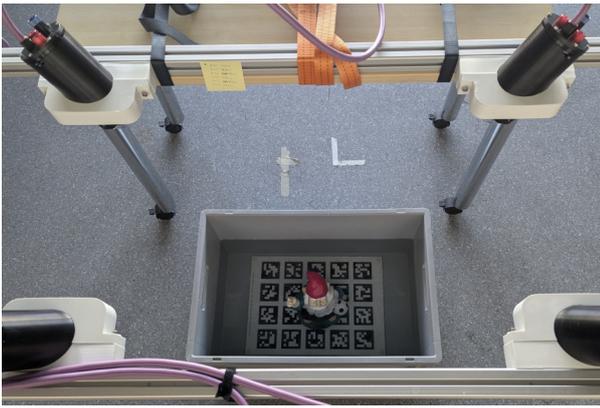


Figure 2. Experimental laboratory setup. Left: Overview of the setup. The cameras are positioned in a rectangle, viewing the scene at oblique angles, at a height of 70 cm above the floor. Right: Individual views of the cameras.



Figure 3. Outdoor setup. Cameras mounted in a square are hung on a crane, looking down towards the water tank.

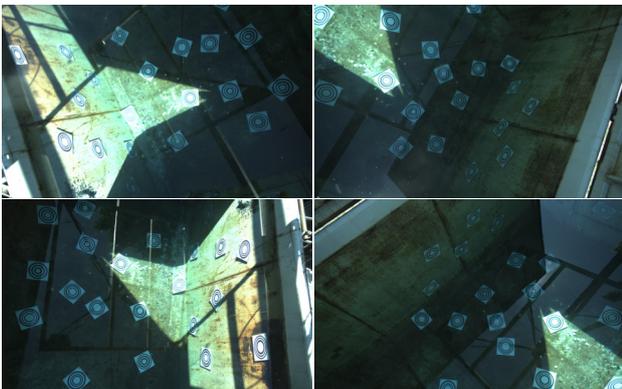


Figure 4. Views of the tank with a nearly flat water surface.



Figure 5. View of one camera with surface distorted by waves, with some fiducials having a distorted appearance.

The underwater surface of the container does not contain enough unique visual structure to simply rely on feature matching. Thus, we add fiducials to the walls and floor of the container. However, the dimensions of the container makes the use of the marker board of Figure 2 unsuitable. To this end, we use CCTag markers (Calvet et al., 2016) printed on aluminium Dibond plates, attached in a quasi-random way to the metal container using magnets. These coded fiducials are originally conceptualized to be detectable under highly challenging conditions. Figure 4 shows the cameras' views of the CCTags. We determine the initial guess similarly to our laboratory setup, but move the support point such that the initial guess lies outside of the water. Although this requires – strictly speaking – knowledge about the position of the true water plane, one could always assume that an initial plane placed closely in front of the camera setup needs to be outside of the water.

To explore our method's robustness against significant distortions of the water surface, we also run our optimizer on images that are distorted by waves. Figure 5 shows one exemplary image where almost all CCTag fiducials are distorted by waves. Still, the tags and their codes remain detectable, presumably because the visual distortion of light refraction by sinusoidal waves is similar enough to high amounts of motion blur – the original application of CCTags.

5. Results

This section presents the results we obtain with our method. To calculate the average water depth from the solution of the Ceres

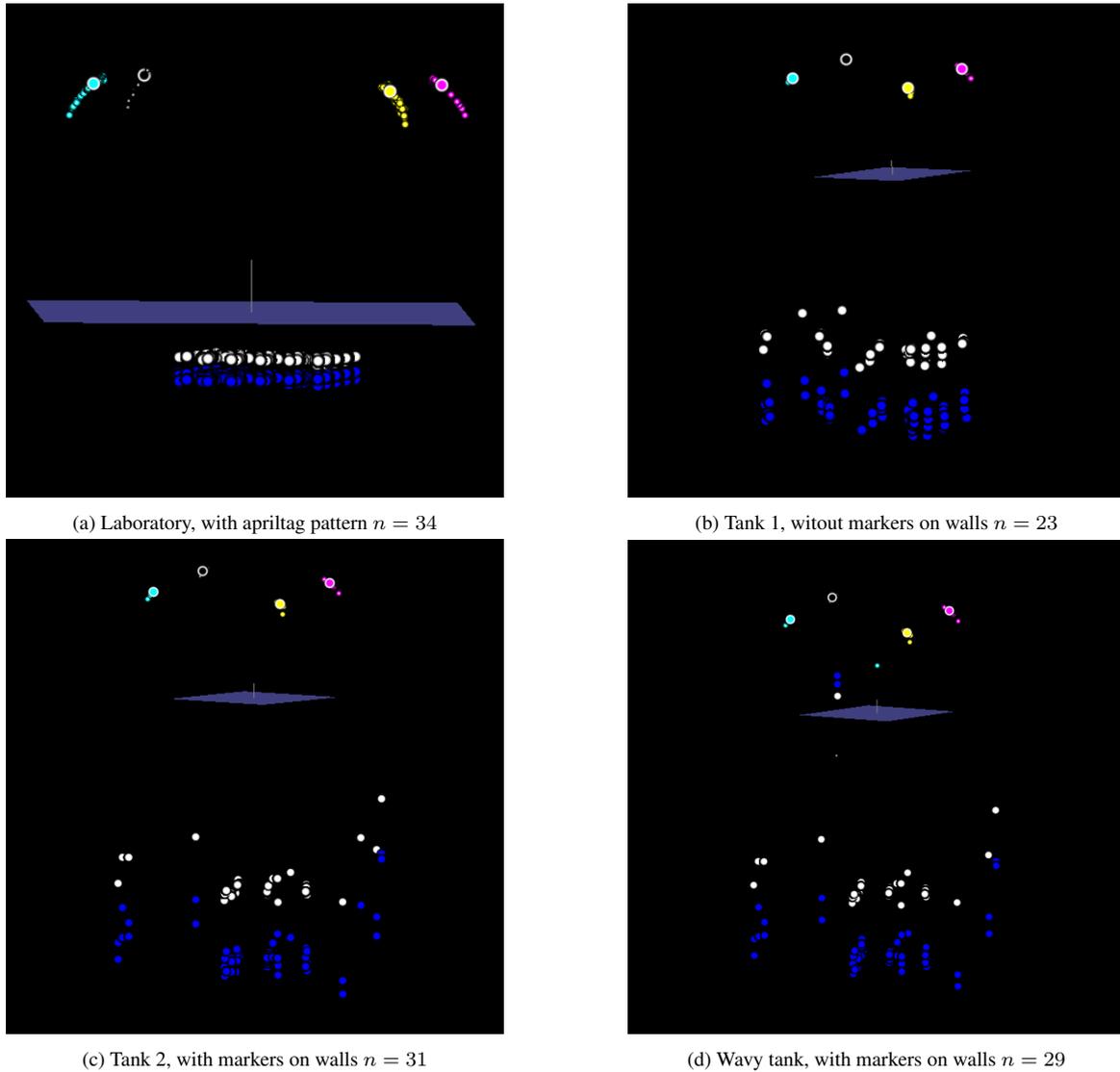


Figure 6. Visualization of converged solutions. Largest black, cyan, yellow and magenta points are camera origins. Scattered points around cameras are the closest distances of backrefractions. Estimated surface in blue, with normal as a white line. White points below the surface show the naive triangulations, while blue points are their corrections.

Setup	α_{br}	d_{PCA}	\bar{d}	$d_{(n+1)/2}$
Lab (April)	0.04	14.67 cm	14.67 cm	14.74 cm
Lab (SIFT)		14.30 cm	14.30 cm	14.29 cm
Tank 1	0.35	2.12 m	2.11 m	2.11 m
Tank 2		2.09 m	2.08 m	2.13 m
Wavy Tank		2.09 m	2.06 m	2.15 m

Table 1. Depth results of converged solutions.

solver, we have multiple options. In the case of the laboratory setup, we can simply decompose the set of corrected underwater points $\{J\}$ and take the distance d_{PCA} from the mean to the solved plane. This is feasible because of two reasons: 1. the low measurement error at a comparatively short measurement distance, and 2. the camera intrinsics being calibrated with images taken at that distance. Both of these reasons result in $J_{n \rightarrow n+1}^i \approx J_{n+1 \rightarrow n}^i \forall i, n$. Or in other words: $\{J\}$ being close to the true points of the apriltag pattern. In the case of the water container, both previous reasons do not strictly apply. Because of that, we compute the mean \bar{d} and median distance $d_{(n+1)/2}$ for each of the points J_i .

Table 1 shows the numeric results visualized in Figure 6. Generally, we find that the parameter α_{br} introduced in Section 3.2 needs to be adjusted on a per-scene basis. Otherwise, the solver either does not converge at all or towards a degenerate solution, caused by the nonlinearity of light refraction that introduces local minima.

In our experiments with the laboratory setup, the optimizer converges towards the true water plane. Measured from the corrected points, the depth evaluates to 14.6 cm with apriltag correspondences, which is within 1 mm of our initial estimate, and 14.3 cm when computed with SIFT correspondences. Since the water plane is close to the pattern in relation to the camera's distance from the surface, visible in Figure 6a, we need to set a low value for α_{br} . The cause for this lies in the way we define our backrefractive error, as the angle between the incident and backrefracted ray amplifies the orthogonal distance of the backrefractive ray to the camera origin. We have to increase this parameter in our watertank setup because of that reason. However, we can keep this value constant, not depending on whether or not we include fiducials on the tank's walls (Figure 6c) or only

have them on its floor (Figure 6b). We also come to the conclusion that our optimizer converges when the image data is disturbed by a wavy water surface. Figure 6c shows the convergence in this case, with one mis-identification where the naive triangulation is above the water surface. The proposed Ceres framework however successfully classifies this as an outlier.

6. Conclusion

This paper presents a mathematical algebra model for multimedia photogrammetry. Using the Ceres framework, we successfully apply nonlinear-least-squares optimization to the nonlinear problem of light refraction. Its goal was to analyze the behaviour of said approach when applied to real-world data. Although more sophisticated methods exist in the literature, our method keeps the number of constraints on the scene small and only requires adjusting one robustification parameter α_{br} in-between measurement scenes. When the ideal case of a flat water plane is present, this method is able to estimate the true water surface. With a wavy water surface, it is possible to determine its average height and the distorted point's average position. In theory, it should be possible to replace the water plane with a more complex surface model. However, this requires that there must be enough points available for the problem to not be underdetermined. Needless to say, a lot of work remains to be done. Future work will concentrate on statistical analysis of our method. For this, more experiments are required: different arrangements of fiducials, wave shapes and magnitudes, camera poses, loss functions and parameters, and water surface models.

Acknowledgement

This work was supported by the project 'Mixed Media Scanning' funded by the German Research Foundation (DFG NU230/14-1). We also acknowledge support from the Elite Network of Bavaria for the academic program Satellite Technology - Advanced Space Systems.

References

- Agarwal, S., Mierle, K., The Ceres Solver Team, 2023. Ceres Solver v2.2. <https://github.com/ceres-solver/ceres-solver>.
- Calvet, L., Gurdjos, P., Griwodz, C., Gasparini, S., 2016. Detection and Accurate Localization of Circular Fiducials under Highly Challenging Conditions. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE.
- Ding, Y., Li, F., Ji, Y., Yu, J., 2011. Dynamic Fluid Surface Acquisition Using a Camera Array. *2011 International Conference on Computer Vision*, 2478–2485.
- Ferreira, R., Costeira, J. P., Santos, J. A., 2005. *Stereo Reconstruction of a Submerged Scene*. Springer Berlin Heidelberg, 102–109.
- Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60, 91–110.
- Maas, H.-G., 1995. New developments in Multimedia Photogrammetry. *Optical 3D measurement techniques III*.
- Morris, N., Kutulakos, K., 2005. Dynamic Refraction Stereo. *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, IEEE, 1573–1580 Vol. 2.
- Mulsow, C., Sardemann, H., Gueguen, L.-A., Mandelburger, G., Maas, H.-G., 2024. Concepts for compensation of wave effects when measuring through water surfaces in photogrammetric applications. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2-2024, 289–295. <https://isprs-archives.copernicus.org/articles/XLVIII-2-2024/289/2024/>.
- Olson, E., 2011. AprilTag: A robust and flexible visual fiducial system. *2011 IEEE International Conference on Robotics and Automation*, 3400–3407.
- Qian, Y., Gong, M., Yang, Y.-H., 2017. Stereo-Based 3D Reconstruction of Dynamic Fluid Surfaces by Global Optimization. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6650–6659.
- Qian, Y., Zheng, Y., Gong, M., Yang, Y.-H., 2018. *Simultaneous 3D Reconstruction for Water Surface and Underwater Scene*. Springer International Publishing, 776–792.
- Rofalski, R., Colson, A., Luhmann, T., 2024. Multimedia Photogrammetry for Automated 3D Monitoring in Archaeological Waterlogged Wood Conservation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2-2024, 355–362. <https://isprs-archives.copernicus.org/articles/XLVIII-2-2024/355/2024/>.
- Sardemann, H., Mulsow, C., Gueguen, L.-A., Mandlburger, G., Maas, H.-G., 2024. Multimedia Photogrammetry with non-planar Water Surfaces – Accuracy Analysis on Simulation Basis. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2-2024, 363–369. <https://isprs-archives.copernicus.org/articles/XLVIII-2-2024/363/2024/>.
- Thapa, S., Li, N., Ye, J., 2020. Dynamic Fluid Surface Reconstruction Using Deep Neural Network. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 21–30.
- van der Lucht, J., Bleier, M., Leutert, F., Schilling, K., Nüchter, A., 2018. Structured-Light based 3D Laser Scanning of Semi-Submerged Structures. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 287–294.
- van der Lucht, J., Bleier, M., Nüchter, A., 2019. A LOW COST UNDERWATER TEST ENVIRONMENT. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W17, 399–404. <https://isprs-archives.copernicus.org/articles/XLII-2-W17/399/2019/>.
- Wang, Z., Yang, W., Cao, J., Hu, Q., Xu, L., Yu, J., Yu, J., 2023. NeReF: Neural Refractive Field for Fluid Surface Reconstruction and Rendering. *2023 IEEE International Conference on Computational Photography (ICCP)*, 1–11.
- Xiong, J., Heidrich, W., 2021. In-the-Wild Single Camera 3D Reconstruction Through Moving Water Surfaces. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, 12538–12547.
- Zhan, Y., Nobuhara, S., Nishino, K., Zheng, Y., 2023. NeRFrac: Neural Radiance Fields through Refractive Surface. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 18356–18366.
- Zhang, M., Lin, X., Gupta, M., Suo, J., Dai, Q., 2014. *Recovering Scene Geometry under Wavy Fluid via Distortion and Defocus Analysis*. Springer International Publishing, 234–250.
- Zhang, Z., 2000. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334.