The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-2/W10-2025
3D Underwater Mapping from Above and Below – 3rd International Workshop, 8–11 July 2025, TU Wien, Vienna, Austria

Automated marker detection and time series analysis for through-water photogrammetry

Michele Russo[1], Luca Martelli[2], Roberta Ravanelli[3], Agnese Pini[2]

[1] Dept. of History, Representation, and Restoration of Architecture, Sapienza University of Rome, Italy – m.russo@uniroma1.it

[2] Dept. of Civil, Constructional, and Environmental Engineering, Sapienza University of Rome, Italy – luca.martelli@uniroma1.it, agnese.pini@uniroma1.it

[3] Dept. of Geography, University of Liège, Belgium – roberta.ravanelli@uliege.be

Keywords: Multimedia Photogrammetry, Stereo-matching, Metrology, Encoded Targets, Time Series Analysis, UCH.

Abstract

Through-water photogrammetry for surveys of Underwater Cultural Heritage (UCH) is a problem addressed by scholars from multiple perspectives. In addition to light attenuation as a function of distance travelled by light radiation through the liquid and variation of water optical characteristics across short lenghts, undulation of the free surface and magnitude of refraction (which varies at every point of each acquired image) also come into account. Our research aims to implement an automatic analysis method to estimate a priori — and correct a posteriori — the camera behaviour, focussing on surveys of objects in shallow waters. We proposed and tested a mathematical model to represent the effects of refraction across different water levels and states of motion. In addition to acquiring image pairs with two consumer-grade cameras, we recorded short video sequences using a high-speed, high-sensitivity monocular camera to enable time-series analysis of the phenomenon.

1. Introduction

Shallow water contexts hinder the use of submerged instruments for surveying (we can take as an example an immovable cultural asset, attached to a shallow seabed and therefore immersed in a volume of water only a few tens of centimetres high). Using optical instruments out of water to detect geometric shapes of submerged or semisubmerged objects requires knowledge of: A) the behaviour of light rays outside and inside the liquid; B) how to correct data recorded as results of the capture of these same rays by the sensors (Agrafiotis et al., 2018). The aim of our investigation (1), which is focussed on Underwater Cultural Heritage (UCH), is to formulate a problem from this need and propose a solution for it, first by addressing some preliminary theoretical and application issues that came to light from experiments in a controlled environment. This is in order to develop and test, at a later stage, data acquisition and processing workflows suitable for *through-water photogrammetry* operations under uncontrolled conditions and performed with cameras placed outside water, such as those mounted on Unmanned Aerial Vehicles (UAVs). Succinctly, the objectives of our study are: I) to analyse by means of encoded target detection the physical behaviour of light rays, passing through two media (air and water) and interacting with one or more digital cameras placed outside the liquid, to which different conditions of motion or free surface heights are imposed; II) to formulate, from the observations made, a mathematical model that allows the prediction of optical distortions in the captured images and can be used to obtain corrective parameters; III) to apply the aforementioned mathematical model and corrective parameters during processing of data acquired with photogrammetric surveys of cultural assets submerged at shallow depths or semisubmerged (Russo et al., 2024b). In this paper we further investigate the formulation and validation of our mathematical model presented earlier (Russo et al., 2024a). We considered three case studies: for the first two, we captured images with two consumer-grade cameras in stereoscopic mode, with a resolution of 350 Dot Per Inch (DPI); for the last one, we took video footage with a high-sensitivity, high-speed monoscopic camera capable of recording up to 500 frames per second (FPS), setting up time series analyses.

(1): The authors participated equally in the experimental phase. In writing this paper, M. R. was responsible for the Introduction and Conclusions, A. P. for Section 2 (State of the Art), R. R. for Section 3 (Methodology) and L. M. for Section 4 (Experiments and Results).

2. State of the Art

In photogrammetry, the computation of spatial geometric properties of a scene relies on images acquired through perspective projections. Consequently, the photogrammetric process requires the reconstruction in 3D space of the bundles of light rays that formed these images. For each camera, knowledge or simultaneous calibration of its intrinsic orientation (IO) and extrinsic orientation (EO) parameters is required to solve the collinearity equations (Luhmann et al., 2019). For both *in-water photogrammetry* and *through-water photogrammetry*, we must bear in mind that the liquid body itself has a 'multimedia behaviour', as its optical characteristics vary as a function of density, pressure, temperature, and salinity. Furthermore, depending on the distance light radiation travels from the source through water, absorption and scattering exponentially reduce its intensity (Mardani Nejad et al., 2021). In the case of *through-water photogrammetry*, additional factors come into play: I) free surface undulation and refraction magnitude — which varies at every point of the acquired image — lead to mathematically unstable solutions of the photogrammetric problem; II) according to Snell's Law, the effect of refraction given by a light ray is a function of the depth it reaches and the angle of incidence α of the ray with respect to the air/water interface. In a situation of shallow waters, a standard camera calibration procedure with a planar target is not sufficient to model refraction effects in the captured images. Researchers are developing several strategies to overcome the issues summarised above. There is a first methodological classification of these approaches, which distinguishes between two main criteria based on survey configuration and how the collinearity equations are applied (Luhmann et al., 2019, Mardani Nejad et al., 2021):

- The *object-invariant interfaces approach*, which assumes that position and orientation of the interfaces between the optical media are constant with respect to the object to be surveyed.
- The *bundle-invariant interfaces approach*, whereby the interfaces between the optical media are constant with respect to the camera.

As a second recognised discriminating criterion, the literature distinguishes five strategies nested in two main approaches, depending on the optical media where IO and possible relative orientation (RO) are determined (Kahmen et al., 2020, Rofallski and Luhmann, 2022):

- *The use of standard software to calibrate one or more cameras, accompanied by implicit modelling of refraction effects* to determine IO parameters and, where possible, RO parameters, can automatically lead to gross scaling errors *if the IO is determined in air* or to acceptable results *if the IO is defined in water* (e.g. by pre-calibration and using any RO to obtain scalar data).
- In the case of multiple cameras, the probability of scaling errors can be reduced by *an explicit modelling of the interfaces between media, which must be associated with water*: a first option is *to define both the IO and the RO in water*; a second strategy is *to determine both the IO and the RO in air*; a third alternative is *to define the IO in air and the possible RO in water*.

A third methodologically relevant classification concerns the way refraction effects are corrected in the acquired data (Maas, 2015, Skarlatos and Agrafiotis, 2018):

- The *analytical methods*, consisting of a modification of collinearity equations within the photogrammetric problem to adapt them to a situation of through-water survey. These in turn are subdivided into the explicit *deterministic analytical method* and the ambiguous, iterative *undefined analytical method*.
- The *image-based methods*, which involve a reprojection of the acquired image.

Most studied *through-water photogrammetry* applications contemplate the generation of Digital Elevation Models (DEMs) of submerged and semisubmerged areas with Bundle Block Adjustment (BBA) or Structure from Motion (SfM), using datasets acquired with passive optical sensors mounted on satellite (Cao et al., 2019) or aircraft platforms, including UAVs (Del Savio et al., 2023). During surveys under uncontrolled conditions, refractive indices and free water surface heights may be measured separately or included as unknowns in the BBA. However, in the presence of waves, schematising the air/water interface as a simple plane predictably leads to geometric errors. This is because agitation provokes quasi-random refraction phenomena. Recently, researchers are developing — through simulations or laboratory experiments — workflows addressing this specific issue: model regular waves of the liquid body in sinusoidal shapes, schematising the free surface as a diffuse or specular surface (Rupnik et al., 2020); compute parameters related to the air/water interface as unknowns in a BBA process (Sardemann et al., 2024); seek, within data relating to images acquired with water in a state of agitation, entities analogous to measurements that can be made through a planar free surface (Mulsow et al., 2024).

3. Methodology

So far, we have conducted seven experiments in the Hydraulics Laboratory of Sapienza University of Rome. The controlled environment consists of an 80 × 80 × 50 cm tank with 1 cm thick transparent glass side walls, placed under a metal frame equipped with a system for millimetric movement of the cameras. The glass walls and bottom of the tank were covered with 0.5 mm thick polyvinyl chloride (PVC) panels, secured with silicone sealant. An adhesive depth gauge was placed at one inner corner of the tank to monitor the water level in real time up to 45 cm. To mitigate the reflection of artificial light, shaped panels were added to both sides of the metal frame (Figure 1).

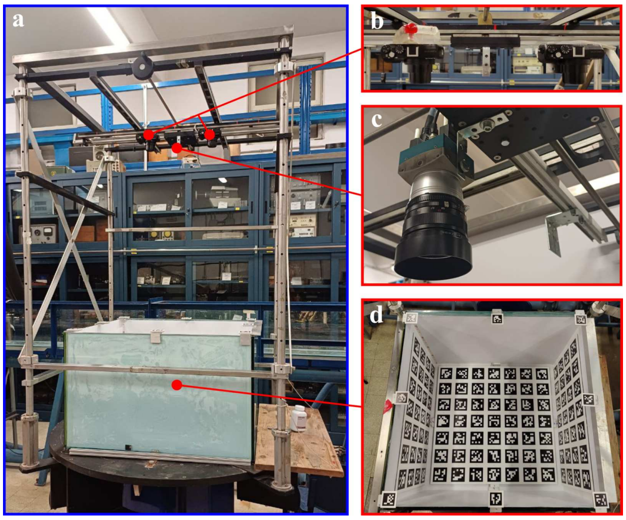

Figure 1. Photos of the equipment in the laboratory: a) view of the entire set-up; b) *Sony DSC-HX60* compact cameras; c) *Mikrotron EoSens CL MC1362* high-sensitivity, high-speed camera (500 FPS maximum); d) tank with encoded targets.

For the seven experiments carried out (Figure 2), we devised three possible acquisition configurations (Figure 3): a) zenithal monoscopy, used only in Tests I and VII, with the optical axis of the camera coinciding with the line perpendicular to the bottom of the tank and therefore a null angle ω between these two lines ($\omega = 0$); b) zenithal stereoscopy, used from Test II to Test VI, with the optical axes of the cameras parallel to the line perpendicular to the bottom of the tank ($\omega = 0$) and 300 mm apart (baseline = 300 mm); c) stereoscopy with converging optical axes, used in Tests II, III, IV, and VI, inclined 15° with respect to the line perpendicular to the bottom of the tank ($\omega = 15°$) and having a 600 mm distance between the barycenters of the sensors (baseline = 600 mm). Parameters and operating modes of the consumer cameras used in Tests I to VI — two *Sony DSC-HX60* cameras — are shown in Table 1, while the corresponding settings are shown in Table 2. The parameters and operating mode of the high-sensitivity, high-speed monocamera used in Test VII — an *EoSens CL MC1362* camera — are shown in Table 3. Before acquisition, each camera was calibrated out of the water (i.e., in air), thus estimating their IO parameters. The encoded target type we chose belongs to the *AprilTag* family, a visual fiducial system commonly used in

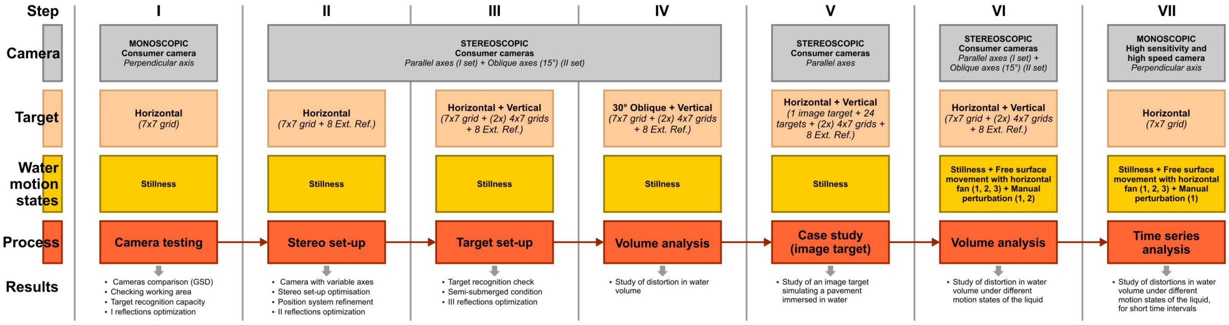

Figure 2. The experiments under consideration within our entire research project.

robotics and designed specifically to be detected effectively by automated algorithms. *AprilTag* markers feature a standard modular design, where a double square frame contains a planar grid of bits arranged in sequence. Each *AprilTag* marker has a unique identification sequence and is unequivocally named with a number (Krogius et al., 2019, Olson, 2019). For this series of experiments, we used the *tag36h11* family and the latest version of the *AprilTag* detection algorithm, *AprilTag 3*, implemented in the open source *pupil-apriltags* library for *Python* programming language (Prietz, 2019, Rauch, 2019).

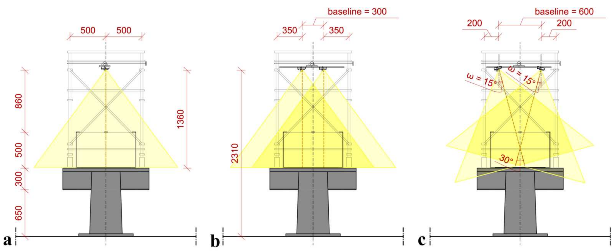

Figure 3. The three acquisition schemes used during the seven tests: a) zenithal monoscopy; b) zenithal stereoscopy; c) stereoscopy with converging axes.

We devised the experimental set-up considering the survey reference system as fixed to the objects to be detected — i.e., the grids of encoded targets in the tank. By imposing variations on the conditions of the water volume in the tank, namely free surface level and state of motion of the liquid, we caused alterations in the trajectories of the light rays captured by the sensors. Consequently, in each image captured in the presence of water, optical deformations of the *Apriltag* markers and thus displacements in comparison to a reference image captured with the tank empty were generated.

Lens	Sony G
Sensor	CMOS Exmor R - 7.76 mm (1/2.3”)
CMOS Size	5184 × 3456 px (6.03 × 4.62 mm)
Working Distance	1360 mm (to the base of the tank)
GSD	0.4 mm

Table 1. *Sony DSC-HX60* consumer-grade cameras parameters and working set-up.

In the stereoscopic tests, we evaluated, for each image, camera positions and orientations (already known to us) by estimating baseline and angle ω on the basis of IO parameters previously calculated with calibration operations and EO parameters determined with the Point-n-Perspective (PnP) approach (Marchand et al., 2016).

Focal Length	Shutter Speed	ISO	Diaphragm
4 mm	1/10	80	3.5

Table 2. *Sony DSC-HX60* consumer-grade cameras settings.

For each image dataset collected during the experiments, and for both cameras in the stereoscopic cases, the optical distortion caused by water refraction was automatically detected by tracking the four corner points of each marker across images captured under varying water levels or motion conditions. Since the marker grids remained fixed throughout the experiments, variations in the coordinates of the marker points, with respect to their positions in the reference image — the image captured with the tank empty, so without refraction — were attributed solely to the optical distotion induced by refraction and used to quantify it.

Sensitivity	25 V/lux.s
CMOS Size	1280 × 1024 px (17.92 × 14.34 mm)
Maximum frame rate at 1280 × 1024 px mode	500 frames per second
Working Distance	1360 mm (to the base of the tank)

Table 3. *EoSens CL MC1362* high-sensitivity, high-speed monocular camera parameters and working set-up.

In particular, we modelled the effects of refraction on target displacements Δ using a radial distortion law, where the displacement modulo increases with distance d from a central point of optical deformation, i.e. distortion centre $C(x_C, y_C)$ (Figure 4). Therefore, target displacements encode deformation caused by refraction of optical rays due to presence of increasing water levels. By marker displacement Δ, we mean the Euclidean distance of segment $\overline{PP'}$ between distorted position $P'(x_{P'}, y_{P'})$ — expressed in terms of image coordinates — of the i-th marker at the k-th filling step of the tank (every water level height equals $5 \times k$ cm with $k \geq 1$) and undistorted position $P(x_P, y_P)$ of the same marker at the first filling step of the tank (no water: water level equals to 0 cm). Thus, we propose the following functional model: $\Delta = K \cdot d^X$, where displacements Δ represent the observations and the four unknown parameters are the distortion centre coordinates (x_C, y_C), and the two parameters K and X. The model just outlined above is not linear, since: 1) coordinates (x_C, y_C) are within distance d, which is given by Pythagoras' Theorem as the square root of the sum of their squares; 2) X is an exponent. If, however, we apply the natural logarithm to both sides of the previous equation, we obtain: $ln(\Delta) = ln(K) + X - ln(d)$, a linearised form of the model that allows for the application of the Least Squares method via Gauss-Newton iterative scheme for the parameter estimation.

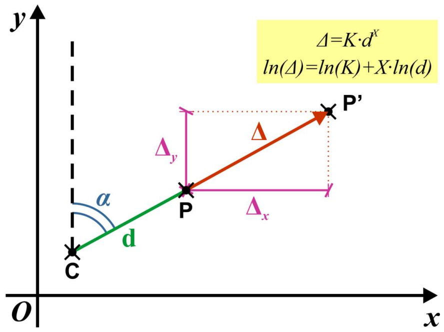

Figure 4. Schematic representation of optical distortion: P denotes the real position of a marker point (no water), while P' represents its apparent (distorted) position caused by water refraction. Displacement vector Δ, from P to P', quantifies the deformation considered in the proposed model.

From Test I to Test VI, the consumer cameras were individually controlled via Wi-Fi using the *Sony Imaging Edge Mobile (IEM)* app for portable digital devices, avoiding movement during the recording phase and attempting to capture frames simultaneously. In Test VII, several 1.5-second takes were recorded with the tank empty and with a water volume of 20 cm in various motion conditions, controlling the high-speed monoscopic camera via cable. We mainly organised the *AprilTag* markers according to known size and layout grids, printed on 3 mm thick *Forex* panels and secured to the tank using methods that were gradually refined (in order to prevent buoyancy, deformation, and unwanted movements). In all experiments, a 7 × 7 grid covering the tank bottom (parallel to the latter or inclined) was always present, while for tests III onwards, we added two 7 × 4 grids — one on the left wall and one on the right wall of the tank inside. The targets of these three grids have a side length of 75 mm. We also placed eight smaller *AprilTag* markers, with a side length of 40 mm, on the tank edge to define the constant reference frame of the survey. For Test V, designed to simulate a shooting from a *UAV*, a target with an orthophoto was printed on a 5 mm thick *Plexiglas* sheet, then glued to the *Forex* panel at the tank bottom with removable adhesive.

4. Experiments and Results

For each camera in every experimental setup dictated by the free surface height or the motion state of water, one (Test V) or three images (Test III, IV, and VI) were captured and processed. Whenever possible, we selected for each water condition the best pair from the three available pairs of images before processing. In Test VII, on the other hand, we recorded a 1.5-second take with the high-speed monoscopic camera at 500 FPS for each of the several water motion conditions considered, collecting frames that were all different from each other. Given the computing power required in this case, we have so far been able to perform partial processing by subsampling the recorded dataset at 25 and 50 FPS. Thanks to the known geometry of the marker grids (Ground Truth), the position and orientation of each camera could be estimated in relation to the targets. For every acquisition, positions of the four corners of each marker detected were automatically tracked, monitoring displacements of target vertices caused by refraction of the optical rays due to progressively higher water levels — Tests III, IV, and V (Figure 5) — or increasingly intense motion — Tests VI (Figure 6) and VII (Figure 7). For each experiment, the files containing the acquired images were sorted and divided according to the respective test iteration number, the source camera (monoscopic, left, or right), and the dataset (parallel axis or converging axis configuration). In Tests VI and VII, we envisaged eleven different water motion conditions, imposed with an electric fan and manual stirring: 1) still water; 2) fan at slow speed; 3) fan at medium speed; 4) fan at maximum speed; 5) fan at maximum speed and additional mild stir; 6) fan at maximum speed and additional strong stir; 7) water undergoing stabilisation after 6 minutes since condition no. 6; 8) water undergoing stabilisation after 9 minutes since condition no. 6; 9) water undergoing stabilisation after 10 minutes since condition no. 6; 10) water undergoing stabilisation after 11 minutes since condition no. 6; 11) water undergoing stabilisation after 12 minutes since condition no. 6. The analysis was carried out using a custom *Python* script developed with the *Open Source Computer Vision Library (OpenCV)* (OpenCV team, 2000). It first read the IO parameters of the cameras, obtained previously through calibration outside water (i.e., in air) using standard software, and reconstructed the Ground Truth of the marker grids by calculating their already known 3D coordinates. To improve efficiency of the marker recognition phase, the script converted each image from RGB colours to greyscale and applied a combination of thresholding techniques: the binary method and Otsu's method (Otsu, 1979). Each *AprilTag* is detected by the dedicated software library when its unique 36-bit sequence (white or black square elements), its black inner frame (with a 1-bit border) and its white outer frame (also with a 1-bit border) are recognisable in the image under consideration (Rauch, 2019). The script then located the four corners of each marker that met the aforementioned conditions and plotted the position of its centre as the intersection between the diagonals of a parallelogram with the same corners (Figures 5b, 5g, 6b, 6g, 7b), before storing the image coordinates of these five significant points in a nested array associated with the analysed image. Once the targets were detected, the script independently calculated the EO parameters for each camera using the PnP approach (Marchand et al., 2016). For each marker grid, the image coordinate sets of the significant points of the targets were divided into Ground Control Points (GCPs) and Check Points (CPs), allowing the average reprojection error to be checked using the estimated EO parameters, thus reprojecting the known 3D coordinates of the markers examined into the reference system of the image under processing. Using the independent EO of the cameras, our script conducted an estimate of every baseline — calculated as the Euclidean distance between the estimated positions of the centres of the two cameras — and a comparison with the values set in the experimental setup. The values of baselines and angles ω (ω being the angle between each optical axis and the line perpendicular to the tank bottom) were then saved in a spreadsheet in .csv format, together with their corresponding expected values. The strategy employed up to this point can be subsumed under *image-based methods* of refraction correction, since EO parameters, baselines, and angles ω of the two cameras were calculated by reprojecting the captured frames (Maas, 2015, Skarlatos and Agrafiotis, 2018). Our deliberate choice — undertaken independently by other researchers as well (Mulsow et al., 2024) — to avoid explicit modelling of the free surface and to calibrate the cameras in air using standard software was

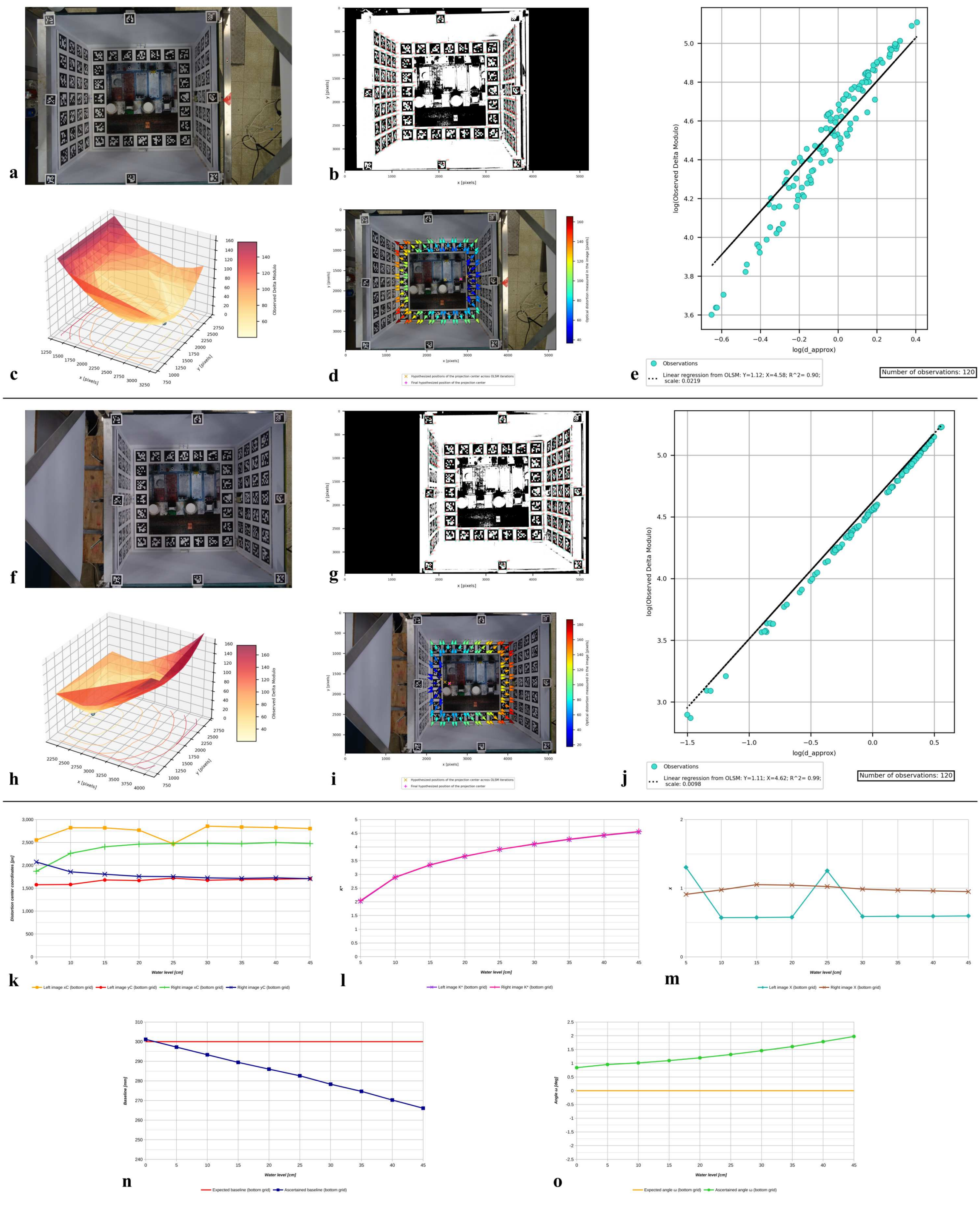

Figure 5. Excerpts from Test V (stereoscopy with parallel axes). Left camera: reference image DSC03393.JPG without water (a) and results from its comparison with image DSC03428.JPG, taken with a 45 cm water level (b-e). Right camera: reference image DSC01591.JPG without water (f) and results from its comparison with image DSC01626.JPG, taken with a 45 cm water level (g-j). Behaviour plots of parameters x_C and y_C (k), K^* (l), X (m), baseline (n), and angle ω (o).

dictated by the need to obtain common IO parameters for all shooting conditions (water levels and motion states). Our method, therefore, falls into neither *object-invariant interfaces* nor *bundle-invariant interfaces approaches* (Luhmann et al., 2019, Mardani Nejad et al., 2021), however, the multimedia photogrammetric problem framing adopted here does not suffer from scaling errors (Kahmen et al., 2020, Rofallski and Luhmann, 2022), thanks to the computation of a set of EO parameters estimated with the PnP approach for each shooting condition (Marchand et al., 2016). To begin the modelling phase of each sequence of n images captured with a given camera in each dataset, the script compared the image coordinate array of significant points of markers detected in the respective reference image — without water (Figures 5a, 5f, 6a, 6f, 7a) — with the corresponding coordinates of the same points in all the other $n-1$ images. For each of these $n-1$ comparisons, the program calculated the geometric distortion field — represented by the displacements of marker significant points — and depicted it as

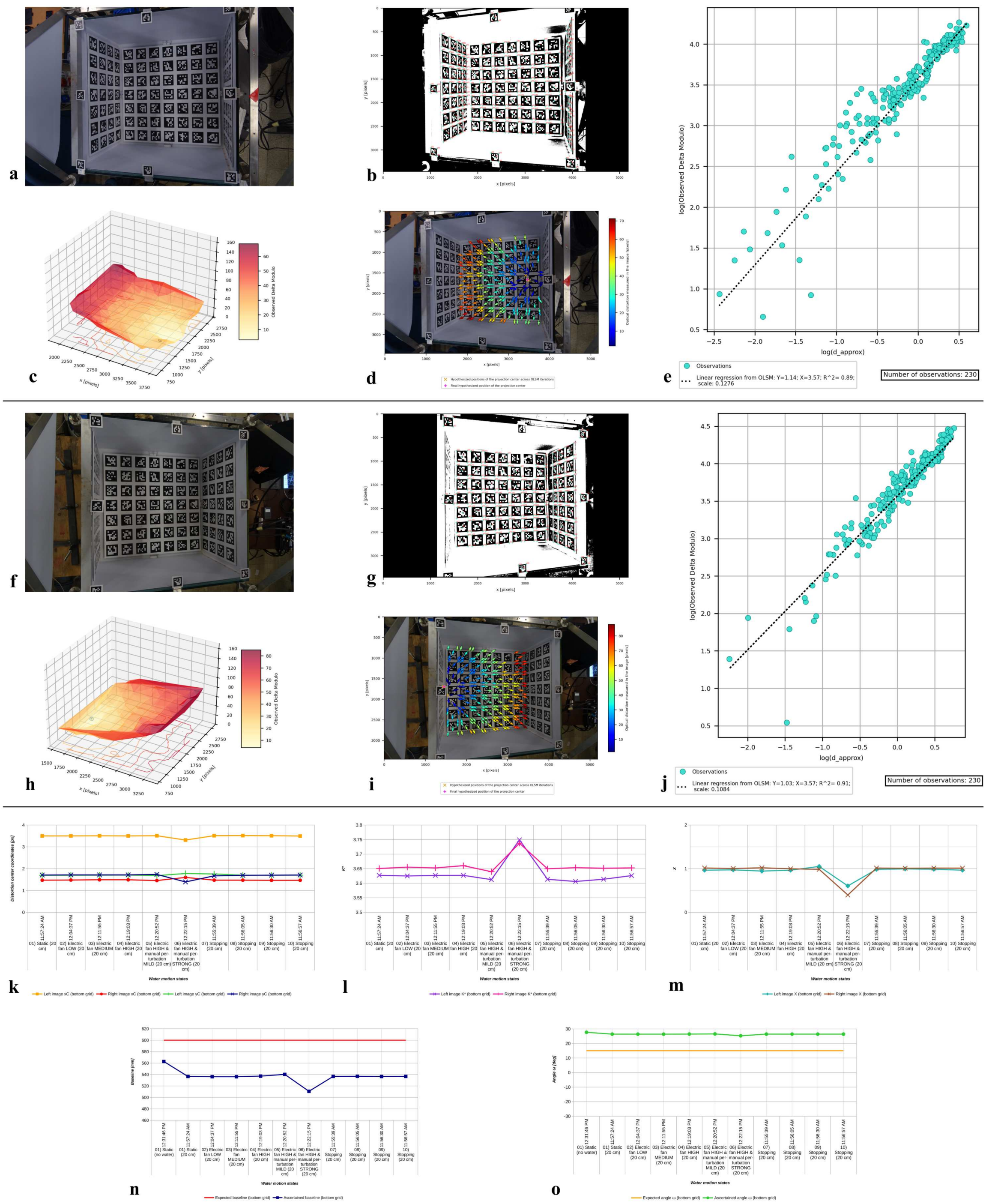

Figure 6. Excerpts from Test VI (stereoscopy with converging axes). Left camera: reference image DSC06998.JPG without water (a) and results from its comparison with image DSC06993.JPG, taken with a 20 cm water level in motion state no. 5 (b-e). Right camera: reference image DSC04426.JPG without water (f) and results from its comparison with image DSC04421.JPG, taken with a 20 cm water level in motion state no. 5 (g-j). Behaviour plots of parameters x_C and y_C (k), K^* (l), X (m), baseline (n), and angle ω (o).

a ruled surface in Cartesian space. In absence of relevant perturbations, the field of optical distortions induced by water refraction tends to take the form of an elliptic pseudo-paraboloid (Figures 5c, 5h), a function of image coordinates x, y and observed values of modulus Δ. The more the water is agitated, the more irregular the 3D diagram of the optical deformation field becomes (Figures 6c, 6h, 7c). In order to estimate the four unknowns of the mathematical model, it was first necessary to apply a regularisation of the marker image coordinates to avoid numerical instability phenomena. The script then computed approximate values of the coordinates of distortion centre C, assuming that it coincides with the minimum point of the displacement vector field, i.e. the point where observed modulus Δ assumes the smallest value in the entire image. To calculate approximate values of unknowns K and X, we first applied the RANdom SAmple Consensus (RANSAC) algorithm (Fischler and Bolles, 1981) to robustly estimate the parameters of the line fitting the observations in $ln - ln$ domain. Secondly, the

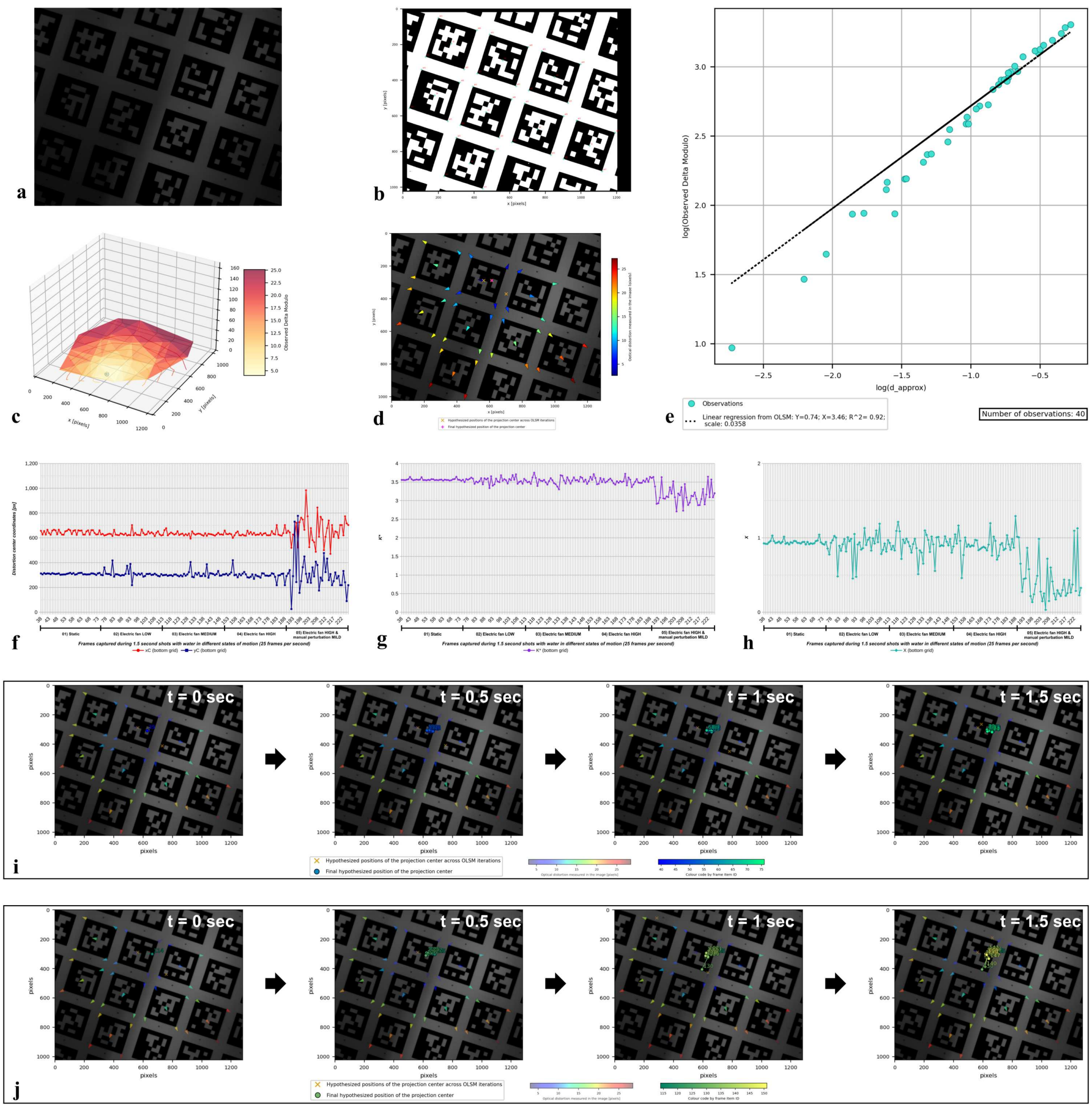

Figure 7. Excerpts from Test VII (Monoscopic, subsampled 25 FPS footage). Reference image A_Set1_00cm_01Static_00000.bmp without water (a) and results from its comparison with image E_Set1_20cm_04EFAhig_00660.bmp, taken with a 20 cm water level in motion state no. 4 (b-e). Behaviour plots of parameters x_C and y_C (f), K^* (g), X (h). Distortion centre C trajectories during observation intervals of 1.5 seconds in motion states no. 1 (i) and 3 (j).

script implemented the non-linear Ordinary Least Squares (OLS) method according to the Gauss-Newton iteration scheme (Fletcher, 1987) until satisfactory coefficients of determination R^2 and scale values σ_0 (sigma zero) were obtained, thus ensuring that the numerical algorithm used during modelling was stable. Once the iterative process was complete, the script determined the optimal values of the four modelling parameters (x_C, y_C, K, X) and plotted two diagrams with $ln - ln$ interpolation lines: the graph of the RANSAC line used to find the approximate values of the parameters K and X as well as the graph of the final optimised line, whose parameters were estimated at the end of the least squares iteration scheme. Predictably, the results with higher R^2 coefficients and lower σ_0 values were those with water in static or stabilising conditions (Figures 5e, 5j), since the more intense the motion, the greater the number of markers that cannot be correctly detected due to ripples, thus lowering the number of observations useful for the calculation (Figures 6e, 6j, 7e). For each of the $n-1$ comparisons, the script also generated a quiver plot, i.e. a diagram representing the optical distortions — the displacements of all significant points of the detected markers — as vectors (Figures 5d, 5i, 6d, 6i, 7d). Finally, for each dataset, the program plotted behaviour diagrams of modelling parameters (x_C, y_C, K, X) (Figures 5k, 5l, 5m, 6k, 6l, 6m, 7f, 7g, 7h) and, when possible, baselines (Figures 5n, 6n) and angles ω (Figures 5o, 6o) across the various shooting conditions (depth levels and water motion states). In Test VII, carried out by recording a 1.5-second video for each water motion condition, these diagrams effectively became historical series and also allowed us to plot the distortion centre C trajectory across time, e.g. in static condition (Figure 7i) or with motion induced by an electric fan at medium speed (Figure 7j).

5. Conclusions

We tested our model under controlled conditions, with free surface heights ranging from 5 to 45 cm and motion induced by fans and manual stir. Empirical data from the time series presented here confirm the robustness of our model as long as water is not excessively agitated. Aside from possible improvements in automatic detection of marker vertices displacements (e.g. through machine learning), the most immediate future developments for our method concern applications to case studies involving non-planar geometric shapes in space (a necessary step before carrying out any test under uncontrolled conditions).

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

The authors would like to thank Sapienza University of Rome for the research funding that supported the experimental path.

References

Agrafiotis, P., Skarlatos, D., Forbes, T., Poullis, C., Skamantzari, M., Georgopoulos, A., 2018. Underwater photogrammetry in very shallow waters: Main challenges and caustics effect removal. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2, 15–22.

Cao, B., Fang, Y., Jiang, Z., Gao, L., Hu, H., 2019. Shallow water bathymetry from WorldView-2 stereo imagery using two-media photogrammetry. *European J. of Remote Sens.*, 52(1), 506–521. 10.1080/22797254.2019.1658542.

Del Savio, A. A., Luna Torres, A., Vergara Olivera, M. A., Llimpe Rojas, S. R., Urday Ibarra, G. T., Neckel, A., 2023. Using UAVs and Photogrammetry in Bathymetric Surveys in Shallow Waters. *Appl. Sci.*, 13(6), 3420. doi.org/10.3390/app13063420.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6), 381–395. doi.org/10.1145/358669.358692.

Fletcher, R., 1987. *Practical Methods of Optimization*. John Wiley and Sons Ltd.

Kahmen, O., Rofallski, R., Luhmann, T., 2020. Impact of Stereo Camera Calibration to Object Accuracy in Multimedia Photogrammetry. *Remote Sensing*, 12(12), 2057. doi.org/10.3390/rs12122057.

Krogius, M., Haggenmiller, A., Olson, E., 2019. Flexible layouts for fiducial tags. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macau, China, 3 – 8 November 2019*, Institute of Electrical and Electronics Engineers, New York, 1898–1903.

Luhmann, T., Robson, S., Kyle, S., Boehm, J., 2019. *Close-range photogrammetry and 3D imaging*. De Gruyter.

Maas, H.-G., 2015. On the accuracy potential in underwater/multimedia photogrammetry. *Sensors*, 15(8), 18140–18152. doi.org/10.3390/s150818140.

Marchand, E., Uchiyama, H., Spindler, F., 2016. Pose estimation for augmented reality: A hands-on survey. *IEEE Transactions on Visualization and Computer Graphics*, 22(12), 2633–2651. doi.org/10.1109/TVCG.2015.2513408.

Mardani Nejad, A., Luhmann, T., Kersten, T. P., Dorschel, B., Purser, A., 2021. An underwater photogrammetric application to generate micro-bathymetry data for benthic habitat mapping and analysis at Arctic Ocean. *IFAC-PapersOnLine*, 54(16), 402–409. doi.org/10.1016/j.ifacol.2021.10.123.

Mulsow, C., Sardemann, H., Gueguen, L.-A., G., M., Maas, H.-G., 2024. Concepts for compensation of wave effects when measuring through water surfaces in photogrammetric applications. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2-2024, 289–295.

Olson, E., 2019. Apriltag: A robust and flexible visual fiducial system. *2011 IEEE International Conference on Robotics and Automation. Shanghai, China, 9 – 13 May 2011*, Institute of Electrical and Electronics Engineers, New York, 3400–3407.

OpenCV team, 2000. Opencv - open computer vision library. 'OpenCV - Open Computer Vision Library'. URL: https://opencv.org/ (8 February 2025).

Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66. doi.org/10.1109/TSMC.1979.4310076.

Prietz, P., 2019. Pupil-apriltags: Python bindings for the apriltags3 library. 'Pupil-apriltags: Python bindings for the Apriltags3 Library', Pupil-Labs. URL: https://github.com/pupil-labs/apriltags (8 February 2025).

Rauch, C., 2019. Apriltag 3. 'AprilTag 3', AprilRobotics. URL: https://github.com/AprilRobotics/apriltag (8 February 2025).

Rofallski, R., Luhmann, T., 2022. An efficient solution to Ray tracing problems in multimedia photogrammetry for flat refractive interfaces. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 90(1), 37–54. doi.org/10.1007/s41064-022-00192-1.

Rupnkik, E., Jansa, J., Pfeifer, N., 2020. Sinusoidal Wave Estimation Using Photogrammetry and Short Video Sequences. *Sensors*, 15(12), 30784–30809. doi.org/10.3390/s151229828.

Russo, M., Martelli, L., Ravanelli, R., A., P., 2024a. Automatic tag detection in multimedia survey condition. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W7-2024, 129–136.

Russo, M., Ravanelli, R., Pini, A., Flenghi, G., Martelli, L., Monti, P., Crespi, M., 2024b. Multimedia photogrammetry 2.0. a first step for underwater cultural heritage application. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W4-2024, 373–380.

Sardemann, H., Mulsow, C., Gueguen, L.-A., Mandelburger, G., Maas, H.-G., 2024. Multimedia photogrammetry with non-planar water surfaces – accuracy analysis on simulation basis. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2-2024, 363–369.

Skarlatos, D., Agrafiotis, P., 2018. A novel iterative water refraction correction algorithm for use in structure from motion photogrammetric pipeline. *J. Mar. Sci. Eng.*, 6(3), 77. doi.org/10.3390/jmse6030077.