The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B2-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission II, 4–11 July 2026, Toronto, Canada

Low-cost stereo vision and deep learning for river water level measurement

Pedro Zamboni[1], Robert Krüger[1], László Bertalan [2], Xabier Blanch[1, 3], Paul Hindorf [1], Anette Eltner[1]

[1] Institute of Photogrammetry and Remote Sensing, TUD Dresden University of Technology, Germany
[2] Department of Physical Geography and Geoinformatics, University of Debrecen, Hungary
[3]Department of Civil and Environmental Engineering, Universitat Politècnica de Catalunya, Spain

Keywords: Low-cost, stereo-photogrametry, deep learning, water level.

Abstract

The increasing frequency of extreme hydrological events, driven by climate change, necessitates dense and scalable water level monitoring networks. We evaluate a low-cost, non-contact, stereo-vision camera system for automated water level estimation. We compare two distinct image-processing pipelines — with and without semantic masking — to determine their camera pose stability and measurement accuracy. Our results show that raw stereo-vision estimates are highly correlated with reference sensor measurements (correlations ranging from 0.70 to 0.77), capturing the overall hydrologic behavior. By implementing a masking technique to isolate static environmental features, we successfully corrected a baseline error (i.e., offset), aligning the system with the true physical geometry. Although masking improves absolute accuracy, it introduces transient instability (i.e., spikes in pose estimation). This study serves as a proof of concept for the deployment of low-cost, edge-based stereo-vision systems in hydrological monitoring.

1. Introduction

Floods are the most common natural disaster to impact global populations (Du et al., 2010), exhibiting an increasing trend in the wake of climate change (Chen et al., 2023). Despite the rise in global flood events, there has been a decline in the number of active gauge stations over recent decades (Global Runoff Data Center, 2016). This has resulted in many river systems being inadequately monitored, ungauged, or operating under non-open-data policies. Nevertheless, in order to support the modelling of water resources and the planning thereof, it is necessary to establish dense networks of hydrological monitoring at an appropriate spatial and temporal scale (Eltner et al., 2021).

Water level is among the most critical hydrological variables necessitating monitoring. Conventionally, pressure gauges are the most popular type of monitoring station for water level measurement (Morgenschweis, 2010). Nevertheless, the expense of such systems is such that their deployment is limited, and so is the establishment of extensive monitoring networks. Furthermore, the loss of these devices during flooding events has been documented (Eltner et al., 2021), with the potential for the data to become unusable if the gauge is overwhelmed during such an event (Vandaele et al., 2021). Remote sensing has emerged as a cost-effective solution for measuring water levels, and it has attracted increasing interest. These systems have the obvious advantage of being safely installed outside the river cross-section. In recent years, camera gauges have gained attention as a low-cost and effective remote sensing technique for hydrological applications.

Several approaches have been explored to extract water levels from images. One approach involves the integration of water lines from images with physical structures such as scaled bars (Kuo and Tai, 2022) and stage boards or vertical rock surfaces (Leduc et al., 2018). Alternative approaches involve the utilisation of full images, including approaches such as landmark-based water-level estimation (LBWL) (Vandaele et al., 2021) and the Static Observer Flooding Index (SOFI) (Moy de Vitry et al., 2019). The LBWL is predicated on identifying landmarks in the image and interpolating two elevation values. Conversely, the SOFI tracks water level trends. Another approach includes the reprojection of segmented water line points into a 3D point cloud to produce centimetre-accurate water level measurements (Eltner et al., 2021, Blanch et al., 2026). The majority of the proposed methods are based on a single camera in conjunction with supplementary information, such as a scale bar or 3D models, for the purpose of accurately measuring water level. Nonetheless, changes in the river cross-section, including increased plant growth or sediment deposition and erosion, may lead to alterations in the geometric configuration of the river cross-section or cause impairment and obstruction of physical scale bars, degrading the accuracy of such methods.

A camera-based approach that has received comparatively less attention involves the utilization of two (or more) cameras. The main advantage of stereo-photogrammetry is that such systems are not dependent on the availability of ancillary information, such as a 3D model of the cross-section. It is possible to direct reproject points directly reproject points from water lines segmented in both images into three-dimensional space. Existing commercial stereo cameras have been utilized in conjunction with inertial measurement units (IMUs) for the estimation of water level and discharge (Hutley et al., 2023). However, these systems are not well suited for outdoor deployment, as they are often costly and feature small baselines (approximately 12 cm), which significantly limit their effective measurement range. Other research has already demonstrated the potential of using multi-camera setups for change detection of (Ulm et al., 2025, Hindorf and Eltner, 2026), but this has not yet been marginally explored for hydrological applications.

Our work introduces a novel stereoscopic system specifically designed for river and water level monitoring. The proposed system combines affordability with adaptability, offering a practical alternative to conventional, high-cost setups while maintaining reliable accuracy and performance. By eliminating the dependence on auxiliary reference objects or pre-existing three-dimensional models, the approach enables flexible deployment across diverse riverine settings. In this study, we evaluate strategies for deriving the external camera orientation, which is a fundamental component of our approach. Furthermore, an initial evaluation was conducted to ascertain the efficacy of the employed methodology in the retrieval of water level values, which were then compared with reference values.

2. Methods

2.1 Study area and system setup

The study area is located along the Hungarian section of the Sajó River. The configuration of the camera system is consistent with the design outlined in earlier studies (Krüger et al., 2024, Blanch et al., 2024). Four low-cost camera systems, based on Raspberry Pis equipped with Raspberry Pi Camera Module V3 sensors, were installed with known internal geometry. Each system is powered by solar panels and capable of storing images locally as well as transmitting data. The distance between each pair of cameras ranges from 8 to 11 meters. We selected two cameras for our experiments, with a baseline of 11.27 meters. We used a subset of 84 images from November 24, 2025, to December 23, 2025 because a larger flood event was captured in that period.

2.2 Deep learning for image segmentation

In order to segment the areas of interest (river and sky), a strategy previously evaluated in the literature (Zamboni et al., 2025) was employed. We used two state-of-the-art foundation models: Grounding DINO (Liu et al., 2024) and Segment Anything Model (SAM) (Kirillov et al., 2023). Grounding DINO is an open-set object detection model that has the capacity to identify objects in images based on textual prompts (e.g., "river" or "sky"). SAM constitutes a foundational model for image segmentation, with the capacity to generate precise segmentation masks from minimal input, including text, points, or bounding boxes. In the present study, Grounding DINO was utilized in the initial phase to detect and generate bounding boxes for the target areas. These bounding boxes were subsequently transferred to SAM to generate the corresponding segmentation masks. The production of sky and river masks was undertaken for the purpose of preventing ambiguous tie points between images. Furthermore, water masks had the additional purpose of identifying river lines for the measurement of water levels.

2.3 Stereoscopic workflow

In the context of a stereo camera system, it is essential to account for both interior and exterior geometry as a prerequisite for the reliable retrieval of water level data from images. Prior to the initiation of the experiment, a calibration process was applied to all cameras to obtain the interior geometry (principal distance, principal point, distortion parameters). Even though the Raspberry Pi Camera Module 3 features an auto-focus system, we assumed that the interior geometry remained constant across all image epochs.

In order to retrieve the exterior geometry of each camera, two strategies were employed: global and epoch optimization. The efficacy of both strategies was determined through empirical testing, utilizing comprehensive image data. We conducted an analysis for both strategies using images with and without river and sky masks. Here, masks were used to reduce potentially ambiguous tie points in water and sky regions. The deep learning feature extractor SuperPoint (DeTone et al., 2018) was utilized in conjunction with the deep learning feature matcher LightGlue (Lindenberger et al., 2023).

The global optimisation was achieved through the implementation of a conventional structure-from-motion approach. Thereby, the matched points in each stereo image pair for each epoch were consolidated into a single set of matched points. Therefore, a single camera pose for each camera of the stereo system was calculated and shared across all epochs. The fundamental premise of this approach is that the cameras remained stable throughout the image acquisition period, an assumption that may not always hold in practice. The method combines all points into a single reconstruction, relying on tie points from well-illuminated images to compensate for those obtained under less favorable conditions. However, this can also propagate errors, as tie points from suboptimal images may still influence the overall reconstruction.

As an alternative approach, during the epoch-based optimization the poses of the individual cameras for each epoch were calculated (Figure 1). In this study, a combination of tie points from cameras C1 and C2 at epoch T with tracked points from each camera at epoch T–1 was employed to achieve a more stable camera pose solution across all epochs. After, we estimate the camera pose for both global and epoch optimization.

2.4 Water height measurement

In order to extract water height values from the stereo image pairs, it is necessary to consider the interior and exterior camera geometry as well as the river shorelines from each image. Water level estimation was performed exclusively using camera poses derived from global optimization in order to rigorously demonstrate the capability of the proposed methodology to achieve accurate and reliable results.

The shoreline was extracted from the upper part of the river water mask in each image of the stereo pair. A subset of water points along the contour was extracted, covering a length of 50 pixels each. As the images did not exhibit complete overlap, only points in the overlapping areas were retained. For each water point in each image, a square grid of seven pixels was created. The SIFT algorithm (Lowe, 2004) was then employed to extract a descriptor for each proposed grid. This strategy was employed to narrow the search window while incorporating additional information from the surrounding area of each water point, thereby providing broader contextual information. Subsequently, L2 norm distances were computed from all grid candidates from the left image to the right image. Finally, the Lowe's ratio test was applied in order to ascertain the matched points.

The reprojection of all matched points into three-dimensional space (in an arbitrary coordinate system) was achieved by utilizing the interior and exterior camera geometries. The camera base information was used to scale down the results. A filter based on the distribution of the retrieved Z-axis values for the water points was undertaken to eliminate potential outliers. The mean Z-axis value and the standard deviation were computed. Note** that** the scaling of our Z values was not conducted to water level or absolute heights. All results were maintained within a scaled relative coordinate system.

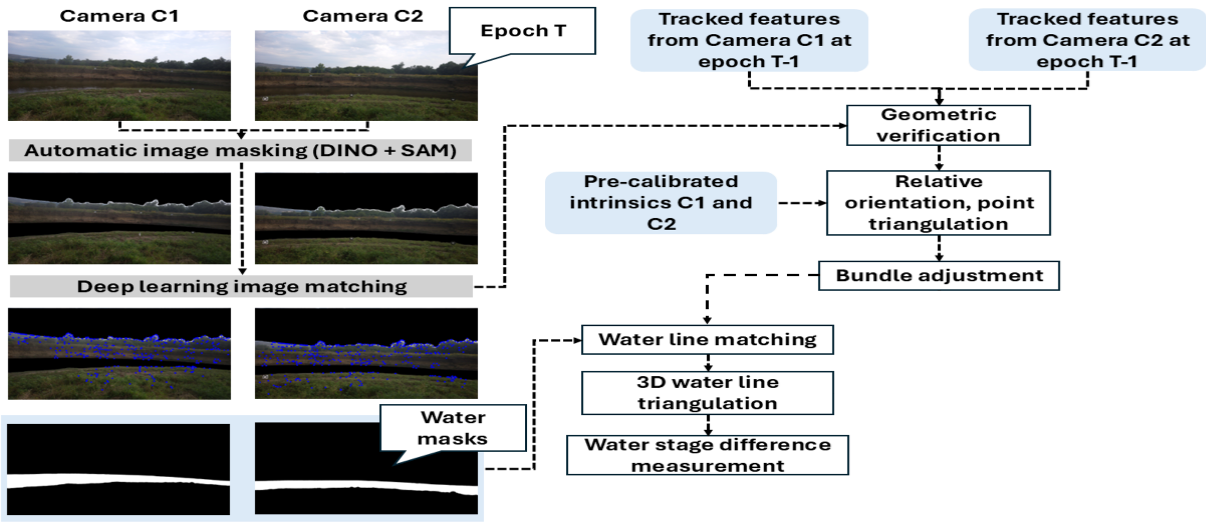

Figure 1. Epoch Stereoscopic 3D reconstruction and water level measurement.

In the absence of a reference at the camera site, the nearest available gauging station was utilized. Upon recognizing that the results were not commensurate, the subsequent analysis was limited to correlation analysis and a visual interpretation of the overall water level trends between the employed methodology and the reference.

3. Results

3.1 Exterior geometry estimation

We compute the exterior geometry for each camera using global and epoch optimization. Afterwards, we compute the camera pose to compare its position over different epochs, with and without the use of masks.

The estimation of the camera poses was based on global and local optimization, with and without the use of masks. In the case of global optimization, the utilization of masks did not result in a significant alteration of the camera pose (Figure 2). In the case of the epoch-based method, a compact cluster of camera poses characterized by reduced temporal drift was observed in the absence of masks when compared to the clusters formed when using masks. The utilization of masks resulted in intermittent substantial deviations, predominantly along the Y and Z axes, which is indicative of a significant portion of the scenery being obscured, particularly during flood events, resulting in an inadequate area of the image being available for effective image matching.

Camera pose deviation for exterior geometry estimates using masks is generally lower than for estimates without masks; however, masked approaches exhibit occasional spikes (Figure 3). This suggests that while masking helps the algorithm converge toward a more accurate “true” pose, it also introduces fragility. When the mask obscures too many reliable features, the estimation can momentarily fail. In contrast, the unmasked solution shows lower variance and no large outliers, indicating a more robust optimization process, albeit one that may be affected by systematic bias. Overall, masking improves precision, whereas the unmasked approach enhances robustness.

For both cameras, the primary source of error is along the Z-axis (Figure 3). For Camera 1, the X and Y components demonstrate stability, suggesting that lateral pose estimation is well constrained. The unmasked approach demonstrates a more refined deviation behavior. For Camera 2, the masked solution exhibits elevated levels of noise, particularly along the Z-axis. In contrast, the X and Y components demonstrate relatively consistent behavior in both cases (with/without mask), with the primary distinction being a consistent offset between the two solutions. With regard to the camera distance, the unmasked approach provides more stable estimates, although it generally underestimates the true distance. This phenomenon can be attributed to a well-established bias-variance trade-off, whereby masking leads to more accurate but less stable scale estimates, while the unmasked approach yields more stable but systematically biased reconstructions.

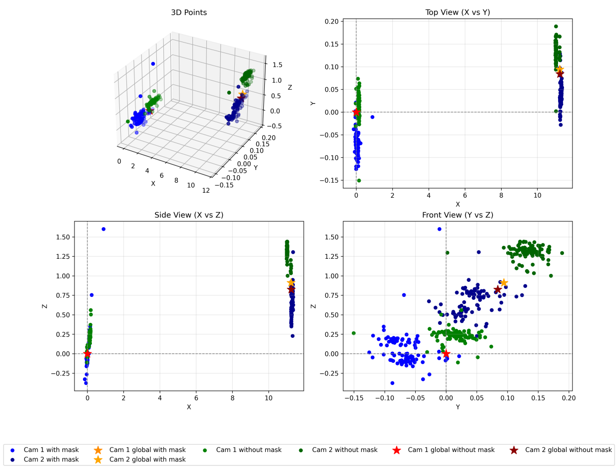

Figure 2. 3D camera positions across all epochs and globally optimized positions.

The camera poses estimated with the unmasked configuration manifest as a flat and stable plateau, whilst the masked solution demonstrates noisier motion patterns (Figure 4). This is further highlighted by the distribution of the camera poses: the unmasked solution follows an approximately normal distribution centered around 0.3 meters, while the masked results appear closer to a bimodal distribution. The masked solution exhibits a reduced median displacement; however, it also displays a wider

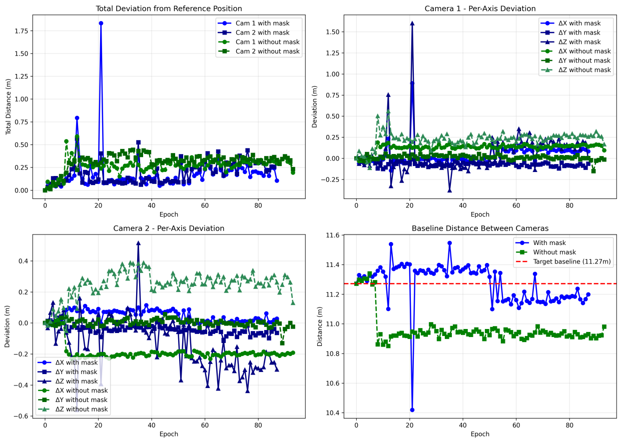

Figure 3. Camera center deviations.

inter-quartile range and numerous high-value outliers. In contrast, the unmasked solution demonstrates a higher median but a more concentrated distribution, with a reduced number of extreme values. These results exemplify a bias-variance trade-off: masking enhances central tendency (accuracy) but reduces reliability, whereas the unmasked approach augments repeatability at the expense of systematic bias.

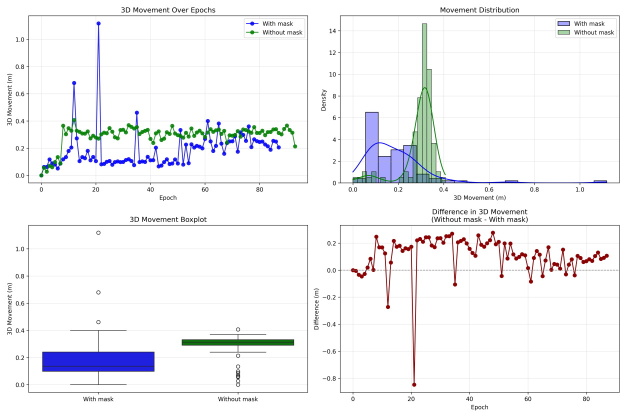

Figure 4. Estimated camera drift across epochs.

3.2 Water height measurement

We present water-level estimates derived from a stereoscopic system computed both with and without masking.

Initial experiments were conducted to assess the optimal grid kernel for water point feature description and subsequent matching. Grids of 5×5, 7×7, and 9×9 were tested, and their impact on water level estimation was evaluated. The results suggested that the 7×7 grid provided the most reliable water level estimates.

The estimated camera poses were comparable across both global configurations (with and without masks); thus, the corresponding water-level estimates also showed minimal variation. In both cases, the estimated values were strongly correlated with in situ measurements (Figure 5). Pearson correlation coefficients ranged from 0.70 to 0.71, while Spearman’s rank correlation coefficient was 0.77 for both configurations. High Spearman and Pearson correlation values indicate that both time series, reference and camera-based, have a strong and consistent monotonic relationship; however, it is not strictly linear (Figure 5), which is to be expected due to different river cross-sections and additional tributaries between the reference and the camera gauge.

Our results showed a decrease in the confidence of larger water level estimates (Figure 5). The camera gauge values were more accurate for lower reference water levels (80 cm – 140 cm), but as the water level increased, the variance increased and the estimation became less certain.

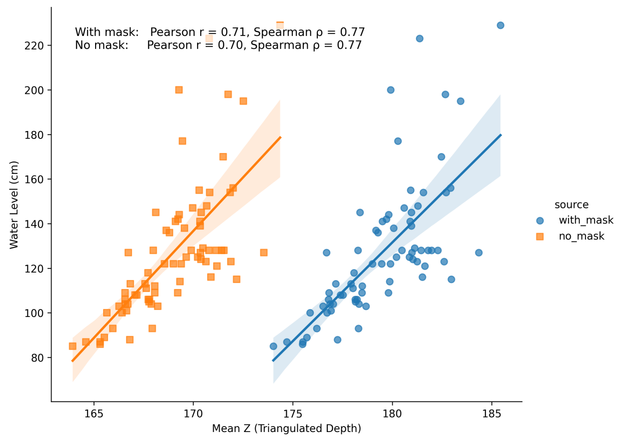

Figure 5. Correlation between estimated and reference water levels.

Both methods — with/without mask — achieved similar water height values, being only shifted by about 10 units (Figure 6). Our approach was highly effective at tracking the trend during a flood event, capturing the dominant hydrological signal observed in the reference data.

The sharp peak observed in the reference data is, however, only partially reproduced by our camera gauge, implying limitations in capturing more challenging conditions, potentially due to unfavorable geometric constraints and low image quality. A strong difference between reference and camera gauge was identified during the period spanning from 9 to 17 December, which was characterized by elevated water levels. Strongly overestimated water levels may be attributable to erroneously segmented water masks and inadequate water point matching. Such issues are indicative of low image quality and suboptimal illumination conditions.

One important aspect to highlight is the different nature of the two measurement approaches. On the one hand, our method captures the water level at a specific instant in time. On the other hand, traditional gauge stations report measurements averaged over predefined intervals — 15 minutes in this case. This factor may also lead to an imperfect comparison between our results and the reference values.

Overall, our findings demonstrate that water-level estimation based on a single exterior geometry per camera and a known base can achieve a high correlation and capture the overall hydrologic behavior of water level trends. While masking introduced a consistent offset in the reconstructed height values, it did not improve correlation with ground truth measurements. Consequently, our results indicate that the choice of masking may be guided by practical considerations rather than performance gains in this specific configuration.

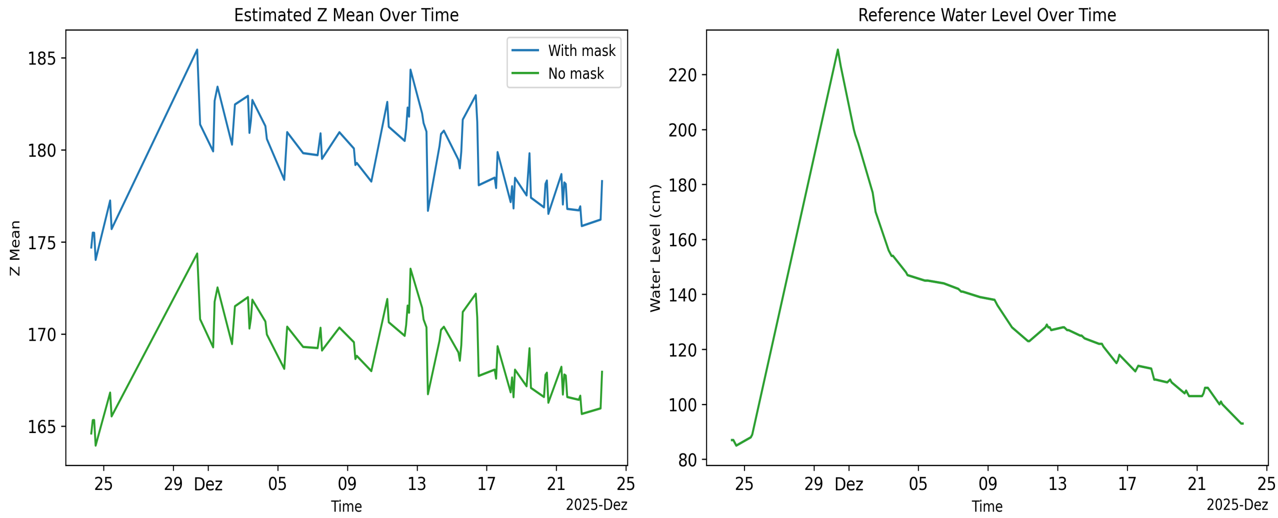

Figure 6. Comparison between estimated and measured water levels over time.

4. Limitations and future works

Notwithstanding the encouraging initial performance of the proposed stereoscopic system, there are several limitations that remain to be addressed. The accuracy of water level estimation is contingent upon the quality of river segmentation, a process that can be adversely affected by challenging environmental conditions, such as low lighting and reflections. Furthermore, the cameras were not installed with the specific purpose of measuring water levels. Instead, they were positioned in such a manner that they would be oriented towards the riverbank in a near normal-stereo configuration to perform 3D change detection. The initial strategy devised was to match the water points from both images by employing the epilines from the left camera to the right camera, in addition to conducting a symmetry check. This was identified as a challenge in the given scenario, where the epilines were found to be in close proximity to each other, resulting in highly ambiguous point matching.

Furthermore, the implementation of the SIFT descriptor to align water points across both images was found to be influenced by challenging illumination conditions. This factor has the potential to result in suboptimal performance outcomes. The camera setup was implemented using the Pi Camera Module V3 sensor, which features auto-focus capabilities. It can thus be concluded that the initial assumption that the geometry of the camera interior remained constant across all images was erroneous. In order to demonstrate the concept, a single camera exterior geometry was employed for each camera across all epochs. Nevertheless, the results obtained demonstrated significant variations in the orientation of the camera exterior over time. Consequently, it is possible that the methodology employed may also result in suboptimal performance.

Due to several limitations in the experimental setup and dataset, the initial objective was to achieve decimeter-level accuracy. However, reaching centimeter-level precision will require further developments. First, fixed-focus cameras specifically designed to operate in a convergent configuration could be employed. This would improve image overlap and help stabilize the interior camera geometry. Additionally, the hypothesis that the stereo camera pair may be misaligned with respect to one or more axes should be investigated, as correcting such misalignment could enhance water surface point matching. Furthermore, a sliding-window optimization approach could be introduced to improve the estimation of exterior camera geometry. This could be combined with a modified bundle adjustment strategy to reduce drift in camera poses across different epochs.

As it stands, camera-based systems rely on water segmentation to extract the river shoreline, particularly through deep image segmentation techniques. However, such methods are known to struggle with accurately delineating object boundaries. Therefore, approaches that provide not just a single estimated shoreline, but a probabilistic representation of the river shoreline position, could be beneficial. Incorporating such uncertainty estimates would improve the overall robustness of the system.

5. Conclusion

We proposed a low-cost stereo camera system for water level estimation. We evaluated the impact of masking on camera pose estimation and its downstream effect on water-level retrieval using a fixed exterior geometry per camera.

Clear differences emerge in the camera pose estimation. A masked image configuration improved accuracy. However, the gain in reduced error is offset by reduced robustness, as evidenced by intermittent large errors, instability in the vertical component, and occasional breakdowns in baseline consistency. In contrast, the unmasked configuration yielded more stable and temporally consistent pose estimates with fewer outliers and more reliable relative camera geometry. However, a systematic bias affecting both camera poses and derived measurements was thereby introduced.

Across all experiments, both masked and unmasked configurations, we were able to successfully capture the dominant temporal dynamics of water-level variation, yielding comparable correlation with reference measurements. This indicates that the predictive capability of the system was primarily governed by temporal consistency rather than absolute geometric accuracy.

Overall, these findings demonstrate that reliable water-level estimation can be achieved even with imperfect camera geometry, provided that temporal coherence is preserved.

6. Acknowledgments

This research was funded by the Federal Ministry of Education and Research (BMBF) and DAAD (project no. 57703453). Paul Hindorf and Xabier Blanch were funded by the German Research Foundation (DFG, CRC/TRR 408 AgiMo project no. 531327426 and AI4Glaciers project no. 536233846).

References

Blanch, X., Grundmann, J., Hedel, R., Eltner, A., 2026. AI image-based method for a robust automatic real-time water level monitoring: a long-term application case. *Hydrology and Earth System Sciences*, 30(3), 797–824. https://hess.copernicus.org/articles/30/797/2026/.

Blanch, X., Guinau, M., Eltner, A., Abellan, A., 2024. A cost-effective image-based system for 3D geomorphic monitoring: An application to rockfalls. *Geomorphology*, 449, 109065.

Chen, J., Shi, X., Gu, L., Wu, G., Su, T., Wang, H.-M., Kim, J.-S., Zhang, L., Xiong, L., 2023. Impacts of climate warming on global floods and their implication to current flood defense standards. *Journal of Hydrology*, 618, 129236.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. Superpoint: Self-supervised interest point detection and description. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 224–236.

Du, W., FitzGerald, G. J., Clark, M., Hou, X.-Y., 2010. Health Impacts of Floods. *Prehospital and Disaster Medicine*, 25(3), 265–272.

Eltner, A., Bressan, P. O., Akiyama, T., Gonçalves, W. N., Marcato Junior, J., 2021. Using Deep Learning for Automatic Water Stage Measurements. *Water Resources Research*, 57(3), e2020WR027608. e2020WR027608 2020WR027608.

Global Runoff Data Center, 2016. Global runoff data base, temporal distribution of available discharge data. [Online; accessed June 20, 2022].

Hindorf, P., Eltner, A., 2026. Kombinierter einsatz von künstlicher intelligenz und time-lapse-photogrammetrie für robustes geomorphologisches monitoring. *Publikationen der DGPF, Band 34*, Darmstadt, Germany, 39. 46. Wissenschaftlich-Technische Jahrestagung der DGPF.

Hutley, N. R., Beecroft, R., Wagenaar, D., Soutar, J., Edwards, B., Deering, N., Grinham, A., Albert, S., 2023. Adaptively monitoring streamflow using a stereo computer vision system. *Hydrology and Earth System Sciences*, 27(10), 2051–2073. https://hess.copernicus.org/articles/27/2051/2023/.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment anything.

Krüger, R., Blanch, X., Grundmann, J., Al-Rawas, G., Eltner, A., 2024. Low-Cost sensors for measuring wadi discharge - a Raspberry Pi based seismometer and time-lapse camera setup. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W8-2024, 243–250. https://isprs-archives.copernicus.org/articles/XLVIII-2-W8-2024/243/2024/.

Kuo, L.-C., Tai, C.-C., 2022. Robust Image-Based Water-Level Estimation Using Single-Camera Monitoring. *IEEE Transactions on Instrumentation and Measurement*, 71, 1-11.

Leduc, P., Ashmore, P., Sjogren, D., 2018. Technical note: Stage and water width measurement of a mountain stream using a simple time-lapse camera. *Hydrology and Earth System Sciences*, 22(1), 1–11. https://hess.copernicus.org/articles/22/1/2018/.

Lindenberger, P., Sarlin, P.-E., Pollefeys, M., 2023. LightGlue: Local Feature Matching at Light Speed. *ICCV*.

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., Zhang, L., 2024. Grounding dino: Marrying dino with grounded pre-training for open-set object detection.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94.

Morgenschweis, G., 2010. *Hydrometrie: Theorie und Praxis der Durchflussmessung in offenen Gerinnen*. Springer-Verlag.

Moy de Vitry, M., Kramer, S., Wegner, J. D., Leitão, J. P., 2019. Scalable flood level trend monitoring with surveillance cameras using a deep convolutional neural network. *Hydrology and Earth System Sciences*, 23(11), 4621–4634. https://hess.copernicus.org/articles/23/4621/2019/.

Ulm, M., Elias, M., Eltner, A., Lotsari, E., Anders, K., 2025. Automated change detection in photogrammetric 4D point clouds – transferability and extension of 4D objects-by-change for monitoring riverbank dynamics using low-cost cameras. *Applied Geomatics*, 17(2).

Vandaele, R., Dance, S. L., Ojha, V., 2021. Deep learning for automated river-level monitoring through river-camera images: an approach based on water segmentation and transfer learning. *Hydrology and Earth System Sciences*, 25(8), 4435–4453. https://hess.copernicus.org/articles/25/4435/2021/.

Zamboni, P. A. P., Blanch, X., Junior, J. M., Gonçalves, W. N., Eltner, A., 2025. Do we need to label large datasets for river water segmentation? Benchmark and stage estimation with minimum to non-labeled image time series. *International Journal of Remote Sensing*, 46(7), 2719–2747. https://doi.org/10.1080/01431161.2025.2457131.