Human Interaction and Emerging Technologies (IHIET-FS 2026), Vol. 235, 2026, 167–176
https://doi.org/10.54941/ahfe1007272

Deep-Learning Assisted Digital Twin of Stereo Camera for Non-Invasive Underwater Fish Biomass Estimation

Hsun-Yu Lan[1], Bo-Cheng Huang[2], and Shyi-Chyi Cheng[2]

[1]Department of Aquaculture, National Taiwan Ocean University, Keelung 202301, Taiwan
[2]Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung 202301, Taiwan

ABSTRACT

Digital twins have become increasingly important in aquaculture, a sector traditionally dependent on manual, subjective, invasive, and labor-intensive practices. This trend is driven by the convergence of precision aquaculture and AIoT technologies, enabling a shift from experience-based practices to intelligent, data-driven systems. Existing AIoT architectures are highly specialized, addressing function-specific requirements. This work proposes a digital twin–enabled smart stereo camera system for monitoring fish growth through non-invasive biomass estimation of freely swimming fish. The digital twin framework includes a processing pipeline consisting of RGB-D video acquisition, 6D (RGB-XYZ) representation generation, 3D point cloud Transformer-based segmentation, and fish biomass regression. First, the stereo camera captures RGB-D video in the aquaculture environment, which is automatically transmitted to a cloud system for further processing. The RGB-D frames are then transformed into 6D (RGB-XYZ) representations for subsequent analysis. A 3D point cloud Transformer is then used to detect and segment fish objects from the 6D representations. Finally, the reconstructed 3D fish objects are used for k-nearest neighbors (KNN) regression to estimate fish biomass. The contributions of this work are as follows. First, the digital twin approach enables the transformation of aquaculture toward intelligent farm management. Second, the 3D computer-vision-based fish biomass estimation scheme is a non-invasive model for understanding the status of fish growth without disturbing fish schools. Third, the proposed 3D point cloud Transformer has low computational complexity and can be deployed on edge-computing platforms with limited GPU resources. Finally, the digital twin model synthesizes fish growth data based on the existing fish growth model to improve the estimation accuracy of fish biomass. To the best of our knowledge, this work presents one of the first digital twin–enabled smart camera systems deployed in real aquaculture environments for real-time fish growth monitoring.

Keywords: Smart camera, Digital twin, Deep learning, Artificial Intelligence of Things (AIoT), Fish biomass estimation, RGB-D video

INTRODUCTION

Aquaculture has long relied on manual sampling and experience-based judgment to assess fish growth status. Related measurements often require capturing and contact-based handling, which are not only labor-intensive but

Received April 19, 2026; Revised May 6, 2026; Accepted May 22, 2026; Available online June 22, 2026

may also impose stress on fish schools and affect their behavior and growth performance. With the development of the Artificial Intelligence of Things (AIoT) and precision aquaculture technologies, aquaculture management is gradually shifting from experience-driven practices to data-based monitoring, making non-invasive fish measurement an important research direction.

A digital twin provides a feasible framework for linking physical farming sites with virtual models. By continuously integrating sensing data and analytical models, a virtual system can reflect the physical state in real time and thereby support monitoring and decision-making (Tao & Zhang, 2017). In recent years, digital twin technology has been increasingly applied in agriculture and aquaculture. Through AIoT-based integration of sensing data and analytical models, it can effectively improve management efficiency and decision quality (Ubina et al., 2023). However, most existing systems focus primarily on environmental monitoring or single-function applications and still lack mechanisms for individual-level growth analysis and measurement of fish bodies.

In non-invasive measurement, computer vision has been widely applied to fish analysis. Stereo vision systems can recover scene depth from image disparity and further establish three-dimensional geometric structures, enabling fish morphology to be measured in spatial space (Scharstein & Szeliski, 2002). With the development of RGB-D sensing devices, image and depth information can be acquired simultaneously, making 3D reconstruction more stable and suitable for real-time use. Based on such multimodal data, integrating image information with spatial coordinates into high-dimensional feature representations, such as RGB-XYZ, has become an important direction in 3D scene understanding (Guo et al., 2021).

For 3D data processing, point cloud representation has become a mainstream way to describe spatial structures, allowing objects to be analyzed and modeled directly in three-dimensional space (Guo et al., 2021). In recent years, deep learning methods have been increasingly applied to 3D data analysis. Among them, Transformer architectures, with their ability to model global features, have been extended to vision tasks and have shown strong performance in handling complex data representations (Khan et al., 2022). Such methods are particularly suitable for high-dimensional point cloud data because they can effectively learn features while preserving spatial structural information. In addition, image segmentation has gradually evolved from 2D semantic segmentation to 3D spatial segmentation, enabling target objects to be identified more robustly in complex backgrounds.

For fish biomass estimation, a key issue is how to infer weight from geometric or feature information. Compared with highly complex models, similarity-based regression methods remain stable and interpretable in scenarios with limited data or real-time computing requirements. Among them, k-nearest neighbors (KNN) is widely used in various estimation tasks because of its simplicity and the fact that it does not require a training process (Cover & Hart, 1967).

Although these technologies have each matured individually, integrating RGB-D imagery, multidimensional feature representation, 3D deep learning models, and biomass estimation into a complete processing pipeline that can operate in real aquaculture environments remains challenging. In particular, under conditions involving freely swimming fish, occlusion, and underwater environmental interference, it remains difficult to perform stable fish segmentation and individual measurement while maintaining computational efficiency and deployment feasibility.

This study proposes a smart stereo camera system integrated with a digital twin framework for non-invasive fish biomass estimation. As shown in Figure 1, the system captures RGB-D fish school imagery through underwater stereo imaging devices and transforms the data into a 6D (RGB-XYZ) representation so that subsequent analysis can be carried out directly in three-dimensional space.

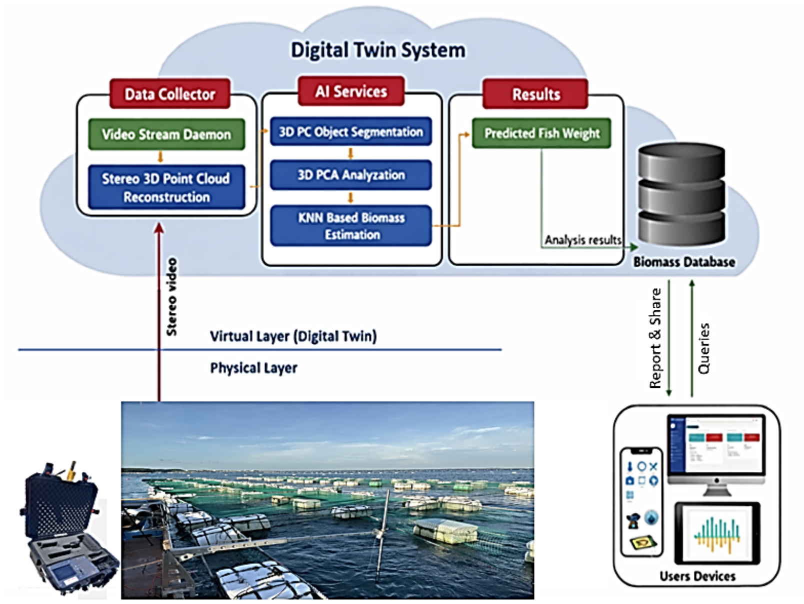

Figure 1: Overall architecture of the proposed digital twin–enabled stereo camera system. RGB-D data are transformed into 6D (RGB-XYZ) representations for 3D fish segmentation and biomass estimation, and the results are fed back to the digital twin model for continuous monitoring.

In terms of methodology, a 3D point cloud Transformer is adopted for fish detection and segmentation, and the segmented fish objects are then input into a k-nearest neighbors regression model to estimate biomass. By integrating image acquisition, data transformation, 3D segmentation, and biomass estimation into a continuous processing pipeline, the system is able to monitor fish growth status in real time and serve as the basis for updating the digital twin model. The main contributions of this study are as follows. First, it constructs a digital twin-based fish growth monitoring framework, allowing fish status in aquaculture environments to be continuously mapped to a virtual system. Second, it establishes a non-invasive fish biomass estimation pipeline by integrating RGB-D sensing with 3D computer vision techniques. Third, it introduces a 3D point cloud Transformer that maintains segmentation performance while also considering computational efficiency, thereby supporting deployment on edge-computing platforms. Finally, it enhances the stability and application value of biomass estimation by integrating fish growth information into the digital twin model.

METHODOLOGY

The system was deployed at an offshore cage farm in Hengchun, Pingtung, and at the aquaculture facility of the Aquaculture Research Center, National Taiwan Ocean University. As shown in Figures 2 and 3, the hardware setup consisted of an underwater stereo camera (ZED 2i) mounted in a waterproof housing with a fixed support frame for long-term image acquisition, while an embedded device (Jetson Nano) was used for data collection and transmission control. The image data were transmitted through an ICT-based architecture to a cloud platform for storage and management, forming a time-series database to support subsequent analysis and digital twin updates. Compared with manual sampling, this approach enables continuous observation without disturbing fish schools.

Figure 2: Hardware architecture of the underwater stereo camera system, including the stereo camera, embedded device, and communication module for continuous data acquisition and cloud-based processing.

Figure 3: Field deployment of the underwater stereo camera system in an offshore cage environment, enabling long-term and non-invasive observation of freely swimming fish.

As shown in Figure 4, the underwater images were first enhanced to improve contrast and visibility, and stereo rectification was then performed using the intrinsic and extrinsic camera parameters so that the left and right images satisfied epipolar geometry. This process constrains the correspondence search to the horizontal direction, thereby improving matching stability, reducing errors, and providing a consistent geometric basis for subsequent depth estimation.

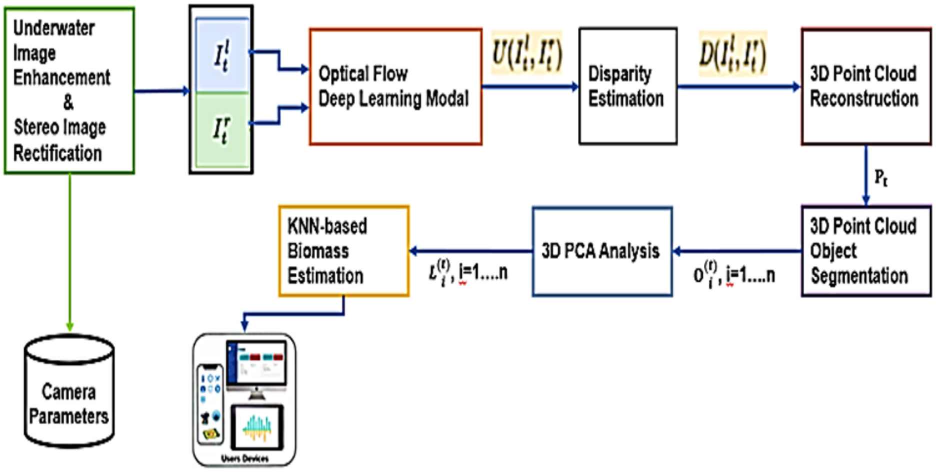

Figure 4: Processing pipeline of the proposed system, including image enhancement, stereo-based depth estimation, 3D point cloud reconstruction, fish segmentation, and KNN-based biomass estimation. In real aquaculture environments, feature inconsistency caused by occlusion and noise may lead to mismatched nearest neighbors in KNN-based estimation.

Disparity estimation was then performed on the rectified images to obtain depth and motion information. The displacement between the left and right images was converted into disparity and combined with the camera parameters to estimate depth, thereby generating a scene depth map. Meanwhile, temporal information describes fish movement and supports subsequent 3D analysis. Compared with conventional methods, this procedure is more robust under low-contrast and noisy conditions.

Based on the estimated depth information, the images were transformed into 3D point clouds and further combined with color information to form a 6D (RGB-XYZ) representation. On this basis, a 3D point cloud Transformer was adopted for fish segmentation, enabling geometric and appearance features to be learned directly in 3D space to handle fish overlap and background interference. Compared with 2D methods, the 3D representation preserves scale and shape information and thus improves segmentation stability.

Geometric features were then extracted from the segmented fish point clouds, and k-nearest neighbors (KNN) regression was applied for weight estimation. This method estimates biomass based on feature similarity and remains stable and interpretable in scenarios with limited data. The final results were written back to the digital twin model to continuously update fish growth status.

To clearly present each processing module of the proposed system and its corresponding technical basis, the related methods and references are summarized in Table 1.

Table 1: Summary of the proposed system pipeline, including data processing modules, their functional roles, and corresponding technical references.

Data Processing Module	Method Description	References
Underwater Image Enhancement	Enhances underwater images by compensating for color distortion and contrast degradation caused by light absorption and scattering, improving feature visibility for subsequent processing.	Ancuti et al., 2018
Stereo Image Rectification	Aligns stereo image pairs using camera intrinsic and extrinsic parameters to enforce epipolar geometry, reducing correspondence search to a one-dimensional horizontal problem.	Scharstein & Szeliski, 2002
Optical Flow Deep Learning Model	Estimates pixel-wise motion between stereo and temporal image pairs using a learning-based model, providing robust correspondence under low-texture and noisy conditions.	Ilg et al., 2017
3D Point Cloud Reconstruction	Converts depth information into 3D spatial coordinates using camera geometry, generating structured point cloud representations of the scene.	Newcombe et al., 2011
3D Point Cloud Object Segmentation	Segments individual fish directly in point cloud space by learning geometric and appearance features using a 3D deep learning model.	Guo et al., 2020; Khan et al., 2022
3D PCA Analysis	Extracts geometric descriptors such as principal axes and shape distribution from point clouds to represent fish body structure.	Abdi & Williams, 2010
KNN-based Biomass Estimation	Estimates fish biomass by mapping geometric features to weight using nearest-neighbor regression based on feature similarity.	Cover & Hart, 1967
Digital Twin Integration	Integrates sensing data and analysis results into a virtual model to continuously update fish growth status and support monitoring and decision-making.	Tao & Zhang, 2017; Ubina et al., 2023

To validate the proposed digital twin architecture for non-invasive fish measurement, underwater stereo image data were collected from the Hengchun offshore cage farm in Pingtung and the A13 pond at National Taiwan Ocean University. A summary of the video dataset is provided in Table 2.

Table 2: Summary of the experimental video dataset collected from offshore cage and pond environments for system validation.

Video ID	Duration
Hengchun Offshore Cage Farm	
Hengchun-Cage-1	23 min 34 s
Hengchun-Cage-2	11 min 47 s
Hengchun-Cage-3	23 min 34 s
Hengchun-Cage-4	30 min 46 s
A13 Pond	
A13-Pond-1	43 min 37 s
A13-Pond-2	38 min 28 s
A13-Pond-3	24 min 59 s
A13-Pond-4	37 min 59 s

Among the two sites, the Hengchun offshore cage farm served as the primary validation environment, while the A13 pond was used as an auxiliary dataset to evaluate preprocessing and segmentation performance under different conditions. In addition, paired manual measurements were collected at the Hengchun site and used as ground truth for evaluating fish size, weight, and biomass estimation.

The experimental results show that image enhancement improves fish contour visibility and segmentation stability, thereby improving the quality of subsequent 3D point cloud reconstruction, as illustrated in Figures 5 and 6.

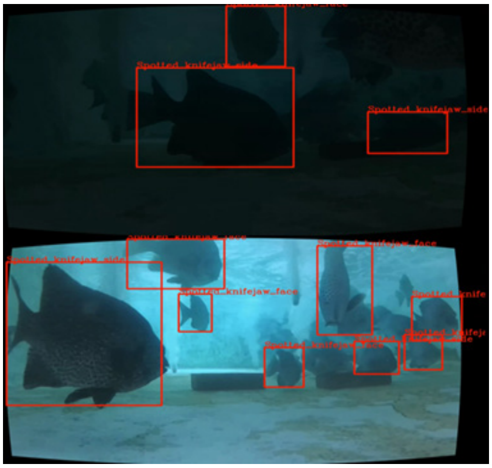

Figure 5: Comparison of fish segmentation results before and after image enhancement, showing improved boundary clarity and more stable separation between fish and background.

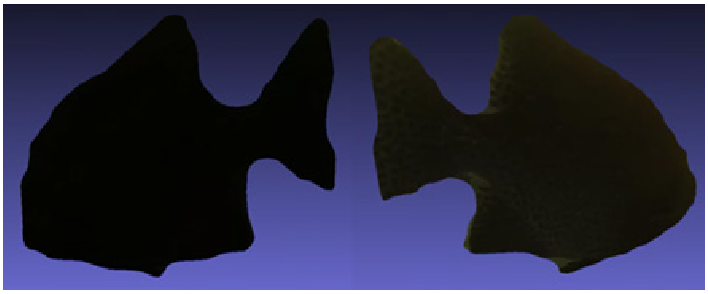

Figure 6: Comparison of reconstructed 3D point clouds before and after image enhancement, demonstrating improved geometric representation for subsequent analysis.

Table 3 presents fish body length estimation results for golden pompano collected from the Hengchun offshore cage farm. When the actual body length was 34.94 cm, the estimated length from the blurred image was 30.75 cm, corresponding to an error of 11.9%. After image enhancement, the estimated length increased to 31.8 cm, and the error decreased to 8.9%. In addition, the distribution of predicted body length values shows a generally consistent trend with the ground truth, indicating that the extracted 3D geometric features effectively capture fish body size information.

Table 3: Body length estimation results before and after image enhancement, illustrating its impact on measurement accuracy.

Image Condition	Actual Body Length (cm)	Estimated Body Length (cm)	Error (%)
Blurred Image	34.94	30.75	11.9
Enhanced Image	34.94	31.8	8.9

For weight estimation, Table 4 shows that the performance of KNN regression is strongly influenced by the choice of K. For a representative sample with an actual body weight of 1209.1 g, the estimated weight reached 1209.1 g when K = 200, with an error of 0.0%. When K = 199 and K = 198, the errors were 0.3% and 0.6%, respectively, indicating stable estimation performance at larger K values.

In contrast, when small K values were used, significant estimation errors were observed. For example, when K = 1, the estimated body weight was 293.11 g, resulting in an error of 75.8%. This behavior can be attributed to feature inconsistency caused by occlusion, incomplete 3D reconstruction, and data imbalance in real aquaculture environments. Under such conditions, the nearest neighbor selected by KNN may not correspond to fish of similar size, leading to mismatched estimation results.

Table 4: KNN-based body weight estimation results under different K values. The results show that small K values may lead to mismatched nearest neighbors due to feature inconsistency, while larger K values provide more stable estimation performance.

K Value	Actual Body Weight (g)	Estimated Body Weight (g)	Error (%)
K = 1	1209.1	293.11	75.8
K = 2	1209.1	299.14	75.3
K = 3	1209.1	302.39	75.0
…	…	…	…
K = 198	1209.1	1200.99	0.6
K = 199	1209.1	1204.5	0.3
K = 200	1209.1	1209.1	0.0

Overall, the combination of 3D geometric features and KNN regression enables the transformation of fish body shape information into weight estimation results and supports biomass analysis. Despite the sensitivity of KNN to parameter selection, stable performance can be achieved when appropriate K values are chosen.

In summary, the proposed pipeline successfully integrates fish extraction, 3D geometric analysis, and biomass estimation from underwater stereo imagery. The results demonstrate its feasibility as a non-invasive measurement method in real aquaculture environments. Furthermore, the estimated fish size, weight, and biomass information can be continuously fed back to the digital twin system, enabling synchronized mapping between the physical farm and the virtual model for fish growth monitoring and management.

CONCLUSION

This study presents a digital twin–enabled stereo camera framework for non-invasive fish biomass estimation in aquaculture. By integrating underwater image enhancement, stereo-based depth estimation, 3D point cloud reconstruction, fish object segmentation, geometric feature extraction, and KNN-based biomass estimation, the proposed system establishes a complete processing pipeline for deriving fish size, weight, and biomass information from underwater stereo imagery without disturbing normal fish behavior.

Experimental results obtained from real aquaculture environments demonstrate the feasibility of the proposed framework for practical fish measurement and biomass analysis. The results show that image enhancement improves segmentation quality and 3D reconstruction accuracy, while the combination of 3D geometric features and KNN regression enables effective estimation of fish body size and weight.

The main contribution of this work lies in integrating these processing modules into a unified digital twin architecture. Rather than providing isolated measurement results, the proposed framework enables fish size, weight, and biomass information to be continuously mapped from the physical environment to the virtual system. This allows fish growth status to be monitored in a continuous and non-invasive manner.

In addition, the use of 3D point cloud representation preserves geometric characteristics of fish bodies, while the adopted 3D point cloud Transformer maintains segmentation performance with relatively low computational complexity. This design supports deployment on edge-computing platforms with limited resources, making the system applicable to real-world aquaculture environments.

For aquaculture applications, the proposed framework reduces reliance on manual sampling and contact-based measurements, minimizes disturbance to fish schools, and improves the continuity of growth monitoring. Overall, this study demonstrates that underwater stereo vision, 3D geometric analysis, biomass estimation, and digital twin concepts can be effectively integrated into a practical system for fish growth monitoring, with potential to support data-driven aquaculture management.

ACKNOWLEDGMENT

This work was supported in part by the National Science and Technology Council, Taiwan, under grant numbers NSTC 114-2221-E-019-022-MY3 and NSTC 114-2637-8-011-004-.

REFERENCES

Abdi, H. and Williams, L.J. (2010) 'Principal component analysis', *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), pp. 433–459.

Ancuti, C.O., Ancuti, C., De Vleeschouwer, C. and Bekaert, P. (2018) 'Color balance and fusion for underwater image enhancement', *IEEE Transactions on Image Processing*, 27(1), pp. 379–393.

Cover, T. M. and Hart, P. E. (1967) 'Nearest neighbor pattern classification', *IEEE Transactions on Information Theory*, 13(1), pp. 21–27.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L. and Bennamoun, M. (2021) 'Deep learning for 3D point clouds: A survey', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), pp. 4338–4364.

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A. and Brox, T. (2017) 'FlowNet 2.0: Evolution of optical flow estimation with deep networks', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2462–2470.

Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S. and Shah, M. (2022) 'Transformers in vision: A survey', *ACM Computing Surveys*, 54(10s), Article 200, pp. 1–41.

Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohli, P., Shotton, J., Hodges, S. and Fitzgibbon, A. (2011) 'KinectFusion: Real-time dense surface mapping and tracking', in *2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 127–136.

Scharstein, D. and Szeliski, R. (2002) 'A taxonomy and evaluation of dense two-frame stereo correspondence algorithms', *International Journal of Computer Vision*, 47(1–3), pp. 7–42.

Tao, F. and Zhang, M. (2017) 'Digital twin shop-floor: A new shop-floor paradigm towards smart manufacturing', *IEEE Access*, 5, pp. 20418–20427.

Ubina, N.A., Lan, H.-Y., Cheng, S.-C., Chang, C.-C., Lin, S.-S., Zhang, K.-X., Lu, H.-Y., Cheng, C.-Y. and Hsieh, Y.-Z. (2023) 'Digital twin-based intelligent fish farming with Artificial Intelligence Internet of Things (AIoT)', *Smart Agricultural Technology*, 5, 100285.