

## Article

# Binocular Camera-Based Depth Recognition for Motion Monitoring and Response Analysis of Flexible Floating Structures for Offshore Photovoltaics

Jijian Lian <sup>1,2</sup>, Shuo Yang <sup>1</sup>, Lirui Yan <sup>1</sup>, Xiuwei Yang <sup>1,\*</sup>, Ye Yao <sup>1,\*</sup> and Chao Liang <sup>2</sup>

<sup>1</sup> School of Ocean Energy, Tianjin University of Technology, Tianjin 300384, China; jjlian@tju.edu.cn (J.L.); shuoyang\_tjut@163.com (S.Y.); liruiyan\_tjut@163.com (L.Y.)

<sup>2</sup> State Key Laboratory of Hydraulic Engineering Intelligent Construction and Operation, Tianjin University, Tianjin 300350, China; chaos@tju.edu.cn (C.L.)

\* Corresponding author. E-mail: yangxw\_tut@163.com (X.Y.); yaoye-111@163.com (Y.Y.)

Received: 24 October 2025; Revised: 10 November 2025; Accepted: 19 November 2025; Available online: 28 November 2025

**ABSTRACT:** Driven by the global goal of carbon neutrality, offshore floating photovoltaic (OFPV) technology has become a primary focus of photovoltaic research. In particular, flexible thin-film structures have become a central focus of research in sustainable energy development. It offers numerous advantages, including light weight, low cost, and strong adaptability to the marine environment. However, traditional experimental methods still face challenges in accurately capturing the motion response of flexible thin films. To address this issue, this study proposes a motion measurement and monitoring framework based on binocular vision. The framework is validated using gyroscope data, and the results demonstrate its high accuracy and real-time performance. The research team conducted experiments on a flexible floating photovoltaic structure in a wave flume, applying the proposed framework to monitor its motion response under wave excitation. The experimental results show that wave height and wave period have significant effects on the acceleration response of the thin film: higher wave heights lead to notably greater accelerations, whereas longer wave periods result in a gradual decrease in acceleration. Overall, the proposed framework provides reliable technical support for the design optimization and safety assessment of flexible thin-film FPV structures.

**Keywords:** Flexible floating photovoltaic structures; YOLOv8; Binocular depth camera; Motion response



© 2025 The authors. This is an open access article under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the continuous growth of global electricity demand and the extensive consumption of fossil fuels, and as the challenges posed by climate change and the increasing frequency of extreme weather events continue to intensify, the development of clean and renewable energy has become an urgent priority [1–3]. However, the rigid land requirements of ground-based photovoltaic (PV) projects have become a major bottleneck restricting their large-scale deployment [4]. As a highly promising solution, offshore floating photovoltaic (OFPV) power plants deploy PV modules on the water surface using floating structures, not only conserving valuable land resources but also making full use of the abundant solar energy available at sea [5]. In recent years, offshore floating photovoltaic (OFPV) systems have gradually become a research focus in both academia and industry [6]. Companies such as Ocean Sun (Norway), Oceans of Energy (The Netherlands), as well as China Resources Power, State Power Investment Corporation, and China National Offshore Oil Corporation have established demonstration projects in locations including the Philippines, the North Sea, Dong ying, the Bohai Sea in Tianjin, and the Shandong Peninsula [7,8].

Currently, offshore photovoltaic (PV) technology is mainly classified into two categories: fixed and floating systems [9]. Among them, floating systems offer significant advantages: they do not require pile foundations, substantially reducing steel consumption and construction costs, while also avoiding disturbances to nearshore ecosystems [10]. In addition, they can be deployed in areas with water depths of 5 m or more and withstand wave forces, thereby increasing installed capacity per unit area by over 20% [11]. Moreover, the natural cooling effect of seawater

can lower module temperatures by 5–8 °C, thereby increasing photovoltaic conversion efficiency by 12–13% [12–14]. The mobility and upgradability of floating photovoltaic systems can also extend their service life, providing a “zero ecological impact” solution that combines technological, economic, and environmental benefits. At the same time, by reducing excessive photosynthesis and algal growth, they contribute to improving water quality [15–17]. Nonetheless, floating systems still face numerous challenges in acquiring motion response data and ensuring structural safety. Therefore, research on accurate motion measurement and tracking of floating bodies holds significant engineering and scientific value.

Measurement and tracking methods can generally be classified into two categories: contact and non-contact approaches. Traditional contact-based methods rely on rulers, potentiometers, and gyroscopes to manually record the motion states of floating structures [18,19]. Such methods are not only labor-intensive and lack real-time capability, but also often fail to meet the precision requirements of experiments. In recent years, sensor systems composed of accelerometers, gyroscopes, and orientation sensors have been developed to measure instantaneous acceleration and angular position at a given point, with displacement estimated through integration of the acceleration data [20–22]. However, such systems are prone to drift errors during long-term monitoring, which can significantly compromise data accuracy. Moreover, offshore floating photovoltaic (OFPV) systems are typically composed of a ring frame and an internal thin film, and the weight of contact sensors can interfere with the natural motion response of the film, thereby limiting measurement accuracy.

Non-contact visual recognition technologies have attracted increasing attention due to their high accuracy, low cost, ease of deployment, and long-distance monitoring capabilities [23]. Visual tracking methods are generally classified into two types: monocular and binocular approaches. Although monocular systems can capture two-dimensional features using a single camera, they cannot directly obtain depth information and are prone to cumulative estimation errors [24]. In contrast, binocular vision mimics the human stereoscopic perception mechanism and can achieve real-time 3D scene reconstruction without prior size information, offering both compact structure and high accuracy [25,26]. However, conventional binocular algorithms require substantial computational resources for feature matching and disparity calculation. In marine environments, strong reflections from the water surface and weak texture features can degrade performance, resulting in unstable depth estimation [27,28].

In recent years, binocular vision-based systems for marine and underwater applications have achieved significant progress. For example, binocular stereo vision has been applied to underwater target detection and ranging, measurement of underwater object dimensions, and identification of defects in underwater pile concrete [29–31]. Beyond the field of construction, further studies in agriculture and biological monitoring have also demonstrated its effectiveness in positioning and target recognition [32–34]. In addition, binocular vision has been employed for road pothole detection, robot navigation, and real-time 3D reconstruction for robots [35–37].

These advances indicate that the simultaneous implementation of motion tracking and stereoscopic information acquisition has become one of the core capabilities of modern machine vision technologies. Binocular depth cameras, with their compact structure and mobility, are particularly suitable for dynamic monitoring of floating engineering structures, offering a highly promising solution for motion tracking and measurement of floating bodies in real marine environments [38,39]. The remainder of this paper is organized as follows: Section 2 provides a detailed description of the experimental site, the physical model, the overall observation system, and the corresponding test conditions. Section 3 explains the working principles of the binocular depth camera used in the experiments, including its technical specifications and relevant mathematical formulations. Section 4 presents the validation and comparative evaluation of the proposed monitoring method, and analyzes the effects of different test conditions on the dynamic response of the model. Finally, Section 5 summarizes the main findings and conclusions of this study.

## 2. Experimental Setup

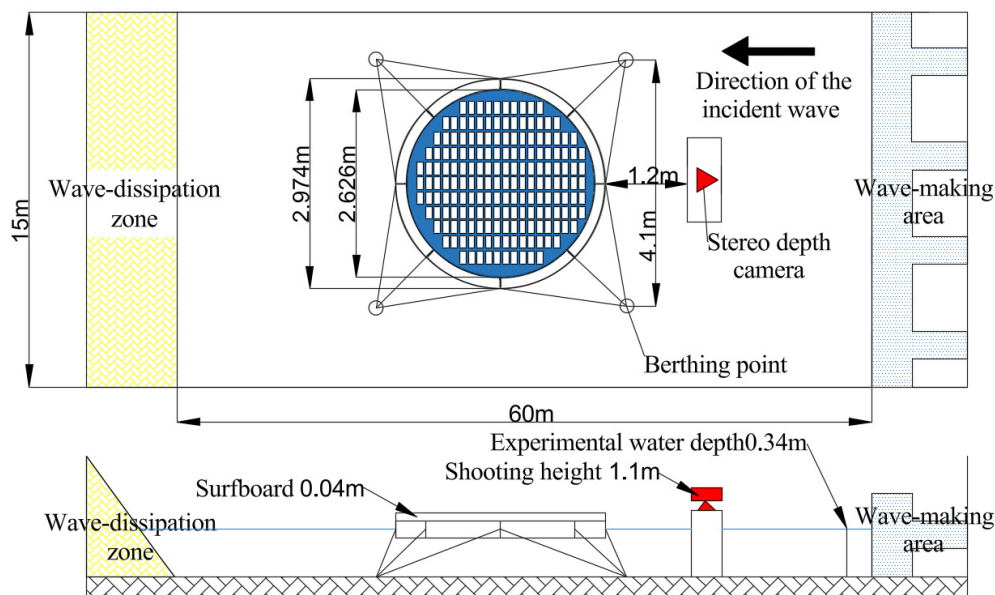
### 2.1. Model Experiment Site

The experiments were conducted in the harbor basin of a key laboratory at Tianjin University. The harbor basin measures 60 m × 15 m × 1.5 m (length × width × depth), with a maximum effective water depth of 0.6 m. Ten wave generators are installed along one side of the basin, of which six were activated in this study. These wave generators can produce both regular and irregular waves, with adjustable wave periods and heights, covering a wave period range from 0.5 s to 5.0 s. To suppress wave reflections, the remaining boundaries of the basin are equipped with breakwaters and wave-absorbing materials.

## 2.2. Experimental Model

The experimental model was fabricated from high-density polyethylene (HDPE) and primarily consists of three components: a breakwater plate, a float box, and side plates. The breakwater plate is arranged in two concentric layers on both sides of the structure, designed to attenuate incident wave forces. The float box serves as the primary buoyant component, while the side plates enable modular connection of smaller components.

The model represents a flexible thin-film floating system, comprising a floating ring, a thin-film structure, and a mooring system. The floating ring is assembled from eight arc-shaped units, each comprising a buoyancy chamber, a breakwater panel, and nine side plates. The buoyancy chamber provides flotation, the breakwater panel reduces wave loads, and the side plates ensure structural connection between adjacent units. The overall structure of the floating body model and the experimental setup are shown in Figure 1.



**Figure 1.** Model Experimental Setup.

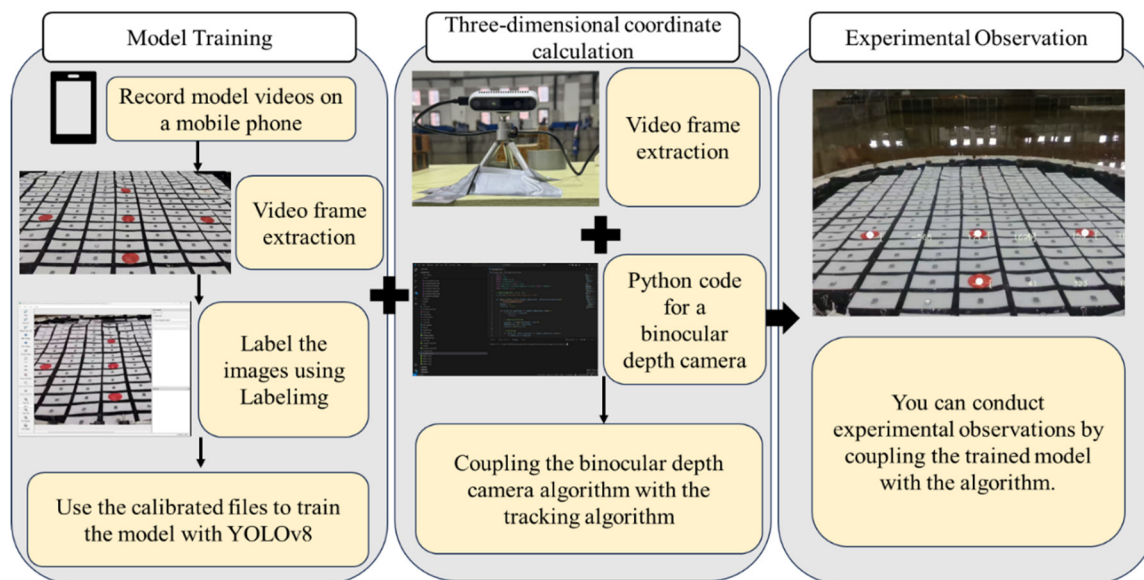
As shown in Figure 1, the experiments were conducted under a water depth of 0.34 m. The single-ring model has an outer diameter of 2.974 m and an inner diameter of 2.626 m, and is equipped with a 0.04 m high wave-guiding plate. The thin-film area is 21 m<sup>2</sup>, and the model was designed according to a geometric similarity ratio of  $\lambda=1:50$ . To satisfy gravitational similarity, a total lead ballast of 2.4386 kg was added to the inner surface of the waterproof membrane, complemented by uniformly distributed weights to simulate the load of the photovoltaic panels. In addition, 32 iron ballast blocks (dimensions: 0.15 m  $\times$  0.025 m  $\times$  0.05 m; mass per block: 1.47 kg) were installed at the outer corners of the float box.

The mooring system consists of four anchors and twelve mooring chains, designed according to the principles of elastic similarity. Linear springs were connected in series along the mooring chains to simulate the linear variation characteristics of mooring stiffness. A total of twelve mooring lugs were installed at eight equally spaced positions along the outer circumference of the ring to connect the mooring chains. Motion tracking was performed using a binocular camera, positioned at a horizontal distance of 1.2 m from the model and mounted at a height of 1.1 m.

## 2.3. Tracking and Measurement Records

The flexible thin-film floating body tracking system, based on a deep learning-driven YOLOv8 algorithm and a stereo depth camera, follows the workflow illustrated in Figure 2.

As shown in Figure 2, the study established a real-time tracking framework by seamlessly integrating the YOLOv8 object detection network with a stereo depth camera system. The overall workflow can be divided into two main stages: model training and optimization, and three-dimensional coordinate reconstruction and tracking.



**Figure 2.** Workflow of the monitoring experiment.

**Model training.** To construct a representative training dataset, video sequences of the experimental environment were first captured with a handheld imaging device under varying lighting and viewing angle conditions. Approximately 1000 frames were uniformly sampled from these sequences to maximize dataset diversity and completeness, ensuring that the trained model possesses strong generalization capability. Each frame was annotated using labeling tools, and the bounding box coordinates and class labels were carefully verified manually to ensure high annotation reliability. After dataset preparation, transfer learning was performed within the YOLOv8 framework, initializing the model with pre-trained weights to accelerate convergence and enhance robustness. During training, convergence was continuously monitored using cross-validation, F1 score, and mAP@0.5 metrics, and the weights achieving the best validation performance were selected as the final detection model for deployment.

For three-dimensional coordinate computation, the stereo image pairs were rectified and aligned in real time using the stereo camera SDK within a Python environment, and subsequently generated dense depth maps. The rectified RGB frames were processed by the trained YOLOv8 model to achieve stable object detection in dynamic scenes. Subsequently, using the camera intrinsic matrix, the centroids of each detected bounding box were back-projected along with their corresponding depth values to reconstruct three-dimensional spatial coordinates with millimeter-level accuracy. To reduce random fluctuations and suppress high-frequency noise, the reconstructed coordinate time series across consecutive frames was processed using a median filter. In addition, the Deep SORT algorithm was employed to assign globally unique tracking IDs, maintaining object identities and ensuring trajectory continuity even under partial occlusion or temporary detection loss.

By tightly coupling the detection and reconstruction stages, the proposed system demonstrates centimeter-level spatial localization accuracy with millisecond-level update rates. This capability not only enables accurate and real-time trajectory monitoring but also provides a high-quality data foundation for advanced tasks such as motion prediction, behavior recognition, and dynamic system analysis.

#### 2.4. Binocular Depth Camera

This study employed an integrated stereo depth camera, which, leveraging advanced stereo vision sensing technology, enables video capture, depth information acquisition, and real-time localization. When the stereo depth camera captures a target object, its two cameras simultaneously record images from different viewpoints. By analyzing the disparity between these two images, the three-dimensional spatial coordinates and distance of the target relative to the camera can be calculated. The stereo depth camera provides high-resolution, high-accuracy depth data over a wide range, making it suitable for various application scenarios, as illustrated in Figure 3.

As shown in Figure 3, the Intel RealSense D325i camera was selected as the core sensing device of the experimental framework, primarily due to its integrated stereo vision and depth perception capabilities. The active infrared stereo depth channel of the camera provides an absolute metric scale for the bounding boxes detected by YOLOv8, thereby enabling precise three-dimensional localization. The D325i camera outputs high-definition dense depth maps with a resolution of  $1280 \times 720$  pixels at a frame rate of 30 frames per second, and an effective operating depth range from

0.16 m to 2.0 m. According to the manufacturer’s specifications and experimental verification, the root mean square (RMS) error of the D325i remains below 2%, ensuring high depth accuracy and frame-to-frame consistency. In addition to the depth data stream, the camera is capable of simultaneously outputting full HD RGB images at a resolution of  $1920 \times 1080$  and a frame rate of 30 frames per second. Although these RGB images are not directly involved in the final 3D coordinate computation, their rich texture information plays a crucial role in enhancing stereo matching performance, particularly in texture-sparse regions.



**Figure 3.** Use of the Intel D325i stereo depth camera in the experiment.

The D325i camera is equipped with an onboard D3 ASIC processor dedicated to real-time depth map computation, effectively offloading intensive processing tasks from the host CPU and GPU to reduce computational load. The RGB-D and MJPEG video streams are transmitted via a USB-C interface with minimal latency, allowing the system’s GPU to fully dedicate its resources to the YOLOv8-based object detection process and achieve optimal performance. In addition, the camera integrates a six-degree-of-freedom (6-DoF) inertial measurement unit (IMU), which, in combination with an external hardware trigger interface, enables timestamp synchronization with microsecond-level precision. This hardware-level timing control establishes a reliable foundation for subsequent spatiotemporal calibration and real-time motion compensation.

In summary, the Intel RealSense D325i camera, with its compact design, hardware-decoupled architecture, and high-precision depth sensing, serves as a stable, scalable data acquisition component within the YOLOv8 + depth 3D perception framework. It is suitable for real-time applications in dynamic, cluttered, or partially occluded environments.

2.5. Experimental Conditions

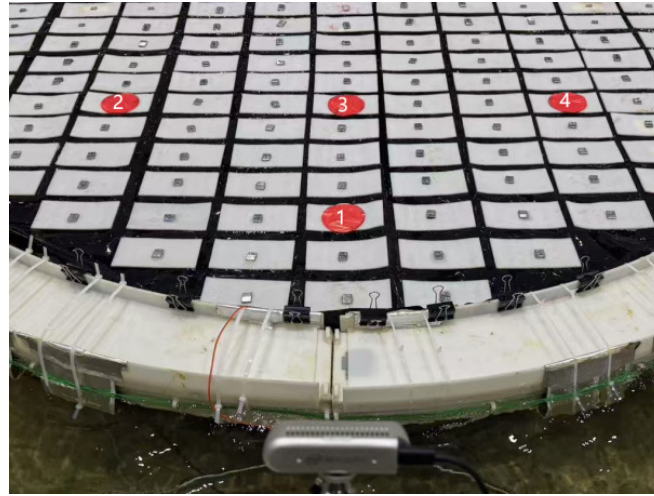
The experimental water depth was set to 0.34 m, corresponding to a prototype depth of 17 m. To simulate irregular sea conditions, this study employed the JONSWAP spectrum and designed corresponding experimental scenarios: Under a fixed wave period, the acceleration of the floating body varies with different wave heights; conversely, under a fixed wave height, the acceleration changes with different wave periods. In the fixed wave period scenario, the wave period was set to 9.9 s, while the wave height was increased from 3 m to 6 m in 1 m increments. In the fixed wave height scenario, the wave height was maintained at 4 m, and the wave period was set successively to 6 s, 8 s, 9.9 s, and 12 s. The experiments were conducted at a geometric scale of 1:50, and the wave maker input parameters were calculated accordingly. The main experimental conditions are summarized in Table 1. Considering the symmetry of the flexible membrane and its mooring system, the wave incidence direction was fixed at 0°.

**Table 1.** Input Conditions for Wave-Making Machine Experiment (Scale 1:50).

|                       |                                     | Actual | Experiment                               |
|-----------------------|-------------------------------------|--------|------------------------------------------|
| Water Depth           |                                     | 17 m   | 0.34 m                                   |
| Operating Condition 1 | 3 m wave height with a 9.9 s period |        | 0.06 m wave height with a 1.4 s period   |
| Operating Condition 2 | 4 m wave height with a 9.9 s period |        | 0.08 m wave height with a 1.4 s period   |
| Operating Condition 3 | 5 m wave height with a 9.9 s period |        | 0.1 m wave height with a 1.4 s period    |
| Operating Condition 4 | 6 m wave height with a 9.9 s period |        | 0.12 m wave height with a 1.4 s period   |
| Operating Condition 5 | 4 m wave height with a 6 s period   |        | 0.08 m wave height with a 0.848 s period |
| Operating Condition 6 | 4 m wave height with an 8 s period  |        | 0.08 m wave height with a 1.131 s period |
| Operating Condition 7 | 4 m wave height with a 12 s period  |        | 0.08 m wave height with a 1.697 s period |



Four measurement points were arranged on the membrane, and the proposed method was employed to monitor these points. The specific layout of the measurement points is shown in Figure 4, while the experimental conditions are detailed in Table 1.



**Figure 4.** Layout of measurement points on the membrane.

Among the measurement points, points 1 and 3 were used to analyze the characteristics during the forward and backward propagation of waves, while points 2, 3, and 4 were employed to examine whether wave attenuation differences occur along the same horizontal line under the given model conditions. For each test scenario, the measurement sampling interval was 0.0557 s, and approximately 1000 data points were collected over the 80-s test duration.

### 3. Principle of Binocular Depth Camera

Stereo depth cameras have become a core sensing technology in the field of 3D perception. Their operating principle is based on geometric triangulation and can be rigorously derived using the classical pinhole camera model. By combining intrinsic and extrinsic parameters, the projection of 3D spatial points onto 2D image coordinates can be analytically described. Simultaneously, depth information can be inferred by estimating the disparity between stereo image pairs. This chapter presents the theoretical foundations of stereo imaging, including the pinhole projection model, the geometric relationships of stereo triangulation, coordinate transformations, and the back-projection mapping. These formulations establish the mathematical basis for subsequent applications in 3D reconstruction, object localization, and motion tracking under marine and engineering scenarios [39–41].

#### 3.1. Pinhole Camera Model

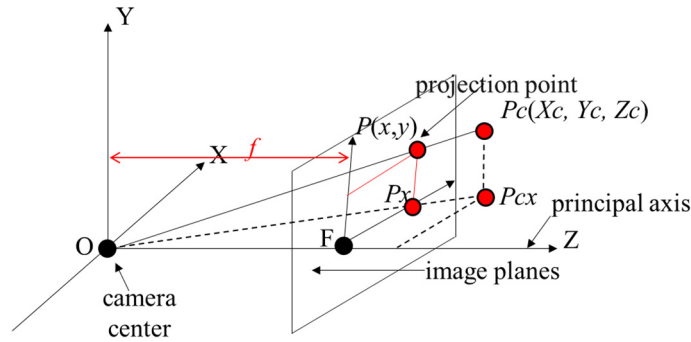
The binocular depth camera operates based on the principle of triangulation, and the classical pinhole camera formulation can rigorously model its imaging process. In this model, the camera is regarded as an ideal perspective projection system, where a spatial point  $P_c(X_c, Y_c, Z_c)$  in the camera coordinate system is projected onto the image plane at  $P(x, y)$ . The projection depends on the intrinsic parameters of the camera (focal length, principal point, pixel scaling factors) and the extrinsic parameters (rotation and translation between the camera and world coordinate systems).

As illustrated in Figure 5, the pinhole camera model, widely adopted in computer vision and photogrammetry, assumes that the camera performs ideal perspective projection, in which three-dimensional points are mapped onto a two-dimensional image plane through an infinitesimally small aperture. The geometric interpretation of the projection is illustrated in Figure 5. The optical center  $O$  is located at the origin of the camera coordinate system, and the principal axis is aligned with the  $Z$ -axis. The image plane is perpendicular to the principal axis and positioned at a distance  $f$ , where  $f$  denotes the focal length. The intersection of the principal axis and the image plane defines the image center  $F$ . Based on the similarity of triangles  $\triangle OPPx$  and  $\triangle OPcPcx$ , the fundamental projection equations can be derived as:

$$\frac{x}{X_c} = \frac{y}{Y_c} = \frac{f}{Z_c} \quad (1)$$

$$x = f \frac{X_c}{Z_c}, \quad y = f \frac{Y_c}{Z_c} \quad (2)$$

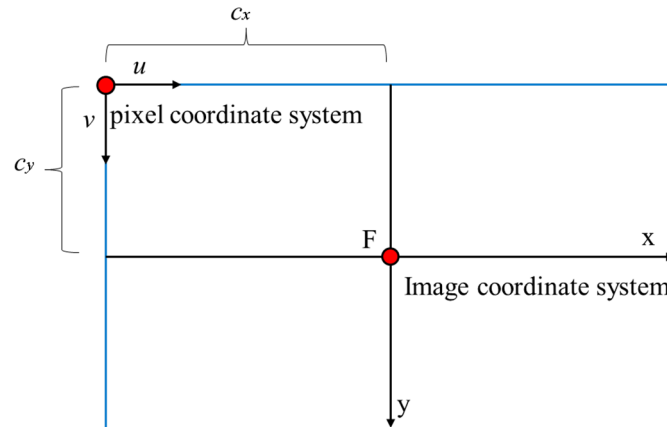
Here,  $(X_c, Y_c, Z_c)$  denote the 3D coordinates of a spatial point  $P$  in the camera coordinate system, and  $(x, y)$  are the corresponding image-plane coordinates.  $f$  represents the focal length of the camera. Equation (2) thus establishes the mapping between 3D spatial coordinates and 2D image coordinates under an ideal perspective projection without lens distortion.



**Figure 5.** Schematic diagram of a simplified pinhole camera model.

### 3.2. Transformation to Pixel Coordinates

To clearly illustrate the geometric relationship between different coordinate systems involved in the imaging process, Figure 6 presents the transformation from the image coordinate system to the pixel coordinate system.



**Figure 6.** Transformation of the image coordinate system to the pixel coordinate system

As illustrated in Figure 6, the image coordinate system  $(x, y)$  is transformed into the pixel coordinate system  $(u, v)$  to match actual pixel measurements. This transformation involves both translation and scaling, and is expressed as:

$$\begin{cases} u = \alpha x + c_x \\ v = \beta y + c_y \end{cases} \quad (3)$$

where  $(c_x, c_y)$  denotes the principal point in the pixel coordinate system (*i.e.*, the image center),  $\alpha$  and  $\beta$  are the horizontal and vertical scaling factors, respectively, which convert physical distances on the image plane into pixel units. Combining Equations (2) and (3), the overall imaging process from a spatial point in the camera coordinate system to the pixel coordinate system can be represented in matrix form as:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha f & 0 & u \\ 0 & \beta f & v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c/Z_c \\ Y_c/Z_c \\ 1 \end{bmatrix} \quad (4)$$

To describe the mapping from the world coordinate system  $(X_w, Y_w, Z_w)$  to the pixel coordinate system  $(u, v)$ , the extrinsic parameters are introduced to represent the transformation between the world and camera coordinate systems. Accordingly, the complete imaging model can be expressed as:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (5)$$

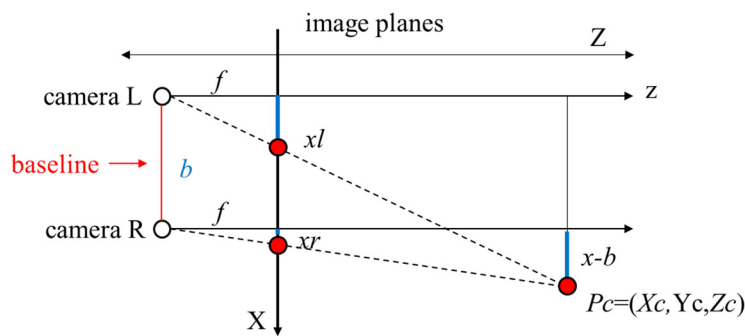
$$K = \begin{bmatrix} \alpha f & 0 & u \\ 0 & \beta f & v \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

where  $s$  is a nonzero scale factor,  $K$  denotes the intrinsic parameter matrix of the camera that incorporates the focal length and pixel scaling factors,  $R \in R^{3 \times 3}$  represents the rotation of the world coordinate system relative to the camera, and  $t \in R^{3 \times 1}$  denotes the translation vector. Equation (5) establishes the complete geometric relationship between the three-dimensional world coordinates and the two-dimensional pixel coordinates, providing the mathematical foundation for binocular vision and three-dimensional reconstruction.

### 3.3. Stereo Triangulation Geometry

Stereo depth cameras employ the principle of triangulation, where disparity information is obtained by matching images captured simultaneously by the left and right cameras, and, combined with intrinsic and extrinsic parameters, back-projection is performed to reconstruct the three-dimensional coordinates of scene pixels.

As illustrated in Figure 7, a typical stereo vision geometry can be described under the assumption of a parallel binocular configuration with perspective projection. Two horizontally aligned cameras (denoted as camera L and camera R) form a baseline system, where the distance between the two optical centers is the baseline length  $b$ , a critical parameter for disparity computation. Each camera has an image plane (depicted as black lines perpendicular to the optical axes), and the distance from the optical center to the image plane corresponds to the focal length  $f$ . A spatial point  $P$  (approximated as  $(x, z)$  in the simplified camera coordinate system) is projected onto the left and right image planes at coordinates  $x_l$  and  $x_r$ , respectively, via optical rays passing through the two camera centers.



**Figure 7.** Geometry model of the binocular stereo vision system in the  $X$ -direction.

For the left camera, the optical center, point  $P$ , and the projection  $x_l$  form similar triangles, yielding the relationship

$$x_l = f \frac{X_w}{Z_w} \quad (7)$$

which can be rearranged to solve for  $z$  (the depth):

$$Z_w = f \cdot \frac{X_w}{x_l} \quad (8)$$

Similarly, for the right camera, point  $P$ , the optical center, and the projection  $x_r$  Satisfy:



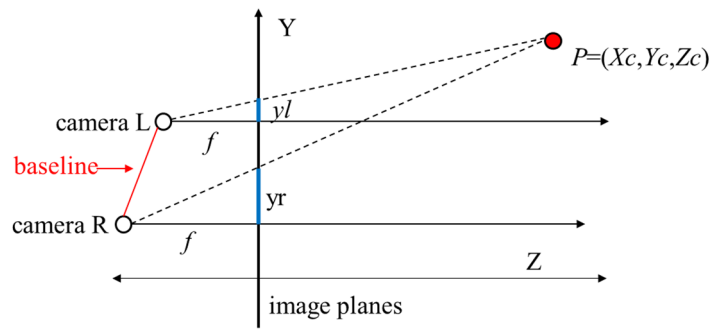
$$x_r = f \frac{X_w - b}{Z_w} \quad (9)$$

Combining Equations (7)–(9) gives the fundamental depth–disparity relationship:

$$Z_w = \frac{f \cdot b}{x_l - x_r}, \quad d = x_l - x_r \quad (10)$$

where  $d = x_l - x_r$  denotes the disparity between corresponding image points in the left and right views. In this formulation,  $f$  represents the focal length of the camera,  $b$  the baseline length (*i.e.*, the distance between the optical centers of the two cameras),  $(x_l, y_l)$  and  $(x_r, y_r)$  the pixel coordinates of the same spatial point projected on the left and right image planes,  $P(X_w, Y_w, Z_w)$  the world coordinates of the corresponding point in three-dimensional space, and  $Z$  the depth of that point from the camera plane. This geometric relationship provides the mathematical foundation for depth recovery in stereo vision systems.

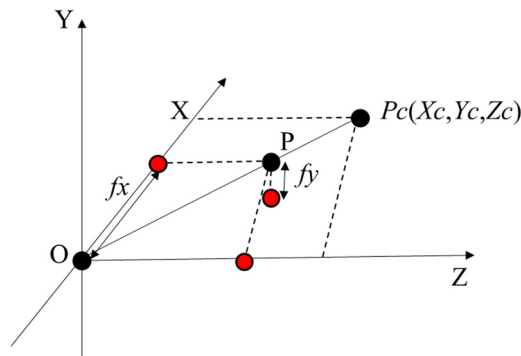
As shown in Figure 8, in a rectified stereo depth camera configuration with row-aligned image planes and negligible lens distortion, the projections of the same 3-D scene point onto the left and right images exhibit identical vertical coordinates, *i.e.*,  $y_l = y_r$ . Consequently, depth estimation relies exclusively on the horizontal; the vertical coordinate does not participate in the triangulation process, and therefore, no additional constraints along the  $Y$ -axis are required in the stereo depth model.



**Figure 8.** Geometry model of the Y-direction binocular stereo vision system.

### 3.4. Inverse Projection Transformation

Once the depth  $Z$  is estimated, the corresponding 3D spatial coordinates  $(X_c, Y_c, Z_c)$  can be reconstructed by applying the inverse projection from Equation (2). As illustrated in Figure 9 the principle of inverse projection describes the transformation from two-dimensional image coordinates to three-dimensional spatial coordinates. The camera center  $O$  is defined as the origin of the camera coordinate system, with the  $Z$ -axis aligned with the optical axis, representing the depth direction. The focal length  $f$  denotes the perpendicular distance between the camera center  $O$  and the image plane, serving as the projection plane scaling factor in the pinhole model. In this framework, a spatial point  $(X_c, Y_c, Z_c)$  is projected onto the image plane at pixel coordinates  $(u, v)$ .



**Figure 9.** Principle of inverse projection transformation.

As illustrated in Figure 9, the principle of inverse projection highlights the transformation from two-dimensional pixel coordinates  $(u, v)$  to three-dimensional spatial coordinates. The camera center  $O$  is defined as the origin of the camera coordinate system, with the  $Z$ -axis aligned with the optical axis, representing the depth direction. The focal length  $f$  denotes the perpendicular distance between the camera center  $O$  and the image plane, serving as the projection scaling factor in the pinhole camera model. In this model, a spatial point  $(X_c, Y_c, Z_c)$  is projected onto the image plane as pixel coordinates  $(u, v)$ . The forward projection can be expressed as:

$$u = \frac{X_c \cdot f_x}{Z_c} + c_x, \quad v = \frac{Y_c \cdot f_y}{Z_c} + c_y \quad (11)$$

Conversely, when the depth value  $Z_c$  is known, the 3D spatial coordinates can be recovered through the inverse projection as:

$$X_c = \frac{(u - c_x) \cdot Z_c}{f_x}, \quad Y_c = \frac{(v - c_y) \cdot Z_c}{f_y}, \quad Z_c = Z_c \quad (12)$$

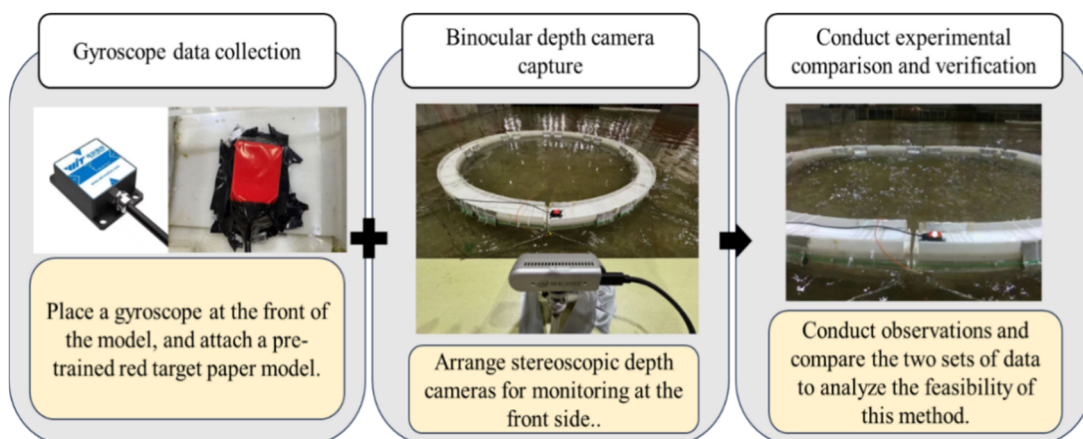
This equation establishes a direct mapping from the image coordinates  $(x, y)$  to the corresponding spatial coordinates  $(X_c, Y_c, Z_c)$ . Here,  $x$  and  $y$  denote the horizontal and vertical image coordinates in physical units (e.g., millimeters or meters), while  $(u, v)$  represent the pixel coordinates on the image plane. The parameters  $f_x$  and  $f_y$  are the focal lengths of the camera, expressed in pixels,  $c_x$  and  $c_y$  represent the principal point coordinates, which define the optical center of the image. The variable  $Z_c$  corresponds to the depth value of the point along the optical axis, which determines the scale of the back-projection from the image plane to the spatial domain. This inverse projection process provides the theoretical foundation for vision-based 3D reconstruction, object localization, and motion tracking, and it plays a crucial role in binocular stereo vision, RGB-D imaging systems, and simultaneous localization and mapping (SLAM) frameworks.

#### 4. Analysis of the Motion Response of Floating Thin-Film Structures

##### 4.1. Verification Test

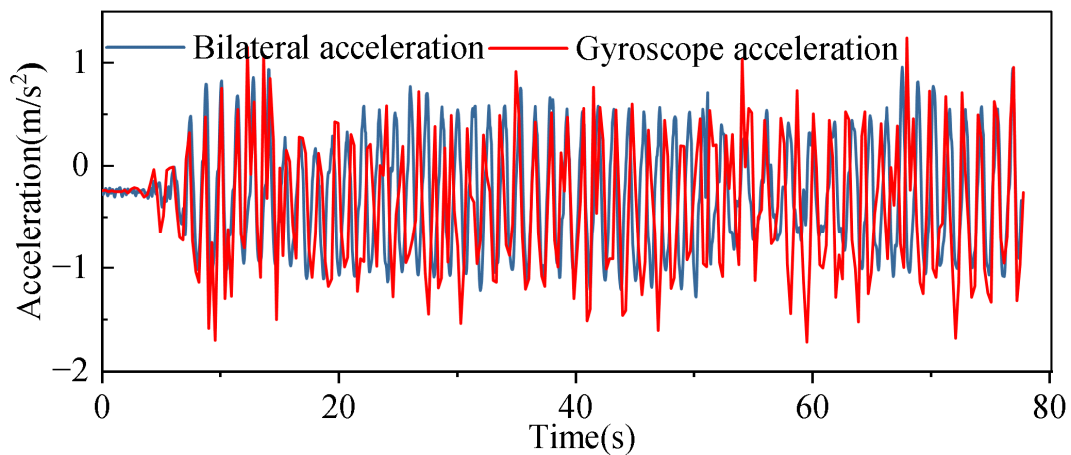
To validate the effectiveness of the proposed flexible membrane floating body tracking system based on the YOLOv8 algorithm and a stereo depth camera, synchronous data acquisition experiments were conducted. The objectives of the experiment were twofold: first, to evaluate whether the system can efficiently and accurately track the target; and second, to verify whether the obtained displacement data can be reliably used for analyzing the membrane's motion response.

As shown in Figure 10, the experimental setup was carefully designed to ensure measurement accuracy and repeatability. A gyroscope was installed at the front section of the structural model as the primary reference sensor. To facilitate stereo vision tracking, a red calibration target was firmly attached to the exposed surface of the gyroscope. To minimize potential optical interference and reflection artifacts, the area surrounding the calibration target was covered with black waterproof tape, thereby reducing background noise and enhancing the accuracy of image-based measurements.

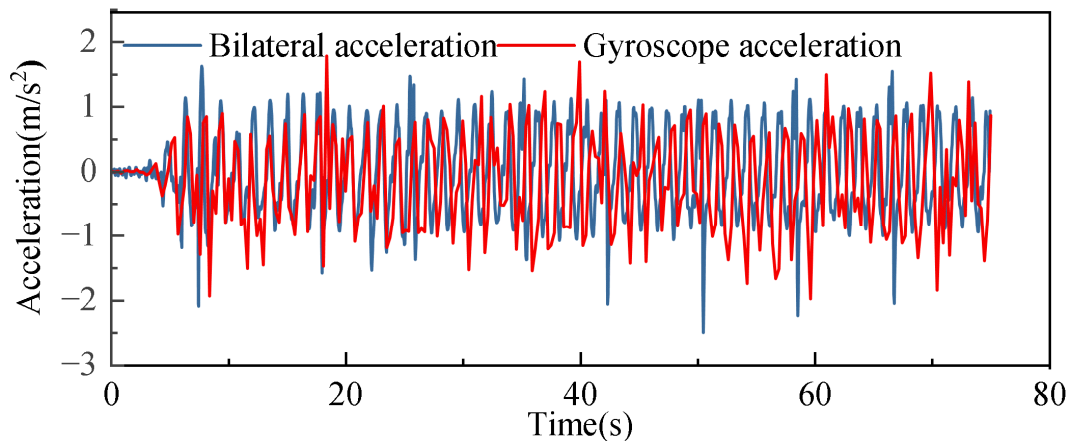


**Figure 10.** Verification Experiment Setup.

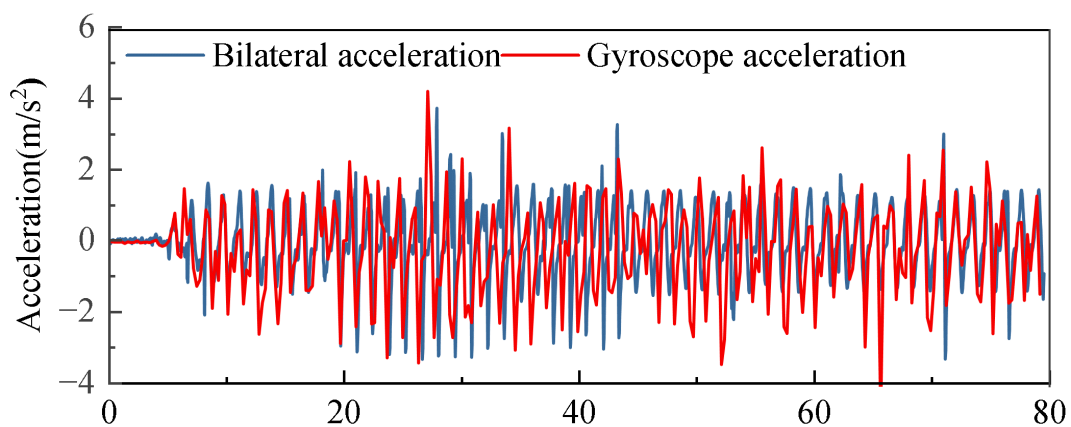
For the purpose of validation, three representative wave conditions were selected, corresponding to incident wave heights of 3 m, 4 m, and 5 m in prototype scale. Different wave conditions were generated by adjusting the wave maker parameters, with input amplitudes set to 0.06 m, 0.08 m, and 0.1 m, while maintaining a constant wave period of 9.9 s to ensure comparability across tests. Under these test conditions, the gyroscope continuously recorded high-resolution acceleration and tilt signals to directly measure the dynamic response of the model. Simultaneously, the stereo camera system captured the 3D displacement of the red calibration target, enabling precise tracking of the model's spatial motion. The combination of these two measurement systems provides a reliable basis for comparative analysis and lays the foundation for validating the stereo vision-based measurement approach. As shown in Figures 11–16, the blue curves represent measurements from the stereo depth camera, while the red curves correspond to data collected by the gyroscope.



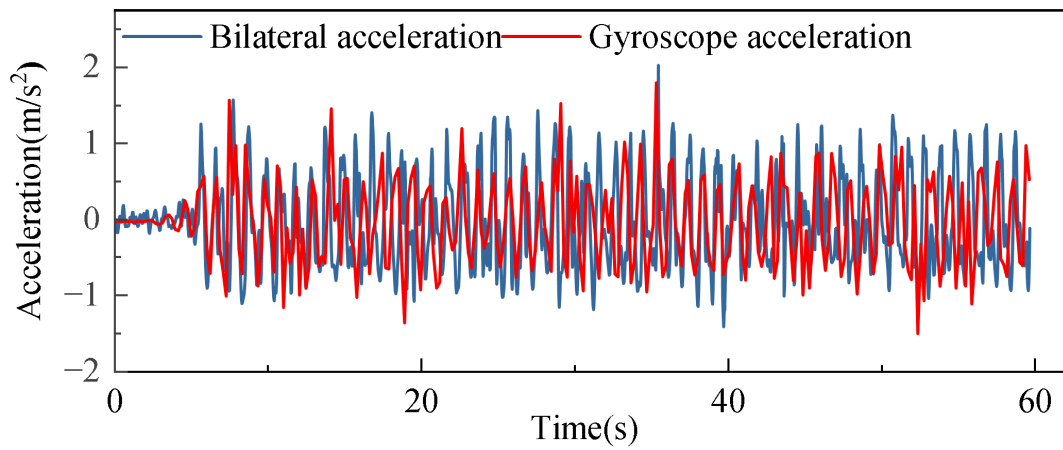
**Figure 11.** Wave height 3 m period 9.9 s comparison in the x direction.



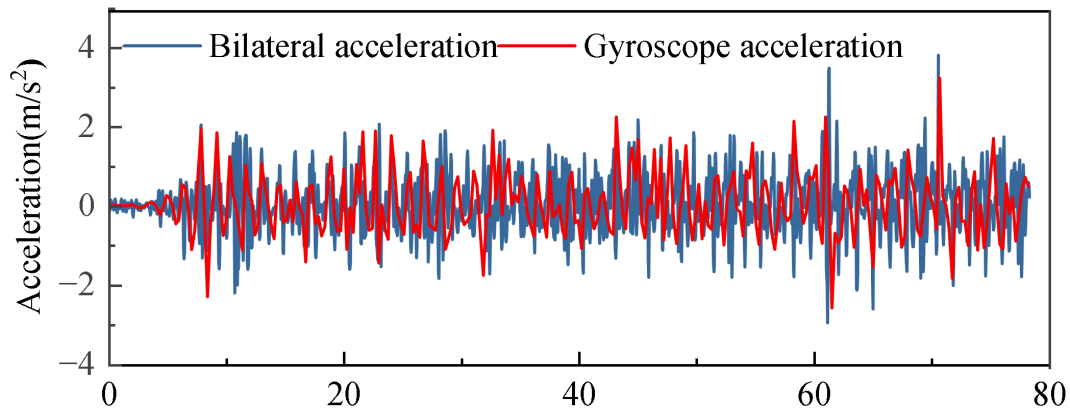
**Figure 12.** Wave height 4 m period 9.9 s comparison in the x direction.



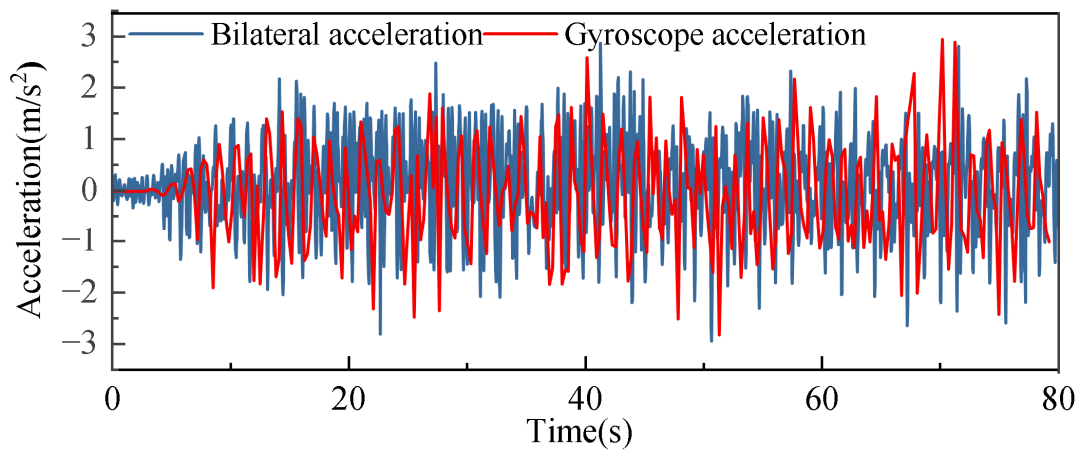
**Figure 13.** Wave height 5 m period 9.9 s comparison in the x direction.



**Figure 14.** Wave height 3 m period 9.9 s comparison in the z direction.

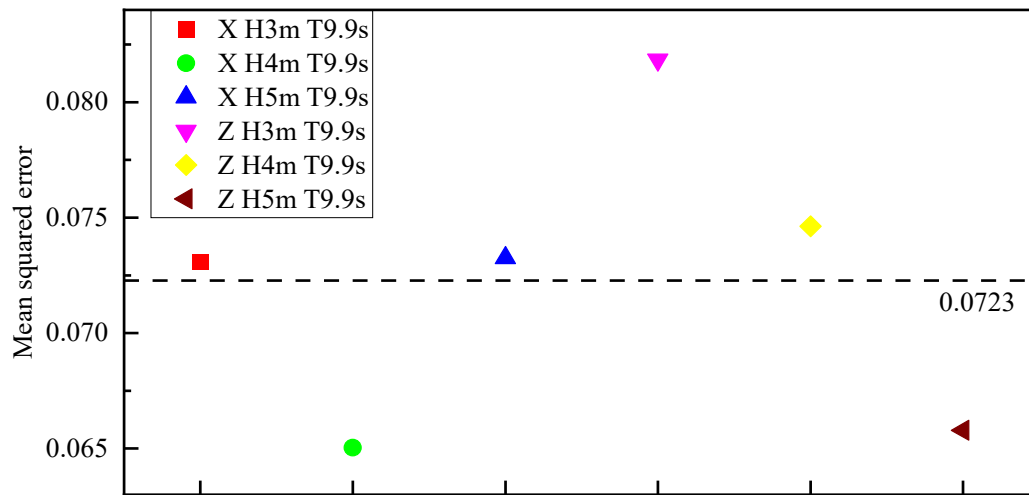


**Figure 15.** Wave height 4 m period 9.9 s comparison in the z direction.



**Figure 16.** Wave height 5 m, period 9.9 s, comparison in the z direction.

The position and timestamp data collected by the stereo camera were first post-processed to derive acceleration, which was then compared with the acceleration data obtained from the gyroscope. The comparison was conducted along the x- and z-directions, as illustrated in Figures 11–16. The results indicate that the acceleration time histories measured by the gyroscope exhibit high consistency with those derived from the stereo camera data in both directions. To quantitatively assess this consistency, Figure 17 presents the mean squared error (MSE) corresponding to the comparisons shown in Figures 11–16, providing a numerical measure of the deviation between the two measurement methods.

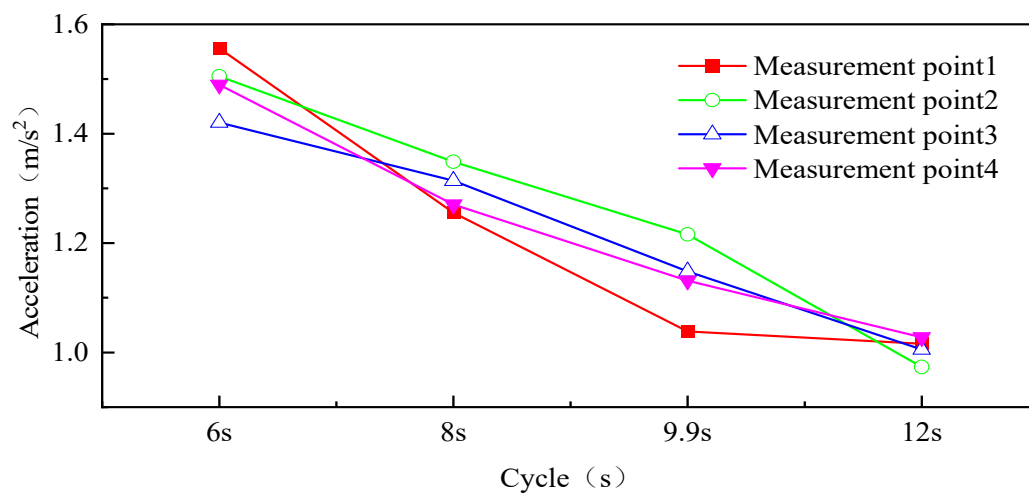


**Figure 17.** Mean square error of data comparison.

As shown in Figure 17, the mean squared error (MSE) values range from 0.065 to 0.082, with an overall average of 0.0723. These results indicate that the numerical deviations between the stereo depth camera and the gyroscope are within an acceptable range, thereby validating the reliability and accuracy of the stereo camera as an alternative measurement method. The observed minor discrepancies may primarily arise from limitations in time synchronization between the two measurement systems, potential noise introduced during the stereo camera image recognition process, and inherent sensor drift of the gyroscope. In addition, slight differences in the spatial positioning of the measurement reference points, as well as variations in the gyroscope's installation height relative to the reference plane, may also contribute to these deviations.

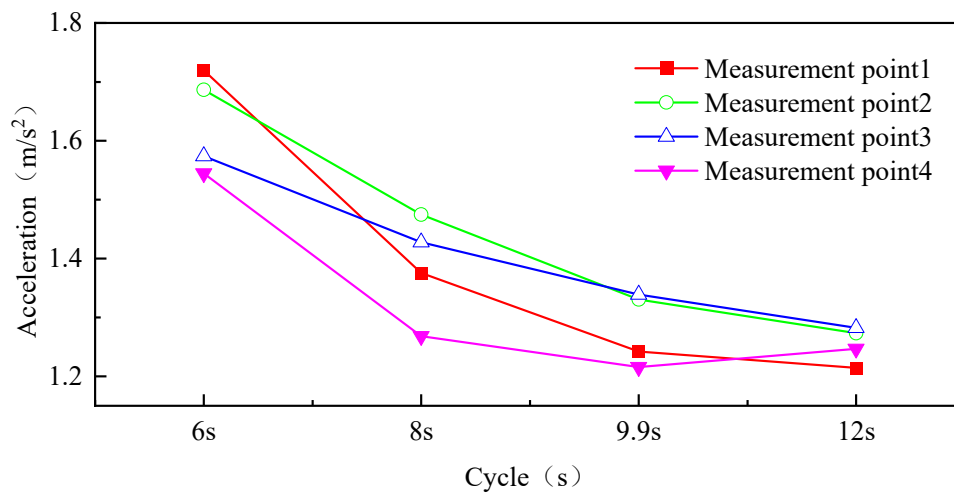
#### 4.2. Analysis of Motion Response Results

As shown in Figure 4, the measurement points were arranged on the structure. Figures 18 and 19 present the acceleration data along the x and z directions for different wave periods (6 s, 8 s, 9.9 s, and 12 s) under the same wave height condition. Data points with different colors and marker shapes represent different measurement locations, and their distributions are shown in Figure 18.



**Figure 18.** Same wave height, different period in the x-direction.



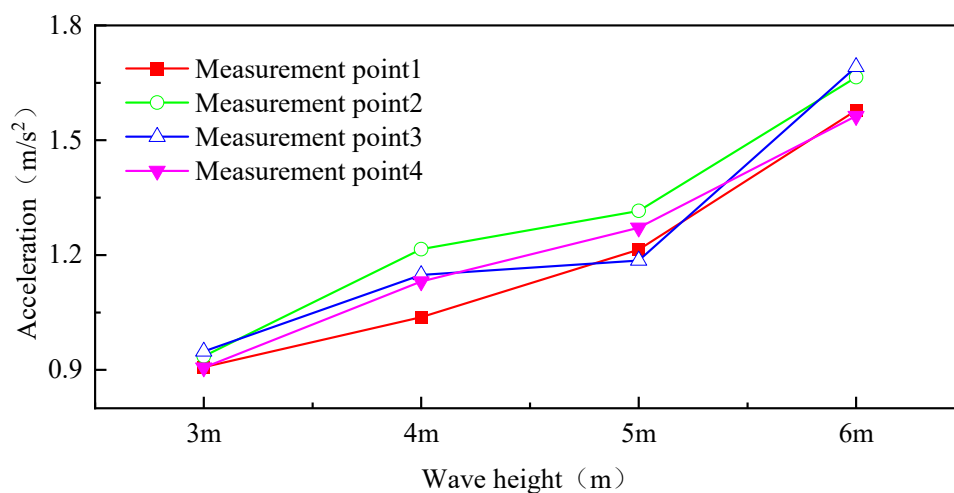


**Figure 19.** Same wave height, different period in the z-direction.

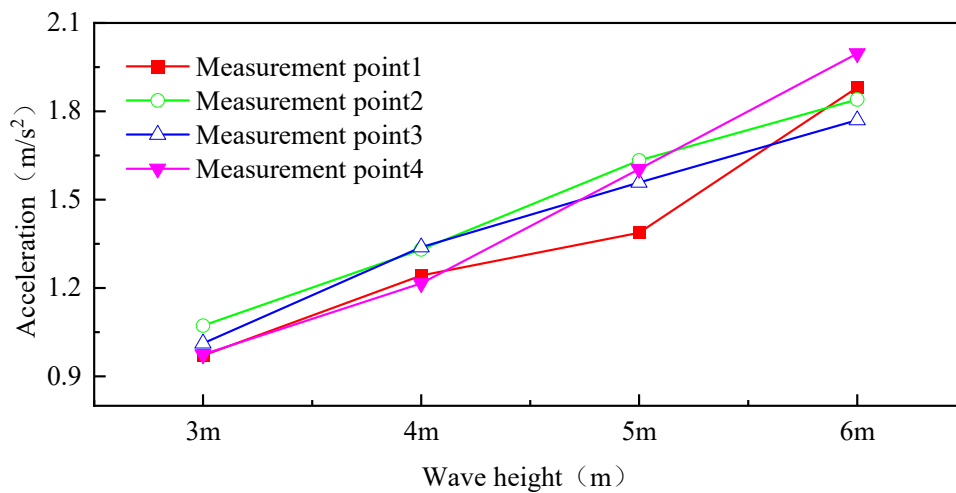
The results shown in Figures 18 and 19 indicate that as the wave period increases, the rate at which the structure's acceleration decreases gradually slows down. Under a constant wave height, as the wave period increases from 6 s to 12 s, the accelerations in both the horizontal (x) and vertical (z) directions exhibit an overall decreasing trend; however, the rate of decrease is not uniform across different intervals. Regarding the horizontal (x) acceleration: when the wave period increases from 6 s to 8 s, the average acceleration decreases by approximately 14.7%; from 8 s to 9.9 s, it decreases by 12.6%; and from 9.9 s to 12 s, the reduction slows further to only 11.7%. The vertical (z) acceleration exhibits a similar but more pronounced trend: a 15.3% decrease from 6–8 s, 7.55% from 8–9.9 s, and a minimal reduction of 2.3% from 9.9–12 s. These results indicate that as the wave period lengthens, the rate of acceleration attenuation gradually diminishes, reflecting a “diminishing marginal effect”.

From a physical perspective, this phenomenon can be explained by the dynamic interaction between wave excitation and structural inertia. As the wave period increases, the corresponding wave frequency decreases, which directly reduces the inertial forces acting on the structure, thereby lowering the acceleration amplitude. However, as the wave period extends further, the energy density of the wave excitation tends to stabilize, resulting in smaller increments of energy input per unit time. This causes the rate of acceleration decay to gradually slow down, ultimately leading to a nonlinear weakening of the structural response.

As shown in Figures 20 and 21, the acceleration responses in the x and z directions are presented for different wave heights (3 m, 4 m, 5 m, and 6 m) under a constant wave period condition. Different colors and marker shapes in the figures represent different measurement points, with their spatial distribution shown in Figure 4.



**Figure 20.** Same period, different wave heights in the x-direction.



**Figure 21.** Same period, different wave heights in the z-direction.

As shown in Figures 20 and 21, under a constant wave period, the acceleration responses of the structure in all directions increase significantly with wave height. This trend is particularly pronounced within the representative wave period range of 6 s to 12 s. Taking the x-direction as an example, when the wave height increases from 3 m to 4 m, the average acceleration rises by approximately 21.22%; from 4 m to 5 m, it further increases by 10%; and from 5 m to 6 m, the increment significantly jumps to 30.4%. The z-direction acceleration exhibits a similar trend, but with slightly higher increments, reaching 27.3%, 20.5%, and 17.43% across the three respective intervals. These results indicate that, compared with wave period, wave height has a more direct and pronounced effect on structural acceleration. However, the increments across different wave height intervals are not uniform, exhibiting a clear stage-wise variation.

Conversely, under a constant wave height, structural acceleration generally decreases as the wave period increases. Statistical analysis indicates that for every 2 s increase in wave period, the average x-direction acceleration decreases by approximately 13%, while the average Z-direction acceleration decreases by about 8.3%. This attenuation pattern aligns with the expected reduction in inertial effects caused by lower wave frequencies, indicating that long-period waves, even at the same amplitude, exert weaker dynamic excitation on the structure.

Overall, under a constant wave period, the acceleration response exhibits an approximately linear relationship with wave height. Specifically, for every 1 m increase in wave height, the average x-direction acceleration increases by 20.54%, while the average z-direction acceleration rises by 21.74%. This nearly linear correlation indicates that wave height is the dominant factor determining the magnitude of structural acceleration, whereas wave period primarily influences the rate of response attenuation.

From the spatial distribution perspective, the acceleration magnitude at measurement point 1 is generally higher than at points 2, 3, and 4, although the differences are relatively small. This indicates that, at the scale of the present study, acceleration decay during wave propagation is not significant, implying a high degree of spatial uniformity in structural acceleration responses and strong spatial correlation of the wave-induced dynamic excitation across the structure.

From an engineering perspective, the key implication of the above findings is that increasing wave height significantly amplifies structural accelerations, thereby intensifying dynamic load effects and imposing stricter requirements on the safety and suitability of offshore structures. Although long-period waves gradually reduce structural accelerations, their attenuation effect is relatively limited, and the long-term impact should not be overlooked. Therefore, in the design and safety assessment of offshore structures, particular attention should be paid to structural responses under high wave heights, while also accounting for the cumulative effects of long-period waves.

## 5. Conclusions

This study proposes a non-contact motion monitoring and response analysis method for flexible floating photovoltaic (FPV) structures, based on binocular vision and deep learning techniques. By integrating a YOLOv8-based object detection algorithm with a binocular depth camera, a real-time monitoring system was developed that enables the extraction of three-dimensional coordinates and the identification of motion parameters. The system can accurately capture the dynamic behavior of flexible FPV structures under wave excitation without altering their hydrodynamic characteristics. The system performance was validated through physical model experiments, and the results were compared with gyroscope measurements, confirming the accuracy and stability of the proposed method. This study

presents a novel experimental approach to investigate the dynamic responses of flexible FPV structures. The main conclusions are summarized as follows:

- (1) The proposed binocular vision monitoring system enables high-precision, non-contact motion tracking of flexible FPV structures. Compared with gyroscope measurements, the system demonstrates excellent consistency, with a root mean square (RMS) error of approximately 0.07. The proposed approach effectively avoids local mass interference associated with traditional contact sensors and offers high accuracy, strong real-time performance, and excellent scalability, providing a new pathway for non-contact dynamic monitoring of flexible structures.
- (2) The results indicate that the motion responses of flexible FPV structures exhibit significant nonlinear characteristics with varying wave parameters. As wave height increases, both vertical displacement and pitch response amplitudes increase accordingly. Resonance amplification occurs when the wave period approaches the natural period of the structure. Under high wave energy conditions, the flexible structure exhibits a degree of self-regulating stability, with response amplitudes tending to stabilize. The temporal and spectral consistency observed across all monitoring points confirms the strong overall coordination of the structure and validates the effectiveness of the flexible design in suppressing excessive responses and enhancing overall stability.
- (3) The binocular vision-based monitoring framework can effectively capture the spatiotemporal evolution of structural motion, providing a novel experimental tool for stability assessment and dynamic analysis of FPV systems. The system offers high accuracy, low cost, and convenient deployment, making it particularly suitable for investigating the dynamic responses and multi-field coupled behaviors of flexible floating structures under varying wave conditions. It thereby provides reliable experimental data for model validation, response prediction, and design optimization of FPV systems.

Although the proposed method performs well in experiments, several limitations remain. The experiments were conducted under controlled laboratory conditions, without fully accounting for environmental factors such as wind, illumination changes, and water surface reflections. Future research will focus on enhancing the robustness and stability of binocular depth cameras in complex wave environments. Algorithmic optimization and lightweight network architectures will be employed to improve depth computation efficiency and coordinate reconstruction accuracy. These efforts aim to enhance the system's reliability in real ocean environments, providing technical support for the safe and long-term operation of FPV systems.

### **Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process**

During the preparation of this work, the authors used ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

### **Author Contributions**

Conceptualization, S.Y.; Methodology, S.Y. and X.Y.; Software, S.Y.; Validation, S.Y. and L.Y.; Formal Analysis, S.Y.; Investigation, S.Y.; Resources, J.L., X.Y., Y.Y. and C.L.; Data Curation, S.Y.; Writing—Original Draft Preparation, S.Y.; Writing—Review & Editing, X.Y.; Visualization, S.Y.; Supervision, J.L., Y.Y. and C.L.; Project Administration, J.L., X.Y. and Y.Y.; Funding Acquisition, J.L., X.Y. and Y.Y.

### **Ethics Statement**

Not applicable. This study did not involve humans or animals.

### **Informed Consent Statement**

Not applicable. This study did not involve human participants.

### **Data Availability Statement**

The data presented in this study are available on reasonable request from the corresponding author. The data are not publicly available due to their large size and subsequent ongoing research.

## Funding

This work was supported by the National Key R&D Program of China under Grant Number 2022YFB4200701, and National Natural Science Foundation of China under Grant Number 52571308, and Natural Science Foundation of Tianjin under Grant Number 24JCYBJC00870.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Rahman T, Hossain Lipu MS, Alom Shovon MM, Alsaduni I, Karim TF, Ansari S. Unveiling the impacts of climate change on the resilience of renewable energy and power systems: Factors, technological advancements, policies, challenges, and solutions. *J. Clean. Prod.* **2025**, *493*, 144933. doi:10.1016/j.jclepro.2025.144933.
2. Wang Z, Liu X, Gu W, Cheng L, Wang H, Liu J, et al. Power-to-Hydrogen-to-Power as a pathway for wind and solar renewable energy utilization in China: Opportunities and challenges. *Renew. Sustain. Energy Rev.* **2026**, *226*, 116195. doi:10.1016/j.rser.2025.116195.
3. Sahu A, Yadav N, Sudhakar K. Floating photovoltaic power plant: A review. *Renew. Sustain. Energy Rev.* **2016**, *66*, 815–824. doi:10.1016/j.rser.2016.08.051.
4. Firoozi AA, Firoozi AA, Maghami MR. Harnessing photovoltaic innovation: Advancements, challenges, and strategic pathways for sustainable global development. *Energy Convers. Manag.* **2025**, *27*, 101058. doi:10.1016/j.ecmx.2025.101058.
5. Bai B, Xiong S, Ma X, Liao X. Assessment of floating solar photovoltaic potential in China. *Renew. Energy* **2024**, *220*, 119572. doi:10.1016/j.renene.2023.119572.
6. Ji Q, Liang R, Yang S, Tang Q, Wang Y, Li K, et al. Potential assessment of floating photovoltaic solar power in China and its environmental effect. *Clean Technol. Environ. Policy* **2023**, *25*, 2263–2285. doi:10.1007/s10098-023-02503-5.
7. Xiong L, Le C, Zhang P, Ding H. Hydrodynamic characteristics of floating photovoltaic systems based on membrane structures in maritime environment. *Ocean. Eng.* **2025**, *315*, 119827. doi:10.1016/j.oceaneng.2024.119827.
8. Fan S, Ma Z, Liu T, Zheng C, Wang H. Innovations and development trends in offshore floating photovoltaic systems: A comprehensive review. *Energy Rep.* **2025**, *13*, 1950–1958. doi:10.1016/j.egyr.2025.01.053.
9. Claus R, López M. Key issues in the design of floating photovoltaic structures for the marine environment. *Renew. Sustain. Energy Rev.* **2022**, *164*, 112502. doi:10.1016/j.rser.2022.112502.
10. Yan J, Koutnik J, Seidel U, Hübner B. Compressible simulation of rotor-stator interaction in pump-turbines. *IOP Conf. Ser. Earth Environ. Sci.* **2010**, *12*, 012008. doi:10.1088/1755-1315/12/1/012008.
11. Lu W, Lian J, Xie H, Li P, Shao N, Zhang G, et al. Experimental study on hydrodynamic response characteristics of a novel floating box array offshore floating photovoltaic structure. *Renew. Energy* **2026**, *256*, 124387. doi:10.1016/j.renene.2025.124387.
12. Ranjbaran P, Yousefi H, Gharehpetian GB, Astaraei FR. A review on floating photovoltaic (FPV) power generation units. *Renew. Sustain. Energy Rev.* **2019**, *110*, 332–347. doi:10.1016/j.rser.2019.05.015.
13. Golroodbari SZ, van Sark W. Simulation of performance differences between offshore and land-based photovoltaic systems. *Prog. Photovolt. Res. Appl.* **2020**, *28*, 873–886. doi:10.1002/pip.3276.
14. Lian J, Zuo L, Wang X, Yu L. Ambient vibration analysis of diversion pipeline in Mount Changlong Pumped-Storage Power Station. *Appl. Sci.* **2024**, *14*, 2196. doi:10.3390/app14052196.
15. Hooper T, Armstrong A, Vlaswinkel B. Environmental impacts and benefits of marine floating solar. *Sol. Energy* **2021**, *219*, 11–14. doi:10.1016/j.solener.2020.10.010.
16. Essak L, Ghosh A. Floating Photovoltaics: A Review. *Clean Technol.* **2022**, *4*, 752–769. doi:10.3390/cleantechnol4030046.
17. Tina GM, Bontempo Scavo F. Energy performance analysis of tracking floating photovoltaic systems. *Heliyon* **2022**, *8*, e10088. doi:10.1016/j.heliyon.2022.e10088.
18. López M, Soto F, Hernández ZA. Assessment of the potential of floating solar photovoltaic panels in bodies of water in mainland Spain. *J. Clean. Prod.* **2022**, *340*, 130752. doi:10.1016/j.jclepro.2022.130752.
19. Liu J, Huang G, Hyypää J, Li J, Gong X, Jiang X. A survey on location and motion tracking technologies, methodologies and applications in precision sports. *Expert Syst. Appl.* **2023**, *229*, 120492. doi:10.1016/j.eswa.2023.120492.
20. Zivanovic M, Vilella I, Iriarte X, Plaza A, Gainza G, Carlosena A. Main shaft instantaneous azimuth estimation for wind turbines. *Mech. Syst. Signal Process.* **2025**, *228*, 112478. doi:10.1016/j.ymssp.2025.112478.
21. Hashim HA. Advances in UAV avionics systems architecture, classification and integration: A comprehensive review and future perspectives. *Results Eng.* **2025**, *25*, 103786. doi:10.1016/j.rineng.2024.103786.

22. Prikhodko IP, Zotov SA, Trusov AA, Shkel AM. Foucault pendulum on a chip: Rate integrating silicon MEMS gyroscope. *Sens. Actuators A Phys.* **2012**, *177*, 67–78. doi:10.1016/j.sna.2012.01.029.
23. Gu Y, Wu J, Liu C. Error analysis and accuracy evaluation method for coordinate measurement in transformed coordinate system. *Measurement* **2025**, *242*, 115860. doi:10.1016/j.measurement.2024.115860.
24. Harris JM, Wilcox LM. The role of monocularly visible regions in depth and surface perception. *Vis. Res.* **2009**, *49*, 2666–2685. doi:10.1016/j.visres.2009.06.021.
25. Liu H, Shen B, Zhang J, Huang Z, Huang M. A two-stage fast stereo matching algorithm for real-time 3D coordinate computation. *Measurement* **2025**, *247*, 116672. doi:10.1016/j.measurement.2025.116672.
26. Lyu Y, Liu Z, Wang J, Jiang Y, Li Y, Li X, et al. A high-precision binocular 3D reconstruction system based on depth-of-field extension and feature point guidance. *Measurement* **2025**, *248*, 116895. doi:10.1016/j.measurement.2025.116895.
27. Yue Z, Huang L, Lin Y, Lei M. Research on image deformation monitoring algorithm based on binocular vision. *Measurement* **2024**, *228*, 114394. doi:10.1016/j.measurement.2024.114394.
28. Liu Y, Wang Y, Cai X, Hu X. The detection effect of pavement 3D texture morphology using improved binocular reconstruction algorithm with laser line constraint. *Measurement* **2020**, *157*, 107638. doi:10.1016/j.measurement.2020.107638.
29. Zhou Y, Li Q, Ye Q, Yu D, Yu Z, Liu Y. A binocular vision-based underwater object size measurement paradigm: Calibration–Detection–Measurement (C-D-M). *Measurement* **2023**, *216*, 112997. doi:10.1016/j.measurement.2023.112997.
30. Chen Q, Liu H, Gan W. A real-time recognition and distance measurement method for underwater dynamic obstacles based on binocular vision. *Measurement* **2025**, *252*, 117329. doi:10.1016/j.measurement.2025.117329.
31. Xu X, Li Q, Du Z, Rong H, Wu T, Wang S, et al. Recognition of concrete imperfections in underwater pile foundation based on binocular vision and YOLOv8. *KSCE J. Civ. Eng.* **2025**, *29*, 100075. doi:10.1016/j.kscej.2024.100075.
32. Long L, Guo J, Chu H, Wang S, Xu S, Deng L. Binocular vision-based pose monitoring technique for assembly alignment of precast concrete components. *Adv. Eng. Inform.* **2025**, *65*, 103205. doi:10.1016/j.aei.2025.103205.
33. Wen Y, Xue J, Sun H, Song Y, Lv P, Liu S, et al. High-precision target ranging in complex orchard scenes by utilizing semantic segmentation results and binocular vision. *Comput. Electron. Agric.* **2023**, *215*, 108440. doi:10.1016/j.compag.2023.108440.
34. Zhai Z, Zhu Z, Du Y, Song Z, Mao E. Multi-crop-row detection algorithm based on binocular vision. *Biosyst. Eng.* **2016**, *150*, 89–103. doi:10.1016/j.biosystemseng.2016.07.009.
35. Xing C, Zheng G, Zhang Y, Deng H, Li M, Zhang L, et al. A lightweight detection method of pavement potholes based on binocular stereo vision and deep learning. *Constr. Build. Mater.* **2024**, *436*, 136733. doi:10.1016/j.conbuildmat.2024.136733.
36. Li M, Lu R. Target ball localization for industrial robots based on binocular stereo vision. *Ind. Robot. Int. J. Robot. Res. Appl.* **2025**, *52*, 600–609. doi:10.1108/IR-07-2024-0317.
37. Cai G, Zhang H, Bai B. *In situ* 3-dimensional measurement of ice shape based on binocular vision. *Measurement* **2025**, *258*, 119033. doi:10.1016/j.measurement.2025.119033.
38. Han J, Fang T, Liu W, Zhang C, Zhu M, Xu J, et al. Applications of machine vision technology for conveyor belt deviation detection: A review and roadmap. *Eng. Appl. Artif. Intell.* **2025**, *161*, 112312. doi:10.1016/j.engappai.2025.112312.
39. Huang MQ, Ninić J, Zhang QB. BIM, machine learning and computer vision techniques in underground construction: Current status and future perspectives. *Tunn. Undergr. Space Technol.* **2021**, *108*, 103677. doi:10.1016/j.tust.2020.103677.
40. Feng G, Liu Y, Shi W, Miao Y. Binocular camera-based visual localization with optimized key point selection and multi-epi polar constraints. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 102228. doi:10.1016/j.jksuci.2024.102228.
41. Rao Z, Yang F, Jiang M. A calibration method of a large FOV binocular vision system for field measurement. *Opt. Lasers Eng.* **2025**, *186*, 108767. doi:10.1016/j.optlaseng.2024.108767.