Hindawi
Structural Control and Health Monitoring
Volume 2023, Article ID 4752072, 19 pages
https://doi.org/10.1155/2023/4752072

Research Article

Bridge Deformation Measurement Using Unmanned Aerial Dual Camera and Learning-Based Tracking Method

Shang Jiang[1] **Jian Zhang**[1,2] **and Chenhao Gao**[1]

[1] *School of Civil Engineering, Southeast University, Nanjing 211189, China*
[2] *Jiangsu Key Laboratory of Engineering Mechanics, Southeast University, Nanjing 211189, China*

Correspondence should be addressed to Jian Zhang; jian@seu.edu.cn

Received 9 September 2022; Revised 15 November 2022; Accepted 1 April 2023; Published 17 April 2023

Academic Editor: Young-Jin Cha

Bridge deformation response data are the basis for calculating the dynamic parameters of the bridge, and it is of great significance to accurately measure the deformation response of the bridge during the load test and service conditions. A bridge deformation measurement method using an unmanned aerial system (UAS) with dual cameras and a deep learning-based object tracking method is proposed to measure the bridge deformation. The contributions are as follows: (1) To address the problem that the movement of the UAS brings error to the deformation measurement results, dual cameras with telephoto and wide-angle lenses are used to simultaneously capture the deformed points and stable points on the bridge, so as to simultaneously measure the deformation of the bridge and the displacement of the UAS, and then the displacement of UAS is eliminated by using the homography relationship between the two cameras. (2) To solve the problem that the traditional digital image correlation-based displacement measurement method is easily disturbed by factors such as light changes and occlusion, a displacement calculation method based on object detection network and target tracking algorithm is proposed to achieve the stable target displacement measurement. Finally, the proposed method was verified in a laboratory test and applied to the deformation measurement of an in-service bridge to verify the practicability of the proposed method.

1. Introduction

The deformation measurement of bridges under vehicle load and environmental effects (such as wind and temperature) is an important content of bridge safety evaluation. Especially, in the load test of bridge, the accurately measured deformation response of bridge can not only be compared with the bridge deformation of finite element analysis but also calculate the deep-level characteristic parameters of bridge (such as structural frequency response function and modal flexibility) according to the theory of the structural dynamics, and these parameters provide basic data for bridge damage identification [1–4]. Bridge deformation measurement can be divided into long-term monitoring and short-term inspection. Long-term monitoring generally requires the installation of fixed sensors on the bridge to measure the deformation of some key areas and give early warning when the bridge suffers extreme load or accident. This method is costly and generally only applied on a few long-span bridges. The short-term deformation measurement is more commonly used in load tests and periodic safety evaluations during service, so many of the existing bridge deformation measurement methods are short-term methods [5].

Most of the existing bridge deformation measurement methods are based on traditional remote sensing equipment, including static global positioning system (GPS), total station, and liquid level connecting pipe. The measurement frequency of this kind of method is low and generally does not exceed a few hertz, so it is only suitable for static measurement [6]. With the development of instrumentation science in recent years, some cutting-edge sensors have been gradually applied to bridge testing. High-sensitivity acceleration sensors are currently the most commonly used sensors in bridge vibration testing. Research on using the quadratic integration of acceleration data to calculate the bridge deformation has been carried out in recent years.

Hester et al. proposed a method to calculate the bridge displacement of mobile trucks through double integration of bridge acceleration. The results show that the observed measurement accuracy can reach 0.5 mm but also pointed out the limitations that this method is only for the conditions of short duration interval and small amplitude displacement [7]. The linear variable differential transformer (LVDT) sensor is a commonly used displacement measurement sensor. It is applied in some studies to measure the deformation of local areas of the bridge, but LVDT needs to be installed on a fixed base point near the measuring point, so it is unable to measure the deformation of long-span bridges, and the measurement range of LVDT is limited. In addition to these contact-based measurement methods, noncontact-based measurement methods are gradually and widely used in bridge deformation inspection. Gentile and Bernardini tried the application of microwave interferometric radar in the measurement of environmental vibration response of concrete bridges [8]. Zhang et al. proposed a method for measuring the deformation and cable force of long-span bridges using microwave interferometric radar [9], and then applied it to the multipoint displacement measurement of a suspension bridge with a main span of 1200 meters. The results show that the measured displacement is in good agreement with those of the total station. Microwave interferometric radar has the advantages of high precision and long effective distance, but it is difficult to locate the measuring points under long range and the equipment is expensive. Garg et al. applied laser doppler vibrometer (LDV) to measure the bridge displacement and proposed a method to compensate for the measurement error caused by the angle and linear movement of the vibration meter, so as to obtain accurate bridge lateral displacement measurement without calibration and reference [10]. However, the measurement distance using LDV is limited and cannot meet the measurement of large bridges.

With the rapid development of computer vision technology, vision-based measurement methods have been applied to structural deformation or vibration inspection [11–14]. These methods are inexpensive, simple in system composition, and have been proven to achieve the required measurement accuracy in the measurement of some civil engineering structures. Yu et al. proposed a vision-based multipoint displacement measurement method for large bridges. The bridge light is used as the target point to calculate the image scaling coefficient, and the camera model is used to correct the vibration interference of the camera. Finally, the proposed method is verified on a long-span, cable-stayed bridge [15]. Jana and Nagarajaiah proposed a handheld, camera-based cable force measurement method [16]. This method first detects the camera movement using the Kanade–Lucas–Tomasi (KLT) feature tracking algorithm and eliminates the camera motion using the affine transformation matrix obtained by the random sample consistency (RANSAC) algorithm, then estimates the cable motion using a phase-based motion estimation technique, and finally determines the cable force by estimating the real-time frequency changes from the displacement of the cable based on the short-time Fourier transform (STFT). Jeon et al. used Cable-ROI Net and Uni-KLT to measure the structural displacement of cables, and the proposed method automatically selects the ROI using a CNN and tracks the feature points in the cable robustly using a modified Kanade–Lucas–Tomasi (KLT) algorithm. The test results show that the proposed method is expected to improve the current methods for measuring cable vibrations [17]. Xu et al. proposed a bridge deformation measurement method against environmental light interference and tested the proposed method in a short-span pedestrian bridge and a long-span highway bridge [18]. Wu et al. proposed a method of tracking laser fringes to analyze bridge vibration in view of the strong dependence of traditional vision-based methods on targets and lighting conditions. By tracking the subpixel position of the centerline of laser fringes in the image, the bridge displacement time history curve can be obtained [19]. Lee et al. proposed a long-term displacement measurement method based on a self-motion compensation dual-camera, in which the primary camera is used to measure the target displacement and the secondary camera measures the self-motion of the dual-camera system so that the error caused by the motion can be estimated and compensated [20]. The abovementioned research proves that the vision-based method has obvious application prospects in bridge deformation measurement, but it also exposes the following two inherent defects of this method: (1) When using vision-based methods to inspect the deformation of long-span bridges, it is difficult to find a suitable camera placement or angle. Because bridges are generally located above the river, the camera can only be tilted towards the bridge to take images. Besides, when measuring the middle part of the bridge span, the measurement distance for long-span bridges may exceed one kilometer. Under this distance, atmospheric interference and microvibration of the camera itself have a great impact on the results. (2) Vision-based methods are generally based on matching between images and thus calculating the motion of the point to be measured in images of time series, but the quality of the images is susceptible to light interference, so the traditional matching methods based on image processing (IP) are very sensitive to interference such as light and occlusion. Therefore, there is an urgent need for responsive methods to solve the abovementioned problems.

The UAV technology, which has grown more extensively and employed in recent years, is predicted to become a breakthrough point in addressing the problem of long measuring distance when traditional vision-based methods are applied to bridge the measurement. Structure damage inspection has been aided by the use of a UAV system [21–26]. Among them, representative studies include Kang and Cha's suggestion to use UAVs to detect bridge damage, especially at the bottom of bridges, and they adopted ultrasonic beacons to provide localization data for UAVs in places with weak GPS signals such as the bottom of bridges, and then a deep learning-based object detection method was used to identify the damage [27, 28]. Based on this, Ali et al. further investigated the use of modified Faster R-CNN to identify multiple types of bridge damage, thus improving the accuracy of disease detection to 93.31% [29, 30]. When a stationary camera is replaced with a UAV to detect the bridge deformation, the measuring distance is considerably reduced, allowing the displacement measurement site to be chosen freely [26, 27]. However, the UAV will inevitably move in flight, resulting in the instability of the measurement base point, which limits the application of the measurement method based on the UAV. Aiming at the problem of base point motion using UAV, Cha and Kang proposed a bridge model vibration measurement method based on homography. The geometric correction of the image is realized by establishing a plane homography transformation with fixed points, and the actual displacement of the bridge model is obtained [28]. Ali et al. estimated the 6-DOF camera motion by tracking the background feature points, so as to obtain the displacement of the UAV itself and combined with the displacement of the UAV and the relative structural displacement to restore the absolute structural displacement [29]. Cha et al. used template matching based on normalized cross-correlation (NCC) to deduce the cubic translation and cubic rotation of UAV, so as to calculate the displacement of the UAV itself from the fixed point of background [30]. Zhang et al. used a motionless laser spot projected from a distance away as a reference; therefore, the video taken by the UAV directly represents the relative displacement of the bridge and the laser projection point, thus avoiding the need to consider the vibration of the UAV [31]. It can be seen that most of the existing studies use the fixed point of the structural background photographed by the UAV as a reference to calculate the displacement of the UAV itself. However, for long-span bridges, because they are generally located on rivers, it is difficult to find fixed points as a reference in the image.

The traditional vision-based methods are susceptible to interference such as light and occlusion, while object detection based on deep learning generally has good anti-interference ability. Deep learning methods have been widely studied in the field of structural damage detection and proved to be robust under the conditions of illumination change, occlusion, and stain erosion [32–34]. Therefore, studies on integrating deep learning methods into vision-based deformation measurement have been considered to have the potential to solve the abovementioned problem. Xu et al. proposed an accurate deformation measurement method of the structure under illumination interference, which integrates Siamese tracking and correlation-based template matching algorithm to enhance the measurement robustness under environmental changes [18]. Chen et al. developed a structural displacement measurement method based on deep learning enhanced vision, which uses convolutional neural network (CNN) and generative adversarial network (GAN) to enhance the target features in the image, so as to obtain more robust measurement results [33]. Yoon et al. proposed a visual tracking algorithm based on reconstructed efficient convolution operator, which uses the multiresolution depth feature framework to efficiently encode the image information representation, and the result shows that the method achieved high accuracy on no target surface [34]. Bai and Yang proposed a digital image correlation (DIC) method based on deep learning, which realizes the direct end-to-end prediction of displacement and strain by designing two convolutional neural networks, a displacement network and a strain network [35]. Han et al. proposed a real-time dynamic displacement measurement method based on virtual point tracking, which uses deep learning to calibrate the region of interest (ROI) and detect the virtual point of interest. The proposed method is applied to measure the lateral displacement of the wheel on the track [36]. These studies effectively applied various types of methods in deep learning to improve the stability of vision-based measurements, but most of them are not developed for bridge deformation measurement, while deep learning methods have obvious selectivity on the training dataset, so further research on such methods in bridge deformation measurement scenarios is needed.

This research offers a solution based on a UAV with dual cameras to address the aforementioned issues. Through the target tracking method based on deep learning, the UAV displacement and bridge displacement are calculated, respectively, so the absolute deformation of the bridge can be obtained using the homography between two cameras. The main contents include (1) the UAV with a dual-camera system is applied to capture both deformation areas (middle span) and stable areas (pier) of the bridge, respectively. The displacement of the dual camera itself is the same while the field of view is different. Therefore, the displacement information measured by the dual camera can be subtracted through the derived transmission relationship, so the measurement error caused by the displacement of the UAV can be eliminated. (2) A displacement measurement framework based on deep learning-based target tracking is proposed. Firstly, the target on the bridge surface is detected and predicted by the Kalman filter to determine the target ID, and then the target position is accurately extracted by subpixel algorithm. Therefore, the stable and accurate displacement of the bridge target can be obtained.

The rest of the paper is organized as follows. Section 2 presents the framework of the proposed method. Section 3 introduces the UAS with dual-camera system. Section 4 introduces the displacement measurement method based on deep learning-based target detection and tracking. Section 5 presents the suggested approach that is tested on an in-service bridge. Section 6 presents the conclusion.

2. Framework of the Proposed Method

The proposed bridge deformation measurement method is a vision-based noncontact measurement method. Unlike the existing methods using standing cameras, the proposed method is based on dual cameras in UAS, so the deformation position of the bridge can be captured at a close range with a parallel angle. It can overcome the problems of air disturbance interference, large image field of view, and difficulty in finding a suitable measurement position caused by using the standing camera. However, the consequent problem of measurement base point movement is a key problem to be solved by the proposed method. In addition, a deep learning-based tracking method is proposed instead of the traditional image processing-based displacement measurement method to overcome the measurement errors caused by light changes and sudden occlusions. The framework for the proposed bridge deformation measurement method is shown in Figure 1. The suggested method is divided into the following two parts: a UAS motion removal method using dual cameras and a displacement calculation method using deep learning-based target tracking.

Aiming at the elimination of the base point vibration of the UAS, the dual-camera system with a telephoto lens and a wide-angle lens is adopted to capture the deformation point of the bridge body and the stable reference point of the pier at the same time. The relationship between the displacements measured by the dual camera, in this case, is deduced theoretically, so as to eliminate the displacement of the UAS. The accuracy of the vision-based displacement measurement method is directly proportional to the field size of the image; therefore, measuring the bridge displacement using a telescopic camera can assure better measurement accuracy than commonly used wide-angle cameras. However, the field size of the telescopic camera is small, so it is difficult to capture the stable points of the bridge such as the piers. While adding stable points as references is the most effective way to eliminate the movement of the UAS from the bridge deformation, a wide-angle camera is added to capture bridge piers as references. The data analysis of dual cameras is divided into three parts: dual camera calibration, target displacement calculation, and UAS movement elimination. The dual camera calibration includes the calibration of the wide-angle camera and telephoto camera, respectively, and the calibration of the homography relationship between the two cameras. Based on this, the displacement relationship between the image coordinate systems of the two cameras can be obtained.

For the calculation of bridge deformation, a displacement calculation framework based on deep learning target tracking is proposed, including target detection using the you only look once (YOLO) v5s network, target tracking based on deep-simple online and real-time tracking (DeepSORT), and subpixel center extraction for ROI. First, the pretrained YOLO v5s network is used to automatically identify the prepasted targets on the bridge to obtain the target bounding boxes. Then, a feature recognition network is used to number the targets in each frame of the image, and the center trajectory of the target is predicted by the Kalman filter to improve the antiinterference ability of trajectory prediction. Finally, the coordinates of the target center point are directly extracted by the subpixel algorithm to obtain the subpixel-level result of the trajectory.

3. UAS Displacement Elimination Method Using Dual Cameras System

3.1. Overview of the UAS Measurement System. Although some studies have applied UAS equipped with a single camera for structural displacement or vibration measurement and verified the feasibility, so far, most of the existing studies have adopted small-scale structures in the laboratory environment or small-span bridges with an obvious background. However, for many of the in-service bridges, it is difficult to find a stable reference point in the background. When a UAS equipped with a single camera is used for bridge deformation inspection, an insurmountable problem arises because the stable reference point of the pier cannot be captured at a short distance, and the large field of view causes measurement accuracy to decline when taking images at a long distance. Therefore, the proposed method adopts the idea of dual cameras to solve the contradiction between a large field of view and high precision. The dual camera used is Zenmuse H20 from DJI company, which integrates a zoom camera, a wide-angle camera, a laser rangefinder, and a three-axis stabilized gimbal. The resolution of the zoom camera is 20 million pixels, the focal length range is 6.83~119.94 mm, and it can take images with 5184 × 3888 pixels and videos with 3840 × 2160 pixels. Since the zoom camera in this study maintains a fixed focal length during testing, the zoom camera will be collectively referred to as the telephoto camera in the subsequent discussion. The resolution of the wide-angle camera is 12 million pixels, the focal length is 4.5 mm and it can take images with 4056 × 3040 pixels and videos with 1920 × 1080 pixels. The video frame rate collected by both cameras is 30 fps. The shaking angle of the camera is less than 0.01° in the X, Y, and Z axes with the help of the gimbal. The dual camera is mounted on the DJI M300 RTK UAS platform, whose hovering accuracy is 0.1 m, and it can provide a flight time of about 45 min. The UAS is shown in Figure 2(a), and the images of different fields of view taken by the dual camera are shown in Figure 2(b).

3.2. Basepoint Displacement Removal with the Dual-Camera System. Camera calibration is the basic step of displacement measurement using vision-based methods. Because the proposed method adopts a dual-camera system, it is necessary to calibrate the two cameras first and then calibrate the dual-camera system. The pinhole model of the camera is the most commonly used camera model, so it is for theoretical derivation here, and Zhang's calibration method is applied to calibrate the two cameras, respectively [30]. For the wide-angle camera, its focal length is 4.5 mm, and only one calibration is needed to obtain the internal parameter matrix of the camera. For the telephoto camera, because its focal length will change before measurement, it is necessary to calibrate the camera under multiple focal lengths. During inspection for in-service bridges, the focal length of the camera changes at a fixed magnification, which is 2, 5, 10, 20, 40, and 80 times of zoom, respectively. Among them, 5, 10, and 20 times of zoom are the most commonly used, so we calibrate the camera under these three times of zoom. The internal parameters of dual cameras calibrated by Zhang's calibration method are wide-angle camera internal parameter K_w and telephoto camera internal parameter $K_{t=5}$, $K_{t=10}, K_{t=20}$. The bridge displacement calculated using the video of the moving target captured by the telephoto camera, the displacement of the UAS calculated using the video of stability reference points of the pier captured by the wide-angle camera, and the final removal of the UAS displacement from the displacement calculated by the telephoto camera are further discussed separately in this study.

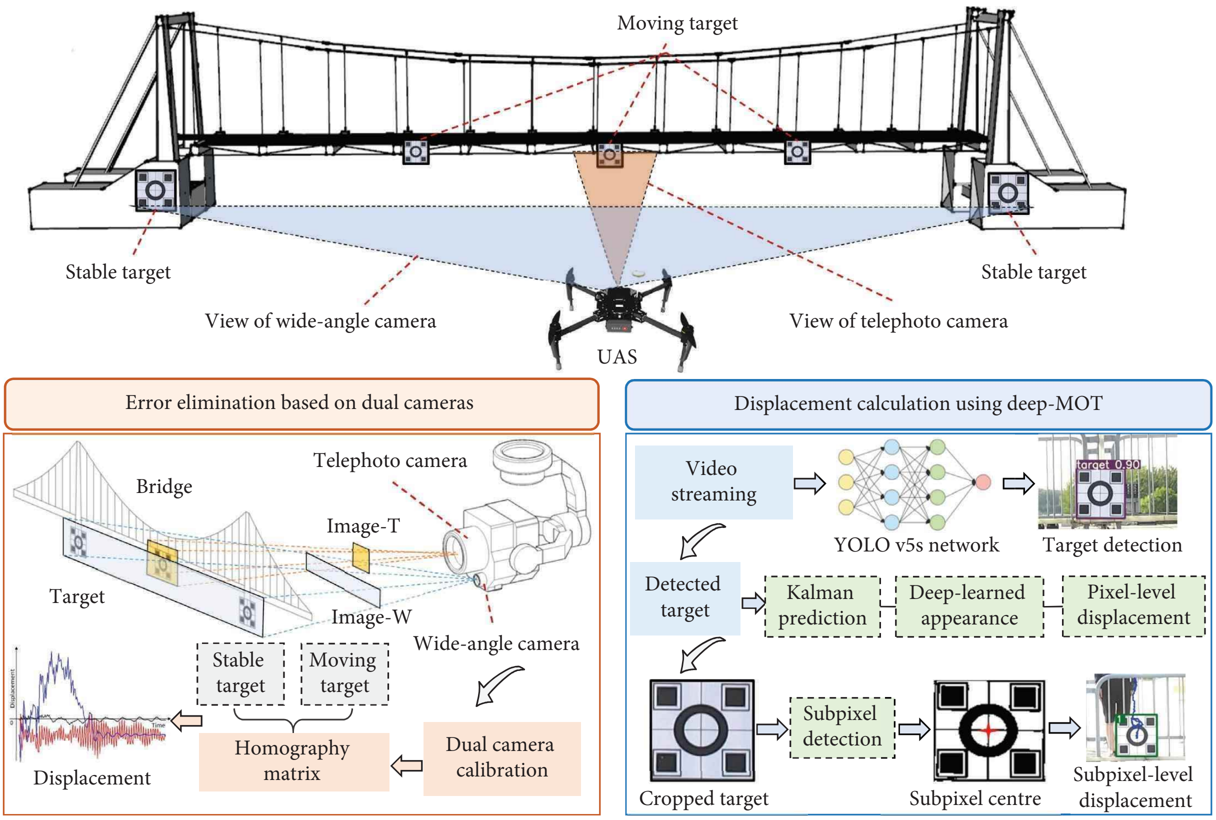
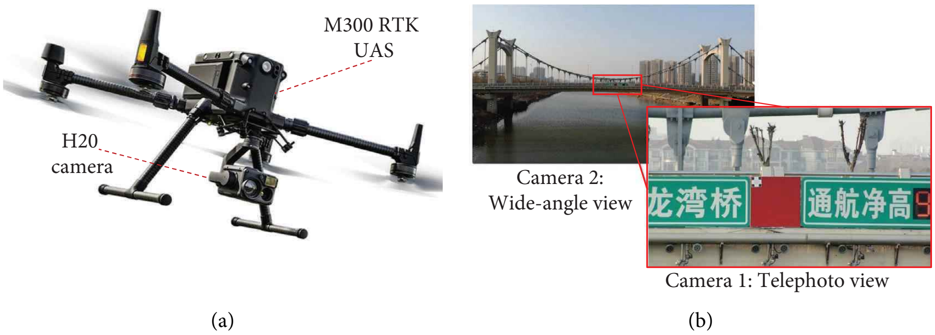


Figure 1: Framework of the proposed method.



(a)

(b)

Figure 2: The applied UAS (a) and field view of dual cameras (b).

For both wide-angle and telephoto cameras, the essence is to use the camera to measure the translation and rotation of the target in three directions in the world coordinate system, so both measurement principles can be represented in Figure 3(a). It can be considered that the camera is fixed and the target changes its position relative to the camera. From time T_0 to T_1, it can be considered that the 3D points in the world coordinate system undergo a rotation of R and a translation of T, and the imaging process of the camera before and after displacement can be expressed as

$$k_i U_i = K_m (RP_i + T), \tag{1}$$

where k_i is the scale coefficient of the target at time T_1, U_i is the coordinate of three-dimensional points in the image coordinate system at time T_1, K_m is the internal parameter matrix of the camera, and P_i is the coordinate of the three-dimensional point in the world coordinate system. The internal parameter matrix of the camera is

$$K_m = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{2}$$

where f_x and f_y are the equivalent focal lengths in both directions and c_x and c_y represent the coordinates of the principal point of the camera. The three-dimensional point coordinates in the world of coordinates and the coordinates under the image coordinate system are expressed as

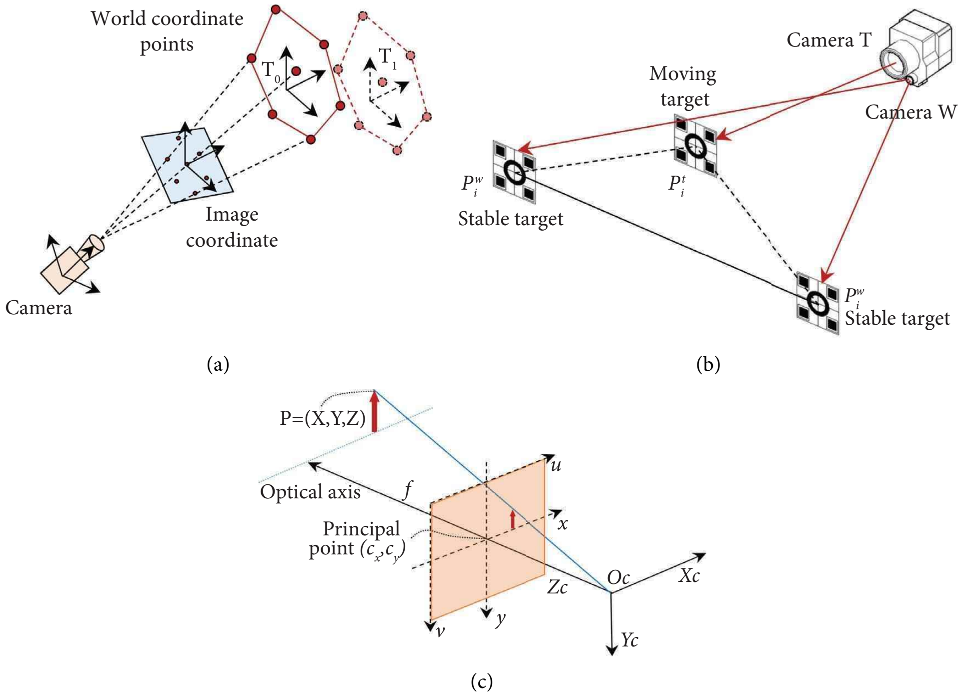


Figure 3: Principle of the camera pose solution (a), dual camera measurement principles (b), and camera model (c).

$[x_i \; y_i \; z_i \; 1]^T$ and $[u_i \; v_i \; 1]^T$, respectively, in homogeneous coordinates, and then formula (1) is expressed as

$$k_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K_m \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix}. \tag{3}$$

The rotation matrix R can be expressed as the rotation of the 3D points around the pitch, roll, and yaw axes, and the translation matrix T can be expressed as the displacement components of the 3D points in three directions: horizontal, vertical, and depth directions. According to the perspective-n-point (PnP) method [37], the rotation matrix R and translation matrix T can be solved by several 3D-2D corresponding points.

These six parameters need to be considered when the camera changes in the attitude of six degrees of freedom. However, since both cameras are installed on the three-axis stable gimbal with an angular resolution of 0.01° and the UAS has the function of locking the heading axis, it can be considered that the three-axis rotation of the camera does not change during the inspection, so only the translation should be considered. In addition, from time T_0 to T_1, the time change is short and it can be considered that the scale parameter k_i of the image remains unchanged. Therefore, only the translation matrix needs to be calculated. Generally, in the deformation measurement of the bridge, only the vertical deformation is concerned, so the vertical displacement in the translation matrix is the required solution result.

The principle of bridge deformation measurement based on dual cameras is shown in Figure 3(b), where the wide-angle camera captures the stable targets at the bridge piers and the telephoto camera captures the moving targets at the bridge body. Assuming that the points in the world coordinate system are (x_w, y_w, z_w), the moving UAV coordinate system is (x_u, y_u, z_u), and the dual camera coordinate system is (x_k, y_k, z_k), where $k = 1, 2$. The change of the camera coordinate is $(\alpha, \beta, \gamma, \Delta x, \Delta y, \Delta z)$, where (α, β, γ) is the rotation angle of the camera around the three-axis and $(\Delta x, \Delta y, \Delta z)$ is the translational component of the camera in the three directions. Setting P_i as the stable target point in the field of view of camera W, p_i is the projection point of the stable point in the image of camera W, where $i = 1, 2, 3, \ldots, n$ when the camera is displaced, that is, P_i is relatively moved, and the displacement component of the projection point p_i in one direction includes the rotation of the camera around the rotation axis in that direction plus the translation component in that direction, which can be expressed as (taking the y axis as an example)

$$\begin{aligned} P_y' = {} & -P_x \sin\gamma^w + P_y \cos\alpha^w \cos\gamma^w \\ & + P_z \sin\alpha^w \cos\gamma^w + \Delta Y^w, \end{aligned} \tag{4}$$

where α^w, β^w, and γ^w are the rotation angles between the camera and the stable target in the three-axis direction, P_x, P_y, and P_z are the coordinates of P_i before the motion occurs, and ΔY^w is the translation component in the y-axis direction. Then, the relative vertical displacement between camera W and the stable target is

$$\begin{aligned} H_w = P_y' - P_y = {} & P_y \left(\cos\alpha^w \cos\gamma^w - 1\right) \\ & + P_z \sin\alpha^w \cos\gamma^w + \Delta Y^w. \end{aligned} \tag{5}$$

Then, the point displacement in the image coordinate system is

$$\Delta p_{\mathrm{iy}} = v_i' - v_i = k_{\mathrm{wi}} H_w, \tag{6}$$

where the scale coefficient k_{wi} can be calculated from the homography relationship between the target and the camera. When the camera plane is parallel to the target, it can also be simply calculated from the object's distance and focal length. Since the motion in the depth direction is small, it can be considered that k_{wi} remains unchanged. Since Δp_{iy} can be calculated directly from the two frames, and α^w, β^w, and γ^w are also small with the stabilization compensation of the 3-axis gimbal, the aforementioned equation can be simplified to

$$\Delta p_{\mathrm{iy}} = k_{\mathrm{wi}} \Delta Y^w. \tag{7}$$

So far, the vertical displacement of the camera during inspection could be obtained.

Camera *W* and camera *T* are fixedly connected, and the relationship between the two cameras is:

$$P_i^t = R^{t,w} P_i^w + T^{t,w}, \tag{8}$$

where $R^{t,w}$ and $T^{t,w}$ are the rotation translation relationship of the two cameras. Therefore, after the rotation translation relationship of the two cameras is obtained by dual-camera calibration, the vertical displacement of camera *W* can also be transmitted to camera *T*. Similar to camera *W*, the vertical displacement of the three-dimensional point captured by camera *T* is

$$\Delta p_{\mathrm{iy}}^t = k_{\mathrm{ti}} \Delta Y^t. \tag{9}$$

Then, the pure deformation of the bridge excluding the base point displacement of the UAS is

$$\Delta P_b = k_{\mathrm{ti}} \Delta Y^t - R^{t,w} k_{\mathrm{wi}} \Delta Y^w. \tag{10}$$

Firstly, Zhang's calibration method is used to calibrate each camera [38], and then stereo camera calibration is carried out. The results are as follows:

$$R_{t,w} = \begin{bmatrix} 0.9999 & -0.0007 & 0.0004 \\ 0.0006 & 0.9999 & 0.0001 \\ -0.0005 & -0.0009 & 0.9999 \end{bmatrix} T_{t,w} = \begin{bmatrix} 20.021 \\ 36.967 \\ 33.115 \end{bmatrix}. \tag{11}$$

The rotation matrix is transformed into the rotation angles in three directions, which are –0.029°, 0.026°, and 0.037°, respectively. It is proved that the two cameras can be considered to be arranged in parallel, and the angle between cameras has little effect on the transmission of displacement. Since the translation matrix will be subtracted in the displacement transfer, the displacement measured by the wide-angle camera in millimeters can be directly subtracted from the displacement measured by the telephoto camera in millimeters according to the previous formula to obtain the absolute displacement of the bridge.

4. Deep Learning-Based Target Tracking Method for Displacement Measurement

Displacement measurement methods based on traditional feature tracking methods such as DIC or optical flow generally involve setting a region of interest for the measurement object in a video or image sequence and then performing feature matching on a frame-by-frame image and calculating the pixel position changes in subregions within the region of interest [26]. These methods generally require the presence of rich textures in the region of interest, such as scattered texture, and require a series of artificially set parameters, which make it easy to lose the target when the lighting condition in the measurement area changes significantly or even when the measurement target is briefly obscured. Deep learning-based target detection methods have unique advantages in terms of automatic acquisition of regions of interest, resistance to interference such as illumination, and the absence of manually set parameters. In fact, for images detected using object detection networks, the bounding boxes of the target in frames can be quickly extracted and the center coordinates of the objects can be calculated. Connecting the center coordinates frame by frame can be considered as a coarse pixel displacement of the object. However, this approach suffers from the same set of problems, including the change of the ID of the target after being occluded and the problem that direct calculation of the centroids only yields pixel-level displacement results. Based on this, the proposed method adds a target tracking method and subpixel refinement extraction method based on the application of the object detection network, and its main processing flow is shown in Figure 4.

4.1. Automatic Target Detection Based on YOLO v5s Network.

The first step of the proposed method is to use the object detection network to quickly and accurately identify the targets on the bridge, and the first to be discussed is the selection of the most suitable object detection network. After years of rapid development, a large number of object detection networks have been proposed. These networks can be divided into the following two categories: anchor-based networks and anchor-free networks. The anchor-based network needs to cluster several groups of anchor sizes from the dataset in advance so that the most appropriate anchor can be chosen to establish the best bounding box with high intersection over union (IoU). The anchor-free method, on the other hand, determines the location of the bounding boxes by dense prediction of feature pixels or key points of the target in the image [39]. One-stage networks and two-stage networks are two types of anchor-based networks. The former regresses the class probability and location coordinate values of objects directly through the backbone network, while the latter first generates a series of sample bounding boxes by clustering algorithms or region proposal network (RPN) and then classifies the samples through a convolutional neural network [40–45]. Object detection is divided into two parts in two-stage networks: producing region suggestions from pictures and generating

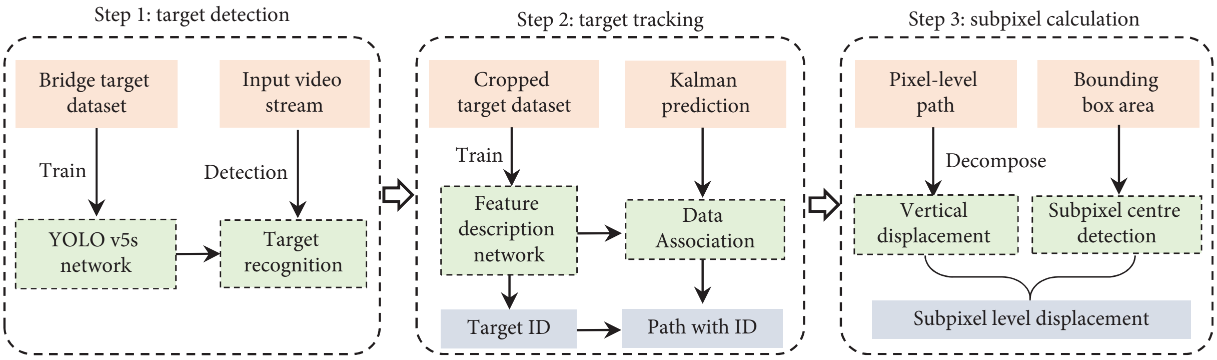


Figure 4: Steps of the displacement calculation method based on deep learning-based tracking.

final object frames from region proposals. Representative networks include faster region-convolutional neural networks (R-CNNs) [39] and feature pyramid network (FPN) [41]. The one-stage network eliminates the requirement for an area proposal stage and generates the object's category probability and location coordinate value directly. Representative networks include the YOLO series network and single shot multibox detector (SSD) network [42]. In contrast, the anchor-free method abandons the preset anchor and completes the detection by determining the key points, such as CenterNet [43] and fully convolutional one-stage object detection (FCOS) network [44].

The YOLO series networks have the highest overall performance and are the most extensively utilized anchor-based networks [45]. Because of their excellent precision and inference speed, YOLO V3 networks are commonly used in the respective industrial detection scenarios. Based on the YOLO V3 network, the YOLO V4 network adopts a series of optimization structures, including cross-stage partial (CSP) structure, spatial pyramid pooling (SPP) structure, and path aggregation net (PA-Net) structure, as well as Mish activation function and complete-intersection over union (C-IOU) loss function, to improve the model training time and obtain higher mean average precision (mAP). In addition, the YOLO v4 network proposes four networks with different depths and widths. The depths and widths of these four networks increase accordingly, and the mAP of the corresponding model increases while the speed decreases, thus enabling developers to choose the most suitable network according to different application scenarios. The latest version of the YOLO series networks is the YOLO v5 series network, which is similar to the YOLO v4 network with four different sizes of subnetworks. It has optimized data enhancement, activation function, and loss function based on YOLO v4, so it has more advantages in terms of speed and platform portability. Considering that the size variation of the target bounding box is small when the UAV shoots with several fixed zoom times, there is no difficulty for the preset anchor to match the size variation of the bounding box of the actual detection scene, so the anchor-free network is not selected here. In addition, the target detection in the proposed method only needs to detect targets with obvious characteristics, so the YOLO v5s network with minimum depth is directly selected as the target detection network.

The next step is to build the bridge target dataset for training the YOLO v5s network. Although some existing methods have been applied for target-free measurements [46], most of these methods require the object to have obvious texture features, and the bridge surface is generally coated with smooth coatings, and many of the existing target-free methods may suffer from target loss. In addition, considering that it is simple and feasible to manually paste some special targets on the bridge deck and that specially designed targets can also provide convenience in determining the scale factor of the images, the proposed method uses a predefined target on the bridge deck and then uses the centre of the measured target as the target point for displacement measurement.

Since the proposed method has been tested on several bridges, most of the images in the dataset are from target videos on multiple types of bridges taken by UAS. This training dataset close to the final engineering application scenario helps to enhance the generalizability and practicality of the proposed method. In addition, considering that the target images may have different degrees of illumination conditions in the actual detection, the established dataset also contains a large number of target images with different illumination conditions to improve the generalizability of the model to different illumination conditions, as shown in Figure 5. The final dataset contains 1,336 target images, and the size of each image is 1920×1080 pixels. The images in the training set are manually labeled with the target locations, and then the dataset is converted into Pascal Voc format for training. After labeling, the k-means clustering was used to cluster nine groups of anchor sizes from them. Since the input size of the model is 640×640 pixels, the clustered anchor sizes were [47, 64], [42, 75], [46, 84], [60, 103], [106, 149], [145, 192], [159, 213], [183, 240], and [271, 342]. The results of clustering show that the nine groups of anchor sizes cover 92.07% of the labeled boxes in the dataset, which proves that the anchor sizes were well clustered.

The model was trained on a computer with i7 11700K CPU, RTX3090 GPU, and 32G RAM, and the framework used for training was Pytorch. The three methods of scaling, color space adjustment, and mosaic enhancement are used to enhance the data in advance to increase the generalization ability and small target recognition ability of the model. The

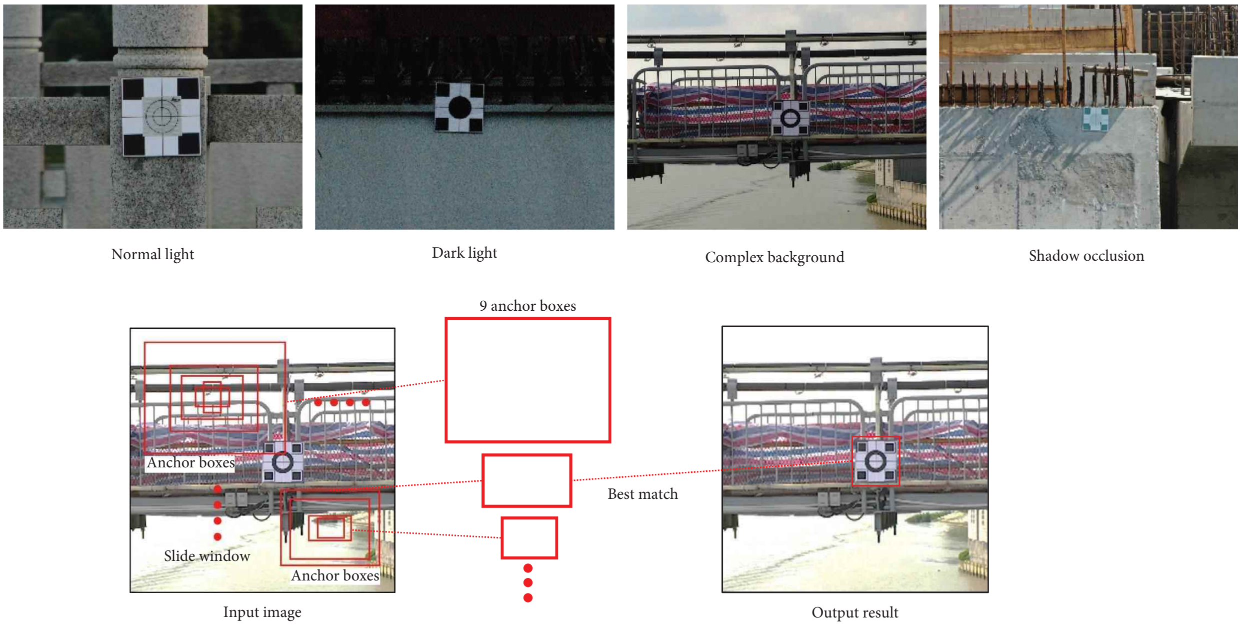


Figure 5: Target image in the dataset and the clustered anchor boxes.

loss curves, AP curves, P-R curves of the training process, and the results of the randomly selected images in the test set after the training are recorded to evaluate whether the model is trained correctly and to obtain the final accuracy of the model, as shown in Figure 6. The trend of the loss curve shows that the network has stabilized after 2,000 steps of training, the accuracy curve shows that the final accuracy is 0.998, and the envelope area of the P-R curve also proves that the trained model achieves high accuracy.

4.2. Target Tracking Using Deep-SORT. After the detection of the target, the center of the bounding box is calculated for each frame, and the rough displacement trajectory of the target is obtained by connecting the centers of the bounding box of each frame. However, there are problems with this method of directly calculating the center of the bounding box; that is, when using deep learning to detect targets in the video, there is a possibility that the targets in some frames are briefly lost due to high frame rate or brief occlusion. Once the target is lost, on the one hand, the obtained displacement misses at certain time points; on the other hand, the target will be reidentified and assigned a new ID, which is not conducive to the analysis of displacement results. Based on this, the method of step 2 in Figure 4 is adopted to deal with the tracking problem.

The core of this approach is using the Deep-SORT algorithm to track targets, and the Deep-SORT algorithm has the advantage of accurately handling the reidentification problem in target movement and the identity transformation problem under large displacement or occlusion [47]. The reidentification problem is to determine the correlation of the same target between the previous image frame and the current image, which is to correlate the feature of the detection areas of the same target in different frames, and then minimize the intraclass difference and maximize the interclass difference. For example, when tracking multiple targets, if target 1 moves to the position of target 2 at the previous time, how to keep the ID of target 1 at the position of target 2 is still recognized as 1. Deep-SORT is an improved algorithm based on SORT [33]; for the original SORT, its method of dealing with judging the interclass difference takes the direct calculation of the distance between marker frames, and the characteristics of the targets inside the marker frames are not considered, so the problem of confusing target IDs is easy to occur in the case of tracking multiple targets with large displacement. Deep-SORT solves this problem in another way, by adding a simple CNN to learn the identified targets in advance, so as to determine the difference between different targets by deep learning-based classification [48]. The identity transformation problem refers to the fact that if two adjacent targets are located close to each other in the video, or even partly overlapped, the target detection network may assign the number of the two targets incorrectly, resulting in errors in the final displacement trajectory. To solve this problem, it is necessary to predict the displacement result at each moment and evaluate whether it obeys the trend of displacement motion. Cha et al proposed a phase-based optical flow algorithm for displacement calculation, where an unscented Kalman filter is used to predict the displacement trend and thus remove the noise from the displacement measurements [49, 50]. Therefore, in this part, the Kalman filter is used to get the predicted trajectory of the current coordinates after getting the coordinates of the recognition frame or target center of the current frame and inputting the predicted trajectory to the target detection of the next frame. If the detection result of the next frame has a good correlation with the predicted trajectory of the next frame, the detection result is considered correct and the process continues according to this flow. In addition, for the problem that the target is not recognized at a certain moment after the predicted trajectory

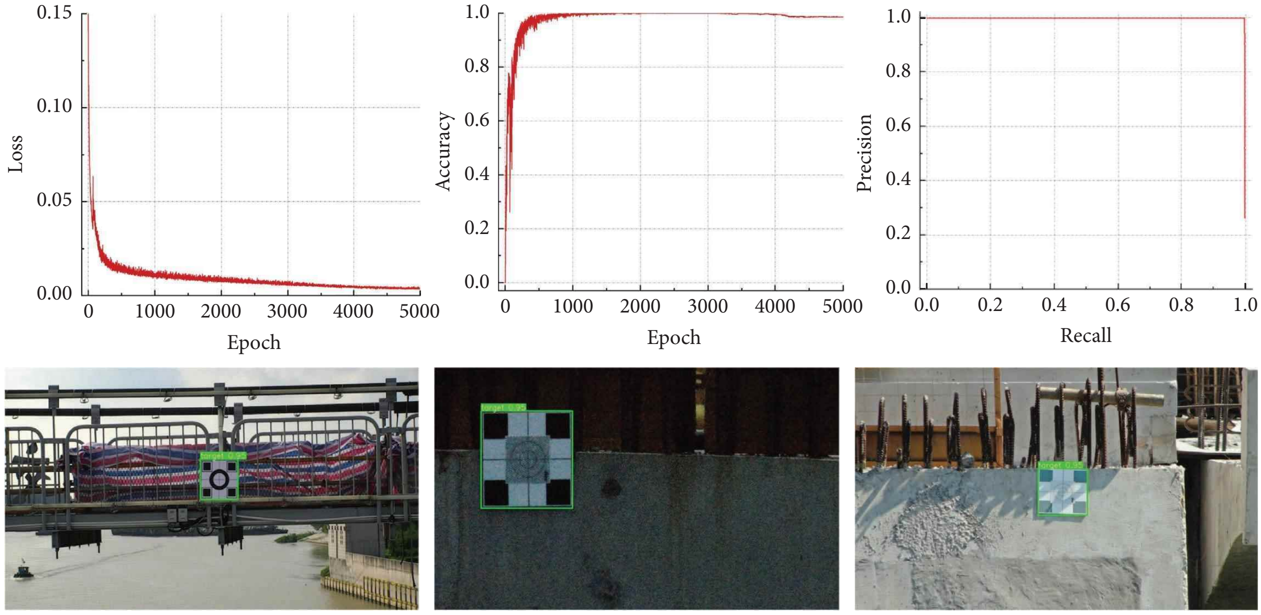


Figure 6: Training results for YOLO v5s network.

is obtained by the Kalman filtering, there is no detection result associated with it at that moment. The processing method for this case is to make IOU matching for the trajectory with the target detected thereafter and continue to depict the trajectory. The whole processing flow is shown in Figure 7.

The specific processing of the adopted method has been illustrated in step 2 in Figure 4. Firstly, for the targets identified and labeled in the first step, the targets can be segmented down in batches according to the range of the bounding boxes using the trained model. The segmented target images are manually classified by classifying the images belonging to the same target, which can be made into a dataset for training the interclass recognition CNNs. The same network as the original study of Deep-SORT is used here for training, and it is a CNN similar to the VGG network, which has an input size of 128 × 128 pixels, so it is necessary to resize all the images in the dataset to a size of 128 × 128 pixels. The computer and framework for training the interclass recognition CNN are the same as those used for training YOLO v5s mentioned above, considering that the target features are simple and the CNN used is also a simple structured network, so the model is only trained for 200 steps and the learning rate is set at 0.1. The loss curve and accuracy curve of the training process are shown in Figure 8, and the trend of the curve indicates that the model is trained correctly. The final training accuracy is 97.33%, and the test accuracy is 82.82%.

4.3. Fine Optimization of Target Displacement Based on Subpixel Detection.

The processing of step 1 and step 2 in Figure 4 can obtain stable pixel-level target center displacement trajectories, but the deformation of bridges under in-service loads is generally small, so step 3 in Figure 4 is a further refinement of the measurement results of step 2. Since the object detection-based target identification results are pixel-level, the target centroid error obtained using this result alone may exceed a dozen pixels, so although the stable and numbered target bounding box and its motion trajectory have been obtained in step 2, it is necessary to further optimize the measurement results. Research on accurately calculating the position of the target in images is the focus of the detection of fine damages such as cracks, and methods have been proposed based on pixel-level image segmentation; for example, Kang and Cha proposed a novel semantic transformer representation network (STRNet) for crack segmentation at the pixel level in complex scenes in real-time, and the model obtained 91.7% accuracy and 49.2 fps in the test [51]. Choi and Cha proposed a semantic damage detection network (SDDNet) for segmenting concrete cracks, which achieves a mIOU of 0.846 and a speed of 36 fps for an input image size of 1025 × 512 pixels [52–54]. However, the pixel-level segmentation result corresponds to the same pixel-level displacement measurement accuracy. In order to improve the coordinate calculation accuracy of the target center as much as possible, this section applies a subpixel detection method to identify the midpoint of a circle and the corner point of a square in the target image, and then the subpixel level displacement trajectory of the target center point can be obtained.

The designed target is shown in Figure 9. Although the targets of different sizes may have slightly different shapes, the targets mainly consist of four rectangles in the corners and a circle in the middle; the point of this design is that firstly, the coordinates of the four rectangles in the corners can be used as known points to calculate the scale parameters of the image. Secondly, the center point formed by the rectangular coordinates of the four corners coincides with the center point coordinates of the middle circle, so it is also possible to calculate the center point coordinates from a small number of corner point coordinates when the target is partially obscured. Specifically, assuming that the center point coordinate of the target is $P_c(x_c, y_c)$, the detected quadrangular corner points are $P_1(x_1, y_1)$, $P_2(x_2, y_2)$,

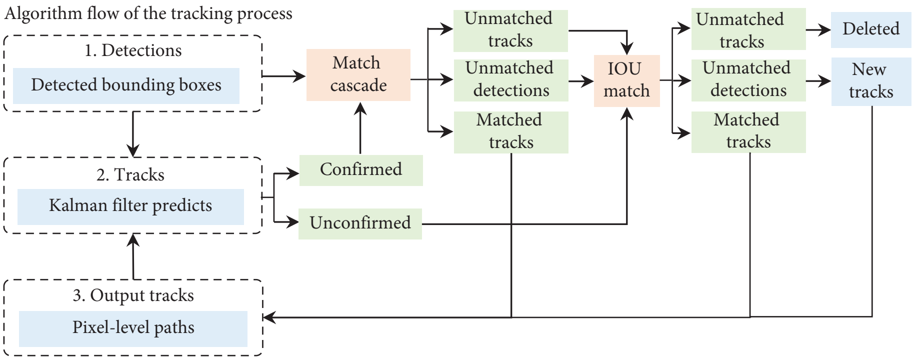


FIGURE 7: Algorithm flow of the tracking process using Deep-SORT.

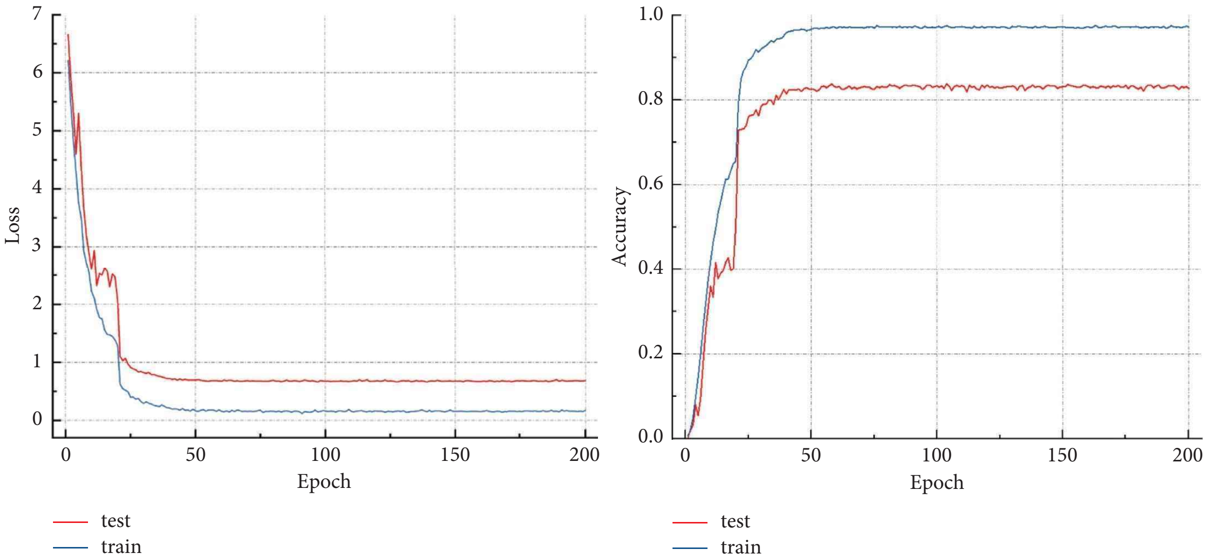


FIGURE 8: Training results of the target classification network.

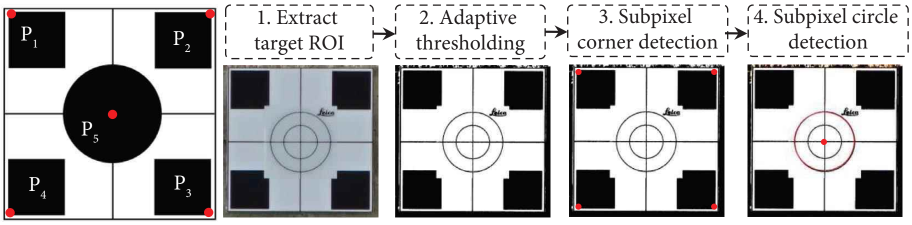


FIGURE 9: The designed target and center calculation method.

$P_3(x_3, y_3)$, and $P_4(x_4, y_4)$ and the center circle point coordinate is $P_5(x_5, y_5)$ and then P_c is calculated by taking the average of these five coordinates. Therefore, the key to this part is to detect the coordinates of the five points with subpixel accuracy. The target is made of flat hard plastic plate. The target pattern is sprayed on the surface of the plastic plate by a precision ink-jet printer and then covered with an antioxidation transparent film to slow down the fading of the target.

Subpixel edge or corner detection methods have been widely used in various measurement and detection fields, and it mainly includes the moment-based method [50], interpolation-based method [53], and fitting-based method [54]. Among them, the moment-based method can be regarded as establishing an edge model for the two-dimensional edge, fitting the actual edge in the image, and then establishing equations to solve the actual edge position of each position. The commonly used moment-based methods include the spatial moment method, Zernike moment method [55], and orthogonal Fourier Marin moment method [56]. The core of the interpolation-based method is to interpolate the gray value or the derivative of the gray value of pixels and add information to realize subpixel edge detection. Among them, the quadratic interpolation method and the B-spline interpolation method are widely used [57]. The fitting-based method is to obtain the subpixel edge location by fitting the gray value of the hypothetical edge model, such as the least square fitting algorithm and Gaussian edge function fitting. Because the fitting-based method does not need numerical differentiation and the fitting is carried out according to the minimum distance from each gray value to the fitting curve, it not only makes reasonable use of the gray value with error but also reduces the influence of gray value error, so the fitting-based method is not sensitive to noise. Considering that the subpixel edge detection function in the OpenCV library adopts the least square method in the fitting method, the inference speed and portability of this function are satisfying, and it is applied to calculate the target center coordinates in this part.

Although the least-squares subpixel detection algorithm can accurately extract the edge and corner point in the target image, it cannot filter out the coordinates of the required five points discussed above, so the corresponding filtering methods need to be set for the rectangles in the target corner and the center circle. The convexity detection method is used to filter the convex graph of the least-squares fitted edges, and then coordinates of the corner points of each rectangle can be obtained. For the center circle, the center coordinates of the circle are detected using the Hough transform [58]. Finally, the subpixel level target center coordinate can be obtained by averaging the extracted five coordinates.

5. Validation Experiments

To validate the effectiveness of the proposed method, a laboratory test on a high-precision electric displacement control platform and a field test on an in-service suspension bridge were conducted to analyze the accuracy of the proposed method.

5.1. Laboratory Validation.

As shown in Figure 10, the fixed industrial camera was used to measure only the moving target fixed on the electric displacement table, and the measured result is used for comparison with those measured by UAS using the proposed method. The telephoto camera of the UAS is used to capture the moving targets, while the wide-angle camera is used to capture stable targets on both sides. The accuracy of the electric displacement stage is 0.03 mm. The displacement stage moved in a preset direction during the measurement, and the target moved 5 mm in each step. The total movement was 11 steps, and the total displacement was 55 mm. The industrial camera is HIKVISION MV-CA050-10GC, the focal length of the camera is 8 mm, the frame rate of the video is 10 Hz, and the image size is 2448 × 2048 pixels. The accuracy of the two methods was analyzed by comparing them with the set displacement curve. Figure 10 shows the testing process, the view of the cameras, and the displacement measurement results of the two methods.

The results in Figure 10(c) show that the displacement trends calculated by the two methods match well and are consistent with the preset displacement mode. However, several fluctuation points are visible in the displacement curves of the fixed camera using the DIC method, and the reason for these points may cause by the shadow obscuration of the image at that moment due to the movement of personnel during the experiment and the possible matching error in the calculation due to the little texture of the target itself. The results calculated by the proposed method are not impressed by these factors, which indicate that the proposed method has the advantage of not being affected by the environment while achieving high accuracy similar to that of the DIC method. Comparing the displacement results calculated by the proposed method with the preset displacement table motion program, the average error of displacement is 0.21 mm and the maximum error is 0.81 mm. In addition, the results in Figure 10(d) show that when the motion of the UAS is not eliminated, the target movement trajectory calculated by the telephoto camera is very different from the preset movement trajectory, which proves that the measurement results cannot be used without eliminating the displacement of the UAS.

5.2. Field Test on an In-Service Bridge.

The tested bridge is shown in Figure 11. The bridge is an urban highway bridge with a total length of about 200 m, of which the middle main span is 100 m, and the bridge type is the suspension bridge. The test time is after 5 pm, so the test time is the peak of traffic flow, which helps to reflect the deformation of the bridge under traffic flow load. In the test process, the fixed camera under the bridge is used as the comparison. The camera used is HIKVISION MV-CA050-10GC, and the focal length is 150 mm.

During the test, the UAS hovered at a distance of about 20 m from the bridge and kept capturing video with the camera facing towards the bridge. The camera of the UAS is set in free mode; that is, the horizontal rotation and vertical rotation angle of the camera remains unchanged and will not rotate due to the movement of the UAS. Therefore, the change of relative displacement caused by the rotation of the camera can be ignored. The focal length of the telephoto camera is set to 240 mm. The fixed camera on the ground aims at the prepasted target on the bridge side. Since there is railing with smooth surface fixed on the bridge, the target is attached to the side of the railing using a strong double-sided adhesive. When the material of the bridge to which the target is attached is rough concrete, the target can also be attached using a fast-setting epoxy resin material glue. The recorded video frame rates for both the fixed camera and the UAS are set to 30 fps. The view of the wide-angle camera, telephoto camera, and fixed camera is shown in Figure 11. The target size pasted on the bridge deck is 20 cm × 20 cm, and three targets are pasted near the middle of the span. Considering that the UAS is far away from the bridge during the inspection, to enable the wide-angle camera to capture the target clearly, the size of the target on the stable bridge pier should be more than 50 cm × 50 cm. Therefore, the texture of the pier is used instead of using the target.

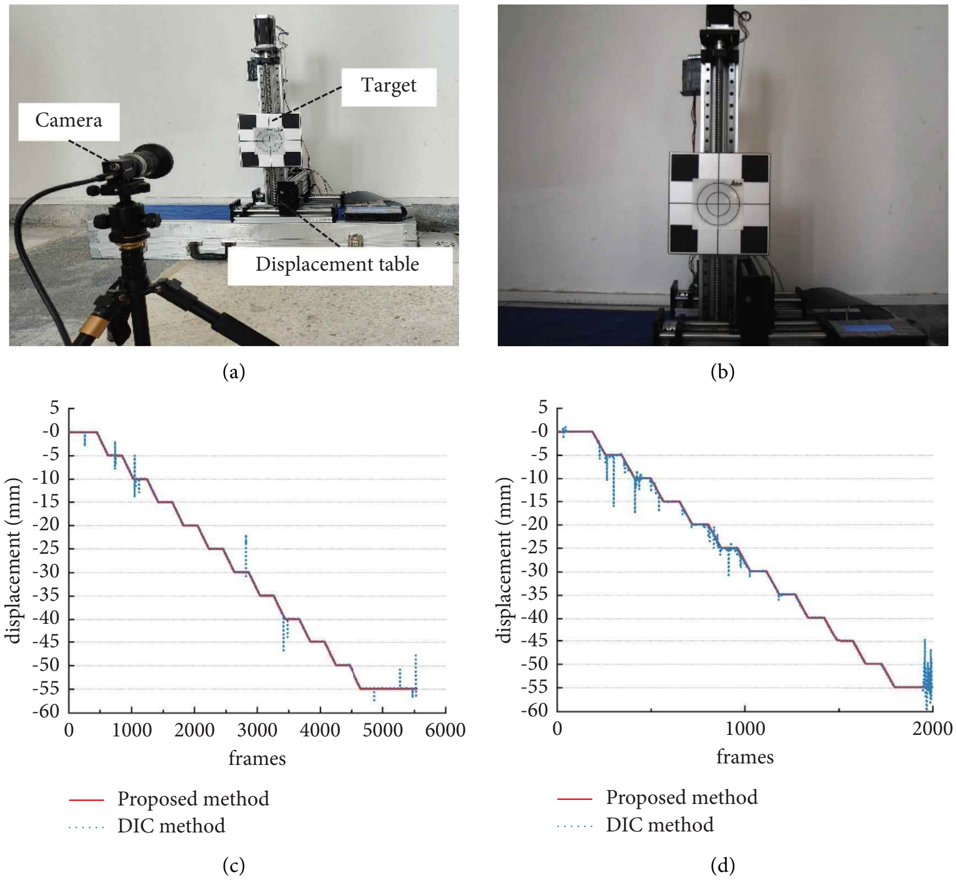


Figure 10: Laboratory test process and test results. (a) Overview of the test. (b) Captured image. (c) Comparison of the results. (d) Results before UAS movement removal.

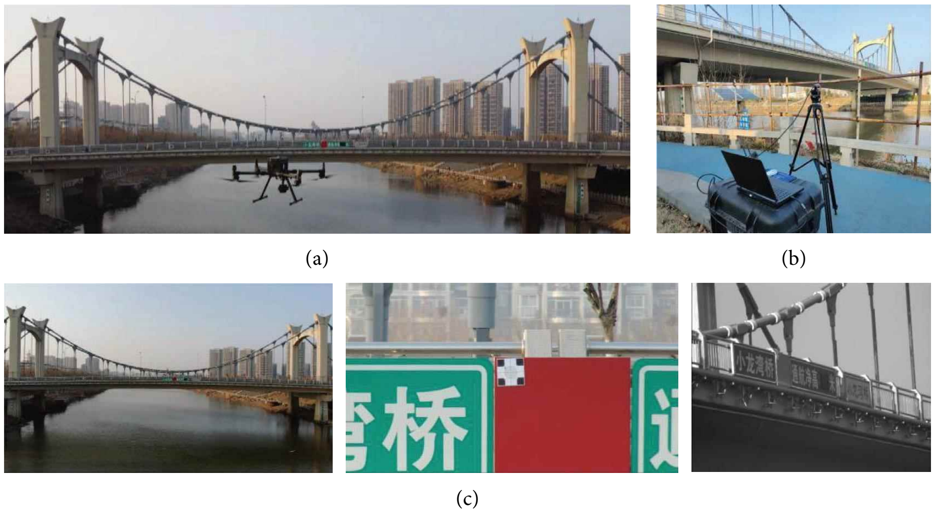


Figure 11: Bridge deformation measurement process. (a) The measurement process. (b) The fixed camera for comparison. (c) Image of wide-angle camera, telephoto camera, and fixed camera.

The three targets in the midspan were recorded separately during the inspection, and each target was recorded for three minutes. Finally, nine videos were collected using the wide-angle camera, telephoto camera, and fixed camera. These videos were analyzed using the following methods: for the fixed camera, the DIC method is applied to calculate the displacement of the target area. Considering that the DIC method is widely used and has been verified in many studies, the results of the fixed camera analysis using the DIC method are regarded as the truth. Besides, the proposed displacement calculation method in Section 4 is also applied to analyze the video of the fixed camera. For the video taken by the UAS, the proposed method is used to process the video of the telephoto camera and the video of the wide-angle camera, respectively. The video of the wide-angle camera takes the piers on both sides as the region of interest. The calculated result is the relative displacement of the UAS to the stable point of the pier. It is subtracted from the target displacement obtained by the video of the telephoto camera, and the real deformation of the bridge can be obtained. The interest area division and interest point selection of the video taken by the fixed camera and wide-angle camera photos are shown in Figures 12(a) and 12(b), and the target center coordinate trajectory description results are measured by the proposed deep learning-based tracking method as shown in Figure 12(c). It can be seen from Figure 12(b) that the bridge piers, which are set as the stationary reference points in the field of view of the wide-angle camera, are located on both sides of the image near the edges. Although the original video captured by the camera will remove the distortion according to the previously calibrated camera parameters before calculating the displacement, the measurement target located at the edge of the image may increase the measurement error. In order to minimize the error of the wide-angle camera, the relative displacement is calculated separately for both sides of the bridge pier to ensure that the difference between the two results is small, and then the average of the two measurements is taken as the measurement result of the wide-angle camera. For the three targets, the results measured by the above methods are shown in Figure 13.

From the results in Figures 13(a)–13(c), the trends of the vertical displacement curves measured by the three methods match and have good overlap, and from the analysis of the curve trends, the displacement errors measured by the proposed UAS-based method and the fixed camera-based method are generally within 1 mm, and the maximum error occurs at the measurement of target 3, which is less than 2 mm. This proves that the proposed method has a similar measurement accuracy and stability to the fixed-camera-based method in most cases, and the correction of the measurement results using dual cameras enables the movement of the UAS to be effectively eliminated. However, for small bridge deformation measurements under daily vehicle loads, the UAS-based method may result in occasional large errors due to the small deformation of the bridge. In contrast, the proposed method is more suitable for deformation measurement during load tests of bridges, where the deformation of the bridge may exceed 10 cm, and for large bridges, the deformation in the span of the bridge during load tests may exceed 1 m. In this case, the accuracy of the UAS-based measurement method will be able to meet the requirements. In addition, comparing the results of the fixed camera using the DIC method and the proposed deep learning-based tracking algorithm, it can be found that the curve trends of the two methods match, but the curve calculated with the DIC method has greater data volatility than the curve calculated with the proposed method. The curve shows that high-frequency spikes are present in the results calculated by the DIC method but not in the results calculated by the proposed deep learning-based method. The proposed method uses deep learning to detect the target and the Kalman filter to predict the displacement in case of temporary occlusion or target loss, so the obtained displacement curve does not have spikes. These results demonstrate that the proposed displacement calculation method can achieve better stability than the DIC method.

From the results in Figures 13(d)–13(f), the measured displacement curve fluctuates greatly before excluding the displacement of the UAS, and the peak value already exceeds 50 mm, which is very different from the curve after excluding the displacement of the UAS. The hovering accuracy of the UAS under real-time kinematic (RTK) positioning is centimeter-level, so the measurement results without excluding the displacement do not match the actual deformation of the bridge and cannot be used for bridge deformation inspection.

For the displacement curves measured by the three methods in Figures 13(a)–13(c), the correlation coefficients were calculated using the Spearman correlation coefficient because the Spearman correlation coefficient is not susceptible to extreme values in the data. For target 1, the correlation between the UAS-based method and the displacement curve measured by the fixed camera using the DIC method is 0.801, and the correlation between the UAS-based method and the displacement curve measured by the fixed camera using the deep learning-based tracking method is 0.848. For target 2, the correlations are 0.918 and 0.924, and for target 3, the correlations are 0.925 and 0.942, respectively, while for the displacement curves before and after excluding the UAS displacement in Figures 13(d)–13(f), the correlations are 0.395, 0.118, and 0.260 for the three targets, respectively. The above results are presented in Table 1. Since the data in this section are positively correlated, the value of Spearman's correlation coefficient ranges from 0 to 1. The closer to 1 proves that the correlation is better. The correlation results also support the above conclusions.

In addition to the accuracy and stability of the measurement, the computational time consumption of the proposed method is also an important indicator. According to the description in Section 4, the proposed method can be divided into three steps: preset target detection, target tracking, and subpixel calculation, and the time consumption of each of the three steps is calculated, and the results are shown in Table 1. The size of the input video here is the same as the video capture size of the camera described previously, which is 3840 × 2160 pixels for the telephoto camera and 1920 × 1080 pixels for the wide-angle camera, both at a frame rate of 30 fps. The time consumed by the whole method to process one frame is 0.074 s, which means that the frame rate of the proposed method is about 13.5 fps when processing real-time video.

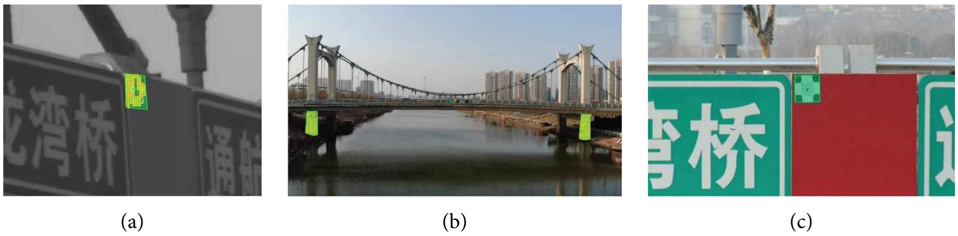


Figure 12: Region of interest (ROI) for images taken by the three cameras. (a) ROI for the fixed camera. (b) ROI for the wide-angle camera. (c) Result for the telephoto camera.

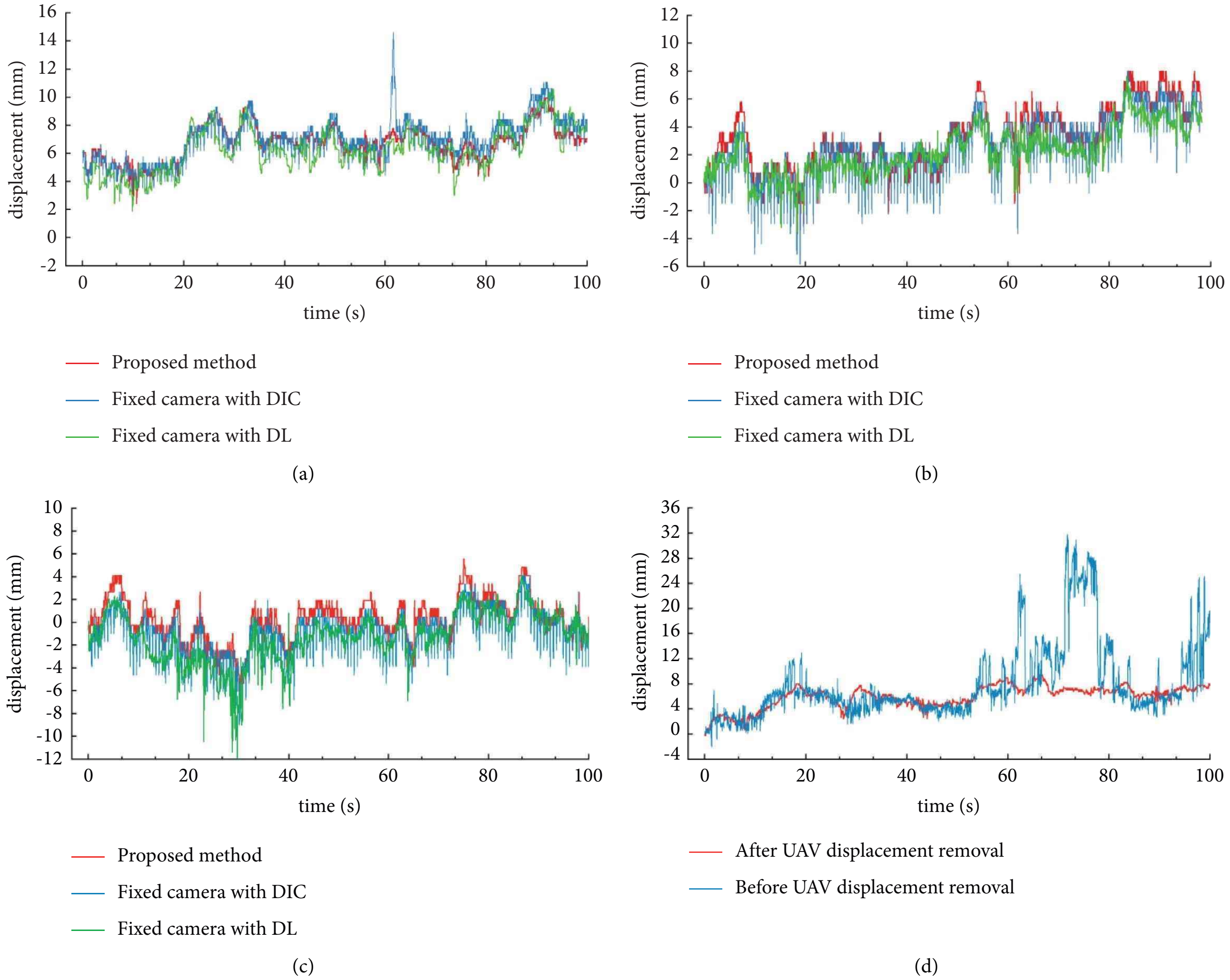
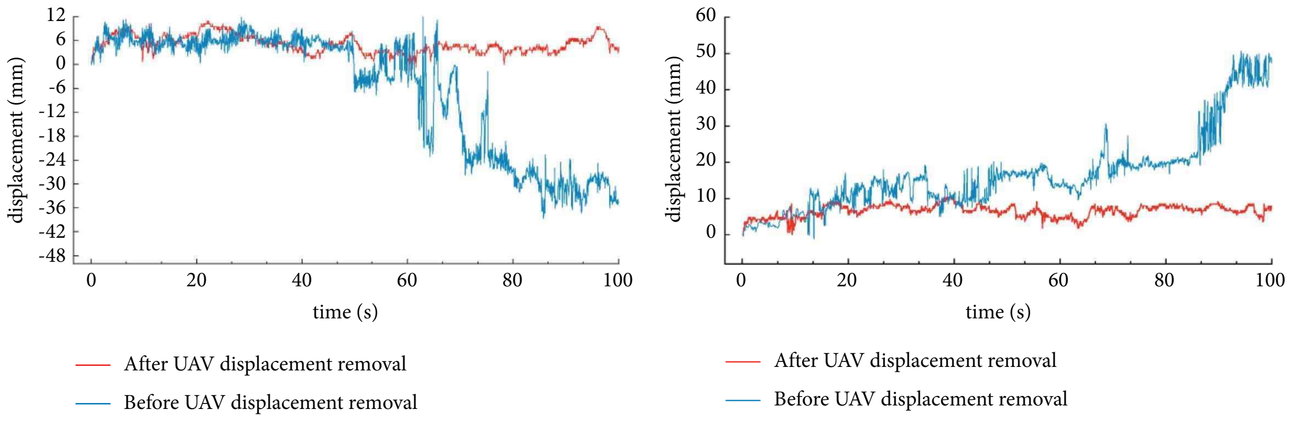


Figure 13: Continued.



(e)

(f)

Figure 13: Comparison of results of the methods. (a) Comparison of the results for target 1. (b) Comparison of the results for target 2. (c) Comparison of the results for target 3. (d) Comparison before and after UAS movement removal for target 1. (e) Comparison before and after UAS movement removal for target 2. (f) Comparison before and after UAS movement removal for target 3.

Table 1: Comparison of displacement measurement results and analysis of calculation time.

Target number	Correlation of the results of the two methods		
	Proposed method and fixed camera with DIC	Proposed method and fixed camera with DL	Before UAV displacement removal and fixed camera with DIC
Target 1	0.801	0.848	0.395
Target 2	0.918	0.924	0.118
Target 3	0.925	0.942	0.260
	Computational time consumption of the proposed method		
Steps	Computational time		Total time consumption
Target detection	0.015 s		0.074 s
Target tracking	0.013 s		
Subpixel calculation	0.046 s		

6. Conclusion

A rapid displacement measurement method based on UAS with a dual-camera system and deep learning-based tracking algorithm is proposed to address the problem of noncontact measurement of dynamic deformation for bridges. The base point displacement caused by the moving of the UAS is removed from the deformation of the bridge using a dual-camera system with a wide-angle lens and telephoto lens, so that both stable points on the bridge pier and deformation points in the middle span can be captured. The problem that the traditional displacement calculation methods are sensitive to lighting changes and accidental occlusion is solved using a deep learning-based tracking method. The main conclusions are as follows:

(1) Through theoretical derivation and experimental verification, it is proved that the UAS equipped with dual cameras can effectively eliminate the movement of the UAS during the measurement of bridge deformation. This method uses both the short-range small field of view (telephoto lens) and long-range wide field of view (wide-angle lens) to capture both stable points and deformed points of the bridge so that the highest accuracy can be observed.

(2) The proposed displacement measurement method based on deep learning tracking integrates the latest object detection network, multitarget tracking algorithm, and subpixel detection algorithm. On the established bridge target dataset, the test accuracy of target detection is 0.998 and the test accuracy of target feature classification is 0.823. It overcomes the problem that the traditional methods are easy to be affected by lighting changes and occlusion of the target. The indoor test shows that the proposed method can effectively avoid data fluctuations while maintaining an accuracy similar to that of the DIC-based method.

(3) The proposed UAS-based deformation measurement method was applied to an in-service bridge, and the results demonstrate the satisfactory accuracy and stability of the proposed method.

In addition to the above beneficial effects, the proposed method has some limitations and needs to be further studied and improved. First of all, the accuracy of the proposed method is expected to be further improved by replacing the wide-angle camera with a freely rotatable telephoto camera. Since the accuracy of the deformation measurement of the proposed method is affected by both the measurement results of the telephoto camera and the wide-angle camera, the telephoto camera can maintain the high resolution of the target object and keep the target in the middle of the field of view for continuous capture, while the large field of view of the wide-angle camera leads to insufficient resolution of the image, although multiple stationary reference points can be selected in the test to calculate the displacement separately and take the average value to reduce the error as much as possible; however, its accuracy still cannot reach the accuracy of the telephoto camera. When applying the proposed method, it is recommended to select the working conditions with large displacements such as on bridges with less stiffness or under heavy load tests. Further study is necessary to be investigated to improve the proposed method. The proposed method demonstrated the required accuracy and robustness in measuring a 100 m-long bridge. However, for deformation measurements of large span bridges, there may still be a problem that the wide-angle camera view is difficult to capture the fixed piers, and one possible method is to replace the coaxial dual camera with two telephoto cameras that can be rotated by themselves, one of which is aimed horizontally forward at the target on the side of the bridge and the other is rotated in the direction aiming at the fixed target on the bridge pier, and then the movement of the UAS can be eliminated according to the calibration of the dual cameras. In this case, the dual cameras are no longer arranged in parallel, and the relative displacement of the stationary point of the bridge pier cannot be directly subtracted from the result calculated from the telephoto camera that captures the moving target of the side of the bridge, but the angle of the two cameras has to be considered, so this part of the theory needs further study.

Another limitation of the proposed method is that the proposed displacement measurement method requires the use of preset targets attached to the side of the bridge. Further work will consider using objects with special shapes on the bridge surface (such as channel indicators and height limit signs) as target detection objects, thus avoiding the disadvantage of requiring manual attachment of targets. In addition to the displacement measurement of the structure, the vibration frequency measurement of the structure is also important for detecting the inherent characteristics of the structure. The existing studies have proved that vision-based measurement methods are feasible in measuring the modalities of high-frequency structures [59, 60], so the proposed method is expected to measure the displacement and frequency of high-frequency vibrating structures after increasing the video acquisition frequency of the camera. Accordingly, optimizing the processing efficiency of the proposed method for the analysis of high-frequency video data is one of the priorities of future research study.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they do not have any conflicts of interest regarding the publication of this paper.

Acknowledgments

The research presented was financially supported by the National Key Research and Development Program of China (no. 2019YFC1511105), the National Key Research and Development Program of China (no. 2020YFC1511900), and the Key Research and Development Program of Jiangsu (no. BE2020094).

References

[1] R. Kromanis and P. Kripakaran, “Data-driven approaches for measurement interpretation: analysing integrated thermal and vehicular response in bridge structural health monitoring,” *Advanced Engineering Informatics*, vol. 34, pp. 46–59, 2017.

[2] Y. D. Tian, J. Zhang, and Y. X. Han, “Structural scaling factor identification from output-only data by a moving mass technique,” *Mechanical Systems and Signal Processing*, vol. 115, pp. 45–59, 2019.

[3] J. Zhang, L. M. Zhou, Y. D. Tian, S. S. Yu, W. J. Zhao, and Y. Y. Cheng, “Vortex-induced vibration measurement of a long-span suspension bridge through noncontact sensing strategies,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 12, pp. 1617–1633, 2021.

[4] F. T. Ni, J. Zhang, and M. N. Noori, “Deep learning for data anomaly detection and data compression of a long-span suspension bridge,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 7, pp. 685–700, 2020.

[5] D. Ribeiro, R. Calcada, J. Ferreira, and T. Martins, “Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system,” *Engineering Structures*, vol. 75, pp. 164–180, 2014.

[6] X. Wang and H. Liu, “Soft sensor based on stacked auto-encoder deep neural network for air preheater rotor deformation prediction,” *Advanced Engineering Informatics*, vol. 36, pp. 112–119, 2018.

[7] D. Hester, J. Brownjohn, M. Bocian, and Y. Xu, “Low cost bridge load test: calculating bridge displacement from acceleration for load assessment calculations,” *Engineering Structures*, vol. 143, pp. 358–374, 2017.

[8] C. Gentile and G. Bernardini, “An interferometric radar for non-contact measurement of deflections on civil engineering structures: laboratory and full-scale tests,” *Structure and Infrastructure Engineering*, vol. 6, no. 5, pp. 521–534, 2010.

[9] G. W. Zhang, Y. L. Wu, W. J. Zhao, and J. Zhang, “Radar-based multipoint displacement measurements of a 1200-m-long suspension bridge,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 167, pp. 71–84, 2020.

[10] P. Garg, F. Moreu, A. Ozdagli, M. R. Taha, and D. Mascareñas, "Noncontact dynamic displacement measurement of structures using a moving laser Doppler vibrometer," *Journal of Bridge Engineering*, vol. 24, no. 9, 2019.

[11] S. Bhowmick, S. Nagarajaiah, and Z. Lai, "Measurement of full-field displacement time history of a vibrating continuous edge from video," *Mechanical Systems and Signal Processing*, vol. 144, Article ID 106847, 2020.

[12] S. Bhowmick and S. Nagarajaiah, "Identification of full-field dynamic modes using continuous displacement response estimated from vibrating edge video," *Journal of Sound and Vibration*, vol. 489, Article ID 115657, 2020.

[13] D. Feng and M. Q. Feng, "Computer vision for SHM of civil infrastructure: from dynamic response measurement to damage detection-A review," *Engineering Structures*, vol. 156, pp. 105–117, 2018.

[14] J. Jiao, J. Guo, K. Fujita, and I. Takewaki, "Displacement measurement and nonlinear structural system identification: a vision-based approach with camera motion correction using planar structures," *Structural Control and Health Monitoring*, vol. 28, no. 8, p. 2761, 2021.

[15] S. S. Yu, Z. F. Xu, Z. Y. Su, and J. Zhang, "Two flexible vision-based methods for remote deflection monitoring of a long-span bridge," *Measurement*, vol. 181, Article ID 109658, 2021.

[16] D. Jana and S. Nagarajaiah, "Computer vision-based real-time cable tension estimation in Dubrovnik cable-stayed bridge using moving handheld video camera," *Structural Control and Health Monitoring*, vol. 28, no. 5, p. 2713, 2021.

[17] G. Jeon, S. Kim, S. Ahn, H. K. Kim, and H. Yoon, "Vision-based automatic cable displacement measurement using Cable-ROI Net and Uni-KLT," *Structural Control and Health Monitoring*, vol. 29, no. 8, p. 2977, 2022.

[18] Y. Xu, J. Zhang, and J. Brownjohn, "An accurate and distraction-free vision-based structural displacement measurement method integrating Siamese network based tracker and correlation-based template matching," *Measurement*, vol. 179, Article ID 109506, 2021.

[19] T. Wu, L. Tang, P. Du, N. A. Liu, Z. X. Zhou, and X. L. Qi, "Non-contact measurement method of beam vibration with laser stripe tracking based on tilt photography," *Measurement*, vol. 187, Article ID 110314, 2022.

[20] J. Lee, K. C. Lee, S. Jeong, Y. J. Lee, and S. H. Sim, "Long-term displacement measurement of full-scale bridges using camera ego-motion compensation," *Mechanical Systems and Signal Processing*, vol. 140, Article ID 106651, 2020.

[21] S. Jiang and J. Zhang, "Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 6, pp. 549–564, 2020.

[22] F. Ni, Z. He, S. Jiang, W. Wang, and J. Zhang, "A Generative adversarial learning strategy for enhanced lightweight crack delineation networks," *Advanced Engineering Informatics*, vol. 52, no. 2022, Article ID 101575, 2022.

[23] W. W. Greenwood, J. P. Lynch, and D. Zekkos, "Applications of UAVs in civil infrastructure," *Journal of Infrastructure Systems*, vol. 25, no. 2, Article ID 04019002, 2019.

[24] S. Bhowmick, S. Nagarajaiah, and A. Veeraraghavan, "Vision and deep learning-based algorithms to detect and quantify cracks on concrete surfaces from UAV videos," *Sensors*, vol. 20, no. 21, p. 6299, 2020.

[25] B. F. Spencer, V. Hoskere, and Y. Narazaki, "Advances in computer vision-based civil infrastructure inspection and monitoring," *Engineering*, vol. 5, no. 2, pp. 199–222, 2019.

[26] Y. J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.

[27] D. Kang and Y. J. Cha, "Autonomous UAVs for structural health monitoring using deep learning and an ultrasonic beacon system with geo-tagging," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 10, pp. 885–902, 2018.

[28] Y. J. Cha and D. Kang, "Damage detection with an autonomous UAV using deep learning," *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018*, vol. 10598, pp. 7–14, 2018.

[29] R. Ali, D. Kang, G. Suh, and Y. J. Cha, "Real-time multiple damage mapping using autonomous UAV and deep faster region-based neural networks for GPS-denied structures," *Automation in Construction*, vol. 130, Article ID 103831, 2021.

[30] Y. J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyüköztürk, "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 9, pp. 731–747, 2018.

[31] C. Zhang, Y. Tian, and J. Zhang, "Complex image background segmentation for cable force estimation of urban bridges with drone-captured video and deep learning," *Structural Control and Health Monitoring*, vol. 29, no. 4, p. 2910, 2022.

[32] Y. D. Tian, C. Zhang, S. Jiang, J. Zhang, and W. H. Duan, "Noncontact cable force estimation with unmanned aerial vehicle and computer vision," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 1, pp. 73–88, 2021.

[33] G. F. Chen, Q. Liang, W. T. Zhong, X. J. Gao, and F. S. Cui, "Homography-based measurement of bridge vibration using UAV and DIC method," *Measurement*, vol. 170, Article ID 108683, 2021.

[34] H. Yoon, J. Shin, and B. F. Spencer, "Structural displacement measurement using an unmanned aerial system," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 3, pp. 183–192, 2018.

[35] X. Bai and M. J. Yang, "UAV based accurate displacement monitoring through automatic filtering out its camera's translations and rotations," *Journal of Building Engineering*, vol. 44, Article ID 102992, 2021.

[36] Y. Han, G. Wu, and D. Feng, "Vision-based displacement measurement using an unmanned aerial vehicle," *Structural Control and Health Monitoring*, vol. 29, no. 10, p. 3025, 2022.

[37] F. T. Ni, J. Zhang, and Z. Q. Chen, "Pixel-level crack delineation in images with convolutional feature fusion," *Structural Control and Health Monitoring*, vol. 26, no. 1, p. 2286, 2019.

[38] M. F. Huang, B. Y. Zhang, W. J. Lou, and A. Kareem, "A deep learning augmented vision-based method for measuring dynamic displacements of structures in harsh environments," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 217, Article ID 104758, 2021.

[39] J. S. Zhu, C. Zhang, Z. Y. Lu, and X. T. Li, "A multi-resolution deep feature framework for dynamic displacement measurement of bridges using vision-based tracking system," *Measurement*, vol. 183, Article ID 109847, 2021.

[40] R. Yang, Y. Li, D. Zeng, P. Guo, and D. I. C. Deep, "Deep DIC: deep learning-based digital image correlation for end-to-end displacement and strain measurement," *Journal of Materials Processing Technology*, vol. 302, Article ID 117474, 2022.

[41] D. C. Shi, E. Sabanovic, L. Rizzetto et al., "Deep learning based virtual point tracking for real-time target-less dynamic displacement measurement in railway applications," *Mechanical*

Systems and Signal Processing, vol. 166, Article ID 108482, 2022.

[42] X. S. Gao, X. R. Hou, J. Tang, and H. F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 930–943, 2003.

[43] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," *Proceedings of the seventh ieee international conference on computer vision*, vol. 1, pp. 666–673, 1999.

[44] S. Q. Ren, K. M. He, R. Girshick, J. Sun, and R.-C. N. N. Faster, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[45] H. Law and J. Deng, "Cornernet: detecting objects as paired keypoints," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 734–750, Glasgow, UK, September 2018.

[46] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: exceeding yolo series in 2021," 2021, https://arxiv.org/abs/2107.08430.

[47] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, Honolulu, HI, USA, July 2017.

[48] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot MultiBox detector," *Computer Vision – ECCV 2016*, vol. 9905, pp. 21–37, 2016.

[49] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649, Beijing, China, September 2017.

[50] Y. J. Cha, J. G. Chen, and O. Büyüköztürk, "Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters," *Engineering Structures*, vol. 132, pp. 300–313, 2017.

[51] D. H. Kang and Y. J. Cha, "Efficient attention-based deep encoder and decoder for automatic crack segmentation," *Structural Health Monitoring*, vol. 21, no. 5, pp. 2190–2205, 2022.

[52] W. Choi and Y. J. Cha, "SDDNet: real-time crack segmentation," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 8016–8025, 2020.

[53] K. W. Duan, S. Bai, L. X. Xie, H. G. Qi, Q. M. Huang, and Q. Tian, "CenterNet: keypoint triplets for object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6568–6577, Seoul, Korea (South), November 2019.

[54] Z. Tian, C. H. Shen, H. Chen, and T. He, "FCOS: fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9626–9635, Seoul, Korea (South), November 2019.

[55] Y. Xin, M. Pawlak, and S. Liao, "Accurate computation of Zernike moments in polar coordinates," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 581–587, 2007.

[56] B. Xiong, Q. Zhang, and V. Baltazart, "On quadratic interpolation of image cross-correlation for subpixel motion extraction," *Sensors*, vol. 22, no. 3, p. 1274, 2022.

[57] L. Zhang, B. Wu, B. Huang, and P. Li, "Nonlinear estimation of subpixel proportion via kernel least square regression," *International Journal of Remote Sensing*, vol. 28, no. 18, pp. 4157–4172, 2007.

[58] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, 1981.

[59] W. Chen, D. Jana, A. Singh et al., "Measurement and identification of the nonlinear dynamics of a jointed structure using full-field data, Part I: measurement of nonlinear dynamics," *Mechanical Systems and Signal Processing*, vol. 166, Article ID 108401, 2022.

[60] J. Javh, J. Slavič, and M. Boltežar, "High frequency modal identification on noisy high-speed camera data," *Mechanical Systems and Signal Processing*, vol. 98, pp. 344–351, 2018.