

## Research Article

# Vision-Based Three-Dimensional Reconstruction and Monitoring of Large-Scale Steel Tubular Structures

Yunchao Tang <sup>1,2</sup> Mingyou Chen,<sup>3</sup> Yunfan Lin,<sup>1,2</sup> Xueyu Huang,<sup>1,2</sup> Kuangyu Huang,<sup>3</sup> Yuxin He,<sup>3</sup> and Lijuan Li <sup>4</sup>

<sup>1</sup>College of Urban and Rural Construction, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

<sup>2</sup>Academy of Smart Agricultural Innovations, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

<sup>3</sup>Key Laboratory of Key Technology on Agricultural Machine and Equipment, Collage of Engineering, South China Agricultural University, Guangzhou 510642, China

<sup>4</sup>School of Civil and Transportation Engineering, Guangdong University of Technology, Guangzhou 510000, China

Correspondence should be addressed to Yunchao Tang; [ryan.twain@gmail.com](mailto:ryan.twain@gmail.com) and Lijuan Li; [lilj@gdut.edu.cn](mailto:lilj@gdut.edu.cn)

Received 24 December 2019; Revised 15 August 2020; Accepted 18 August 2020; Published 18 September 2020

Academic Editor: Xuemei Liu

Copyright © 2020 Yunchao Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A four-ocular vision system is proposed for the three-dimensional (3D) reconstruction of large-scale concrete-filled steel tube (CFST) under complex testing conditions. These measurements are vitally important for evaluating the seismic performance and 3D deformation of large-scale specimens. A four-ocular vision system is constructed to sample the large-scale CFST; then point cloud acquisition, point cloud filtering, and point cloud stitching algorithms are applied to obtain a 3D point cloud of the specimen surface. A point cloud correction algorithm based on geometric features and a deep learning algorithm are utilized, respectively, to correct the coordinates of the stitched point cloud. This enhances the vision measurement accuracy in complex environments and therefore yields a higher-accuracy 3D model for the purposes of real-time complex surface monitoring. The performance indicators of the two algorithms are evaluated on actual tasks. The cross-sectional diameters at specific heights in the reconstructed models are calculated and compared against laser rangefinder data to test the performance of the proposed algorithms. A visual tracking test on a CFST under cyclic loading shows that the reconstructed output well reflects the complex 3D surface after correction and meets the requirements for dynamic monitoring. The proposed methodology is applicable to complex environments featuring dynamic movement, mechanical vibration, and continuously changing features.

## 1. Introduction

Three-dimensional (3D) visual information is the most intuitive data available to an intelligent machine as it attempts to sense the external world [1–3]. Vision 3D reconstruction technology can be utilized to acquire the spatial information of target objects for efficient and accurate noncontact measurement [4, 5]. It is an effective approach to tasks such as real-time target tracking, quality monitoring, and surface data acquisition; further, it is the key to realizing automatic, intelligent, and safe machine operations [6–15].

In the field of civil engineering, researchers struggle to reveal the failure mechanisms of certain materials or structures in seeking the exact properties of composite

materials. Traditional contact measurement methods rely on strain gauges, displacement meters, or other technologies, which may be inconvenient and inefficient. A vision sensor can comprehensively reveal the optical information of the target surface, allow the user to develop a highly targeted measurement scheme for different targets, and achieve high precision and noncontact measurement.

The construction of the vision system and its working process differ slightly across different measurement distances and types of target. For large-scale targets at long distances, the comprehensiveness of sampling and the improvement of the ability of the vision system to resist long-distance interference are the primary design considerations. Multiple cameras and even other types of sensors can be

used together to enhance the system's stability [16–18]. For close-range targets (e.g., within 1–2 m), the measurement accuracy of the vision system is an important consideration. Appropriate imaging models and distortion correction models are necessary to build high-performance visual frameworks that achieve specific high-precision measurement and inspection tasks.

At close range, a measured object can be global or local [19–21]. Structural monitoring tasks rarely require the use of visual systems with very small measurement distances, as the focus of attention in the field of civil engineering tends to be large structures. The real-time performance of the vision system must be further improved to suite deformed structures. The robustness of the 3D reconstruction algorithm also should be strengthened, as the geometric parameters vary throughout the deformation process [22, 23]. Researchers and developers in the computer vision field also struggle to effectively track and measure dynamic surface-deformed objects with stereo vision. The 3D reconstruction of curved surfaces under large fields of view (FOVs) is particularly challenging in terms of full-field dynamic tracking [24]. The core algorithm is the key component of any tracking system. Problematic core algorithms restrict the application of 3D visual reconstruction technology including omnidirectional sampling and high-quality point cloud stitching.

Existing target surface monitoring techniques based on 3D reconstruction include monocular vision, binocular vision, and multivision methods [15, 21, 25, 26]. The monocular vision method cannot directly reveal 3D visual information; it must be restored through Structure from Motion (SFM) technology [27]. SFM works by extracting and matching the feature points of the images taken by a single camera at different positions, so as to correlate the images of each frame and calculate the geometric relationship between the cameras at each position to triangulate the spatial points. The monocular vision method has a simple hardware structure and is easily operated but is disadvantaged by the instability of the available feature point extraction algorithms.

The binocular vision method is also based on feature point matching and triangulation techniques. However, the positional relationship of the cameras in the binocular vision system is fixed, so the geometric relationship between the cameras can be obtained offline with high-precision calibration objects. This results in better measurement performance in complex surface monitoring tasks than the monocular vision approach. Both methods are limited to their narrow FOVs, however, and do not allow users to sample large-scale information [28, 29], which is not conducive to high-quality structural monitoring.

The construction of a multivision system, supported by model solutions and error analysis methodology under coordinate correlation theory, is the key to successful omnidirectional sampling. The concept is similar to that of the facial soft tissue measurement method. The multiangle information of the target is sampled before the 3D reconstruction is completed [22]. However, for real-time structural monitoring tasks, the visual system also must

complete accurate size measurement. Candau et al. [30], for example, correlated two independent binocular vision systems with a calibration object while applying a spray on an elastic target object as a random marker for the dynamic mechanical analysis of an elastomer. Zhou et al. [31] binary-encoded a computer-generated standard sinusoidal fringe pattern. Shen et al. [32] conducted 3D profilometric reconstruction via flexible sensing integral imaging with object recognition and automatic occlusion removal. Liu et al. [29] automatically reconstructed a real, 3D human body in motion as captured by multiple RGB-D (depth) cameras in the form of a polygonal mesh; this method could, in practice, help users to navigate virtual worlds or even collaborative immersive environments. Malesa et al. [33] used two strategies for the spatial stitching of data obtained by multicamera digital image correlation (DIC) systems for engineering failure analysis: one with overlapping FOVs of 3D DIC setups and another with distributed 3D DIC setups that have not-necessarily-overlapping FOVs.

The above point cloud stitching applications transform the point cloud into a uniform coordinate system by coordinate correlation. In demanding situations, the precision of the stitched point cloud may be decisive, while the raw output of the coordinate correlation is ineffective. Persistent issues with dynamic deformation, illumination, vibration, and characteristic changes as well as visual measurement error caused by equipment and instrument deviations yet restrict the efficacy of high-precision 3D reconstruction applications. To this effect, there is demand for new techniques to analyze point cloud stitching error and for designing novel correction methods.

In addition to classical geometric methods and optical methods, deep neural networks have also received increasing attention in the 3D vision field due to their robustness. Sinha et al. [34] obtained the topology and structure of 3D shape by means of coding, so that convolutional neural network (CNN) could be directly used to learn 3D shapes and therefore perform 3D reconstruction; Li et al. [35] combined structured light techniques and deep learning to calculate the depth of targets. These methods alone outperform traditional methods on occlusion and untextured areas; they can also be used as a complement to traditional methods. Zhang et al. [36] proposed a method for measuring the distance of a given target using deep learning and binocular vision methods, where target detection network methodology and geometric measurement theory were combined to obtain 3D target information. Sun et al. [37] designed a CNN and established a multiview system for high-precision attitude determination, which effectively mitigates the lack of reliable visual features in visual measurement. Yang et al. [38] established a binocular stereo vision system combined with online deep learning technology for semantic segmentation and ultimately generated an outdoor large-scale 3D dense semantic map.

Our research team has conducted several studies combining machine vision and 3D reconstruction in various attempts to address the above problems related to omnidirectional sampling, high-accuracy measurement, and robust calculation [39–56]. In the present study, we focus on

stitching error and the recovery of multiview point clouds for the high-accuracy monitoring of large-scale CFST specimens. Existing multiview point cloud stitching technology is further improved here via traditional and deep learning methods. Our goal is to examine the structure and material properties in a noncontact manner with high-accuracy, stable, and robust performance. We ran a construction-and-error analysis of a multivision model followed by improvements to the high-quality point cloud correction algorithms to achieve real-time correction and reconstruction of large-scale CFST structures. The point cloud correction algorithms, which center on the stitching error of the multiview point cloud, are implemented via geometry-based algorithm and deep-learning-based algorithm. Since the proposed point cloud correction algorithms are constructed based on the geometric and spatial characteristics of the target, they are adaptive and efficient under complex conditions featuring dynamic movement, mechanical vibration, and continuously changing features. We hope that the observations discussed below will provide theoretical and technical support in improving the monitoring performance of current visual methods for CFSTs and others.

## 2. Materials and Methods

The algorithms and seismic test operations used in this study for obtaining the dynamic CFST surfaces by four-ocular vision system are described in this section. The dynamic specimen surfaces were obtained, and the relevant geometric parameters were extracted by means of stereo rectification, point cloud acquisition, point cloud filtering, point cloud stitching, and point cloud correction algorithms. A flow chart of this process is shown in Figure 1.

**2.1. Stereo Rectification.** A 2D circle grid with 0.3 mm manufacturing error was placed in 20 different poses to provide independent optical information for calibration, as shown in Figure 2.

We applied camera calibration based on Zhang's method [57] to determine the matrices of each individual camera:

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{M}[\mathbf{R} : \mathbf{T}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (1)$$

where  $[u \ v \ 1]^T$  is the image coordinate of the circle center,  $[X_w \ Y_w \ Z_w]^T$  is the world coordinate of the circle center,  $Z_c$  is the depth of the circle center, and  $M$  and  $[\mathbf{R} \ \mathbf{T}]$  are the camera matrix and extrinsic matrix of the single camera, respectively. Structural parameters of the binocular cameras were calculated based on the extrinsic matrices to realize stereo calibration:

$$\begin{cases} \mathbf{R}_T = \mathbf{R}_r \mathbf{R}_l^{-1}, \\ \mathbf{T}_T = \mathbf{T}_r - \mathbf{R}_r \mathbf{R}_l^{-1} \mathbf{T}_l, \end{cases} \quad (2)$$

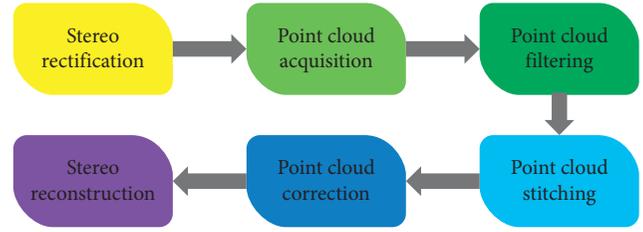


FIGURE 1: Algorithm flow overview.

where  $[\mathbf{R}_T \ \mathbf{T}_T]$  is the structural parameter of binocular cameras and  $[\mathbf{R}_l \ \mathbf{T}_l]$  and  $[\mathbf{R}_r \ \mathbf{T}_r]$  are the extrinsic matrices of the left and right cameras, respectively. The camera calibration results are discussed in Section 3.

After stereo calibration, we used the classical Bouguet's method [58] to implement stereo rectification. Corresponding points were placed from left and right images into the same row as shown in Figure 3.

**2.2. Point Cloud Acquisition.** The camera in our setup samples images after stereo rectification. It is necessary to obtain as much target surface information as possible to perform 3D reconstruction and obtain accurate geometric parameters of the target, so we sought to generate as dense a 3D point cloud as possible. The triangulation-based calculation for dense 3D points is as follows:

$$\begin{cases} X_C = \frac{x_l T_x}{d}, \\ Y_C = \frac{y_l T_x}{d}, \\ Z_C = \frac{f T_x}{d}, \end{cases} \quad (3)$$

where  $[X_C \ Y_C \ Z_C]^T$  is the camera coordinate of the target,  $[x_l \ y_l]^T$  is the left imaging plane coordinate of the target,  $T_x$  is the length of the baseline,  $f$  is the focal length of the camera, and  $d = x_l - x_r$  is the disparity. Here, we used a classical 3D stereo matching algorithm [59] to generate a dense disparity map from each pixel in the images.

The point cloud acquisition process is shown in Figure 4.

**2.3. Point Cloud Filtering.** Point cloud filtering is one of the key steps in point cloud postprocessing. It involves denoising and structural optimization according to the geometric features and spatial distribution of the point cloud, which yields a relatively compact and streamlined point cloud structure for further postprocessing. The point cloud filtering process also involves the removal of large-scale and small-scale noise. Large-scale noise is created by scattering outside the main structure of the point cloud over a large range; small-scale noise is caused by small fluctuations adhering to the vicinity of the point cloud structure. Large-scale noise is generally attributed to noninteresting objects and mismatches, while small-scale noise relates to the spatial resolution limitations of the vision system. A flow

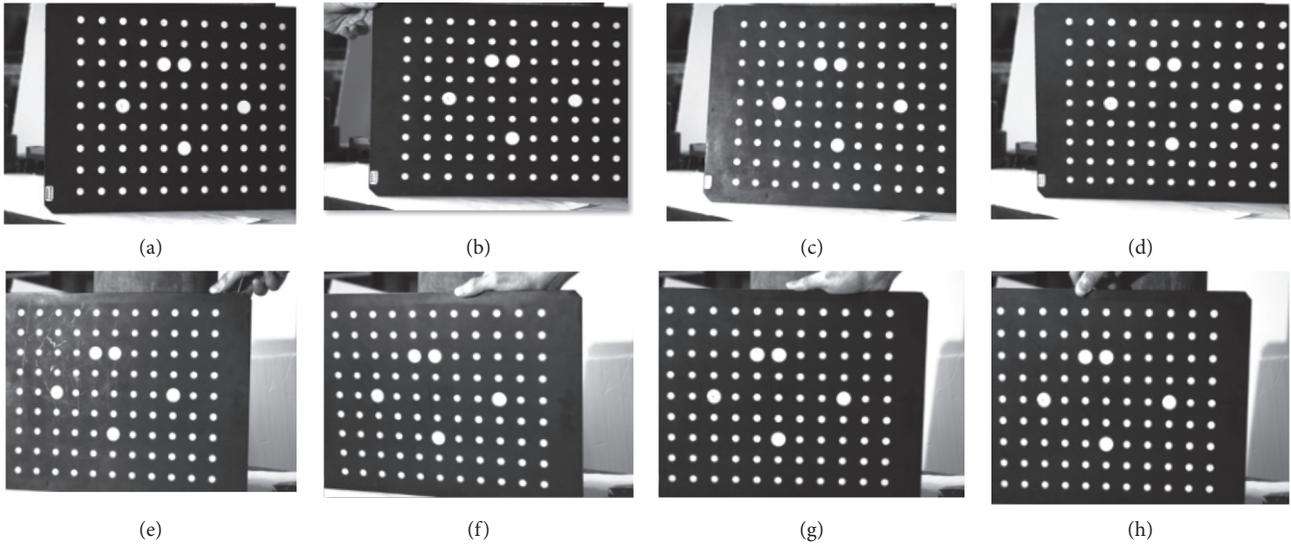


FIGURE 2: Circle grid: various poses for stereo calibration.

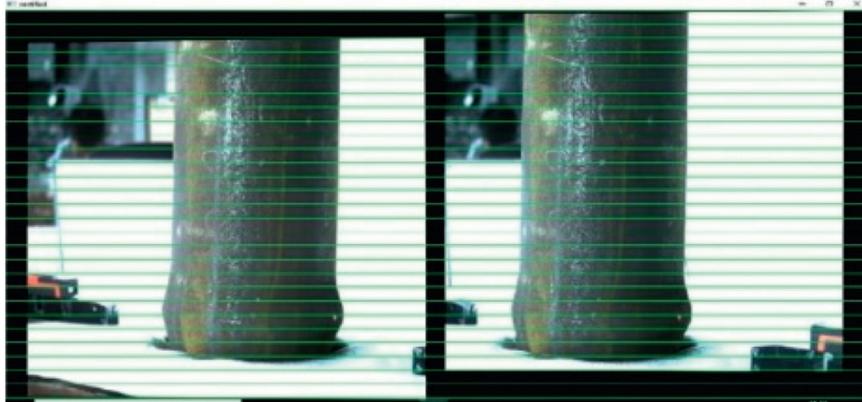


FIGURE 3: Stereo rectification.

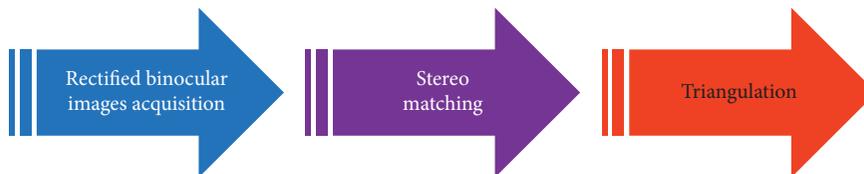


FIGURE 4: Flow chart of point cloud acquisition.

chart of the point cloud filtering process is shown in Figure 5.

**2.3.1. Pass-Through Filtering.** The pass-through filter [60] defines inner and outer points according to a cuboid with edges parallel to three coordinate axes. Points that fall outside of the cuboid are deleted. The cuboid is expressed as follows:

$$\{(X_C, Y_C, Z_C) \mid X_{\min} \leq X_C \leq X_{\max}, Y_{\min} \leq Y_C \leq Y_{\max}, Z_{\min} \leq Z_C \leq Z_{\max}\}, \quad (4)$$

where  $[X_C \ Y_C \ Z_C]^T$  is the inlier defined using the pass-through filter,  $X_{\min}$ ,  $Y_{\min}$ , and  $Z_{\min}$  are the minimum cut-off thresholds in the directions of  $X_C$ ,  $Y_C$ , and  $Z_C$  axes, respectively, and  $X_{\max}$ ,  $Y_{\max}$ , and  $Z_{\max}$  are the maximum cut-off thresholds in the directions of  $X_C$ ,  $Y_C$ , and  $Z_C$  axes, respectively.

The pass-through filter can be used to roughly obtain the main point cloud structure and minimize the cost of calculation. The cut-off threshold should be conservative enough to ensure that the cuboid is consistently larger than expected, which prevents the accidental deletion of the main structure of the point cloud. A fixed and reliable cut-off

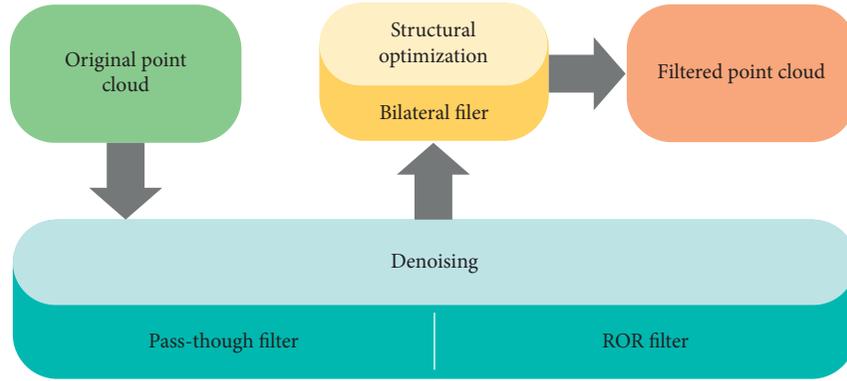


FIGURE 5: Flow chart of point cloud filter.

threshold can be determined by strictly controlling the relative position between the camera and the specimen.

In the task of CFST deformation monitoring, the relative position of the specimen and the optical system can be determined in advance so that the threshold in each direction can be determined accurately. It is worth noting that although the specimen discussed here is cylindrical, it had a large amplitude swing during the test; the cuboid area has better adaptability than the cylindrical area in the filtering task for this reason. A diagram of the pass-through filter is shown in Figure 6.

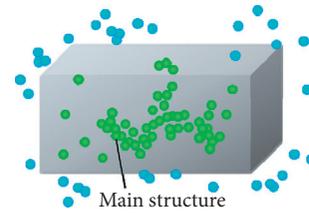


FIGURE 6: Schematic diagram of pass-through filter.

**2.3.2. Bilateral Filtering.** The bilateral filter [61] can remove small-scale noise, that is, anomalous 3D points that are intermingled with the surface of the point cloud. The noise is closely attached to the surface and readily causes fluctuations that can drive down the accuracy of the reconstructed 3D model. The primary purpose of bilateral filtering is to move this superficial noise along the normal direction of the model and gradually correct its coordinates and position. Bilateral filtering smooths the point cloud surface while retaining the necessary edge information. Noise moves along the following direction:

$$p'_i = p_i + \alpha \cdot n, \quad (5)$$

where  $p'_i$  is the filtered point,  $p_i$  is the original point,  $\alpha$  is the weight of bilateral filtering, and  $n$  is the normal vector of  $p_i$ . Detailed descriptions can be found in [60].

**2.3.3. ROR Filtering.** The radius outlier removal (ROR) filter [62] distinguishes inliers from outliers based on the number of neighbors of each element in the point cloud. For any point  $P_i$  in point cloud  $\{P\}$ , a sphere is constructed with a radius of  $r$  centered on it:

$$\{S_i \mid \|S_i - P_i\| \leq r\}. \quad (6)$$

If the number of elements in  $\{S_i\}$  is less than  $N$ , then the ROR filter regards  $P_i$  as noise, as shown in Figure 7. The threshold  $N$  is selected according to a certain ratio of the number of elements in the point cloud:

$$N = c \cdot k, \quad (7)$$

where  $k$  is the number of elements of the point cloud and  $c$  is a scale factor. Generally,  $c$  is 2%–5%.

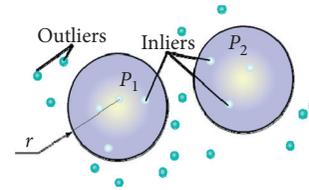


FIGURE 7: Schematic diagram of ROR filtering process.

The original and filtered point clouds are shown in Figure 8.

**2.4. Point Cloud Stitching.** Multiple cameras can be employed in the sampling task to gather as much information as possible about the target. As shown in Figure 9, four cameras constitute two pairs of binocular cameras in our experimental setup. The two left camera coordinate systems are denoted as “CCS-A” and “CCS-B,” respectively, and the world coordinate system is denoted as “WCS.” The principle of point cloud stitching is to solve the coordinate transformation matrices  ${}^C_A T_W$  and  ${}^C_B T_W$  within WCS to CCS-A and CCS-B and then transfer the point cloud from CCS-A and CCS-B to WCS to complete the coordinate correlation. The WCS can be established via a high-precision calibration board with known parameters.

Our calibration board has 99 circular patterns. The coordinates of all the centers on CCS-A, CCS-B, and WCS were combined column-by-column to obtain coordinate data matrices  $M_{CA}$ ,  $M_{CB}$ , and  $M_W$  (sized  $3 \times 99$ ):

$$\begin{aligned} M_{CA} &= {}^C_A T_W M_W, \\ M_{CB} &= {}^C_B T_W M_W. \end{aligned} \quad (8)$$

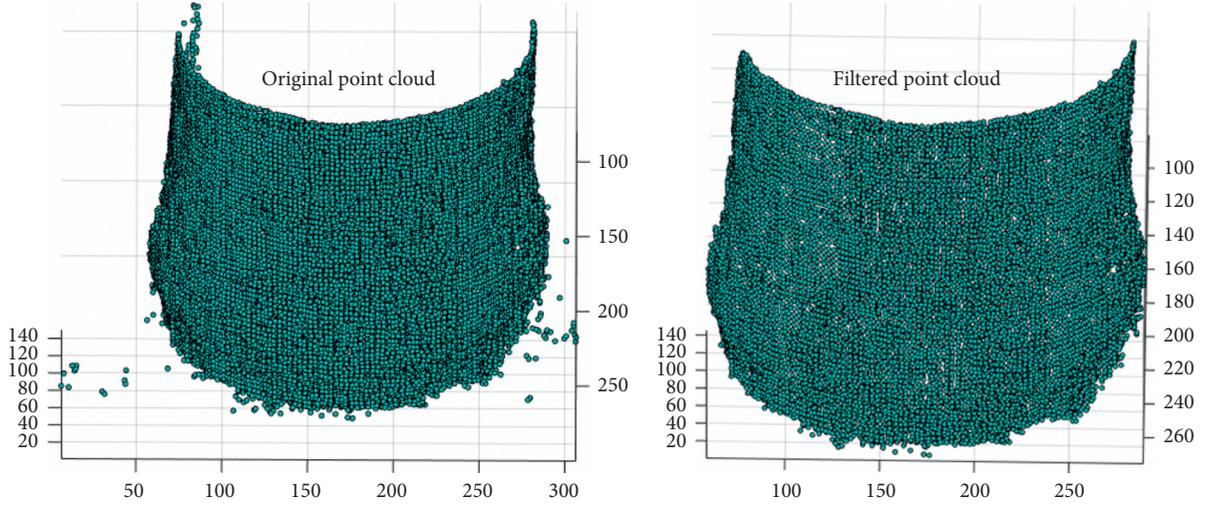


FIGURE 8: Original and filtered point cloud.

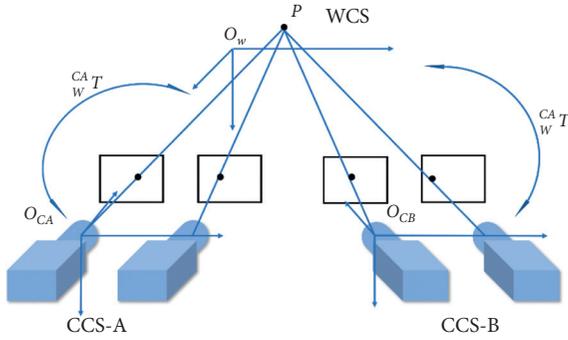


FIGURE 9: Coordinate transformation of multicamera system.

The following is obtained through matrix transformation according to the principal of least squares:

$$\begin{aligned} {}_W^C A^T &= M_{CA} M_W^T (M_W M_W^T)^{-1}, \\ {}_W^C B^T &= M_{CB} M_W^T (M_W M_W^T)^{-1}. \end{aligned} \quad (9)$$

The coordinate data matrices  $Q_{CA}$  and  $Q_{CB}$  of the target relative to CCS-A and CCS-B were calculated according to equation (3) and then converted into the WCS as follows:

$$\begin{aligned} Q_{WA} &= {}_W^C A^T Q_{CA}, \\ Q_{WB} &= {}_W^C B^T Q_{CB}. \end{aligned} \quad (10)$$

The point cloud stitching process is illustrated in Figure 10.

## 2.5. Point Cloud Correction

**2.5.1. Stitching Correction.** The stress state of the devices in this setup is very complex due to the combination of large axial pressure, cyclic tension, and fastening force. Random, difficult-to-measure shocks and vibrations often occur during such tests, which cause the camera frames to move slightly on the ground, although they are fixed in advance

and are far from the specimen (about 1 m). The effect of the slight movement of the camera frames may cause the stitched point clouds to appear staggered, as shown in Figure 11, which can be destructive in demanding monitoring missions. Assume that the translation vector and rotation matrix of two staggered point clouds caused by the complex loading conditions are  $P_M$  and  $R_M$ .

We designed geometry-based and deep-learning-based algorithms in this study to correct two staggered point clouds. We assume that the optical axes of the cameras remain parallel to the (horizontal) ground surface after they move, so the moving direction of the point cloud is perpendicular to the  $Y_W$  axis and  $p_y = 0$ . There is no rotation between the two point clouds:

$$\begin{aligned} P_M &= [p_x \ 0 \ p_z]^T, \\ R_M &= E, \end{aligned} \quad (11)$$

where  $E$  refers to the unit matrix. The translation vector mentioned above was acquired by nonlinear least squares; then the staggered point cloud was translated accordingly. Deformation of the upper part of the specimen was not severe at any point in the test, so it is close to an ideal cylinder and even the bottom presents obvious deformation. In order to simplify the calculation of  $P_M$ , the upper part of the specimen can be regarded as a standard cylinder.

We set a plane  $Y_W = h'_0$  parallel to the  $X_W O_W Z_W$  plane to intercept the upper part of the specimen and obtain a circular arc. The expression of the arc in 3D space is

$$\begin{cases} F_{\text{circle}}(X_W, Z_W) = (X_W - a)^2 + (Z_W - b)^2 - r^2 = 0, \\ Y_W = h'_0, \end{cases} \quad (12)$$

where  $[a \ b]^T$  is the projection of the circle center on the  $X_W O_W Z_W$  plane and  $r$  is the radius of the circle. The optimization target of nonlinear least squares is

$$\min S_{\text{circle}}(a, b, r) = \sum_{i=1}^n F_{\text{circle}}^2(X_W^{(i)}, Z_W^{(i)}), \quad (13)$$

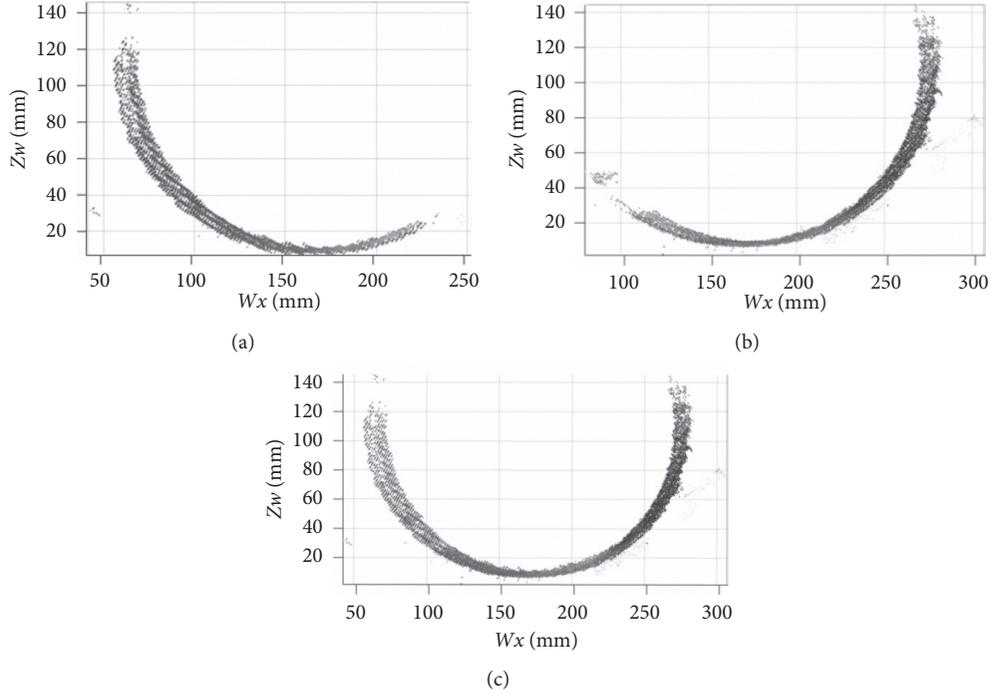


FIGURE 10: Top views of stitched point cloud: (a) point cloud acquired from CCS-A; (b) point cloud acquired from CCS-B; (c) stitched point cloud in WCS.

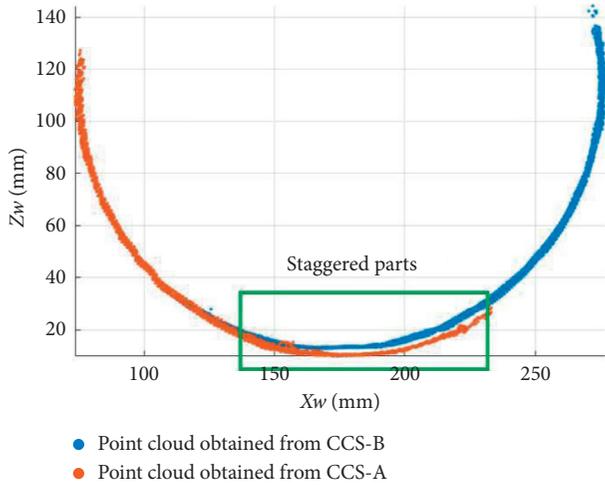


FIGURE 11: Visualization of point cloud stitching problem.

where  $S_{\text{circle}}(a, b, r)$  is the objective function of nonlinear least squares and  $[X_W^{(i)} \ Z_W^{(i)}]^T$  is the  $i$ -th sample point in the section. The optimal values  $\hat{a}$ ,  $\hat{b}$ , and  $\hat{r}$  were iteratively obtained by the Levenberg–Marquardt (LM) algorithm [63] as follows:

$$\omega_{k+1} = \omega_k - [J_{F_{\text{circle}}}(\omega_k)^T J_{F_{\text{circle}}}(\omega_k) + \mu_1 I]^{-1} \cdot g_{F_{\text{circle}}}(\omega_k), \quad (14)$$

where  $\omega_k = [a_k \ b_k \ r_k]^T$  is the value of the model parameter at the  $k$ -th iteration,  $J_{F_{\text{circle}}}(\omega_k)$  is the Jacobian matrix of the objective function at the  $k$ -th iteration,  $g_{F_{\text{circle}}}(\omega_k)$  is the gradient of the objective function at the  $k$ -th iteration,  $\mu_1$

is the damping coefficient, and  $I$  is the identity matrix. When  $h_0$  is around 150 mm, it is basically guaranteed that the edge of the section is an approximate arc.

The respective estimated values  $c_A = [\hat{a}_A, \hat{b}_A]^T$  and  $c_B = [\hat{a}_B, \hat{b}_B]^T$  for the height  $h'_0$  can be obtained according to formula (14). Figure 12 shows the stitching correction process:

$$\mathbf{P}_M = c_A - c_B. \quad (15)$$

Next, we used PointNet++ [64], a robust 3D semantic segmentation network, to extract the common parts of the point cloud and performed ICP point cloud registration [65] on them to correct the relative positions of the two clouds. This method has no stronger assumptions than the geometry-based method:

$$\begin{aligned} P_M &= [p_x \ p_y \ p_z]^T, \\ R_M &= \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \end{aligned} \quad (16)$$

where  $P_M$  and  $R_M$  are not necessarily equal to 0 and form a unit matrix.

As shown in Figure 13, we manually marked the common parts of two staggered point clouds in a large number of samples and fed them to the network for training. After the training was complete, we could use the network to identify and extract the common parts of the two given point clouds.

The blue-shaded area in Figure 14 represents the common parts extracted by the PointNet++ network, which were used to implement ICP registration for point cloud correction.

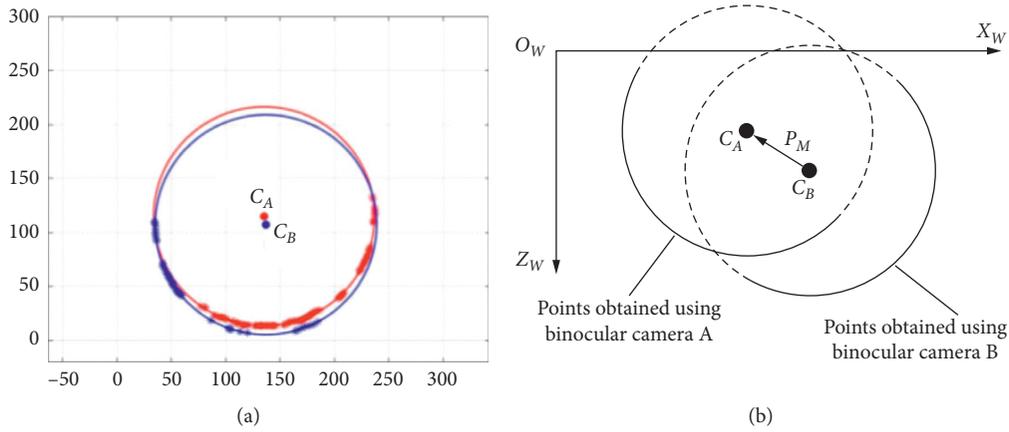


FIGURE 12: Stitching correction process: (a) calculating  $C_A$  and  $C_B$  by circle fitting; (b) calculating  $P_M$  according to  $C_A$  and  $C_B$ .

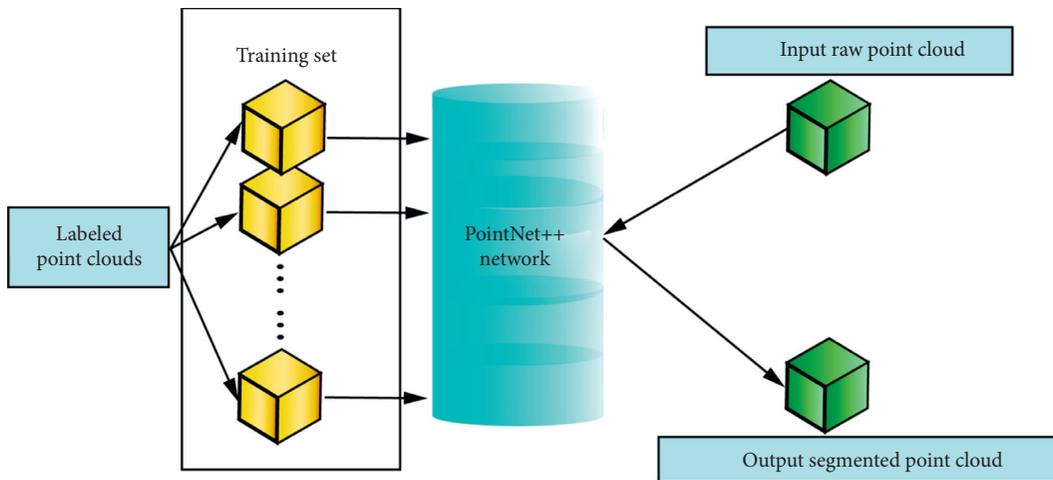


FIGURE 13: PointNet++ network training and application.

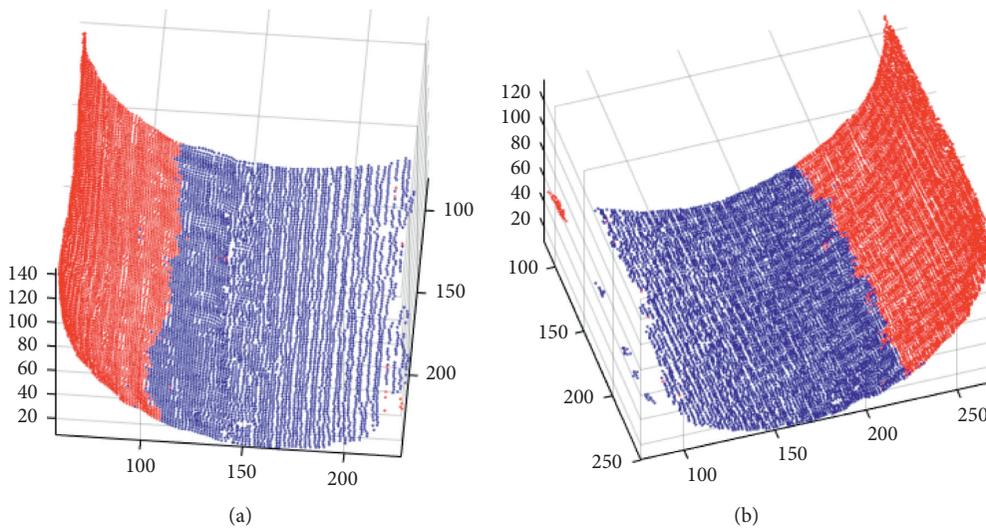


FIGURE 14: Common parts extracted by PointNet++.

2.5.2. *Establishing the Specimen Coordinate System.* In the point cloud stitching step, there is always a small angle  $A$  between the plane of the calibration board and the axis of the

specimen (Figure 15(a)). Indeed, the section in the  $Y_W$  direction of the WCS is not the actual cross section of the test piece but rather an oblique section (red line, Figure 15(a)).

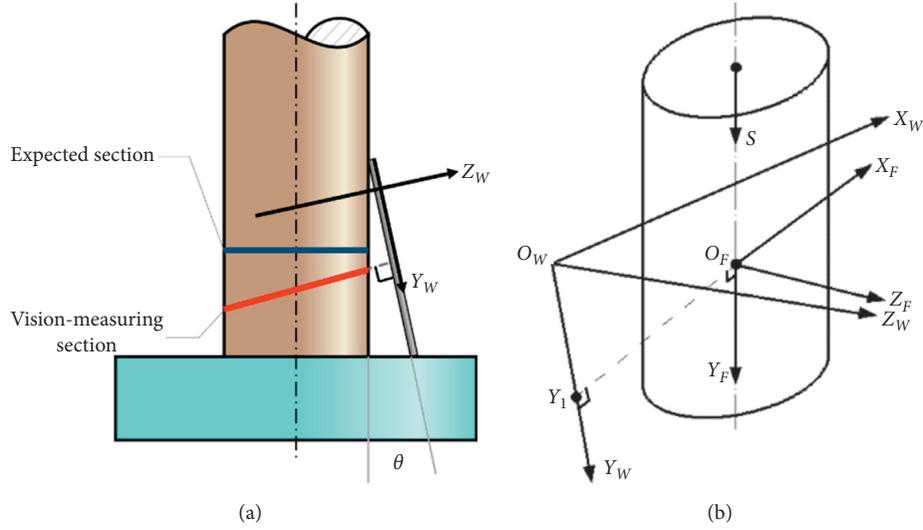


FIGURE 15: Coordinate systems involved in measurement: (a) positional relationship between world coordinate system and specimen; (b) schematics of specimen coordinate system as established.

The vision measurement technique is based on the WCS, so it is difficult to determine the actual measuring position in this setup.

To address this problem, we considered a coordinate system  $O_F - X_F Y_F Z_F$  (specimen coordinate system) fixed to the specimen. As shown in Figure 15(b), the  $Y_F$  axis coincides with the axis of the specimen. If a point cloud is transformed from the WCS to the specimen coordinate system, it is guaranteed that the section in the  $Y_F$  direction actually corresponds to the real cross section of the specimen (green line, Figure 15(a)). Therefore, the point cloud based on the specimen coordinate system is meaningful.

The specimen coordinate system was established as discussed below. The equation for an arbitrarily posed cylinder in the WCS is

$$\begin{aligned} & [A(Z_W - z_0) - C(X_W - x_0)]^2 + [A(Y_W - y_0) - B(X_W - x_0)]^2 \\ & + [B(Z_W - z_0) - C(Y_W - y_0)]^2 - R^2 = 0, \end{aligned} \quad (17)$$

where  $s = [A \ B \ C]^T$  is the unit vector of the axis of the cylinder,  $O_F = [x_0 \ y_0 \ z_0]^T$  is a 3D point on the axis of the cylinder, and  $R$  is the radius of the cylinder.

Similar to the circle-fitting process discussed above, the upper half of the cylinder was fitted using nonlinear least squares. The left side of equation (17) is denoted as  $F_{\text{cylinder}}(X_W, Y_W, Z_W)$ ; then the objective function of the cylinder fitting is

$$\begin{aligned} \min S_{\text{cylinder}}(x_0, y_0, z_0, A, B, C, R) &= \sum_{i=1}^n F_{\text{cylinder}}^2(X_W^{(i)}, Y_W^{(i)}, Z_W^{(i)}), \\ \text{subjected to } & A^2 + B^2 + C^2 = 1, \end{aligned} \quad (18)$$

where  $S_{\text{cylinder}}(x_0, y_0, z_0, A, B, C, R)$  is the objective function of nonlinear least squares and  $[X_W^{(i)} \ Y_W^{(i)} \ Z_W^{(i)}]^T$  is the  $i$ -th sample point. Similarly, equation (18) can be solved by LM algorithm:

$$w_{k+1} = w_k - \left[ J_{F_{\text{cylinder}}}(w_k)^T J_{F_{\text{cylinder}}}(w_k) + \mu_2 I \right]^{-1} \cdot g_{F_{\text{cylinder}}}(w_k), \quad (19)$$

where  $w_k = [x_{0_k} \ y_{0_k} \ z_{0_k} \ A_k \ B_k \ C_k \ R_k]^T$  is the value of the model parameter at the  $k$ -th iteration,  $J_{F_{\text{cylinder}}}(w_k)$  is the Jacobian matrix of the objective function at the  $k$ -th iteration,  $g_{F_{\text{cylinder}}}(w_k)$  is the gradient of the objective function at

the  $k$ -th iteration,  $\mu_2$  is the damping coefficient, and  $I$  is the unit matrix. According to equation (19), the optimal values of  $s$  and  $O_F$  can be iteratively obtained as follows:

$$\begin{aligned} s &= [\hat{A} \ \hat{B} \ \hat{C}]^T, \\ O_F &= [\hat{x}_0 \ \hat{y}_0 \ \hat{z}_0]^T. \end{aligned} \quad (20)$$

The direction of  $Y_W$  is different from that of the axis and  $O_F$  is located on the axis, so a unique point  $Y_1 = [0 \ y_1 \ 0]^T$  for  $\overrightarrow{O_F Y_1} \cdot s = 0$  can be determined on the  $Y_W$  axis:

$$y_1 = \frac{\hat{A}\hat{x}_0 + \hat{C}\hat{z}_0}{\hat{B}} + \hat{y}_0. \quad (21)$$

Taking  $O_F$  as the origin, a directional cylinder axis can be set in the direction of axis  $Y_F$ ; the direction of axis  $X_F$  is the same as  $\overrightarrow{O_F Y_1}$  and the direction of  $Z_F$  is perpendicular to the plane  $X_F O_F Y_F$ . Thus, the specimen coordinate system  $O_F - X_F Y_F Z_F$  is established:

$$\begin{bmatrix} e_{X_F} \\ e_{Y_F} \\ e_{Z_F} \end{bmatrix} = \begin{bmatrix} \left\| \overrightarrow{O_F Y_1} \right\|_2^{-1} \cdot \overrightarrow{O_F Y_1} \\ s \\ e_{X_F} \times e_{Y_F} \end{bmatrix}. \quad (22)$$

Next, the rotation matrix and translation vector of the specimen coordinate system to the WCS can be calculated as follows:

$$\begin{aligned} {}^W_F R &= [e_{X_F} \ e_{Y_F} \ e_{Z_F}], \\ {}^W_F P &= O_F. \end{aligned} \quad (23)$$

According to  ${}^W P = {}^W_F R {}^F P + {}^W_F P$ , the point cloud from the WCS is transformed to the specimen coordinate system as follows:

$${}^F P = {}^W_F R^{-1} ({}^W P - {}^W_F P). \quad (24)$$

**2.6. Surface Reconstruction.** After the corrected point cloud is obtained, the target surface reconstruction is completed via Poisson surface reconstruction algorithm [66]. The Poisson surface reconstruction algorithm is a triangular mesh reconstruction algorithm based on implicit functions, which approximates the surface by optimizing the interpolation of 3D point cloud data. Its stepwise implementation is discussed below.

Point cloud data containing the normal information is used as the input data of the algorithm and recorded as a sample set  $S$ . Each  $S$  includes a plurality of  $P_i$  and corresponding normal  $N_i$ . Assuming that all points are located on or near the surface of an unknown model  $M$ , the indication function of  $M$  is estimated and the Poisson equation is established with the intrinsic relationship of the gradient. The contour surface is then extracted via Marching Cubes algorithm and the reconstruction of the surface is complete. Finally, the surface model data is output.

The Poisson surface reconstruction algorithm process involves octree segmentation, vector field calculation, solving the Poisson equation, surface model generation, and surface extraction. A flow chart of this process is given in Figure 16.

### 3. Experiment

A four-ocular vision system was used to perform 3D surface tracking experiments on a CSFT column under axial and cyclic radial loads. We focused on the accuracy of two selected sample points on the 3D model, which approximately reflects the precision of the reconstructed 3D surface. Diameters of specimen cross sections were measured by laser rangefinders as standard values; then the visual measurements were compared against the standard values. Finally,

indicators for error evaluation were calculated to validate the proposed method. Three CFST specimens were selected for this experiment.

Camera calibration and stereo rectification were implemented prior to the formal initiation of the experiment. Since strong vibrations can cause subtle changes in the structural parameters of the multivision system, the system was recalibrated before each column was tested to ensure accurate 3D reconstruction. Only the parameters corresponding to Specimen 1 are presented here to illustrate the calibration target (Table 1).

During the test, an axial load was applied to the top of the specimen at a constant 10 kN. A cyclic radial load was applied along the horizontal direction with 20 kN added in each cycle. The axial load from the top varied slightly in each cycle as the specimen swung to the side under the cyclic load. To account for this, we continuously fine-tuned the axial pressure during the experiment so that the axial load of the specimen was always 10 kN. After the specimen yielded, the horizontal displacement of the press increased in each cycle; the increment is an integral multiple of the yield displacement of the specimen. Each run of the experiment ended once significant deformation or damage was observed in the bottom of the specimen.

As shown in Figure 17, four calibrated MV-EM200C industrial cameras (1600 \* 1200 resolution) were placed in front of the specimen and two laser rangefinders (1 mm accuracy) were placed symmetrically on its left and right sides. The cameras and laser rangefinders sampled data once per minute. A lighting source was used to ensure constant lighting so that all surface information could be obtained by the vision system. A press produced axial and radial loads on the specimen throughout the experiment to ultimately generate convex deformation at the bottom of the specimen. Figure 18 shows a dynamic 3D surface reconstructed according to the proposed algorithm.

As shown in Figure 19(a), the laser rangefinders and vision system collected dynamic measurements  $d_i^L$  and  $d_i^V$  once per minute. The distance of the laser point to the base  $h_0$  was determined in advance as a standard measurement. In the reconstructed surface model, a cross section of  $h_0$  height was taken to obtain a visual measurement as shown in Figure 19(b). To determine this section on the 3D model, the point with the smallest  $Y_W$  value in the model was targeted; then the section with height  $h_0$  above it was taken as the target section. The visual measurement was taken according to the distance between the leftmost and rightmost points in this section.

The initial diameters of the specimens were about 203 mm, and the tolerance was between IT12 and IT16. Line graphs of the measured values of each specimen are shown in Figure 20(a)–20(c). We used a personal computer (i3-4150 CPU, Nvidia GTX 750Ti GPU, and 8 GB RAM) to accomplish the calculation. The black lines in Figure 19 are standard measurements obtained by the laser rangefinder. The red lines are visual measurements without point cloud correction, the green lines are visual measurements after deep-learning-based correction, and the blue lines are visual measurements after geometry-based correction.

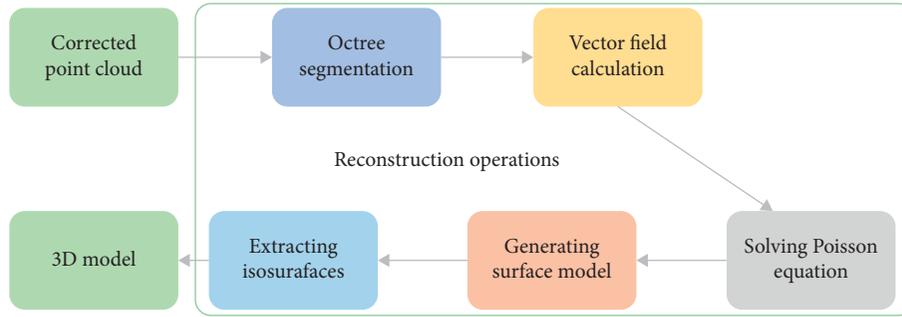


FIGURE 16: Poisson surface reconstruction algorithm.

TABLE 1: Specimen 1 calibration parameters.

Calibration parameters	Binocular cameras in CCS-A	Binocular cameras in CCS-B	Transformation matrix from CCS-A to WCS	Transformation matrix from CCS-B to WCS
$R_T$	$\begin{bmatrix} 1 & 0 & 0.01 \\ 0 & 1 & 0.03 \\ 0 & -0.03 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.01 & -0.03 \\ 0.01 & 1 & -0.02 \\ 0.04 & 0.02 & 1 \end{bmatrix}$	—	—
$T_T$	$[-105.66 \quad -0.69 \quad -3.80]^T$	$[-105.80 \quad -2.03 \quad -1.19]^T$	—	—
${}^C_W T$	—	—	$\begin{bmatrix} 0.78 & -0.01 & 0.62 & -47.45 \\ 0.04 & 1 & -0.02 & -131.01 \\ -0.62 & 0.04 & 0.78 & 1127.72 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.86 & -0.01 & -0.52 & -42.88 \\ 0.01 & 1 & 0 & -122.55 \\ 0.52 & 0 & 0.86 & 960.43 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

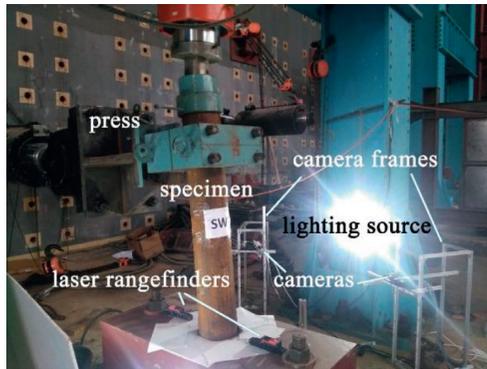


FIGURE 17: Equipment settings.



FIGURE 18: Reconstructed surface.

We collected the calculation times for all sample points as shown in Figure 20. The average calculation times of each type of correction method are also listed in Table 2. Each method took under 2.5 s; their respective averages were 1.75 s, 2.28 s, and 1.87 s. The time described here includes the time necessary to calculate a 3D point cloud from a 2D image.

In order to evaluate the performance of the vision system and the correction algorithm, the maximum absolute error ( $M$ ), mean absolute error (MAE), mean relative error (MRE), and root mean square error (RMSE) were used to evaluate the dynamic measurement error. The calculated indicators are shown in Figure 21.

Figure 21 show that the point cloud correction algorithm effectively reduces visual measurement error; it can compensate for any error caused by vibration of the press and inaccurate establishment of the WCS to provide accurate visual measurements of the 3D model. The dynamic indicator values after point cloud correction are significantly smaller than the uncorrected values, which also indicates that the proposed point cloud correction algorithm is effective.

For the deep-learning-based algorithm, the average  $M$  of each specimen is 3.00 mm, the average MAE is 1.11 mm, the average MRE is 0.52%, and the average RMSE is 1.84 mm. For the geometry-based correction algorithm, the average  $M$  of each specimen is 3.21 mm, the average MAE is 1.23 mm, the average MRE is 0.58%, and the average RMSE is 2.34 mm. These values altogether satisfy the requirements for high-accuracy measurement.



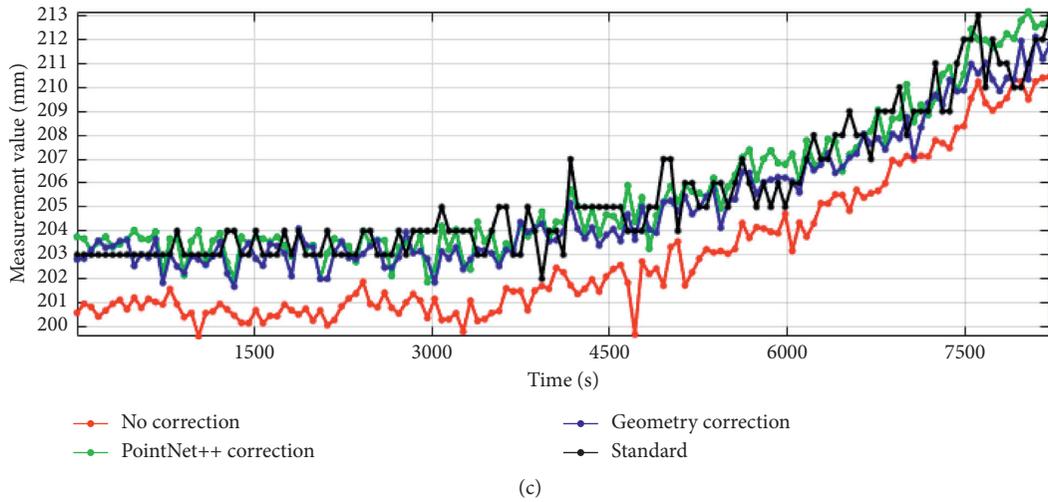


FIGURE 20: Real-time measurements: (a) real-time measurement of specimen 1; (b) real-time measurement of specimen 2; (c) real-time measurement of specimen 3.

TABLE 2: Average 3D reconstruction calculation time.

	No correction (s)	With PointNet++ based correction (s)	With geometry-based correction (s)
$n_1$ specimen 1	1.73	2.35	1.84
$n_2$ specimen 2	1.73	2.22	1.97
$n_3$ specimen 3	1.80	2.27	1.81
Average	1.75	2.28	1.87

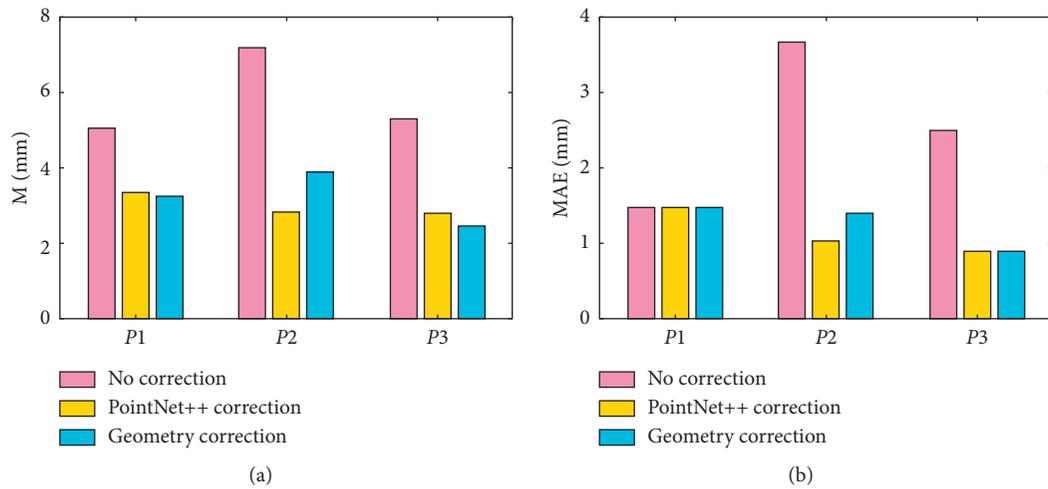


FIGURE 21: Continued.

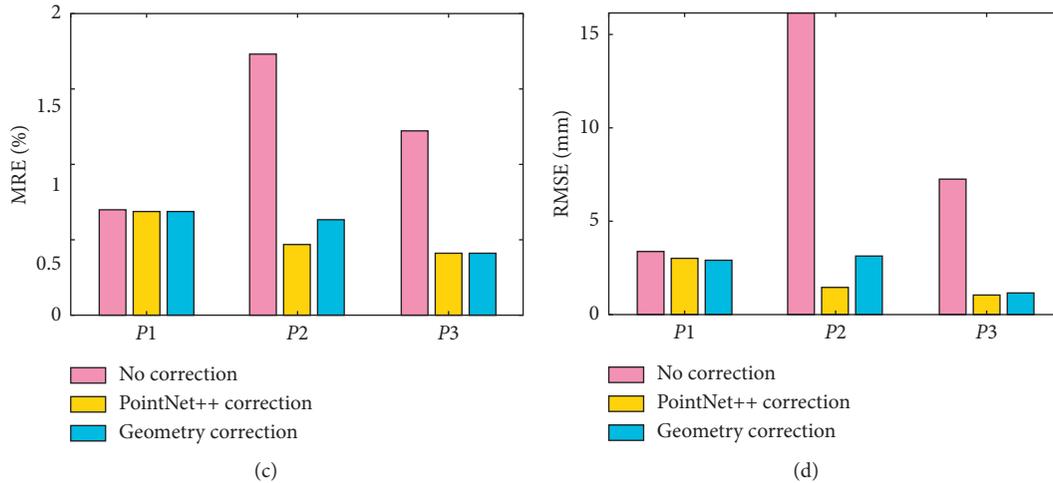


FIGURE 21: Critical indicators of vision measurement: (a)  $M$ ; (b) MAE; (c) MRE; (d) RMSE.

Both of the algorithms we developed in this study can effectively correct the spatial position of point clouds. Their effects do not significantly differ. Compared with the geometry-based algorithm, the deep-learning-based algorithm relies on weaker assumptions and thus is more general. However, it takes longer to run. When the object is irregular, the former algorithm is well applicable. When the object is cylindrical, the latter is a better choice as it is more computationally effective.

#### 4. Conclusion

In this study, we focused on a series of unfavorable factors that degrade the accuracy of surface reconstruction tasks. A point cloud correction algorithm was proposed to manage the unexpected shocks and vibration which occur under actual testing conditions and to correct the stitched point cloud obtained by multivision systems. The essential geometric parameters of the reconstructed surface were measured; then stereo rectification, point cloud acquisition, point cloud filtering, and point cloud stitching were applied to obtain a 3D model of a complex dynamic surface. In this process, a deep-learning-based algorithm and geometry-based algorithm were deployed to compensate for the stitching error of multiview point clouds and secure high-accuracy 3D structures of the target objects.

Geometric analysis and coordinate transformation were applied to design the geometry-based point cloud correction algorithm. This method is based on strong mathematical assumptions, so it has a fast calculation speed with satisfactory correction accuracy. By contrast, the deep-learning-based algorithm relies on a large number of training samples, and the forward propagation of the network is more computationally complicated than the geometry-based algorithm; it takes a longer time to accomplish point cloud correction. However, since the applicable object of the network is determined by the type of objects in the training set, it does not rely on manually designed geometric assumptions and is thus much more generalizable to different types of 3D objects.

The proposed point cloud correction algorithms make full use of the geometric and spatial characteristics of targets for error compensation, so both are more adaptive and efficient than standard-marker-based correction frameworks. They effectively enhance the accuracy of point cloud stitching over traditional methods and their effects do not significantly differ. The deep-learning-based algorithm is highly versatile, while the geometry-based algorithm is more computationally effective for cylindrical objects. They may serve as a reference for improving the accuracy of multiview, high-accuracy, and dynamic 3D reconstructions for CFSTs and other large-scale structures under complex conditions. They are also workable as is for completing tasks such as structural monitoring and data collection.

The two proposed algorithms have their limitations, and neither of them can achieve good output while ensuring satisfactory real-time performance. In the future, we will consider optimizing the structure of the PointNet++ network to make it more suitable for specific tasks and to improve its computing efficiency. We will also extract more general geometric features to improve the geometry-based algorithm, thereby improving its robustness. The advantages of these two algorithms can be combined to form a new algorithm if neither is eliminated, which depends on the specific tasks to which the method is applied.

#### Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

#### Conflicts of Interest

The authors declare that they have no conflicts of interest.

#### Authors' Contributions

Yunchao Tang and Mingyou Chen contributed equally to this work.

## Acknowledgments

This work was supported by the Key-Area Research and Development Program of Guangdong Province (2019B020223003), the Scientific and Technological Research Project of Guangdong Province (2016B090912005), and Science and Technology Planning Project of Guangdong Province (2019A050510035).

## References

- [1] B. Pan, "Thermal error analysis and compensation for digital image/volume correlation," *Optics and Lasers in Engineering*, vol. 101, pp. 1–15, 2018.
- [2] K. Genovese, Y. Chi, and B. Pan, "Stereo-camera calibration for large-scale DIC measurements with active phase targets and planar mirrors," *Optics Express*, vol. 27, no. 6, pp. 9040–9053, 2019.
- [3] Y. Dong and B. Pan, "In-situ 3D shape and recession measurements of ablative materials in an arc-heated wind tunnel by UV stereo-digital image correlation," *Optics and Lasers in Engineering*, vol. 116, pp. 75–81, 2019.
- [4] H. Fathi, F. Dai, and M. Lourakis, "Automated as-built 3D reconstruction of civil infrastructure using computer vision: achievements, opportunities, and challenges," *Advanced Engineering Informatics*, vol. 29, no. 2, pp. 149–161, 2015.
- [5] H. Kim, S. Leutenegger, and A. J. Davison, "Real-time 3D reconstruction and 6-DoF tracking with an event camera," in *Proceedings of the European Conference on Computer Vision*, Springer, Amsterdam, Netherlands, October 2016.
- [6] G. Munda, C. Reinbacher, and T. Pock, "Real-time intensity-image reconstruction for event cameras using manifold regularisation," *International Journal of Computer Vision*, vol. 126, no. 12, pp. 1381–1393, 2018.
- [7] 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.
- [8] D. Feng and M. Q. Feng, "Vision-based multipoint displacement measurement for structural health monitoring," *Structural Control and Health Monitoring*, vol. 23, no. 5, pp. 876–890, 2016.
- [9] Z. Cai, X. Liu, A. Li, Q. Tang, X. Peng, and B. Z. Gao, "Phase-3D mapping method developed from back-projection stereovision model for fringe projection profilometry," *Optics Express*, vol. 25, no. 2, pp. 1262–1277, 2017.
- [10] J.-S. Hyun, G. T.-C. Chiu, and S. Zhang, "High-speed and high-accuracy 3D surface measurement using a mechanical projector," *Optics Express*, vol. 26, no. 2, p. 1474, 2018.
- [11] R. Ali, D. L. Gopal, and Y.-J. Cha, "Vision-based concrete crack detection technique using cascade features," in *Proceedings of the Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, International Society for Optics and Photonics, Denver CO, USA, March 2018.
- [12] C. M. Yeum and S. J. Dyke, "Vision-based automated crack detection for bridge inspection," *Computer-Aided Civil and Infrastructure Engineering*, vol. 30, no. 10, pp. 759–770, 2015.
- [13] X. Yang, H. Li, Y. Yu, X. Luo, T. Huang, and X. Yang, "Automatic pixel-level crack detection and measurement using fully convolutional network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1090–1109, 2018.
- [14] R. Huňady and M. Hagara, "A new procedure of modal parameter estimation for high-speed digital image correlation," *Mechanical Systems and Signal Processing*, vol. 93, pp. 66–79, 2017.
- [15] R. Huňady, P. Pavelka, and P. Lengvarský, "Vibration and modal analysis of a rotating disc using high-speed 3D digital image correlation," *Mechanical Systems and Signal Processing*, vol. 121, pp. 201–214, 2019.
- [16] X.-W. Ye, C. Dong, and T. Liu, "A review of machine vision-based structural health monitoring: methodologies and applications," *Journal of Sensors*, vol. 2016, Article ID 7103039, 10 pages, 2016.
- [17] J. Xiong, S.-d. Zhong, Y. Liu, and L.-f. Tu, "Automatic three-dimensional reconstruction based on four-view stereo vision using checkerboard pattern," *Journal of Central South University*, vol. 24, no. 5, pp. 1063–1072, 2017.
- [18] Y.-Q. Ni, W. You-Wu, L. Wei-Yang, and C. Wei-Huan, "A vision-based system for long-distance remote monitoring of dynamic displacement: experimental verification on a supertall structure," *Smart Structures and Systems*, vol. 24, no. 6, pp. 769–781, 2019.
- [19] C. Ricolfe-Viala and A.-J. Sánchez-Salmerón, "Robust metric calibration of non-linear camera lens distortion," *Pattern Recognition*, vol. 43, no. 4, pp. 1688–1699, 2010.
- [20] Y.-J. Cha, K. You, and W. Choi, "Vision-based detection of loosened bolts using the Hough transform and support vector machines," *Automation in Construction*, vol. 71, pp. 181–188, 2016.
- [21] Z. Ma and S. Liu, "A review of 3D reconstruction techniques in civil engineering and their applications," *Advanced Engineering Informatics*, vol. 37, pp. 163–174, 2018.
- [22] R. Deli, L. M. Galantucci, A. Laino et al., "Three-dimensional methodology for photogrammetric acquisition of the soft tissues of the face: a new clinical-instrumental protocol," *Progress in Orthodontics*, vol. 14, no. 1, p. 32, 2013.
- [23] A. Agudo, F. Moreno-Noguer, B. Calvo, and J. M. M. Montiel, "Real-time 3D reconstruction of non-rigid shapes with a single moving camera," *Computer Vision and Image Understanding*, vol. 153, pp. 37–54, 2016.
- [24] Y. Tang, L. Li, C. Wang et al., "Real-time detection of surface deformation and strain in recycled aggregate concrete-filled steel tubular columns via four-ocular vision," *Robotics and Computer-Integrated Manufacturing*, vol. 59, pp. 36–46, 2019.
- [25] H. Kim and H. Kim, "3D reconstruction of a concrete mixer truck for training object detectors," *Automation in Construction*, vol. 88, pp. 23–30, 2018.
- [26] L. Sun, V. Abolhasannejad, L. Gao, and Y. Li, "Non-contact optical sensing of asphalt mixture deformation using 3D stereo vision," *Measurement*, vol. 85, pp. 100–117, 2016.
- [27] R. Mlambo, I. Woodhouse, F. Gerard, and K. Anderson, "Structure from motion (SfM) photogrammetry with drone data: a low cost method for monitoring greenhouse gas emissions from forests in developing countries," *Forests*, vol. 8, no. 3, p. 68, 2017.
- [28] Y. Liu, J. Yang, Q. Meng, Z. Lv, Z. Song, and Z. Gao, "Stereoscopic image quality assessment method based on binocular combination saliency model," *Signal Processing*, vol. 125, pp. 237–248, 2016.
- [29] Z. Liu, H. Qin, S. Bu et al., "3D real human reconstruction via multiple low-cost depth cameras," *Signal Processing*, vol. 112, pp. 162–179, 2015.
- [30] N. Candau, C. Pradille, J.-L. Bouvard, and N. Billon, "On the use of a four-cameras stereovision system to characterize large 3D deformation in elastomers," *Polymer Testing*, vol. 56, pp. 314–320, 2016.

- [31] P. Zhou, J. Zhu, X. Su et al., "Experimental study of temporal-spatial binary pattern projection for 3D shape acquisition," *Applied Optics*, vol. 56, no. 11, pp. 2995–3003, 2017.
- [32] X. Shen, A. Markman, and B. Javidi, "Three-dimensional profilometric reconstruction using flexible sensing integral imaging and occlusion removal," *Applied Optics*, vol. 56, no. 9, pp. D151–D157, 2017.
- [33] M. Malesa, K. Malowany, J. Pawlicki et al., "Non-destructive testing of industrial structures with the use of multi-camera digital image correlation method," *Engineering Failure Analysis*, vol. 69, pp. 122–134, 2016.
- [34] A. Sinha, J. Bai, and K. Ramani, "Deep learning 3D shape surfaces using geometry images," in *Proceedings of the European Conference on Computer Vision*, Springer, Amsterdam, Netherlands, October 2016.
- [35] F. Li, Q. Li, T. Zhang, Y. Niu, and G. Shi, "Depth acquisition with the combination of structured light and deep learning stereo matching," *Signal Processing: Image Communication*, vol. 75, pp. 111–117, 2019.
- [36] J. Zhang, S. Hu, and H. Shi, "Deep learning based object distance measurement method for binocular stereo vision blind area," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 9, p. 1, 2018.
- [37] S. Sun, L. Rongke, P. Yu, D. Qiuchen, S. Shuqiao, and S. Han, "Pose determination from multi-view image using deep learning," in *Proceedings of the 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, IEEE, Tangier, Morocco, June 2019.
- [38] Y. Yang, F. Qiu, H. Li, L. Zhang, M.-L. Wang, and M.-Y. Fu, "Large-scale 3D semantic mapping using stereo vision," *International Journal of Automation and Computing*, vol. 15, no. 2, pp. 194–206, 2018.
- [39] C. Wang, X. Zou, Y. Tang, L. Luo, and W. Feng, "Localisation of litchi in an unstructured environment using binocular stereo vision," *Biosystems Engineering*, vol. 145, pp. 39–51, 2016.
- [40] C. Wang, Y. Tang, X. Zou, W. SiTu, and W. Feng, "A robust fruit image segmentation algorithm against varying illumination for vision system of fruit harvesting robot," *Optik*, vol. 131, pp. 626–631, 2017.
- [41] C. Wang, T. Luo, L. Zhao, Y. Tang, and X. Zou, "Window zooming-based localization algorithm of fruit and vegetable for harvesting robot," *IEEE Access*, vol. 7, pp. 103639–103649, 2019.
- [42] Y.-C. Tang, L.-J. Li, W.-X. Feng, F. Liu, X.-J. Zou, and M.-Y. Chen, "Binocular vision measurement and its application in full-field convex deformation of concrete-filled steel tubular columns," *Measurement*, vol. 130, pp. 372–383, 2018.
- [43] Y. Tang, S. Fang, J. Chen et al., "Axial compression behavior of recycled-aggregate-concrete-filled GFRP-steel composite tube columns," *Engineering Structures*, vol. 216, p. 110676, 2020.
- [44] L. Fu, J. Duan, X. Zou et al., "Banana detection based on color and texture features in the natural environment," *Computers and Electronics in Agriculture*, vol. 167, p. 105057, 2019.
- [45] Y. Tang, M. Chen, C. Wang et al., "Recognition and localization methods for vision-based fruit picking robots: a review," *Frontiers in Plant Science*, vol. 11, no. 510, 2020.
- [46] L. Luo, Y. Tang, X. Zou, M. Ye, W. Feng, and G. Li, "Vision-based extraction of spatial information in grape clusters for harvesting robots," *Biosystems Engineering*, vol. 151, pp. 90–104, 2016.
- [47] L. Luo, Y. Tang, X. Zou, C. Wang, P. Zhang, and W. Feng, "Robust grape cluster detection in a vineyard by combining the AdaBoost framework and multiple color components," *Sensors*, vol. 16, no. 12, p. 2098, 2016.
- [48] L. Luo, Y. Tang, Q. Lu, X. Chen, P. Zhang, and X. Zou, "A vision methodology for harvesting robot to detect cutting points on peduncles of double overlapping grape clusters in a vineyard," *Computers in Industry*, vol. 99, pp. 130–139, 2018.
- [49] G. Lin, Y. Tang, X. Zou, J. Xiong, and J. Li, "Guava detection and pose estimation using a low-cost RGB-D sensor in the field," *Sensors*, vol. 19, no. 2, p. 428, 2019.
- [50] G. Lin, Y. Tang, X. Zou, J. Xiong, and Y. Fang, "Color-, depth-, and shape-based 3D fruit detection," *Precision Agriculture*, vol. 21, no. 1, pp. 1–17, 2020.
- [51] G. Lin, Y. Tang, X. Zou, J. Li, and J. Xiong, "In-field citrus detection and localisation based on RGB-D image analysis," *Biosystems Engineering*, vol. 186, pp. 34–44, 2019.
- [52] G. Lin, Y. Tang, X. Zou, J. Cheng, and J. Xiong, "Fruit detection in natural environment using partial shape matching and probabilistic Hough transform," *Precision Agriculture*, vol. 21, no. 1, pp. 160–177, 2020.
- [53] J. Li, Y. Tang, X. Zou, G. Lin, and H. Wang, "Detection of fruit-bearing branches and localization of litchi clusters for vision-based harvesting robots," *IEEE Access*, vol. 8, pp. 117746–117758, 2020.
- [54] M. Chen, Y. Tang, X. Zou, K. Huang, L. Li, and Y. He, "High-accuracy multi-camera reconstruction enhanced by adaptive point cloud correction algorithm," *Optics and Lasers in Engineering*, vol. 122, pp. 170–183, 2019.
- [55] M. Chen, Y. Tang, X. Zou et al., "Three-dimensional perception of orchard banana central stock enhanced by adaptive multi-vision technology," *Computers and Electronics in Agriculture*, vol. 174, p. 105508, 2020.
- [56] Y. Tang, Y. Lin, X. Huang et al., "Grand challenges of machine-vision technology in civil structural health monitoring," *Artificial Intelligence Evolution*, vol. 1, no. 1, pp. 8–16, 2020.
- [57] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, 2000.
- [58] M. Sereewattana, M. Ruchanurucks, and S. Siddhichai, "Depth estimation of markers for UAV automatic landing control using stereo vision with a single camera," in *Proceedings of the International Conference on Information and Communication Technology for Embedded System*, 2014.
- [59] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, IEEE, San Diego, CA, USA, June 2005.
- [60] R. A. Zeineldin and N. A. El-Fishawy, "Fast and accurate ground plane detection for the visually impaired from 3D organized point clouds," in *Proceedings of the 2016 SAI Computing Conference (SAI)*, IEEE, London, UK, July 2016.
- [61] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the Sixth International Conference on Computer Vision*, Bombay, India, January 1998.
- [62] B. Skinner, T. Vidal-Calleja, J. V. Miro, F. D. Bruijn, and R. Falque, "3D point cloud upsampling for accurate reconstruction of dense 2.5 D thickness maps," in *Proceedings of the Australasian Conference on Robotics and Automation*, ACRA, Melbourne, Australia, December 2014.
- [63] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.
- [64] C. R. Qi, Y. Li, S. Hao, and L. J. Guibas, "Pointnet++: deep hierarchical feature learning on point sets in a metric space,"

in *Proceedings of the Conference on Neural Information Processing Systems (NIPS) 2017*, Long Beach, CA, USA, December 2017.

- [65] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Sensor Fusion IV: Control Paradigms and Data Structures* International Society for Optics and Photonics, Bellingham, WA, USA, 1992.
- [66] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, Cagliari, Sardinia, June 2006.