



Article ID 1007-1202(2024)04-0301-14 DOI <https://doi.org/10.1051/wujns/2024294301>

Cite this article: GAO Rong, SUN Zhaoyun, GUO Jianxing, *et al.* 3D Surface Reconstruction of Coarse Aggregate Particles from Occlusion-Free Multi-View Images[J]. *Wuhan Univ J of Nat Sci*, 2024, 29(4): 301-314.

3D Surface Reconstruction of Coarse Aggregate Particles from Occlusion-Free Multi-View Images

□ GAO Rong¹, SUN Zhaoyun^{1†}, GUO Jianxing^{2†}, LI Wei¹, YANG Ming¹, HAO Xueli¹, YAO Bobin³, WANG Huifeng³

1. School of Information Engineering, Chang'an University, Xi'an 710064, Shaanxi, China;

2. Office of Infrastructure, Chang'an University, Xi'an 710064, Shaanxi, China;

3. School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, Shaanxi, China

Abstract: Rapidly and accurately assessing the geometric characteristics of coarse aggregate particles is crucial for ensuring pavement performance in highway engineering. This article introduces an innovative system for the three-dimensional (3D) surface reconstruction of coarse aggregate particles using occlusion-free multi-view imaging. The system captures synchronized images of particles in free fall, employing a matte sphere and a nonlinear optimization approach to estimate the camera projection matrices. A pre-trained segmentation model is utilized to eliminate the background of the images. The Shape from Silhouettes (SfS) algorithm is then applied to generate 3D voxel data, followed by the Marching Cubes algorithm to construct the 3D surface contour. Validation against standard parts and diverse coarse aggregate particles confirms the method's high accuracy, with an average measurement precision of 0.434 mm and a significant increase in scanning and reconstruction efficiency.

Key words: 3D shape reconstruction; multi-view imaging; coarse aggregate particles; shape from Silhouettes; multi-camera calibration

CLC number: TP23

0 Introduction

The morphological attributes of coarse aggregate particles are pivotal to the performance of asphalt mixtures, influencing their resistance to high-temperature deformation, low-temperature cracking, and fatigue. These characteristics are also crucial for the gradation of asphalt mixtures and are integral to the quality assurance processes^[1,2]. Traditional sieving methods are laborious and may not provide the timely and precise feedback

necessary for optimal production, potentially compromising the mixture's quality. Thus, developing automated, non-destructive testing methods for accurately characterizing coarse aggregate particles is essential for enhancing asphalt mixture performance and ensuring quality.

In response to the growing demand for accurate and direct analysis, there has been a notable shift from two-dimensional (2D) to three-dimensional (3D) analysis, aiming to capture more comprehensive and precise information on particle morphologies. 3D data of coarse ag-

Received date: 2024-01-28 © Wuhan University 2024

Foundation item: Supported by the Key R&D Projects in Shaanxi Province (2022JBGS3-08)

Biography: GAO Rong, male, Senior engineer, research direction: instrumentation and automation. E-mail: gr@chd.edu.cn

† Corresponding author. E-mail: zhaoyunsun@126.com (SUN Zhaoyun), gjxing@chd.edu.cn (GUO Jianxing)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

gregate particles were collected through methods such as structured light 3D scanning^[3-9], binocular stereo imaging^[10], and computed tomography (CT) scanning^[11-13]. These data were then evaluated for morphological characteristics of coarse aggregate particles, including particle grade, shape, edges and corners, textures, and flaky shape^[6,14], through 3D point cloud processing^[3], machine learning^[4,15], virtual simulation^[13,16,17], and other methods. These approaches have achieved significant results. However, the existing 3D surface data acquisition equipment has certain limitations regarding integrity, scanning efficiency, and equipment cost when characterizing coarse aggregate particle shapes. These limitations restrict the broader application of 3D shape characterization in coarse aggregate particles.

Due to light occlusion, structured light 3D scanning can only scan the local geometric shape of particles^[4], necessitating multiple scans and stitching to obtain the complete 3D contour of coarse aggregate particles. This process is time-consuming and susceptible to errors, resulting in inaccurate 3D mesh. When dealing with texture-less targets such as coarse aggregate particles, binocular stereo vision faces matching difficulties and significant depth estimation errors, preventing accurate acquisition of the target's 3D information. CT is a widely used imaging technology that can quantitatively and non-destructively scan samples. However, its expensive equipment, high experimental costs, slow scanning and reconstruction speed, and radiation production during the scanning process hinder its promotion and application in studying coarse aggregate morphology.

Multi-view 3D reconstruction technology is an important research topic in the field of computer vision. By using multiple low-cost visible light cameras from different views to simultaneously capture images of the same object and fuse the image information through 3D reconstruction algorithms, it is possible to obtain the 3D contour of the object, including surface texture and color information. Multi-view 3D reconstruction technology is widely used in industrial production^[18-21], medical treatment^[22], cultural heritage protection^[23-25] and other fields^[26,27]. This study aims to apply multi-view 3D reconstruction to scan and reconstruct complete 3D contours of coarse aggregate particles. To ensure the accuracy and integrity of the 3D reconstruction results, coarse aggregate particles must be imaged from multiple views, covering all areas of the particle surface as much as possible. Unlike traditional rotating disc or conveyor belt scanning methods, this study utilizes the principle that

there are no occlusions between cameras and the coarse aggregate particles during freefall. A synchronous image acquisition system was designed and constructed to capture unobstructed images during the particles' freefall from multiple different views, achieving comprehensive particle imaging.

Accurate calibration of multi-view imaging systems is vital in obtaining high-quality 3D reconstruction results. This work proposes a novel method for calibrating the multi-view camera system. The semi-enclosed cavity structure refers to a setup where the cameras are placed inside a partially enclosed space, introducing unique challenges for calibration. Traditional calibration methods, such as Zhang's algorithm^[28] and Structure from Motion (SfM)^[29] techniques, have been extensively employed for multi-camera calibration. These methods typically rely on known calibration patterns or feature correspondences in multiple images to estimate the intrinsic and extrinsic parameters of the cameras. While these techniques have proven successful in various scenarios, they may encounter limitations when dealing with complex camera setups or challenging imaging conditions, particularly when objects have limited surface texture or are restricted access to calibration patterns^[30].

We introduce a calibration approach based on geometric error optimization to address these challenges. By formulating the calibration problem as a nonlinear optimization task, we aim to minimize the geometric errors between the projected points from the estimated camera projection matrix and the corresponding ground truth points. This approach allows us to refine the camera projection matrix and achieve accurate calibration results using an easily accessible matte sphere as the calibration object. This method eliminates the need for complex calibration patterns or specialized equipment.

After the calibration process, we proceed with the 3D reconstruction stage. Unlike traditional stereo matching^[31] methods, our approach eliminates the need to find matching points on the images of the scanned object. Instead, it directly reconstructs the multi-view images that contain calibration pose information using the Shape from Silhouette (SfS)^[32] technique. This method enables the generation of 3D voxels representing the coarse aggregate particles. To extract the surface of the particles, we employ the Marching Cubes algorithm^[33], which efficiently creates a mesh representation of the 3D contour of the particles.

By comparing and quantitatively analyzing the calibration of cameras and conducting 3D reconstruction ex-

periments on standard parts and coarse aggregate particle samples, we validate the effectiveness of the proposed multi-view imaging system calibration and SfS 3D reconstruction method. This comprehensive evaluation demonstrates the accuracy and reliability of our approach in reconstructing the 3D shape of coarse aggregate particles.

In summary, this study makes the following main contributions:

(1) **Design and Construction of a Multi-View, Unobstructed Coarse Aggregate Particle Image Synchronization Capture System:** The system incorporates 16 industrial cameras and ensures unobstructed imaging from multiple views during particle freefall. The systematic design of camera parameters, layout, and lighting compensation was implemented to optimize the image capture process.

(2) **High-Precision Calibration of the Multi-View Camera System:** A unique calibration approach based on geometric error optimization was developed for the multi-view camera system with a semi-enclosed cavity structure. The approach utilized multiple ball-dropping experiments to calibrate the camera projection matrices, eliminating the need to calibrate the intrinsic parameters of cameras individually.

(3) **3D Reconstruction Experiments and Quantitative Analysis:** The calibrated multi-view imaging system was used to conduct 3D reconstruction experiments on standard parts and coarse aggregate particles. The reconstructed shapes were compared and quantitatively analyzed, particularly compared to high-precision structured light scanning results.

The structure of this article is as follows. Section 1 describes the procedures involving the acquisition of multi-view coarse aggregate particle images, the calibration approach based on geometric error optimization, the

SfS-based 3D voxel reconstruction, and the Marching Cubes algorithm-based mesh extraction. Section 2 presents the results of the proposed method and discusses these results. Finally, Section 3 concludes the article.

1 Methodology

1.1 The Framework of the Proposed Method

A novel method for the surface reconstruction of coarse aggregate particles using multi-view images is proposed in this work. Following a design similar to a previous study^[16], a multi-view image acquisition device is established to capture coarse aggregate particles during their free-falling process. The device enables the synchronized acquisition of unobstructed multi-view images during the falling process of the target particles. Although the device has a uniform illumination shell, the internal light intensity remains uneven, resulting in background interference in the captured raw images. Therefore, before the subsequent calibration and reconstruction processes, accurate segmentation of the foreground objects in the original multi-view images is achieved using a segmentation network model U^2 -Net^[17]. The camera projection matrices are calibrated using a nonlinear optimization method based on geometric error. The SfS algorithm is then employed with the calibrated results and multi-view images to reconstruct the particle voxel space. Finally, the Marching Cubes algorithm is applied to obtain the mesh data representing the complete 3D contour of the particles. The framework of this method is shown in Fig. 1.

Step 1 Multi-view image acquisition: In this step, a synchronized acquisition system is designed and constructed to capture unobstructed multi-view images during the falling process of the target particles. The device

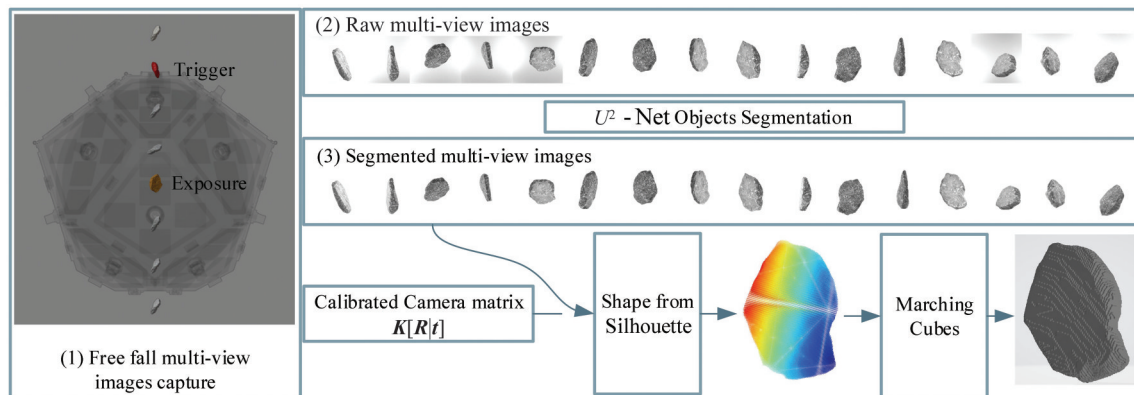


Fig. 1 Framework of the proposed multi-view 3D reconstruction system

ensures the simultaneous capture of images from different views, providing comprehensive coverage of the particle surface.

Step 2 Multi-view imaging system calibration: A novel calibration method is proposed for multi-view camera systems with a semi-enclosed cavity structure. The camera parameters are refined through a nonlinear optimization process that minimizes the geometric errors between the projected points and the corresponding ground truth points. This optimization approach ensures accurate calibration results, enabling precise 3D reconstruction of coarse aggregate particle surfaces. The method utilizes an easily accessible matte sphere as the calibration object, eliminating the need for complex patterns or specialized equipment.

Step 3 Surface reconstruction: The surface reconstruction step utilizes the calibrated camera projection matrices and the multi-view images to reconstruct the voxel space representing the coarse aggregate particles. The SfS algorithm is employed to generate the 3D voxel data of the particles, capturing their surface contours. The Marching Cubes algorithm is applied to convert the voxel data into a mesh representation, producing a complete 3D contour of the particles. This reconstructed surface provides a detailed characterization of the coarse aggregate particles, enabling further analysis and evaluation of their morphological characteristics.

Further details of these three steps are presented in the following sections.

1.2 Multi-View Image Acquisition

The main structure of the multi-view imaging device consists of a polyhedron assembled from 18 irregular carbon fiber panels, as illustrated in Fig. 2. This struc-

ture forms a 180° line connecting the centers of the top and bottom square holes for the entry and exit of coarse aggregate particles. The cameras are arranged in four layers, each spaced at 36° intervals, with four cameras evenly distributed within each layer. Cameras in adjacent layers are staggered at a 45° angle, and all cameras are oriented toward the sphere's center. This layout ensures that the 16 cameras uniformly cover the surface of the coarse aggregate particles with no overlapping fields of view, thus minimizing interference during subsequent target segmentation.

The coarse aggregate particles start to freely fall from the top square hole of the device in a stationary state and eventually exit from the bottom square hole. A grating fall detection sensor is installed at the top hole, which generates a trigger signal when a particle passes through it. This trigger signal is connected to the external trigger ports of the 16 cameras, and the delay trigger duration of the cameras is set to the time it takes for the particles to fall from the top window to the center of the structure. When the particles reach the center of the device during their descent, all cameras capture images simultaneously. With a diameter of 0.6 m, the device operates on the free fall principle, with each particle taking approximately 247.4 ms to travel from the top opening to the center of the structure. The system can capture subsequent particles immediately after completing the current capture sequence, achieving a maximum frame rate of 4 fps for multi-image acquisition.

The internal lighting within the enclosed structure of the device is weak, and the coarse aggregate particles are in a state of free-fall motion. To ensure clear imaging of the particles, the camera's exposure time must be sufficiently short while allowing enough light to reach the

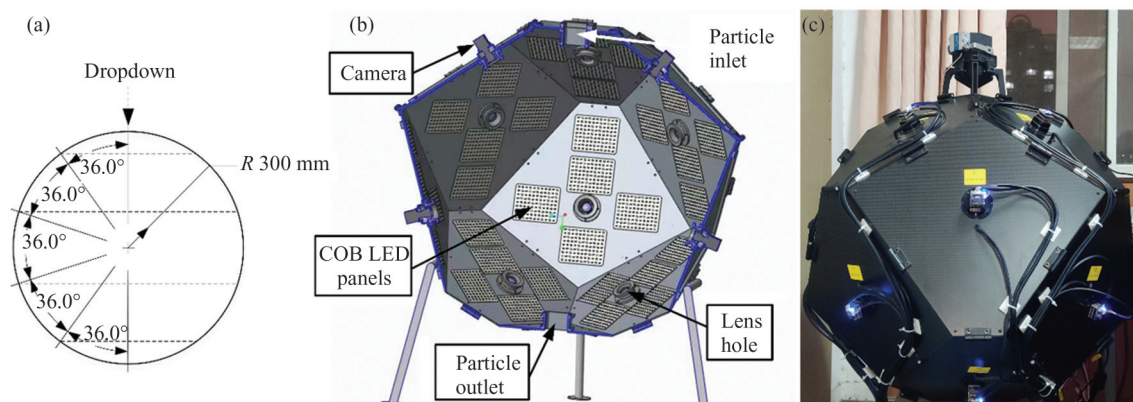


Fig. 2 The main structure and photo of the multi-view imaging device

(a) Schematic diagram of the camera layout. (b) Cross-sectional design diagram of the structure. (c) Photograph of the assembled multi-view imaging system

camera. Bright COB LED light panels are installed on the inner walls of the device to provide sufficient illumination. A 4 mm thick, 600 mm diameter white acrylic diffuse sphere is embedded within the device to achieve uniform lighting. All light panels flicker under the control of the camera flash signal, providing sufficient illumination for the multi-view imaging system.

The experimental setup comprises 16 identical global shutter monochrome industrial cameras with a CMOS SC130GS imaging module. The cameras have a sensor size of 1/2.7", a pixel size of 4.0 μm , and a resolution 1 280×1 024. Each camera has a 25 mm focal length lens with an aperture of F2.0. The cameras are focused on the center position of the device, approximately at a distance of 300 mm, providing a field of view size of 60 mm×45 mm. Each camera is connected to a high-speed switch, model H3C S1226FX, via a Gigabit Ethernet (GigE) interface. The switch has 24 Gigabit ports and 2×10 Gb/s fiber uplink ports connected to the server. The bandwidth is shared among the cameras so they can upload the captured images to the server for subsequent processing. The multi-view images of a single round-shaped coarse aggregate particle collected by the proposed system are shown in Fig. 3. It can be observed that the textures and edges of the captured coarse aggregate particles are clear.

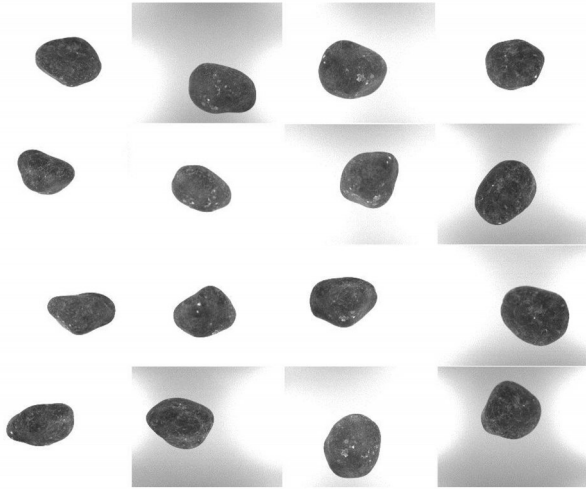


Fig. 3 The multi-view images of a single round-shaped coarse aggregate particle collected by the proposed system

1.3 Camera Parameter Calibration for the Multi-View Imaging System

Camera calibration is an essential procedure for extracting 3D measurements from 2D images. The classical pinhole camera model encapsulates the process of

camera imaging. In this model, a point in 3D space is represented by its homogeneous coordinates $X = [x, y, z, 1]^T$, and its corresponding projection on the camera's image plane is denoted as $x = [u, v, 1]^T$. The relationship between these coordinates is articulated by Eq. (1):

$$\omega x = K_{3 \times 3} [R | t]_{3 \times 4} X \quad (1)$$

where ω is the scale factor, $[R | t]$ is the extrinsic parameter matrix of the camera, consisting of the rotation matrix R and the translation vector t , which encapsulates the rotation and translation from the world coordinate system to the camera coordinate system. K is the intrinsic parameter matrix of the camera, detailed in Equation (2) as follows:

$$K = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

Here u_0 and v_0 represent the pixel coordinates of the camera's principal point, while α and β denote the focal lengths along the x and y axes, respectively. The skew coefficient γ accounts for the skewness between the imaging device's axes, ensuring an accurate representation of spatial relationships. The process of projecting a point from the world coordinate system to the pixel coordinate system is facilitated by the camera's projection matrix P , elegantly encapsulated in Eq. (3):

$$P_{3 \times 4} = K_{3 \times 3} [R | t]_{3 \times 4} \quad (3)$$

Traditional stereo vision systems typically calibrate their camera parameters by capturing patterns from a chessboard or similar calibration objects, employing Zhang's method^[28] or Structure from Motion (SfM) techniques. However, the multi-view imaging system in this study, comprising 16 cameras, presents unique calibration challenges due to its semi-enclosed spherical structure with limited access points, restricting the size of the calibration object. The calibration must also contend with capturing images of an object in free fall, complicating manual intervention and pose control. Moreover, while the cameras are focused toward the center of the structure, their varied distribution across the surface renders traditional chessboard patterns unsuitable for calibration. To surmount these challenges, the paper describes a series of ball-drop experiments yield data on 3D spatial points and their corresponding 2D projections. The paper achieves a comprehensive calibration of the camera parameter matrix by harnessing the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm^[34] for the

nonlinear optimization of geometric errors.

1.4 Geometric Errors Optimization for Camera Projection Matrix Calibration

Let us consider the index k ranging from 1 to N , inclusive. Given two sets of vectors $\mathbf{x}_k = (u_k \times \omega, v_k \times \omega, \omega)^\top$ and $\mathbf{X}_k = [X_k, Y_k, Z_k, 1]^\top$. Here, $\mathbf{X}_k$ represents a point in the world coordinate system, while $\mathbf{x}_k$ denotes the coordinates of the projection of $\mathbf{X}_k$ onto the camera image pixel coordinate. We aim to determine a 3×4 projection matrix $\mathbf{P}$ that satisfies the Eq. (4):

$$\mathbf{x}_k = \mathbf{P}\mathbf{X}_k, \quad (4)$$

$$\begin{bmatrix} u_k \times \omega \\ v_k \times \omega \\ \omega \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X_k \\ Y_k \\ Z_k \\ 1 \end{bmatrix}$$

In this equation, $\omega \neq 0$ represents the unknown scale factor associated with the index k . In pursuit of the projection matrix $\mathbf{P}$, We initiate our analysis by addressing the elimination of the indeterminate scale factor ω , culminating in Equation (5):

$$\begin{cases} u_k = \frac{p_{11}X_k + p_{12}Y_k + p_{13}Z_k + p_{14}}{p_{31}X_k + p_{32}Y_k + p_{33}Z_k + p_{34}} \\ v_k = \frac{p_{21}X_k + p_{22}Y_k + p_{23}Z_k + p_{24}}{p_{31}X_k + p_{32}Y_k + p_{33}Z_k + p_{34}} \end{cases} \quad (5)$$

The elements of the matrix $\mathbf{P}$ can initially be estimated using the Direct Linear Transformation (DLT) method. However, to address the nonlinear errors and uncertainties inherent in the camera imaging process, such as radial distortion, we opt for a nonlinear optimization approach to refine the parameters of the matrix $\mathbf{P}$. This refinement hinges on the calculation of geometric error E , as depicted in Eq. (6):

$$E = \sum_{k=1}^N d(\mathbf{x}_k, \mathbf{x}'_k) \quad (6)$$

where E represents the discrepancy between the true projected value $\mathbf{x}_k$ and the predicted projection $\mathbf{x}'_k = \mathbf{P}\mathbf{Y}_k$, quantified by the Euclidean distance function d . Our multi-view imaging system is pivotal in capturing multiple sets of 3D spatial points and their corresponding 2D projections. Leveraging this system's structural and camera parameters, we employ the open-source software Blender to derive the initial camera matrix $\mathbf{P}$ for each camera's starting configuration. Subsequently, the BFGS nonlinear optimizer from the SciPy library is engaged to perform the optimization, ensuring that the matrix $\mathbf{P}$ is finely tuned. The BFGS algorithm is a quasi-Newton optimization method that efficiently handles unconstrained optimization problems. Quasi-Newton methods use first-

order derivatives (gradients) to approximate the second-order derivatives (Hessian matrix), thereby accelerating the optimization process. The BFGS algorithm optimizes the objective function by iteratively updating an approximate inverse Hessian matrix, achieving a high convergence rate and stability. In each iteration, the BFGS algorithm computes the gradient based on the current camera matrix and determines the optimal step size through line search, updating the camera matrix to reduce the geometric error.

1.5 Ball-Falling Method for Data Acquisition in Calibration

For data acquisition in calibration, we employ the ball-falling method, utilizing readily available black matte spheres with a diameter of 10 mm as calibration objects. As the sphere descends, the system captures a series of images. From these images, the centroid coordinates of the projected images are ascertained for each distinct viewpoint. Ideally, rays emanating from the camera's optical center through the centroid of the ball's image on the imaging plane should intersect at a single point, signifying the centroid position of the ball in the world coordinate system.

However, variations in the manufacturing and assembly of the system's components inevitably introduce deviations from the ideal parameters. These deviations result in the rays not converging perfectly to a single point. During the calibration process, the spatial coordinates that lie closest to the convergence of all rays are calculated, representing the estimated center of the falling ball within the world coordinate system. This estimation serves as a pivotal reference point, subsequently employed to refine the camera's projection matrix, as elegantly depicted in Fig.4.

It is starting from the centroid coordinates p_i of the projected circular pattern of the falling ball as observed by the i -th camera. A ray $\mathbf{r}'_i = \{\mathbf{r}_i = \mathbf{c}_i + t\mathbf{d}_i | t \in \mathbb{R} \geq 0\}$ is defined by its path through the optical center $\mathbf{c}_i$ of the camera, with $\mathbf{d}_i$ representing the normalized direction vector. We aim to identify a point $\mathbf{p}'$ in the world coordinate system that minimizes the cumulative distances to all camera projection rays. This optimization problem is elegantly expressed in Eq. (7):

$$\mathbf{p}' = \arg \min_{\mathbf{x} \in \mathbb{R}^3} \sum_{i=1}^n \delta(\mathbf{x}, \mathbf{r}_i) \quad (7)$$

In this context, n symbolizes the total number of cameras within our multi-view imaging system, which is

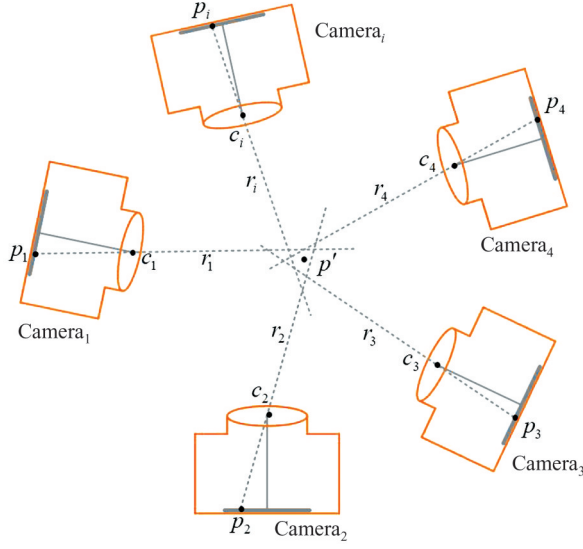


Fig. 4 Multi-rays to estimate the ball's position in the world coordinate system

set at 16. The distance of a point x in 3D space to a camera projection ray is captured by Eq. (8):

$$\delta(x, r_i)^2 = (x - c_i)^T [I - d_i d_i^T] (x - c_i) \quad (8)$$

The closed-form solution for p' is presented in Eq. (9):

$$p' = \left[\sum_{i=1}^n [I - d_i d_i^T] \right]^{-1} \sum_{i=1}^n [I - d_i d_i^T] c_i \quad (9)$$

We collected a rich dataset of 3D spatial coordinates p' and their corresponding 2D-pixel coordinates through ball drop experiments. Additionally, it is essential to note that the structural design parameters of the multi-view imaging system are known. We have imported these parameters, along with the camera positions and orientations, into Blender to generate the ideal extrinsic parameter matrices for each camera. The intrinsic parameter matrices of the cameras are calculated based on the ideal values derived from the lens and photosensitive element parameters:

$$K = \begin{bmatrix} 6250 & 0 & 640 \\ 0 & 6250 & 480 \\ 0 & 0 & 1 \end{bmatrix} \quad (10)$$

where 6250 is the pixel focal length, calculated by dividing the camera lens focal length of 25 mm by the pixel size of the photosensitive device 4 μm . The lens distortion is initialized to zero, and the camera principal point is assumed to be located at the center of the photosensitive element. The camera projection matrix calibration algorithm for multi-view imaging system are listed in Algorithm 1.

Algorithm 1 Camera projection matrix calibration

for $k \leftarrow 1, \dots, T$ **do**

 Collect a sample of falling ball data S_k ;

 Capture one image per camera that includes the projection of the falling ball S_k ;

for $i \leftarrow 1, \dots, N$ **do**

 Object segmentation on the image from i -th camera and extract the centroid coordinates of the projected circle p_i ;

 Use the current camera matrix parameters to obtain c_i .

 Calculate d_i and normalize it;

end

 Calculate p'_k using equation (9)

 Update p'_k as the optimized estimated coordinate of the k -th ball falling experiment out of a total of T trials;

end

for $i \leftarrow 1, \dots, N$ **do**

 Calculate the predicted projection values using the initial camera projection matrix $x'_k = P_i p'_k$;

 Calculate the geometric error E ;

 Optimize the camera projection matrix for the i -th camera

P_i using the BFGS algorithm.

end

1.6 Shape from Silhouette Voxel Reconstruction

Shape from Silhouette is a method used to reconstruct 3D shapes using 2D contour information of the target object. The SfS algorithm requires two essential components: the projection matrices of each camera corresponding to the viewpoints and the segmentation results of the foreground object region in each camera view. The basic principle of SfS is illustrated in Fig. 1. Multiple images of a falling particle are captured using a multi-view imaging system. These images are then fed into a model U^2 -Net, such as the one based on the U -Net and ResNet architectures, to perform object segmentation on multi-view images. This segmentation process identifies the foreground regions corresponding to the projected particle in each viewpoint. For each viewpoint, starting from the camera's optical center, rays pass through the pixels occupied by the foreground object in the image, forming visual cones in 3D space. The segmentation results determine the shape of the visual cones, while the camera calibration results determine their direction. Ultimately, the reconstructed result of the target object is obtained by finding the intersection of the visual cones from all viewpoints. Figure 5(a) illus-

trates the projection of visual cones from three viewpoints, and Fig. 5(b) shows the projection of all 16 visual cones in 3D space. By taking the intersection of the visual cones from all viewpoints, the SfS result is obtained, as shown in Fig. 5(c), which represents the reconstructed voxel set of a ball.

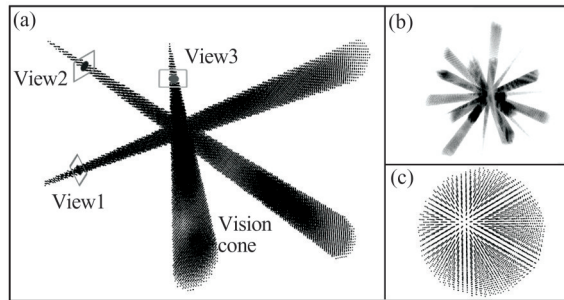


Fig. 5 Schematic illustration of the principle of SfS

(a) Union of visual cones from 3 viewpoints. (b) Union of visual cones from 16 viewpoints. (c) Intersection of visual cones from 16 viewpoints, resulting in the reconstructed object, a sphere

Similar to pixels being the smallest unit in 3D images, voxels are the smallest units in 3D space. Voxels are widely used in fields such as 3D imaging and medical imaging. The aforementioned 3D SfS also utilizes the concept of voxels. There are three key points in voxel initialization:

- 1) The voxel should represent the physical space scale of the actual object, determined based on the maximum physical size of the target object. In the designed multi-view imaging system in this study, the maximum field of view is 60 mm×45 mm. Considering the rotational nature of the target object and leaving some margin at the image edges, the longest side of the target object in the experiment is no more than 40 mm. Therefore, a voxel with a side length of 50 mm is chosen.

- 2) The coordinate of the voxel cube's center in the world coordinate system is selected. In this study, the origin of the world coordinate system is chosen as the center of the voxel cube, which is also the center of the multi-view imaging system.

- 3) The number of voxels on each side of the voxel cube is determined. The voxel cube contains N voxels on each side, and the larger the value of N , the more finely the cubic space is divided by voxels, representing finer physical spaces. The selection of N should balance computational complexity and reconstruction accuracy and can be adjusted based on different requirements for reconstruction accuracy.

1.7 Surface Reconstruction of Volumetric Data

Volumetric data consists of voxels, the basic elements, each represented as $\langle x, y, z, S \rangle$, where x, y, z denote the voxel's coordinates in a 3D grid, and S signifies a specific associated value. In the SfS process, initial voxel data is manipulated using the intersection of view cones from multiple perspectives, creating a solid representation of volumetric data that uniformly encompasses all points within the object of interest. However, this approach can introduce significant redundancy, complicating further processing. To address this, the Marching Cubes algorithm is utilized in this study to extract the isosurface from the volumetric data produced by SfS. Through surface fitting, the data is transformed into a triangular mesh that represents the object's surface, establishing a basis for subsequent analyses, including volume estimation, particle size evaluation, and geometric morphology assessment of coarse aggregate particles.

Within the volumetric data yielded by SfS, the value linked to each voxel, indicated as S , corresponds to the frequency of the voxel's appearance within the view cones formed by multiple perspectives. Its maximum value matches the number of cameras, N , in the multi-view imaging system, signifying that the voxel is present in all view cones. An isosurface is defined as the collection of voxels in the volumetric data sharing identical S values. The Marching Cubes algorithm processes each voxel within the volumetric data, performing operations such as isosurface extraction, vertex lookup, and edge intersection point interpolation. This generates a triangular mesh representation of the object's contour, resulting in surface data.

Marching Cubes is an algorithm used to create polygonal surfaces from 3D isosurfaces. It operates by sliding through a 3D regular grid, with each cube defined by 8 adjacent voxels as its vertices. The algorithm traverses all voxels in the volume data. For each cube, based on a set isovalue threshold called isolevel, the algorithm classifies voxels with $S < \text{isolevel}$ inside the cube as interior and those with $S \geq \text{isolevel}$ outside as exterior, resulting in a binary marking (0 or 1) for each vertex of the cube. With the 8 vertices of the cube ordered, there are a total of $2^8 = 256$ possible vertex configurations.

The essence of the Marching Cubes algorithm is to ascertain whether polygons that form the 3D contour of the object exist within each cube based on the vertex configurations. This is achieved by using a lookup table

to efficiently identify which cube edges are intersected by the isosurface. Figure 6 illustrates the principle of the Marching Cubes algorithm. By assessing whether vertex 3 lies inside or outside the isosurface is relative to the isovalue threshold, the cube's vertex configuration is identified as index 8. Utilizing the precomputed lookup table, the algorithm swiftly generates an edge index array $\text{edge}^{[12]}$ for the cube, indicating the indices of the intersected edges. A value of -1 in the array signifies that the corresponding edge is not intersected. Finally, linear interpolation is used to determine the precise positions of the intersection points P on each edge:

$$P = P_1 + (\text{isovalue} - S_1)(P_2 - P_1)/(S_2 - S_1) \quad (11)$$

where, P_1 and P_2 are the positions of the two vertices on the intersected edge, and S_1 and S_2 are the S values associated with the corresponding voxels of the vertices.

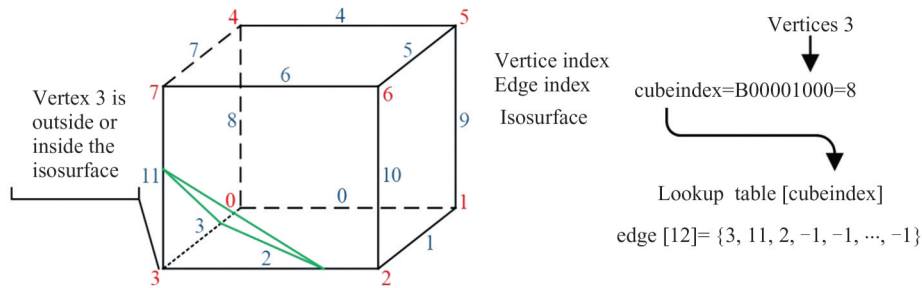


Fig. 6 Schematic diagram of Marching Cubes algorithm

2 Results and Discussion

In this section, a series of comparative experiments was conducted to validate the performance of the proposed method. The experiments were conducted on a PC with the following specifications: CPU: Intel i7-10700F @ 2.9GHz, RAM: 16GB, NVIDIA GeForce RTX 3060. In the task of foreground segmentation for coarse aggregate particles, we utilized the official weight parameters of the U^2 -Net model, which enabled a processing rate of 4.95 images per second for images of coarse aggregate particles at a resolution of 1280×1024 . In this study, the voxel space resolution utilized is $256 \times 256 \times 256$. Using the algorithm proposed in this paper for the 3D reconstruction of coarse aggregate particles, the time required for a single particle is 600 ms.

Table 1 provides a comparative analysis of common 3D scanning technologies and the new method proposed in this study. The proposed system shows a slight decrease in accuracy compared to laser-based 3D scan-

ners and X-ray CT scanners. However, it significantly outperforms these technologies in terms of equipment cost, scanning speed, and field of view. Despite its minor reduction in precision, the proposed system's advantages in cost, speed, and coverage make it particularly suitable for certain applications. This balance of benefits suggests the proposed method could be highly effective in contexts where these factors are prioritized over absolute accuracy. First, the calibration results of the multi-ball experiments were quantitatively analyzed. Then, the calibrated system was used to scan and reconstruct three kinds of standard parts: a sphere with a diameter of 15 mm, and a cylinder with a height of 15 mm, and a cube with a diameter of 15 mm. The accuracy of the proposed 3D reconstruction method was quantitatively analyzed. Finally, coarse aggregate particles of various sizes and shapes were collected and scanned using the proposed system. The reconstructed results were compared and quantitatively analyzed against the 3D data obtained

Table 1 Comparison between different 3D scanning systems

Item	3D optical scanner	X-ray CT	Proposed multi-view scanner
Equipment cost/ \$	15 000	150 000	5 000
Highest resolution/ μm	10	5	100
Time per particle/ s	240	20-40	<10
Scanning scope/ mm	100×100	$D\leq 170$	624×468

from a high-precision structured light 3D scanner. The results showed that the reconstructed surfaces of the proposed method exhibited a high level of similarity with the results obtained from the professional 3D scanner regarding scanning accuracy and completeness.

2.1 Camera Parameter Calibration

A matte black spherical object was manually placed multiple times into the multi-view image acquisition system proposed in this paper. The captured multi-view images were then subjected to image segmentation. Next, contour extraction is performed to obtain the centroid of the projected sphere in the pixel coordinates of each camera. The number of ball drop experiments is an empirical value; in this experiment, a total of 8 ball drop experiments were conducted. The pixel coordinates of the sphere’s centroid, calculated from the multi-view images, are indicated by the green circles in Fig.7. For each set of ball drop experiments, the 3D coordinates of the ball in the world coordinate system can be estimated using Eq. (9) in Section 1. The 3D world coordinates were projected onto the image pixel coordinate system using the camera projection matrix (pre-calibration) exported from Blender software, resulting in the red bounding box in Fig.7. The blue bounding box represents the pixel

coordinates obtained by projecting the 3D world coordinates using the calibrated camera projection matrix. It can be observed that the pixel coordinates obtained using the calibrated camera projection matrix closely match the ideal values. Figure 8 shows the root mean square error between the pixel coordinates of the ball centroid obtained from the camera projection matrices before and after the calibration procedure. It can be seen that the projection error significantly decreases after calibration.

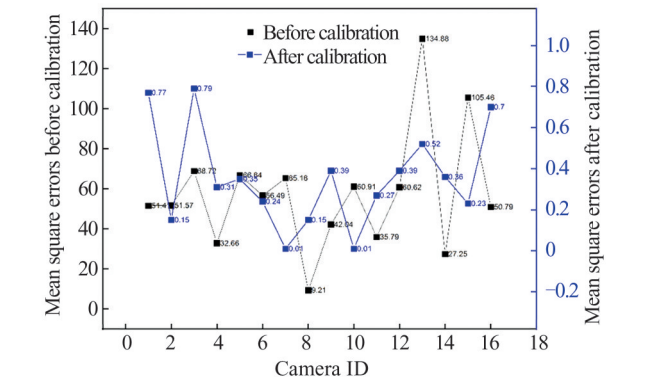


Fig. 8 Pixel root mean square errors comparison

2.2 Standard Parts Scanning and Reconstruction

To quantitatively analyze the accuracy of the system described in this paper, three kinds of standard parts made of black nylon materials were manufactured, as shown in Fig.9. They include a sphere with a diameter of 15 mm, a cylinder with both a base diameter and height of 15 mm and a cube with an edge length of 15 mm. These standard parts were scanned and reconstructed using the multi-view 3D reconstruction system developed in this paper, and the results were compared and analyzed against the ideal contours drawn using Blender software. The scalar field graph, obtained through mesh alignment and distance calculation with Cloud Compare software, is depicted in the second row of Fig.9. Collectively, the 3D contours of the standard parts reconstructed by the system exhibit a high degree

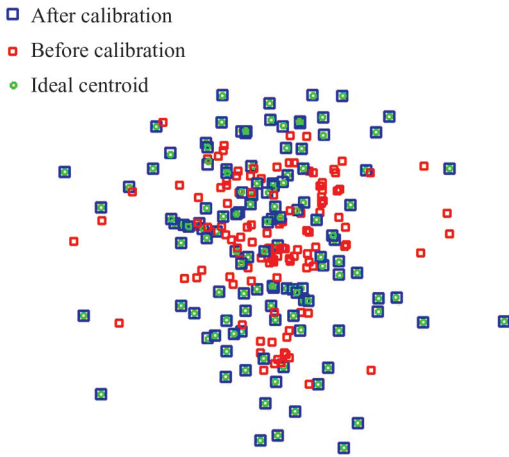


Fig. 7 Comparison of the pixel coordinates of the projected sphere with the ideal values, both before and after calibration

of similarity to their ideal contours. The mean and maximum distances from the ideal contour for the sphere are 0.233 mm and 0.655 mm, respectively. For the cylinder, these values are 0.268 mm and 0.773 mm, respectively. The cube's mean and maximum distances are 0.289 mm and 0.374 mm, respectively. The corresponding numerical values are detailed in Table 2.

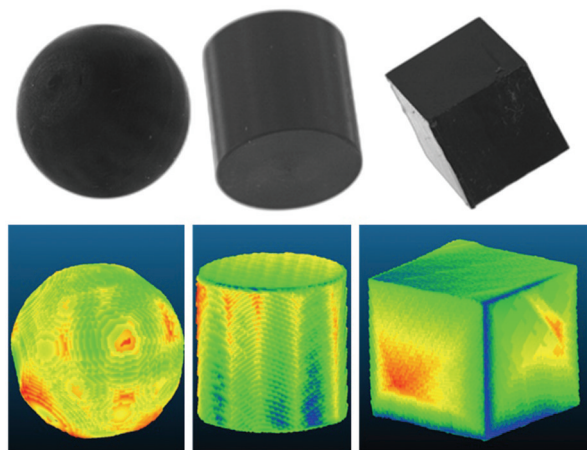


Fig.9 Standard test objects and the scalar field graph of mesh distances

Table 2 Calculation of the distance between the reconstructed mesh and its ideal contour

Item	Max distance	Avg distance	mm
Ball	0.655	0.233	
Cylinder	0.773	0.268	
Cube	0.374	0.289	

2.3 Coarse Aggregate Particles Scanning and Reconstruction

A comparative and quantitative analysis of the accuracy of the proposed system and a high-precision commercial 3D scanner is presented. For this analysis, one coarse aggregate particle of each shape—round, irregular, angular, flaky, and elongated—was selected and underwent scanning and reconstruction using both the proposed system and the high-precision 3D scanner. The AutoScan Inspec system, manufactured by SHINING 3D, was used for comparison. It is a structured light 3D scanner with a fine accuracy of $\leq 10 \mu\text{m}$. The photo of the AutoScan Inspec 3D scanner is shown in Fig. 10. Nevertheless, the scanning and post-processing procedure, performed with the scanner, demands around 4 minutes for each aggregate, proving to be rather time-consuming.

Figure 11 illustrates the scanning and reconstruction



Fig.10 The AutoScan Inspec 3D scanner

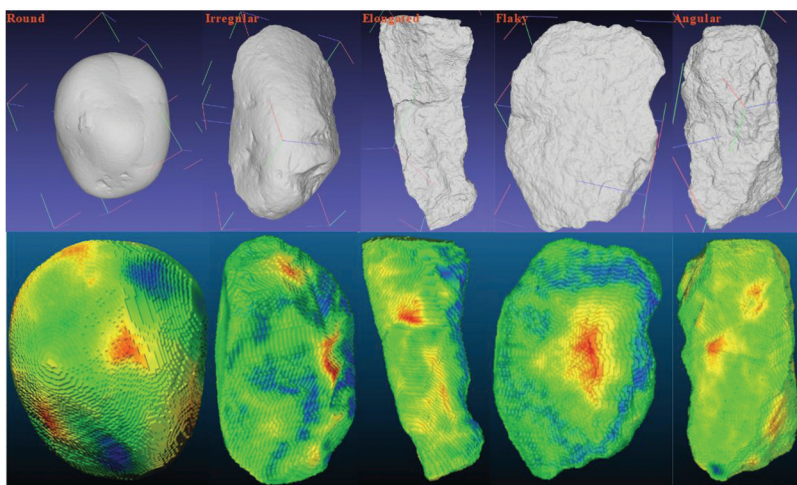


Fig. 11 The 3D contours of coarse aggregate particles

The first row exhibits results obtained by the structured light 3D scanner, whereas the second row shows the corresponding contours obtained by the proposed method and colored with the scalar field graph of mesh distances

tion contours of five typical shapes of coarse aggregates: round, irregular, angular, flaky, and elongated particles. The reconstructed contours were obtained by the multi-view 3D reconstruction system proposed in this paper and compared against the high-precision structured light 3D scanner as a reference. The distance scalar field color maps, which highlight the differences between the two systems, are displayed in Fig.11. It can be observed that the proposed system can perform effective 3D scanning and reconstruction for coarse aggregates of various shapes, with the flaky and elongated aggregates being particularly well reconstructed. In terms of measurement errors, aside from locally large measurement errors where the distance error in some regions reached up to the millimeter level, the average distance measurement error for the five coarse aggregates was 0.434 mm. The corresponding numerical values are detailed in Table 3.

Table 3 Calculation of the distance between the reconstructed mesh of the coarse aggregate particles and its contour scanned by commercial 3D scanner

Shapes	Max dist.	Avg dist.	Box dimensions
Round	1.209 19	0.389 450	25.5×25.4×16.8
Irregular	1.659 24	0.360 095	25.3×20.1×19.2
Elongated	1.388 52	0.415 688	28.1×28.8×15.1
Flaky	1.974 16	0.477 008	27.9×22.6×24.1
Angular	2.040 53	0.525 525	27.6×36.2×22.1
Average	1.654 00	0.434 000	—

3 Conclusion

This study introduces a multi-view imaging system specifically designed for capturing unobstructed images of free-falling coarse aggregate particles. The system incorporates 16 industrial cameras, and the calibration of this multi-view system is achieved through a unique approach based on geometric error optimization. Multiple ball-dropping experiments were conducted to refine the camera projection parameters and enhance calibration precision. The SfS algorithm was employed to generate 3D voxel data from multi-view images containing calibration directly pose information. The Marching Cubes algorithm was then utilized to extract the surface mesh representation of the particles.

Experimental validation was performed through 3D reconstruction experiments on standard parts and coarse aggregate particles of various shapes and sizes. The re-

constructed shapes were compared and quantitatively analyzed. The average distance measurement error for the coarse aggregate samples was 0.434 mm, demonstrating the accuracy and reliability of the proposed method.

Compared with existing 3D scanning methods, the system developed in this study offers several advantages. Firstly, it features a simple operation, making it user-friendly and accessible to a wider range of users. The scanning speed is also significantly faster, allowing for efficient data acquisition. The resulting 3D reconstructions exhibit a high level of completeness, capturing the detailed surface information of the particles. These findings highlight the effectiveness and efficiency of the proposed system for coarse aggregate particle surface reconstruction. Lastly, the system can be implemented at a moderate cost, making it a cost-effective solution for various applications.

Although the proposed method performs satisfactorily with the current configuration, some issues remain unaddressed. Firstly, it does not reconstruct the surface texture of coarse aggregate particles. Secondly, the reconstruction results exhibit local geometric errors. Hence, further research using alternative methods, such as neural networks, could be conducted to improve the geometric accuracy of 3D reconstruction and enable the realistic reconstruction of the surface texture of particles.

References

[1] Gong F Y, Liu Y, You Z P, *et al.* Characterization and evaluation of morphological features for aggregate in asphalt mixture: A review[J]. *Construction and Building Materials*, 2021, **273**: 121989.

[2] Bessa I S, Branco V T F C, Soares J B, *et al.* Aggregate shape properties and their influence on the behavior of hot-mix asphalt[J]. *Journal of Materials in Civil Engineering*, 2015, **27**(7): 04014212 .

[3] Sun Z Y, Wang C F, Hao X L, *et al.* Quantitative evaluation for shape characteristics of aggregate particles based on 3D point cloud data[J]. *Construction and Building Materials*, 2020, **263**: 120156.

[4] Sun Z Y, Liu H Y, Ju H Y, *et al.* Assessment of importance-based machine learning feature selection methods for aggregate size distribution measurement in a 3D binocular vision system[J]. *Construction and Building Materials*, 2021, **306**: 124894.

- [5] Ge H T, Sha A M, Han Z Q, *et al.* Three-dimensional characterization of morphology and abrasion decay laws for coarse aggregates[J]. *Construction and Building Materials*, 2018, **188**: 58-67.
- [6] Anochie-Boateng J K, Komba J J, Mvelase G M. Three-dimensional laser scanning technique to quantify aggregate and ballast shape properties[J]. *Construction and Building Materials*, 2013, **43**: 389-398.
- [7] Li Q J, Zhan Y, Yang G W, *et al.* 3D characterization of aggregates for pavement skid resistance[J]. *Journal of Transportation Engineering, Part B: Pavements*, 2019, **145**(2): 04019002.
- [8] Wang H F, Wu Z J, He Z C, *et al.* Detection of HF-ERW process by 3D bead shape measurement with line-structured laser vision[J]. *IEEE Sensors Journal*, 2021, **21**(6): 7681-7690.
- [9] Wang H F, Wang Y F, Zhang J J, *et al.* Laser stripe center detection under the condition of uneven scattering metal surface for geometric measurement[J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(5): 2182-2192.
- [10] Tuan N M, Kim Y, Lee J Y, *et al.* Automatic stereo vision-based inspection system for particle shape analysis of coarse aggregates[J]. *Journal of Computing in Civil Engineering*, 2022, **36**(2): 04021034.
- [11] Su D, Yan W M. 3D characterization of general-shape sand particles using microfocus X-ray computed tomography and spherical harmonic functions, and particle regeneration using multivariate random vector[J]. *Powder Technology*, 2018, **323**: 8-23.
- [12] Liu B, Fan H M, Jiang Y, *et al.* Evaluation of soil macro-aggregate characteristics in response to soil macropore characteristics investigated by X-ray computed tomography under freeze-thaw effects[J]. *Soil and Tillage Research*, 2023, **225**: 105559.
- [13] Zhao L H, Zhang S H, Huang D L, *et al.* 3D shape quantification and random packing simulation of rock aggregates using photogrammetry-based reconstruction and discrete element method[J]. *Construction and Building Materials*, 2020, **262**: 119986.
- [14] Chen Z Q, Jia Y S, Wang S Q, *et al.* Image-based methods for automatic identification of elongated and flat aggregate particles[J]. *Construction and Building Materials*, 2023, **382**: 131187.
- [15] Pei L L, Sun Z Y, Yu T, *et al.* Pavement aggregate shape classification based on extreme gradient boosting[J]. *Construction and Building Materials*, 2020, **256**: 119356.
- [16] Jin C, Wang S L, Liu P F, *et al.* Virtual modeling of asphalt mixture beam using density and distributional controls of aggregate contact[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2023, **38**(16): 2242-2256.
- [17] Jin C, Wang P S, Yang X, *et al.* Analysis on gradation parameters of asphalt mixture based on 3D virtual measurement[J]. *Journal of Highway and Transportation Research and Development*, 2019, **36**(8): 1-8(Ch).
- [18] Peng Y P, Wu Z B, Cao G Z, *et al.* Three-dimensional reconstruction of wear particles by multi-view contour fitting and dense point-cloud interpolation[J]. *Measurement*, 2021, **181**: 109638.
- [19] Liu C Q, Li J, Gao J, *et al.* Three-dimensional texture measurement using deep learning and multi-view pavement images[J]. *Measurement*, 2021, **172**: 108828.
- [20] Wu X B, Wang J S, Li J J, *et al.* Retrieval of siltation 3D properties in artificially created water conveyance tunnels using image-based 3D reconstruction[J]. *Measurement*, 2023, **211**: 112586.
- [21] Wang Y L, Deng N, Xin B J. Investigation of 3D surface profile reconstruction technology for automatic evaluation of fabric smoothness appearance[J]. *Measurement*, 2020, **166**: 108264.
- [22] Ju X Y, Henseler H, Peng M J Q, *et al.* Multi-view stereophotogrammetry for post-mastectomy breast reconstruction [J]. *Medical & Biological Engineering & Computing*, 2016, **54**(2): 475-484.
- [23] Carvajal-Ramírez F, Navarro-Ortega A D, Agüera-Vega F, *et al.* Virtual reconstruction of damaged archaeological sites based on Unmanned Aerial Vehicle Photogrammetry and 3D modelling. Study case of a southeastern Iberia production area in the Bronze Age[J]. *Measurement*, 2019, **136**: 225-236.
- [24] Li T Y, Liu S C, Bolkart T, *et al.* Topologically consistent multi-view face inference using volumetric sampling[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2021: 3804-3814.
- [25] Xiong J, Zhong S, Zheng L. An automatic 3D reconstruction method based on multi-view stereo vision for the mogao grottoes[J]. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2015, XL-4/W5: 171-176.
- [26] Zhao J, Xu P, Huang S L, *et al.* Underwater 3D reconstruction based on multi-view stereo[C]//*Ocean Optics and Information Technology*. New York: SPIE, 2018: 117-123.
- [27] Ito K, Ito T, Aoki T. PM-MVS: PatchMatch multi-view stereo[J]. *Machine Vision and Applications*, 2023, **34**(2): 32.
- [28] Zhang Z. A flexible new technique for camera calibration system[J]. *IEEE Transactions on Pattern Analysis and Machine*, 2000, **22**(1): 1330-1334.
- [29] Schönberger J L, Frahm J M. Structure-from-motion revis-

- ited[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 4104-4113.
- [30] Perez A J, Perez-Soler J, Perez-Cortes J C, *et al.* Improving multi-view camera calibration using precise location of sphere center projection[J]. *Computers*, 2022, **11**(6): 84.
- [31] Furukawa Y, Hernández C. Multi-view stereo: A tutorial[J]. *Foundations and Trends® in Computer Graphics and Vision*, 2015, **9**(1/2): 1-148.
- [32] Perez A J, Perez-Soler J, Perez-Cortes J C, *et al.* Alignment and improvement of shape-from-silhouette reconstructed 3D objects[J]. *IEEE Access*, 2024, **12**: 76975-76985.
- [33] Lorensen W E. History of the marching cubes algorithm[J]. *IEEE Computer Graphics and Applications*, 2020, **40**(2): 8-15.
- [34] Zhao W. A Broyden-Fletcher-Goldfarb-Shanno algorithm for reliability-based design optimization[J]. *Applied Mathematical Modelling*, 2021, **92**: 447-465.

无遮挡多视角成像粗集料颗粒三维表面重建方法

高荣¹, 孙朝云^{1†}, 郭建兴^{2†}, 李伟¹, 杨明¹, 郝雪莉¹, 姚博彬³, 王会峰³

1. 长安大学 信息工程学院, 陕西 西安 710064

2. 长安大学 基建处, 陕西 西安 710064

3. 长安大学 电子与控制工程学院, 陕西 西安 710064

摘要: 快速准确地评估粗集料颗粒的几何特性对保证公路工程中路面性能至关重要。本文构建了一种基于无遮挡多视角成像技术的粗集料颗粒三维表面轮廓重建系统,能够同步捕捉自由下落状态下粗集料的多视角图像。采用黑色哑光球体和非线性优化方法实现对多视角成像系统中相机投影矩阵的标定。使用预训练的分割模型去除多视角图像中的背景干扰,应用剪影建模算法生成粗集料颗粒的三维体素数据,使用行进立方体算法构建其三维表面轮廓。对标准部件和不同的粗集料颗粒进行了测量和对比分析,结果表明,本系统对粗集料颗粒的平均测量精度达到0.434毫米,实现了亚毫米精度测量,并且显著提高了扫描和重建效率。

关键词: 三维重建; 多视角成像; 粗集料颗粒; 剪影建模; 多视角相机标定

