

## *Retraction*

# **Retracted: Sculpture 3D Modeling Method Based on Image Sequence**

### **Complexity**

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Complexity. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] X. Liu, "Sculpture 3D Modeling Method Based on Image Sequence," *Complexity*, vol. 2021, Article ID 9916725, 13 pages, 2021.

## Research Article

# Sculpture 3D Modeling Method Based on Image Sequence

**Xiaofei Liu** 

*Art College, Xi'an University, Xi'an, Shaanxi 710065, China*

Correspondence should be addressed to Xiaofei Liu; [1019750301@xawl.edu.cn](mailto:1019750301@xawl.edu.cn)

Received 10 March 2021; Revised 8 April 2021; Accepted 15 April 2021; Published 22 April 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Xiaofei Liu. 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.

This thesis first introduces the basic principles of model-based image sequence coding technology, then discusses in detail the specific steps in various implementation algorithms, and proposes a basic feature point calibration required in three-dimensional motion and structure estimation. This is a simple and effective solution. Aiming at the monocular video image sequence obtained by only one camera, this paper introduces the 3D model of the sculpture building into the pose tracking framework to provide initial depth information. The whole posture tracking framework can be divided into three parts, namely, the construction of the initial sculpture model, the posture tracking between frames, and the robustness processing during continuous tracking. In order to reduce the complexity of calculation, this paper proposes a new three-dimensional mesh model and a moving image restoration algorithm based on this model. At the same time, the influence of the intensity and direction factors in the scene is added, the simulation results are given, and the next step is discussed. The optimization work that needs to be done.

## 1. Introduction

The data acquisition and 3D modeling of sculpture buildings are important contents of the construction of digital cities and smart cities. The 3D modeling method based on close-range image sequences uses ordinary digital cameras as image acquisition equipment, which has low cost, high efficiency, and low labor intensity. It reflects the geometric details of the building surface and has realistic texture information, which provides an effective way to quickly, accurately, and truly reproduce 3D city information [1]. However, the ground close-up images often use wide baselines and large angles of intersection, resulting in serious occlusion, large perspective distortion between images, and problems such as single building texture and multiple repetitions, which bring difficulties to the extraction and matching of image information. The accuracy and automation of 3D reconstruction of buildings need to be improved [2]. In the image information, the point feature is accurately positioned, which can effectively restore the mapping relationship between the two-dimensional image and the three-dimensional geometry, and the straight line feature is the main contour feature of the building, which controls the entire building structure [3].

In the process of point cloud reconstruction of sparse image sequence images, automatic sculpture point cloud generation technology is used to improve the ability to distinguish and recognize sparse images of multilayer sculpture points. Through image information processing and analysis methods, analyze sparse images of multilayer sculpture points, and improve the ability of reconstruction and output detection of sparse images of multilayer sculpture points. Research on the 3D reconstruction method of sparse image sequence, combined with image adaptive feature reconstruction and the point cloud data analysis method, realizes the adaptive structure reorganization of sparse image sequence and improves the automatic recognition and detection ability of sparse image. In traditional methods, the sparse image contour reconstruction techniques of sculpture points mainly include the following, which are based on RGB image reconstruction method, block size reconstruction method, point distribution model method, etc. [4]. The traditional three-dimensional 3D model theory can accurately restore the 3D model of the object and can also adjust the light intensity and the selection of texture features artificially, so it has been widely used in mechanical manufacturing, construction engineering, game animation, etc. [5]. However, using this technology requires

people who are very mature in the technology to perform manual operations, and the operability is poor, and it takes a long time to obtain three-dimensional information. When obtaining three-dimensional information of many irregular objects, the results obtained are often true. The information differs greatly, and the expected results are not achieved [6]. The 3D model scanning equipment is expensive and limited for personal use and can only reconstruct a single object. For images with irregular characteristics such as remote sensing images, the 3D reconstruction effect is very poor, and the object is often not obtained. The color distribution information of the surface has been explained. However, multiple images of different angles of the object are used to recover the three-dimensional information of the object through computer vision theory [7]. From the beginning of the research, it has attracted people's attention. This method only requires multiple images of different angles to recover the three-dimensional model. The traditional analysis method carries out the 3D reconstruction of the sparse sculptural point image to improve the feature resolution ability of the image, but this method has the problems of too large computational cost and poor feature resolving ability to reconstruct the sparse image contour of the sculpture point [8].

This paper studies several key technical issues involved in the process of building 3D modeling based on the point and line features of the close-range image sequence. The main work and research results include the following: focusing on the characteristics caused by the large change in the perspective of the close-range image and the single texture of the building. Point matching is difficult, focusing on the local affine invariant features that remain unchanged for viewing angle changes. Based on the analysis of the advantages and disadvantages of the current local affine invariant feature detection methods, a feature point with complementary advantages of multiple features is proposed. The detection and matching method can obtain high accuracy and large number of matching point pairs, which lays a good foundation for the subsequent camera parameter estimation. The robustness of camera parameter estimation is based on the principle of beam adjustment optimization in which camera parameters and three-dimensional point coordinates are solved synchronously. According to the characteristics of anisotropic error distribution caused by the perspective distortion of characteristic points of the close-range image, the objective function is constructed as an entry point, and a method for solving camera parameters and three-dimensional point coordinates that takes into account the anisotropy of feature point errors is proposed. Aiming at the problems of traditional methods, a method for automatic 3D reconstruction of sculpture points based on sparse image sequence is proposed, and a 3D reconstruction method based on sparse scattered points and sharp template feature matching method is constructed for image 3D reconstruction. Three-dimensional corner detection and edge contour feature extraction methods are used to detect the three-dimensional point cloud feature of the sculpture point sparse image and perform information fusion processing on the detected sculpture point sparse image point cloud data; use

the gradient operation method to perform feature decomposition, combined with local mean denoising methods. Purify and filter the image to improve the ability of sparse image contour reconstruction of sculpture points; use sharp template feature matching and block segmentation technology to realize automatic 3D reconstruction of sculpture point cloud, and conduct simulation experiment analysis. In view of the importance of straight line features to the three-dimensional geometric structure of buildings, a strategy of extraction, matching, and reconstruction of straight line features is designed. On the basis of linear feature extraction, a multiconstrained linear feature matching method is proposed to solve the matching difficulties caused by broken lines and incomplete extraction. The feature of this method is to use the local affine invariance between points and lines. Searching for candidate matching sets improves the accuracy of matching straight line search. Combine epipolar constraint and straight line angle and the similarity measure of support domain grayscale to perform straight line feature matching. Experimental results show that this method has a higher correct matching rate.

By converting the three-dimensional color information of the real world into digital signals that can be directly processed by the computer, it can provide an effective means for the digitization of physical objects. It is very different from traditional flatbed scanners, cameras, and graphic capture cards. The scanned object is not a flat pattern, but a three-dimensional object. Secondly, through scanning, the three-dimensional space coordinates of each sampling point on the surface of the object can be obtained, and the color of each sampling point can also be obtained by color scanning. Some scanning devices can even obtain the internal structure data of the object. Finally, it does not output two-dimensional images but contains a digital model file of the three-dimensional space coordinates and the color of each sampling point on the surface. This can be used directly for CAD or 3D animation. The color scanner can also output the color texture map of the surface of the early coordinate measuring machine (CMM) used for three-dimensional measurement. The probe is installed on a servo device with three degrees of freedom (or more degrees of freedom), and the driving device drives the probe to move in three directions. When the probe touches the surface of the object and measures its movement in three directions, the three-dimensional coordinates of the point on the surface of the object can be known. Control the movement and touch of the probe on the surface of the object to complete the three-dimensional measurement of the entire surface. Its advantages are high measurement accuracy; its disadvantages are expensive, complicated control when the object shape is complex, slow speed, and no color information. The rangefinder sends the signal to the surface of the measured object and can estimate the spatial position of the surface of the object according to the reflection time or phase change of the signal.

## 2. Related Work

The so-called image-based 3D reconstruction generally refers to the use of images to restore the 3D geometric model of

an object. For the reconstruction of geometric models, it can be summed up as follows: single-image-based methods, stereo vision methods, motion restoration structure methods, silhouette contour-based methods, etc. [9].

Adamopoulos et al. [10] proposed that the camera projection transformation is a projective transformation from a three-dimensional space to a two-dimensional plane. It is impossible to obtain its geometric structure from a single image of a space scene. Although the shape is restored from light and dark, the method supports the reconstruction of a three-dimensional model from a single image. However, this is always an ill-conditioned problem. The assumption that the surface reflectance of the object is known is usually not true. Gherardini et al. [11] studied the shading method and optimized it in combination with the image contour. It is mainly aimed at specific illumination, with relatively simple structure and ignoring texture. Feature images are difficult to apply to actual images. The general method of 3D reconstruction based on a single image is to use the known geometric characteristics of parallel, vertical, coplanar, vanishing point, and vanishing line in the scene and realize the 3D reconstruction of the object in an interactive manner. Lanteri et al. [12] interactively marked the key points and parallel lines of the objects in the scene for a single image with at least two vanishing points and used the vanishing points to calculate camera parameters to establish a geometric model. Barrile et al. [13] proposed to a method of measuring distance from a single image of the shape. The three-dimensional reconstruction method based on stereo vision is used to reconstruct the geometric model of the scene from two or more calibrated images. For the biological vision system, when two eyes are used to observe an object, there will be a sense of distance or depth. Stereoscopic movies imitate the principle of seeing the world with the human eye, producing a realistic sense of depth. The research of Suciati et al. [14] showed that when shooting, the two cameras shoot at the same time, imitating the perspective of the human eye, and during projection, the images captured by the two cameras are projected onto the same screen at the same time, and the principle of polarized light is used, so that each eye can only see the corresponding polarized light image; that is, the left eye can only see the image taken by the left camera, and the right eye can only see the image taken by the right camera, so that people feel the real three-dimensional scene. Adamopoulos et al. [15] in computer vision systems and digital photogrammetry systems also simulate the way that human eyes observe the scene. The stereo vision method generally obtains two different perspective images of the same space scene by a stereo camera system and calculates the points in the space. The time difference between the two images is used to obtain the three-dimensional coordinates. The stereo vision method focuses on how to accurately determine the dense pixel correspondence between the matching views. Dipietra et al. [16] started from the matching similarity measurement model, epipolar constraint, parallax constraint, etc. and studied the method of dense matching of stereo pairs. Liang Bojun used the matching diffusion method to generate the dense disparity map and then get dense depth information.

Zhang et al. [17] proposed a strategy for camera calibration and single-image modeling through prior knowledge of planes, parallel lines, and angular structures and merged the local models corresponding to each image in multiple image sequences. Complete the unification of multiscene models. Traditional 3D reconstruction methods all need to calibrate the camera in advance. In recent years, more and more researchers have devoted themselves to the research of camera self-calibration work, only through two or more uncalibrated images to recover the geometric structure of the scene and the motion parameters of the camera at the same time. For the 3D reconstruction of uncalibrated images, the most commonly used method is the motion recovery structure method. This method uses numerical methods to detect and match feature point sets in the image, at the same time, recover the camera motion parameters and scene geometry, and obtain the 3D model of the object [18]. When shooting the target, you can freely move the position of the camera and adjust the focal length of the camera as needed; since there is no need to calibrate the camera parameters in advance, the reconstruction result will not be affected by factors such as inaccurate calibration information or slight changes in camera parameters during the shooting process. Because the method calculates only the three-dimensional coordinates of all matching feature points, only sparse three-dimensional spatial feature points can be reconstructed, so it is generally necessary to further reconstruct the three-dimensional model with the help of unique matching between images [19].

Although some domestic research teams in the field of 3D reconstruction have only achieved theoretical breakthroughs and have not fully translated into practical applications, more and more institutions are still investing a lot of energy and funds to achieve expectations breakthrough. Correspondingly, the content studied in this paper is based on consulting a large number of domestic and foreign documents, improving related algorithms, and centering on the three-dimensional reconstruction algorithm based on sequence images and in-depth research based on two images and the multiple reconstruction algorithm. Based on the modeling method of silhouette contour line, the three-dimensional model of the object is obtained by analyzing the contour image or silhouette contour line of the object in multiple perspectives. When observing a space object from multiple perspectives, a silhouette contour line of the object will be projected on the screen of each perspective. This silhouette contour line and the corresponding perspective projection center together determine a cone in the three-dimensional space. The outline of the object can be obtained by viewing the shell. When enough contour images are used, the visual shell is considered a reasonable approximation of the three-dimensional model of the object. Since the intersection calculation of three-dimensional cone shells is a problem of intersecting complex polyhedrons in three-dimensional space, the calculation complexity is high, so the contour method mainly solves the problem of rapid intersection of three-dimensional cone shells. The earliest method of using contours for 3D reconstruction is to discretize the 3D space where the object is located into voxels,

except for the voxels projected outside the contour area to obtain a 3D model of the object [20, 21]. The modeling method based on silhouette contour line is mainly suitable for convex objects, but some concave areas on the surface of the object cannot be obtained from the silhouette contour information, so this method is not suitable for all shapes of entities [22–25]. The more silhouette contours, the closer the generated visual shell is to reality. Therefore, contour-based methods generally require a larger number of pictures and have higher algorithm complexity. The experimental results in this paper show that the point cloud reconstructed by the image 3D reconstruction algorithm based on region growth is dense enough, and the reconstructed target object has a strong sense of reality and can fully restore the detailed features, indicating that the reconstruction of the image target has a strong practicality performance and can effectively reduce data. The improved algorithm can eliminate mismatched points very well and can also speed up the optimization speed of reconstruction results.

### 3. Construction of Sculpture 3D Model Based on Image Sequence

*3.1. Sequence Mapping of Multiview Geometric Reconstruction Model.* Each digital image is stored in the computer in the form of an array. Each element in the row and column of the image corresponds to a pixel in the image, and its value is the brightness or grayscale of the image pixel, in the image pixel coordinate system, taking the upper left corner of the image as the origin of the coordinate system, increasing in the horizontal direction to the right, and increasing in the vertical direction. The coordinates of each pixel are the row and column number of the pixel in the array. The pixel coordinate system only indicates the number of rows and columns of the pixel in the array, while the image coordinate system uses physical units (such as millimeters) to indicate that the pixel is in the image. The origin of the image coordinate system is the intersection point of the optical axis of the camera and the image plane. This point is generally located at the center of the image. If the coordinates are in the pixel coordinate system, then in the image coordinate system, the size of each pixel in the axial direction can be obtained. The formulae (1)-(2) are as follows:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{1}{t_x} & 0 & 0 \\ 0 & \frac{1}{t_y} & 0 \\ 0 & 0 & \frac{1}{t_z} \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}, \quad (1)$$

$$\begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix} [x_u \ y_u \ z_u]^{-1}. \quad (2)$$

Let  $F$  be a domain, the members of the vector space  $V$  are called vectors, and the members of  $F$  are called scalars. If  $F$  is a real number field  $R$ ,  $V$  is called a real vector space; if  $F$  is a complex number field  $C$ ,  $V$  is called a complex vector space; if  $F$  is a finite field,  $V$  is called a finite field vector space. The vector space  $V$  is a set on  $F$ . Let  $0$  be the zero vector in the vector space  $V$ , and the  $k$  vectors in  $V$  are linearly related, if and only if there are  $k$  scalars that are not all zeros satisfying. As shown in the following formulae,

$$f(x, y, z) = u(x, y, kz) - u(x, y, z), \quad (3)$$

$$g(x, y, z) = \sum f(x, y, z) * f(x, y, z)^T. \quad (4)$$

Among them,  $x$  represents the grid cell density,  $y$  represents the optimal threshold of the density histogram,  $z$  represents the relative density of the two cells, and  $f$  represents the pixel point in the histogram.

The collected original 3D sculpture point sparse image adopts the local mean value noise reduction method to separate the noise, as shown in formula (5), determine the threshold value of the feature point, and perform the noise separation processing according to the threshold judgment result, and the formed feature point set is

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \chi \end{pmatrix}. \quad (5)$$

On the reconstructed surface of key feature points, the Harris corner detection method is used to smooth the sparse image of sculpture points. As shown in formula (6),  $f(x, y)$  represents the gray value of the feature point clustering of the sparse image of sculpture points at  $(x, y)$ ,  $g(x, y, z)$  represents the Taubin smoothing operator, and  $h(x, y, z)$  is the seed point array distribution of the table image. The block matching method is used for feature registration, and the degree of clustering coincidence is calculated. Determine the main direction of the current seed point  $P$  of the sparse image of the sculpture point, and perform the spatial neighborhood decomposition according to the Euclidean distance  $E(d(x, y))$  of the corresponding feature vector to obtain the spatial neighborhood decomposition coefficient.  $p$  represents the decomposition scale of the approximately convex part of the sparse image of the sculpture point. The three-dimensional point cloud model is fused and clustered by the principal component segmentation method, and the distance between the cluster centers is  $d(x, y)$ . Based on the voxel feature segmentation method, the feature segmentation output is obtained. According to the above analysis, set the vector quantization feature quantity of the sparse image of sculpture points, extract the gray pheromone of the image, and obtain the first  $i$ -dimensional feature template of the sparse image of sculpture points:

$$d(x, y, z) = |f(x, d(x, y), y) - f'(x, y)|^2. \quad (6)$$

In the formula,  $f$  represents the reference template image and  $f'$  represents the image to be reconstructed. According

to the above analysis, information fusion processing is performed on the point cloud data of the detected sparse images of sculpture points, and the gradient operation method is used for feature decomposition to realize the information enhancement and fusion filtering of the sparse images of sculpture points.

### 3.2. Detection of Feature Points of Image Sculpture Buildings.

Because the feature points are accurately positioned and can provide effective three-dimensional information, establishing the correspondence between feature points in different views has become a necessary prerequisite for restoring the three-dimensional structure of the target object. As for the ground close-up images of buildings, because the wide baseline and large angle of intersection are often used, problems such as serious occlusion, brightness changes between images, and large perspective distortion will occur. In addition, the surface texture of the building is single and repetitive. These make it difficult to automatically recognize and correct the local shape of the corresponding neighborhood window. Therefore, it is necessary to develop a feature point detection and matching strategy for the building ground close-up image. The common feature point detection and matching algorithms are summarized. The specific 3D model hierarchy process is shown in Figure 1, and after analyzing their advantages and disadvantages, combined with the characteristics of the building ground close-up image, a multifeature complementary local affine invariant feature point detection and matching method is proposed. In the process of 3D reconstruction, it is necessary to acquire a large number of points in the image, but not all of these points can be helpful to the reconstruction during reconstruction. Only the points that are significantly different from the surroundings are the points that are needed. These points are called feature points. If the two-dimensional coordinates of these feature points can be obtained in the two images, respectively, then three-dimensional reconstruction can be carried out. The traditional algorithm needs to locate the local extreme points in the scale space and the two-dimensional plane space at the same time to determine the location of the local extreme points, and use the selected local extreme points as the key points. Finally, according to the key points The position information of the key point is calculated, and the gradient direction of all points in the neighborhood of the key point is calculated to determine the main direction of the key point, and then the invariance of the operator to the geometric transformation and rotation can be completed. The standardization of multi-functional local areas can be further realized by feature point local area description and feature point matching and verified by comparison experiments with other algorithms.

After the target is imaged, for each point on the target, the internal and external parameters of the camera are the same, which can be regarded as a known constant, but because the coefficient matrix of the affine transformation is related to the sum, the pixels of the two images do not obey the same affine transformation model shows that the geometric relationship between all pixels of two images can be

represented by the same affine transformation model. Since the two cameras, respectively, represent the axis coordinates of the same target along the respective optical axis directions, that is, the depth in the projection direction, for practical applications, the following cases are constant or can be approximated as constant. As shown in formula (7), the second-order central moment is a more generalized concept of covariance. The second-order central moment based on the vector representation is constructed using the pixel information of the irregular feature area, which is used to adjust the irregular feature area. The irregular area can be adjusted to an elliptical area using the following formula:

$$\sum (u) = \frac{1}{|t|} \sum_u (u - f(u))^T (u - f(u)). \quad (7)$$

The affine transformation between images can be understood as the mapping between sets, and the affine invariant feature can be considered as an invariant subspace in the feature space that is not affected by this mapping. Affine invariant features are divided into global affine invariant feature extraction methods and local affine invariant feature extraction methods. Among them, because of local affine invariant features, only the information of the local area of the target is used. Since these local regions may be scattered in different positions of the target, when features are extracted independently for each region, even if there is some occlusion in the environment of the target, partial information of the target can still be obtained through local feature extraction, so as to realize the recognition and recognition of the target. The most stable extreme value area sets a threshold for the image, analyzes the relationship between the gray value of the image pixel point and the given threshold, and constructs a four-connected area. The gray value of the pixels inside the area is all grayer than the gray value of the boundary pixels. The degree value is large (maximum value area), or the gray value of the pixels inside the area is smaller than the gray value of the boundary pixels (minimum value area). The method is specifically described as follows: set a threshold for a given image, and set the points whose grayscale is less than the threshold to "black," otherwise set to "white." The image is initially set to completely white, then black appears. When gradually increasing, the "black" points representing the local minimum value appear one after another. In this way, as new local points are gradually produced, the old local "black" points gradually merge. When the maximum is reached, the image becomes completely black. Similarly, if the process is reversed, the image will change from all black at the beginning to all white at the end. In the threshold change process, the continuous area composed of the local "black" points (or the "white" points in the reverse process) that have appeared is called the extreme value area. The gray values of the pixels in this area are all less than (or greater than) the gray value of boundary pixels. The detection operator detects homogeneous regions in the image, which depends on the structure of the image itself, so it has the characteristics of covariation with the image affine transformation and linear illumination transformation. The basic principle is to start from the extreme

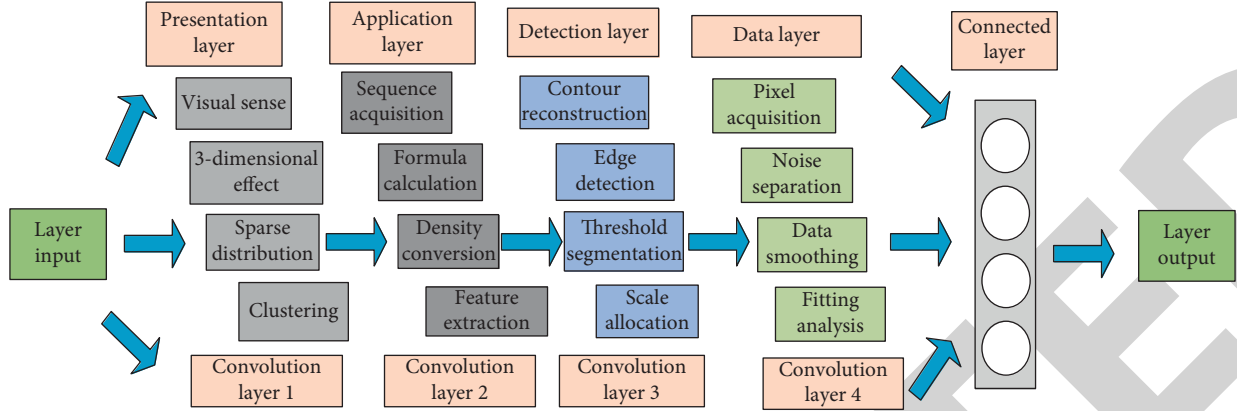


FIGURE 1: Flowchart of characteristic layers of image sculpture buildings.

gray value of the image and find the boundary point outward along the direction of the fortune ray. The boundary point is the pixel point that has the largest grayscale difference from the extreme point and is closest to the extreme point, and then the grayscales of the boundary points of the extreme neighborhood are sequentially connected to form an irregular area. Therefore, it is called the affine invariant region. Among them, the boundary points can be a gray value function is obtained by seeking extreme values. The geometric positioning accuracy is high, the number of detected features is the largest, and a good match can still be obtained in the presence of occlusion, but the performance of the detection operator for illumination and viewing angle changes is not good; although the performance is the best, the detected features are less, and too few features will cause instability in the matching result due to mismatching.

Through the principle of the above local feature detection algorithm, it can be seen that the detected features are stable local blocks, which are more suitable for scenes with obvious texture features, while feature detection based on corner areas is obvious for structures such as buildings. The specific point algorithm of the sculpture model is shown in Figure 2. The scene performs better, and the two features have fewer regions of the same feature. The feature region itself has scale and affine invariance. After the descriptor is used to describe the feature region in vector, the resulting descriptor adds rotation invariance. Since the algorithm makes full use of the neighborhood information when calculating the direction of feature points and uses the idea of histogram statistics and Gaussian weighting when calculating the gradient direction at the feature points, this provides better matching of feature points with positioning deviations. The description operator first takes the neighborhood with the feature point as the center as the sampling window, calculates the gradient direction of each pixel, then uses the gradient histogram (column) to calculate the main direction of the window, and uses the center of the feature area as the origin of the coordinate axis. Rotation is the direction of the feature point to ensure rotation invariance. Divide the feature area into subregions, and calculate the gradient direction histogram of each direction of each subregion. In order to highlight the role played by the center of the region, Gaussian weighting is performed on the center

of the region to make the closer to the center. The greater the contribution of the pixel gradient direction information, the final one-dimensional feature vector is obtained.

**3.3. Image Sequence Parameter Estimation and Three-Dimensional Coordinate Calculation.** In the process of 3D reconstruction of target based on image sequence, camera parameter estimation is a necessary condition for establishing the mapping relationship between 2D image and 3D target, and it is also the core of motion recovery structure. The narrow sense of camera parameter estimation only refers to the internal parameter matrix. The camera parameter estimation in this article refers to the generalized camera parameters. In the form of a projection matrix, it can directly establish a correspondence between the three-dimensional world coordinate system and the two-dimensional image coordinate system. The relationship between is the implicit representation of the camera's internal and external parameters. The use of the projection matrix can further realize the calculation of the three-dimensional space point coordinates, and its calculation accuracy directly determines the accuracy of the model. The camera parameters and three-dimensional space point solution method are based on the beam adjustment optimization model. First, the singular value decomposition of the essential matrix is used to obtain the parameters of the "quasi" European model of the two images (the initial camera internal and external parameters and the three-dimensional space coordinates of the characteristic points), as the initial parameters of the beam adjustment; then, use the beam adjustment model to nonlinearly optimize the internal and external camera parameters of the two images and the spatial coordinates of the points to be fixed; for each new image, according to the calculated three-dimensional points, add the corresponding relationship of the image feature points, solve the projection matrix of the image and solve the matching feature point pairs that have not been reconstructed, and optimize all the current images using the beam adjustment model; finally, the global beam adjustment optimization is performed to obtain the camera. The optimal solution of parameters and three-dimensional point coordinates is obtained. The distribution of image sequence

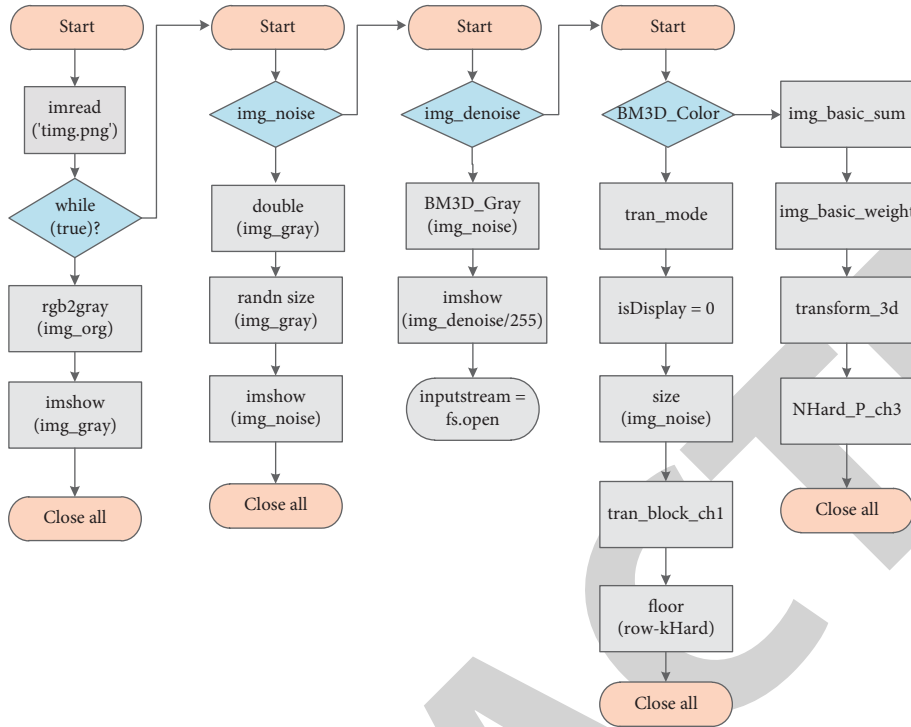


FIGURE 2: Feature point detection algorithm for image sculpture building.

parameter estimates is shown in Figure 3. It can be seen that several sets of samples all show good accuracy. After modeling and optimization, the estimated values of image sequence parameters in the data set show higher correlation, and their distribution is more uniform.

In the three-dimensional coordinate calculation, the antiprojection line of the matching point and the baseline of the two cameras form a triangle. The vertex of this triangle is the intersection of the optical centers of the two cameras and the two antiprojection lines. This intersection is the space point to be determined. However, in practical applications, due to the existence of various noises, for a certain spatial point, the back-projection rays of the imaging point generally do not exactly intersect, as shown in the figure, so it is necessary to define an appropriate cost function and use the beam level. The difference optimization model estimates the best “intersection point” of the spatial point coordinates. The beam adjustment optimization model uses the camera projection matrix to back-project a given initial three-dimensional point cloud. Generally, the sum of the squares of the distance error between the back-projected point and the original image point is minimized to optimize the point cloud structure and camera parameters. The pinhole camera model is simply described as follows: among them, there is the three-dimensional point in the world coordinate system and the point projected on the image by the three-dimensional point, representing each image and its corresponding camera parameters. The standard beam adjustment algorithm contains a series of input information.

The iterative method is used, by constructing the objective function, and then the optimization theory is used to minimize the objective function to obtain the optimal

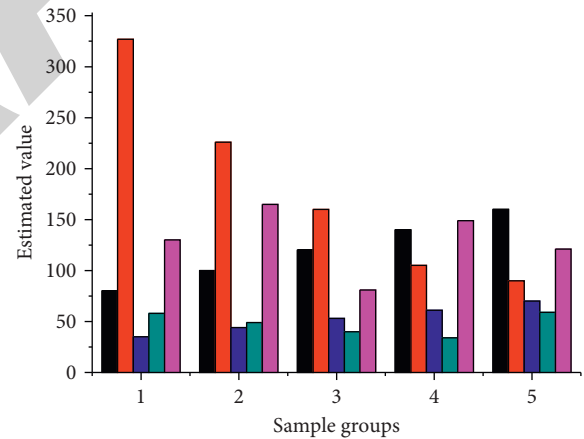


FIGURE 3: The distribution of image sequence parameter estimates.

solution. Due to the large change in the viewing angle of the close-up image, a large distortion occurs at the feature point. Because the extraction accuracy of feature points depends on the gray information of the neighborhood, because each view has different perspective distortions, the neighborhood gray mode of the feature points of the same target 3D point on different images is very different, and it has a strong directionality; assuming that the noise at the feature points of the image is isotropic, the solution obtained by optimizing the objective function is also the maximum likelihood optimal estimation under this assumption; that is, the solution of the model is only available under this assumption. Considering the influence and effect of the neighborhood distortion of feature points on the objective function, the anisotropy of noise is incorporated into the reprojection



error by using the principle of affine transformation between images, and the objective function that takes into account the local anisotropy is constructed to make the feature points. When the noise is anisotropic, the camera parameter solution obtained by minimizing the objective function is the optimal solution in this case.

The affine transformation is essentially the transformation between geometric coordinates, which changes the spatial change of the grayscale mode of the feature point neighborhood, and does not affect the overall grayscale information of the feature neighborhood. Therefore, affine geometry can be used to find the affine. The stable feature points under transformation are affine invariant feature points. However, the gray mode changes in the neighborhood of different feature points and the gray mode changes in the neighborhood of the same pair of feature points have great differences on different images, making the error distribution an anisotropic distribution. The anisotropic distribution of errors can be described by an elliptical local affine invariant region centered on a feature point. The specific 3D model architecture is shown in Figure 4. The first step is to construct a Gaussian pyramid of the image to detect local extrema in the scale space. When a set of 5-layer Gaussian pyramids is built, another set of Gaussian pyramids underneath is built using the following method: first downsample the third layer of the 5-layer Gaussian pyramid, and then use the downsampling results. Take it as the base layer image in the next group, then use this image to perform Gaussian filtering, and finally, get other layer images. When determining the spatial local extremum, it is necessary to compare each pixel with all neighboring points around it, and the highest and lowest point in the scale space should be excluded first; observe the difference between it and its image domain and scale domain. The system can be divided into 3 modules, which are connected to each other. For different types of affine transformations, the affine invariant region is an ellipse with the center of the circle. The direction and length of the minor axis, respectively, represent the direction and size of the uncertainty of the error distribution at this point.

#### 4. Application and Analysis of Sculpture 3D Modeling Based on Image Sequence

*4.1. Image Expression of Three-Dimensional Model of Sculpture Building.* In order to verify the effectiveness and robustness of the method in this paper, four sets of test images published by a machine vision research group are selected for experiments. These four sets of test images are all close-range image pairs related to buildings, with rotation, scale, illumination, and affine transformation, as shown in the figure. In the process of surface reconstruction based on two views, the weighted texture extraction method based on the center of gravity coordinates is adopted. By mapping the refined discrete grid to the image, the texture information of a certain pixel of the four adjacent pixels is obtained. The texture information of the points is converted by a specific formula to form a texture image. The texture information of the image is stored in a rectangular manner. The discrete

grid of the object surface can be used to project the grid points on it into the image, and the texture information of this point in the image is extracted, so that a texture image in the form of a grid is formed. However, because these grid points are too sparse and the texture information obtained is not enough, the resulting texture image is too blurred. Therefore, the discrete grid needs to be refined first to obtain more grid points. The grid point corresponds to a pixel of the texture image, so that there is enough texture information.

In accordance with the technical requirements of first-class wire and second-class leveling, the Leica TS30 model automatic total station and the Tianbao DiNi11 model electronic level are used to accurately measure the plane and elevation coordinates of the measuring points of the ground sculpture building. The measurement accuracy is better than  $\pm 3$  mm. The control point sign is made of a flag cloth with a length of 1 m. A black and white radial pattern and number are printed on the flag cloth. A hole with a diameter of 3 mm is reserved in the center of the pattern, which is convenient for aligning the center experiment of the stainless steel sign. In the process, the algorithm is used to match the feature points of each close-range image pair, and the straight line extraction algorithm in this paper is used to extract the straight line. The measuring points (control points and check points) are measured by Swiss Leica TS30 automatic total station. The nominal accuracy of precise ranging is  $0.6 \text{ mm} + 1 \text{ ppm}$ , and the nominal accuracy of prism-free ranging is  $2 \text{ mm} + 2 \text{ ppm}$ . The nominal accuracy of the angle measurement is  $\pm 0.5''$ . And the Trimble DiNi12 electronic level is used to recheck the elevation of the measuring point, and its nominal accuracy is  $\pm 0.3$  mm for the standard error of the round-trip observation per kilometer. For the evaluation of the matching result, the correct matching pair of the straight line feature is adopted. As shown in Figure 5, the matching results under different geometric changes are counted and compared with the matching results of representative algorithms to compare the pros and cons of the method in this paper. Due to the speed global smoothing constraint, this leads to the propagation of errors, and when multiple objects in the scene move in different directions, there will be an average speed offset problem. Therefore, in the program, this article divides the image into small areas in the horizontal and vertical directions and restricts the speed smoothing constraints to blocks. Although the occlusion or incomplete line extraction is caused by the discontinuity of the image, these factors lead to inconsistent line attributes, and the method in this paper can still be used to obtain the correct matching. At the same time, for the extracted line breaks, the method in this paper will cause "one more match." For the real image sequence sampled by the camera, this article cannot accurately know in advance the speed of the object in the sampled scene moving on the projection plane, so there is no way to compare the calculated velocity field with the real velocity field. The average variance method is used to analyze the test results. In this experiment, the average variances corresponding to the optical flow constraint and the velocity smoothing constraint are calculated: gray variance and velocity variance (because the velocity is a vector, it is represented by the modulus variance and the

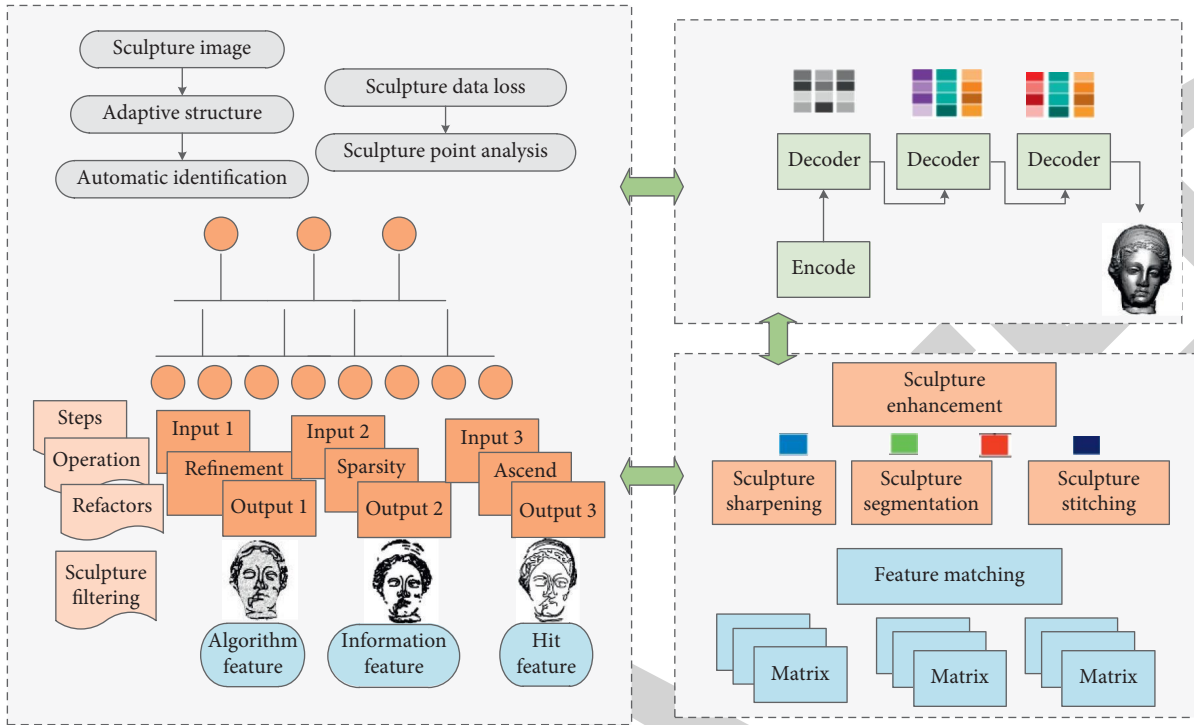


FIGURE 4: Sculpture 3D model hierarchy based on image sequence.

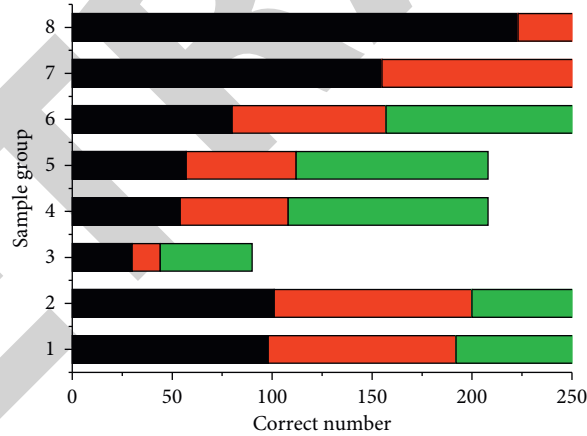


FIGURE 5: The distribution of the correct rate of image sequences in different experimental groups.

angle variance, respectively. The test uses standard video, and the image size of each frame is  $352 \times 288$ . From the experimental data in Figure 6, it can be seen that the recursive refinement of the grayscale change and the calculation method make the displacement size and direction more accurate.

It can be seen from the comparison that the improved algorithm is much better than the original algorithm of the calculation result, which shows the stability and accuracy of

the velocity field. The effect of the LK algorithm and the improved algorithm is roughly the same, but the LK algorithm will have calculation errors when the object moves a large distance. For example, the rolling of the ball at the bottom right of the image does not perform well. Through many experiments, it is shown that the LK algorithm analyzes optical flow only when the projection of the object moves less than one pixel on the projection surface, and the improved HS algorithm through grayscale interpolation can

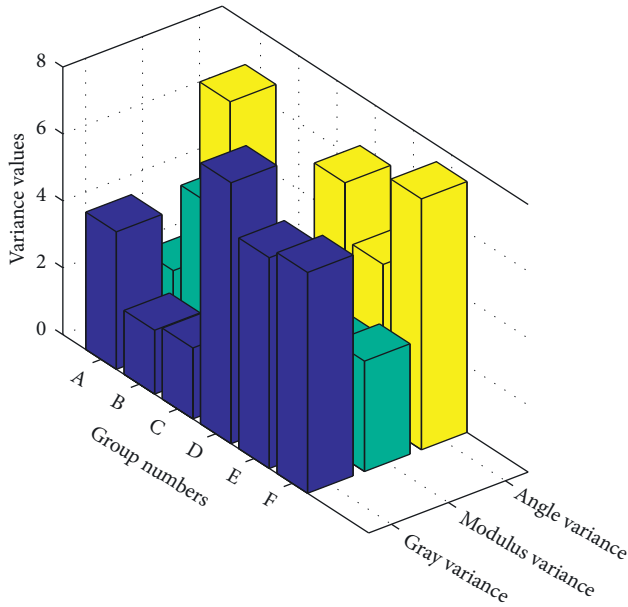


FIGURE 6: Variance distribution of image sequence in different experimental groups.

more accurately analyze the projection movement distance less than two pixels. The movement of each pixel makes the image sequence more accurate.

#### 4.2. Accuracy Analysis of the 3D Model of Sculpture Building.

In the environment of Matlab R2009a, this paper simulates the calibration algorithm of the flat template camera based on the homography matrix. The main functions of the program include the following: feature point extraction, calibration, error analysis, and scene recovery. Sculpture buildings are used as the calibration template, the camera position is fixed, the calibration templates are placed in 12 different positions to shoot them, and 12 images of the calibration template are obtained, which are used as the input of the algorithm. The program first extracts the characteristic points of the calibration points of each image and then obtains the internal and external parameters of the camera by solving linear equations and nonlinear optimization problems. On a noisy image, search for similar blocks in a local area and stack them. In the transform domain (DCT domain and FFT domain), use the hard threshold denoising method to denoise the stacked image blocks to obtain the estimated value of the stacked similar blocks. Perform aggregation according to the mean weight, then obtain the preliminary estimated image, and perform aggregation of similar blocks on the preliminary estimated image; then, use Wiener collaborative filtering to perform image denoising, thereby obtaining the final denoising result and the resulting image. The sequence simulation situation is shown in Figure 7.

Control points and check points can be divided into two groups. The first group is distributed on the ground in the  $300\text{ m} \times 300\text{ m}$  experimental sculpture area, with a total of 25 measuring points used for the construction of the 3D model

in the area and its accuracy evaluation; the second group is distributed in the area. There are 98 measuring points on the exterior surface of the sculpture building, which are used for the construction of the refined model of the single building and its accuracy evaluation. The first group of 25 measuring points are evenly distributed in the test area of  $300\text{ m} \times 300\text{ m}$ , with a diameter of 1 cm and a length of 1 cm. A measurement mark was made on 7 cm stainless steel. We have designed 4 different 3D model construction schemes. According to the principle of uniform distribution control, 3, 6, 9, and 12 measuring points are selected as control points, and the remaining points are accuracy check points. The points are evenly distributed in the test area. The 98 measuring points in the second group are distributed on the outer surface of the library building, arranged in layers at 3 different heights of 10 m, 20 m, and 30 m, taken at the heights of 10 m and 30 m on the 4 sides of the sculpture building. 2 measuring points are used as control points, a total of 16 control points, and the remaining measuring points are used as check points. Select the points with strong reflectivity and obvious characteristics on the exterior surface of the sculpture building as the measuring points. There is no need to paste additional reflective signs. The prism-free function of Leica TS30 automatic total station is used for three-dimensional coordinate measurement, and the accuracy of plane and elevation measurement is better than  $\pm 5\text{ mm}$ . For the 3D model of the  $300\text{ m} \times 300\text{ m}$  survey area, 5 flying heights and 4 control points were set up in this test, and a total of 20 3D models were reconstructed. For the refined model of the sculptured single building, this test uses flying A refined 3D model was constructed for an image sequence with a height of 80 m, and a 3D model was reconstructed. First, image preprocessing, image area network joint adjustment, control points, and coordinates were imported; then, dense 3D points were generated according to aerial triangulation. A 3D white cloud model was generated, the surface mesh was reconstructed to generate a triangulation, and texture mapping was performed to generate a real 3D model.

The accuracy of the 3D model constructed under different number of control points is shown in Figure 8. When the number of control points is 3, the plane root mean square corresponds to different flying heights (80 m, 100 m, 120 m, 140 m, and 160 m). The errors are 5.8 cm, 9.0 cm, 10.1 cm, 10.9 cm, and 14.9 cm, and the root mean square errors of elevation are 12.1 cm, 14.4 cm, 15.8 cm, 16.4 cm, and 16.5 cm. When the number of control points is 12, the error will be different for different flying heights. The root mean square error of the plane corresponding to the height is 3.4 cm, 4.8 cm, 6.2 cm, 6.8 cm, and 10.6 cm, and the root mean square error of the elevation is 3.1 cm, 5.4 cm, 6.7 cm, 6.9 cm, and 7.0 cm. With the control point as the number increases, the plane and elevation errors of the 3D model are gradually reduced, and the accuracy is getting higher and higher. Therefore, the more the number of control points, the higher the plane and elevation accuracy of the model. When the number of control points increases from 3 to 6 and when the height root mean square error is significantly reduced, the number of control points has a significant

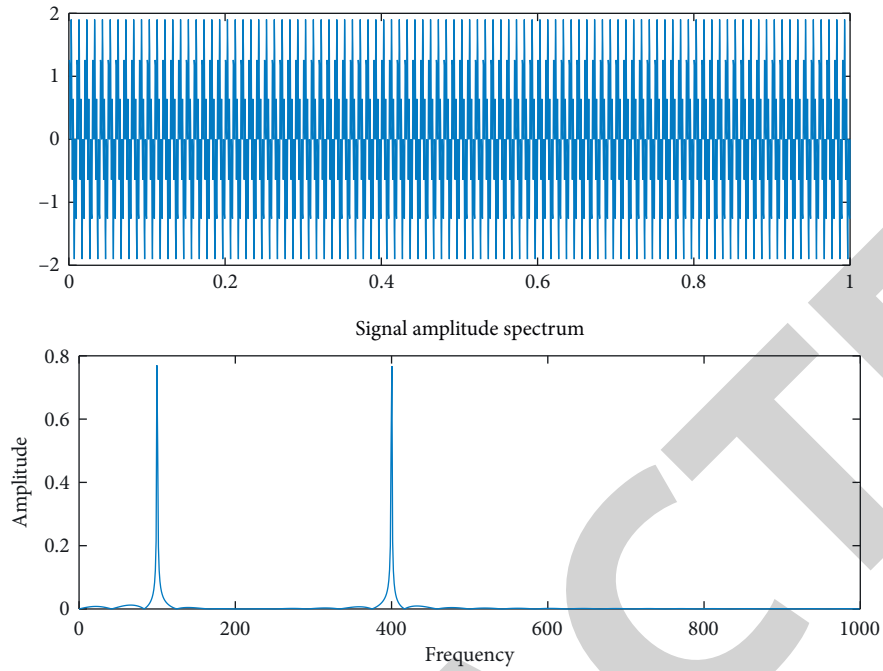


FIGURE 7: Amplitude frequency response of image sequence signal after filtering.

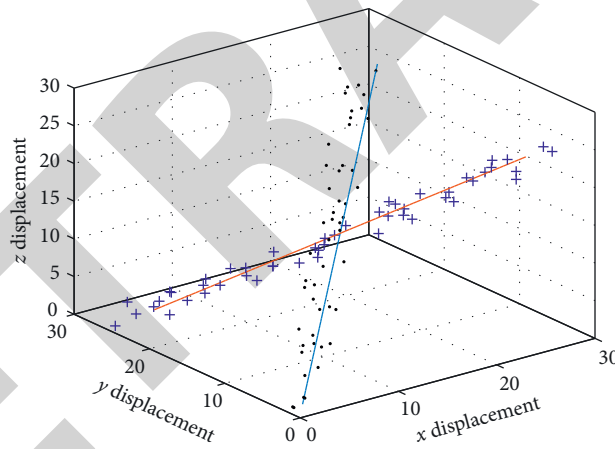


FIGURE 8: 3D fitting curve of control points and check points in image sequence.

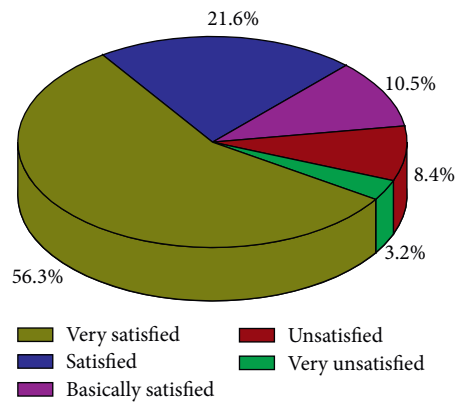


FIGURE 9: Satisfaction survey of the sculpture 3D model based on image sequence.

impact on the model elevation accuracy. When the flying height is 80 m and the number of control points is 12, the plane root mean square error is 3.4 cm, and the height root mean square error is 3.1 cm the plane accuracy and elevation accuracy of the 3D model reach the highest level. The level and elevation accuracy of all 3D models meet the requirements of refined model accuracy, so the number of control points can be minimized when the model accuracy is met (but at least 3), or increase the flight altitude.

In the experiment, we tested four groups of representative experimental data and calculated the error of the tracking system by comparing with the captured real posture. In addition, in the experiment, two models were used to compare each set of experimental data, namely, the real sculpture model captured by a stereo camera and the geometric model constructed from the geometric image sequence. As shown in Figure 9, the sculpture 3D model based on the image sequence shows good satisfaction, and more than half of the data show very satisfactory results. The specific data distribution is shown in the figure. Through the numerical comparison of the posture tracking error of the two, it provides a reference for the selection of the model in the image sequence tracking process.

## 5. Conclusion

This paper proposes a method of 3D reconstruction of sculpture based on image sequence, combined with image adaptive feature reconstruction and the point cloud data analysis method, to realize the adaptive structure reorganization of sparse image sequence and improve the automatic recognition and detection of sparse image ability. Based on the 3D reconstruction of sparse scattered points and the feature matching method of sharpened template, the image 3D reconstruction is carried out. The improved algorithm is used to perform feature point matching. Compared with the traditional method, this method can quickly solve the eigenspace and reduce the computational complexity and does not need to occupy more storage space, and the improved algorithm is more stable, accurate, and fast; the reconstruction results are more prominent in details, more realistic, and closer to real objects. At the same time, 3D corner detection and edge contour feature extraction methods are used to detect the 3D point cloud feature of the sparse image of sculpture points, and the detected sculpture points are sparse. The image point cloud data are processed by information fusion to realize the information enhancement and fusion filtering of the sparse image of the sculpture point; the sharp template feature matching and block segmentation technology are used to realize the automatic 3D reconstruction of the sculpture point cloud. The surface reconstruction method based on two views has been researched and explored, and the deficiencies of its linear constraint mechanism have been analyzed and discussed. On this basis, it has been improved and optimized, and better results have been verified by experiments.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

## References

- [1] J. Du, "Research on optimization of portrait sculpture data based on 3D image and mobile edge computing," *IEEE Access*, vol. 52, pp. 32–47, 2020.
- [2] J. Park and J. Lee, "A study on 3D modeling data acquisition method for sculpture scan," in *Proceedings of the Korean Institute of Information and Communication Sciences Conference*, pp. 612–614, Jeju, Korea, May 2018.
- [3] O. Wiles and A. Zisserman, "Learning to predict 3D surfaces of sculptures from single and multiple views," *International Journal of Computer Vision*, vol. 127, no. 11–12, pp. 1780–1800, 2019.
- [4] J. Yang, Y. Zhao, J. Liu et al., "No reference quality assessment for screen content images using stacked autoencoders in pictorial and textual regions," *IEEE Transactions on Cybernetics*, vol. 6, pp. 3–12, 2020.
- [5] E. Adamopoulos, A. Bovero, and F. Rinaudo, "Image-based metric heritage modeling in the near-infrared spectrum," *Heritage Science*, vol. 8, pp. 1–12, 2020.
- [6] S. Hong, Y. H. Jo, Y. M. Kwon et al., "Restoration of the damaged Korean stone seated bodhisattva sculpture using three-dimensional virtual modeling and printing technologies," in *Proceedings of the 2018 3rd Digital Heritage International Congress (DigitalHERITAGE) Held Jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*, pp. 1–3, IEEE, San Francisco, CA, USA, October 2018.
- [7] N. Tosa, P. Yunian, R. Nakatsu, A. Yamada, T. Suzuki, and K. Yamamoto, "3D modeling and 3D materialization of fluid art that occurs in very short time," in *Lecture Notes in Computer Science*, pp. 409–421, Springer, Cham, Switzerland, 2020.
- [8] K. Dai, Y. Wang, and Q. Song, "Real-time object tracking with template tracking and foreground detection network," *Sensors*, vol. 19, no. 18, p. 3945, 2019.
- [9] D. A. L. Carvajal, M. M. Morita, and G. M. Bilmes, "Virtual museums. captured reality and 3D modeling," *Journal of Cultural Heritage*, vol. 45, pp. 234–239, 2020.
- [10] E. Adamopoulos and F. Rinaudo, "Near-infrared modeling and enhanced visualization, as a novel approach for 3D decay mapping of stone sculptures," *Archaeological and Anthropological Sciences*, vol. 12, no. 7, pp. 1–12, 2020.
- [11] F. Gherardini, M. Santachiara, and F. Leali, "3D virtual reconstruction and augmented reality visualization of damaged stone sculptures," *IOP Conference Series: Materials Science and Engineering*, vol. 364, no. 1, Article ID 012018, 2018.
- [12] L. Lanteri, G. Agresti, and C. Pelosi, "A new practical approach for 3D documentation in ultraviolet fluorescence and infrared reflectography of polychromatic sculptures as fundamental step in restoration," *Heritage*, vol. 2, no. 1, pp. 207–215, 2019.

- [13] V. Barrile and G. Bilotta, "Computer vision in 3D modeling of cultural heritage: the riace bronzes," *Advanced Science Letters*, vol. 24, no. 1, pp. 581–586, 2018.
- [14] N. Suciati, M. T. Baihaqi, H. Tjandrasa et al., "Converting image into bas reliefs using image processing techniques," *Journal of Physics: Conference Series*, vol. 1196, no. 1, Article ID 012037, 2019.
- [15] E. Adamopoulos and F. Rinaudo, "Documenting the state of preservation of historical stone sculptures in three dimensions with digital tools," in *Pattern Recognition*, A. Del Bimbo, Ed., ICPR International Workshops and Challenges, ICPR 2021, Lecture Notes in Computer Science, vol. 12663, Springer, Cham, Switzerland, 2021.
- [16] V. Di Pietra, E. Donadio, D. Picchi, L. Sambuelli, and A. Spanò, "Multi-source 3D models supporting ultrasonic test to investigate an Egyptian sculpture of the archaeological museum in bologna," *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W3, pp. 259–266, 2017.
- [17] Y. Zhang, D. Dong, and Y. Guo, "3D shadow art sculpture based on real items," in *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, pp. 243–246, Singapore, Asia, June 2017.
- [18] K. A. Kolobova, A. Y. Fedorchenko, A. Y. Fedorchenko et al., "The use of 3D-modeling for reconstructing the appearance and function of non-utilitarian items (the case of anthropomorphic figurines from tourist-2)," *Archaeology, Ethnology and Anthropology of Eurasia (Russian-Language)*, vol. 47, no. 4, pp. 66–76, 2019.
- [19] E. Adamopoulos, A. Bovero, and F. Rinaudo, "Image-based metric heritage modeling in the near-infrared spectrum," *Heritage Science*, vol. 8, no. 53, 2020.
- [20] D. Shi, Y. Zhao, and J. Li, "Reconstructing part-level 3D models from a single image," vol. 42, pp. 65–79, in *Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME)*, vol. 42, pp. 65–79, IEEE, London, UK, July 2020.
- [21] R. Yang, M. Xu, T. Liu et al., "Enhancing quality for HEVC compressed videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 7, pp. 2039–2054, 2018.
- [22] Z. Guan, Q. Xing, M. Xu, R. Yang, T. Liu, and Z. Wang, "MFQE 2.0: a new approach for multi-frame quality enhancement on compressed video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 3, pp. 949–963, 2021.
- [23] Y. Zhou, L. Tian, C. Zhu et al., "Video coding optimization for virtual reality 360-degree source," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 118–129, 2019.
- [24] Q. Jiang, F. Shao, W. Gao et al., "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1866–1881, 2018.
- [25] S. H. Mousavi, M. Kanshari, and R. Rahmani, "A fully scalable big data framework for Botnet detection based on network traffic analysis," *Information Sciences*, vol. 512, pp. 629–640, 2020.