

International Journal of Computational Intelligence Studies

ISSN online: 1755-4985 - ISSN print: 1755-4977

<https://www.inderscience.com/ijcistudies>

Application of binocular image reconstruction method in the construction of 3D model of wooden arch corridor bridge structure

Hua Deng

DOI: [10.1504/IJCISTUDIES.2023.10058108](https://doi.org/10.1504/IJCISTUDIES.2023.10058108)

Article History:

Received:	19 September 2022
Last revised:	10 February 2023
Accepted:	10 March 2023
Published online:	05 April 2024

Application of binocular image reconstruction method in the construction of 3D model of wooden arch corridor bridge structure

Hua Deng

Department of Management,
Chengyi College,
Jimei University,
Xiamen, 361021, China
Email: huadengju@yandex.com

Abstract: In order to better understand the structure of wooden-arch corridor bridges and preserve cultural buildings, the study proposes a three-dimensional model construction method based on binocular image reconstruction method. Facing the problem of poor image edge feature extraction, the study uses Laplace combined with the Sobel operator (Laplace-Sobel) for image edge feature extraction. The SIFT structural feature matching algorithm is improved to eliminate the mismatching points in stereo matching. The results show that as the active window size increases, the number of detected edge feature points gradually increases and the number of feature stereo matching pairs also increases, but the number of matching pairs is lower than the number of feature points; the number of features is searched for the most and the matching cost is the least when the neighbourhood range of pixels is 15×15 . The results show that the improved RANSAC-SIFT matching algorithm has an average root-mean-square geometric error of 1.2482 and a high matching accuracy, with an average running time of 94 ms. Compared with the unimproved algorithm, the performance has been significantly improved and can be better applied in binocular image reconstruction, reducing the difficulty of constructing 3D models of wooden arch bridges.

Keywords: binocular image reconstruction method; three-dimensional model; wooden arch gallery bridge structure; matching algorithm.

Reference to this paper should be made as follows: Deng, H. (2023) 'Application of binocular image reconstruction method in the construction of 3D model of wooden arch corridor bridge structure', *Int. J. Computational Intelligence Studies*, Vol. 12, Nos. 3/4, pp.223–237.

Biographical notes: Hua Deng obtained his BE in the Taiyuan University of Technology in 2005. He obtained his ME in Civil Engineering from the Xiamen University in 2010. Presently, he is working as a Lecturer in the Department of Management, Chengyi Collge, Jimei University. His areas of interest are protection of ancient architectural, application of BIM and CNC technology, and digital architecture.

1 Introduction

The wooden arch bridge has a unique and exquisite shape and contains a profound historical culture. Its charming cultural connotation has attracted the attention of many scholars. However, with the development of urbanisation, the widespread use of cement, asphalt and other transportation facilities has gradually replaced the traffic role of the wooden arch bridge, making many traditional buildings gradually fade into people's vision; due to the influence of human and natural factors such as fire, many wooden arch bridges are facing the situation of dilapidation and extinction (Liu, 2020). The protection of the covered bridge has attracted great attention in the architectural field. Many maintenance personnel do not have enough understanding of the wooden arch covered bridge structure, and their maintenance technology is also facing the risk of loss. In order to understand the covered bridge structure on a deeper level, it is a better scheme to construct a three-dimensional model (Bailer and Martin, 2019). The conservation of the bridges has attracted a great deal of attention, but many repairers do not know enough about the structure of wooden arched bridges, and their repair techniques are at risk of being lost. However, the complex structure of the bridge makes the construction of the model difficult. The binocular image reconstruction method is based on binocular vision and the construction of a transformation matrix, which can help the system accurately obtain the coordinate data of the object (Wang et al., 2020). The most important technique in binocular image reconstruction is stereo image matching. Currently, there are problems in the matching algorithm, such as inappropriate calibration methods for binocular cameras, unreasonable selection of feature points and low accuracy of matching (Yin et al., 2020). In order to solve the problems of improper calibration of binocular camera, unreasonable selection of feature points and low matching accuracy existing in the current matching algorithm, and to achieve the accurate construction of the three-dimensional model of the auxiliary wooden arch bridge, a binocular image reconstruction technology is proposed. In the image pre-processing part, the combined edge feature extraction algorithm of Laplace and Sobel is used. In addition, the improved RANSAC algorithm is used to establish stereo feature matching, provide technical support for the accurate construction of the three-dimensional model of the wooden arch bridge. The innovation of the research is that the stereo matching module is studied more deeply. In order to improve the accuracy of the matching set, the RANSAC algorithm is improved using SIFT.

2 Literature review

In response to the difficulty of constructing 3D geological models, Jessell et al. (2021) investigated an automated information processing technique that automatically extracts available information from digital maps to construct 3D geological models. By building a model of the Australian Basin, the results show that the method significantly improves the speed of information acquisition and automates the process from data acquisition to model construction. Jin et al. (2021) found that automated CAD design is difficult to use in complex and variable building environments, and proposed a method for automatically generating 3D models based on geomorphological features in order to save creation time and solve a similar process. Through extensive practice, it was verified that the method can significantly reduce labour costs as well as time costs and has a high experience in

practical applications. Ammad and Ramli (2020) found that problems such as locking holes and branching always occur in the process of constructing 3D terrain models from 2D contours, and in order to solve these problems, they proposed a method to deal with branching and locking holes. The experimental results found that the method could retain the data points on the surface well, and the spatial model constructed by this method effectively avoided problems such as locking holes and branching. Cai et al. (2020) proposed an intelligent algorithm for automatically constructing a 3D brick model, which can identify a given object as a multi-layered planar vector polygon, and fill the plane by bricks. A series of experiments with sculptural and architectural models show that the algorithm can successfully construct brick visualisation models, which can help engineers to improve the efficiency of building large 3D architectural models. Casado et al. (2021) found that the usage of natural materials is not high when building 3D models, and proposed a construction method based on basket technology in order to improve the usage of natural materials. The method involves simulating and optimising the elastic properties of the material in a 3D digital model and applying this material to the geometry construction. Experimental results show that the architectural cultural heritage can be well preserved using the basket technique.

Binocular image reconstruction, (i.e., binocular stereo vision) is widely used in several fields. Ma et al. (2021) found that current inspection systems are unable to automatically detect circuit components and proposed a drone combined with binocular vision perception technology to detect the condition of electrical equipment in real-time. For the cluttered aerial background, the detection and localisation of insulators became a challenge and a spatial localisation method of binocular stereo vision combined with a localisation system was investigated. The results of real-time detection of the equipment show that the method has good robustness and high accuracy in practical applications. Wang's (2021) research team proposed a two-point method to quickly estimate the parameters of a rotating camera in order to obtain high accuracy shots of the camera, while a movement-based target matching algorithm was investigated in order to allow the camera to focus automatically, and the results, evaluated by real data, show that the method proposed in the study can obtain high accuracy 3D coordinates. Li and Zhang (2021) research scholars developed a non-contact displacement measurement system using wireless sensing combined with binocular stereo vision technology, which can analyse athletes' movement patterns and provide a theoretical basis for sports training. The cantilever beam vibration displacement measurement results show that the system has high accuracy and stability. Li et al. (2021a, 2021b) found that appropriate illumination can improve the recognition effect of the camera. In order to control the illumination, the study proposed a method of illumination measurement based on binocular stereo vision, established an illumination analysis model and obtained the relevant theoretical formulae. Experimental results show that the method has good and efficient measurement performance, and the detection results are reliable with a relative error of less than 3.8% (Guo et al., 2022). The researchers investigated a binocular stereo panoramic vision localisation algorithm based on an improved sparse method to address the problems of few image road signs and poor localisation in vision sensors, and used the algorithm to obtain panoramic images of the surroundings of unmanned ground vehicles. The experimental results found that the algorithm can significantly improve the accuracy and reliability of localisation with low computational effort, and has promising applications (Guo et al., 2022). Lu et al. (2018) developed a fall detection method based

on three-dimensional convolutional neural network (3-D CNN) in order to achieve timely detection and quickly deliver medical services to the injured, so as to avoid the requirements of deep learning solutions for large fall datasets. The experimental results show that the method is verified by the fall detection benchmark, and the accuracy rate is 100%. It also achieves excellent performance on other active databases. Hauptmann et al. (2018) proposed a depth neural network, which is specially designed to provide high-resolution three-dimensional images from restricted photoacoustic measurements. The experimental results confirmed the effectiveness of its application in lung computed tomography.

Through a brief description of the research results of domestic and foreign scholars, it is found that the acquisition of model parameters is extremely important in the construction of 3D models, and binocular image reconstruction can use stereo vision to measure relevant parameters, which is an important part of the 3D model reconstruction process. Therefore, it is of great significance to study the techniques in binocular image reconstruction for the construction of three-dimensional models of wooden arch bridges.

3 Design of 3D model of wooden arch bridge based on binocular image reconstruction method

3.1 Edge feature extraction based on Laplace Sobel and construction of improved SIFT stereo matching algorithm

The wooden arch bridge is a model of ancient bridges in China. Its construction technology is superb, but it is extremely vulnerable to natural erosion and man-made damage. Its manufacturing process is also facing the problem of loss. Therefore, it is urgent to use digital technology to protect the wooden arch bridge (Lyu et al., 2020). The external features of the image are extracted, and the accuracy of the image matching is reduced by the differences between the images obtained by the binocular cameras due to the many interfering factors when the images are taken (Sun et al., 2019). In order to reduce the interference factors in image shooting, the image needs to be pre-processed first. Median filtering can not only remove the noise in the image, but also restore the image to the maximum extent without destroying the edge information in the image. At the same time, the operation process is not demanding on the system, and it is simple to use. Therefore, median filtering is selected for image pre-processing. Median filtering uses the median of the points in the target neighbourhood of the image as the target point, i.e., the active window, to adjust the pixel greyscale of the other images. The two-dimensional median filtering window A is defined by equation (1).

$$f(x, y) = Mde\{f(x+s, y+t), (s, t) \in A\} \quad (1)$$

In the adjustment process, the shape of the active window determines the scientific rationality of the median calculation. The size of the window contains the number of neighbourhood points, so the determination of the active window determines the filtering effect. In order to improve the operation efficiency, the important information representing the image features is retained, and the feature information is used to match two images in 3D, and the image edge features are selected. In order to reduce the matching error rate, the extracted image features need to have strong correspondence and high similarity with the original image features; the feature points should not be too

sparse and have a certain density; the extracted feature points should have image characteristics to avoid matching errors or even multiple images matching the same features (Dandapani et al., 2020). General image inflection points, edge points, etc. are commonly used feature extraction points, and the image edge direction is consistent with the direction of the point, the study takes the edge feature extraction method. The effect of edge feature extraction is judged by the edge accuracy and noise immunity, which are mutually constrained. When the edge accuracy is high, the extracted features will produce pseudo-edges; when the noise immunity is high; the obtained edge contours are unclear or even deviate from the original image. In order to neutralise the metrics, extract effective edge features and reduce the false match rate, the study proposes an image edge feature extraction algorithm by introducing the Sobel operator to the Laplace operator, which is an edge detection operator based on a two-dimensional image function and deriving its parameters to obtain image edges, defined by equation (2).

$$\Delta^2 f(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (2)$$

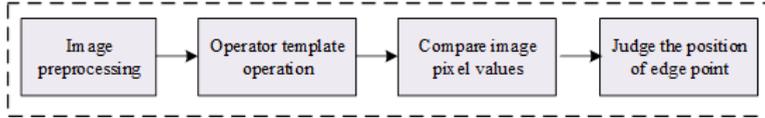
At the edge of the image, the grey value of the pixel is easy to change significantly, resulting in an increase in the error matching rate. The common solution is Laplace operator and Sobel operator. Although the Laplace operator has a small amount of calculation and its calculation result is also independent of the orientation, it is very sensitive to noise and is easy to lose edge feature information. Sobel operator noise suppression effect is very good, but the image edge information obtained is not high, and the accuracy is very low. Therefore, Laplace operator and Sobel operator are selected to extract image feature information. When acquiring an image, there is often a lot of noise, which will cause difficulties in the extraction process of edge feature information and reduce the accuracy of information. Therefore, it is necessary to take denoising before extraction (Chen et al., 2019). The study uses median filtering to remove the image noise and then combines the Sobel operator in the Laplace operator to extract the image feature information. The Laplace Sobel operator combination is decomposed in the horizontal and vertical directions to obtain the rows and columns of the image corresponding to the array respectively. Then in order to get more edge feature information, four diagonal directions of northeast, southeast, northwest and southwest are added to the detection direction when extracting information. The added diagonal direction detection template expression is formula (3).

$$D = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -1 & 1 \\ -1 & 4 & -1 \\ 1 & -1 & 0 \end{bmatrix} \quad (3)$$

As the Laplace operator is a second order differential equation, the choice of threshold needs to be based on the characteristics of the different image edges. When the pixel grey value of the edge of the image is in the form of a leap, the change curve is a first order function, and the threshold is set at $\lambda_{\max} > 0$. If the value of the function is greater than the threshold, the point is the edge point. When the pixel grey value of the edge is in a decaying roof shape, its change curve is a normal function, and the set threshold is the minimum value and $\lambda_{\max} < 0$, if the function value is less than the threshold, the point is the edge point. The specific process of the improved image edge feature extraction

algorithm is shown in Figure 3. Firstly, the image is pre-processed using median filtering, and the image function obtained is $h(x, y)$; the corresponding functions are obtained by computing them with the D , D_1 and D_2 templates; the pixel grey values in the functions are compared to determine the image edge features, and the edge point is obtained by comparing the set threshold.

Figure 1 Improve the flow of image edge feature extraction algorithm (see online version for colours)



In order to achieve 3D model construction, matching feature information one by one to a stereo match in 3D space is one of the most important aspects. Any matching algorithm needs to solve several problems, among which the selection of matching primitives, (i.e., image feature information) is related to the accuracy of matching (Qiu and Wang, 2019). The matching criterion is a rule that the matching primitives must follow in order to ensure that the information selected is an objective and realistic representation of the target’s characteristics. The matching algorithm is designed to check whether the selection of the matching primitives is reasonable. Among the many methods, the feature attributes of the image are relatively stable and the matching is relatively simple. The use of scale invariant feature transform (scale invariant feature transformation, SIFT) method can ensure that the feature information has better stability after normalisation, and the study proposes the stereo matching algorithm with SIFT structural features.

The construction of a scale space allows for easier access to the essential features of an image, and the scale space allows for the generation of a single image to a sequence of images at different scales. The Gaussian convolution kernel can implement the image scale transformation, and its defining equation is shown in equation (4).

$$G(x, y, \delta) = \frac{1}{2\pi\delta^2} e^{-(x^2+y^2)/2\delta^2} \tag{4}$$

In equation (4), δ is the scale parameter. The Gaussian convolution of any image will result in a scale space of different scales, and the scale space expression is shown in equation (5).

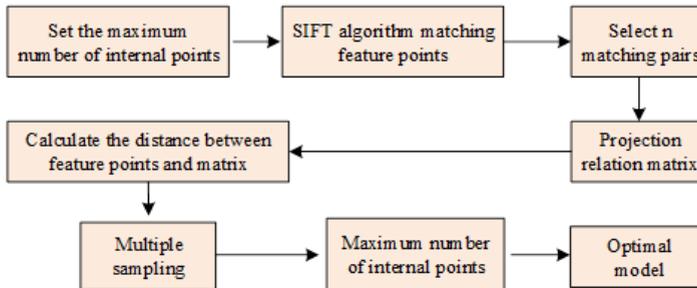
$$N(x, y, \delta) = G(x, y, \delta) \cdot E(x, y) \tag{5}$$

The smaller the $E(x, y)$ value, the higher the resolution of the image; conversely, the lower the resolution of the image. In order to ensure good stability of feature points and low matching difficulty, the feature information of the image needs to be detected, which is done by Gaussian difference formula. However, in order to retain more detection points, the Gaussian smoothing will be reduced, resulting in the loss of information from the image. To improve computational efficiency, a Gaussian difference pyramid is usually used. The direction of each feature key point needs to be determined in a way that ensures scale invariance and high operational adaptability, with the operational expression in equation (6).

$$\begin{cases} m(x, y) = \sqrt{(N(x+1, y) - N(x-1, y))^2 - (N(x, y+1) - N(x, y-1))^2} \\ \theta(x, y) = \arctan \frac{N(x, y+1) - N(x, y-1)}{N(x+1, y) - N(x-1, y)} \end{cases} \quad (6)$$

Equation (6) describes the gradient value and direction of the spatial points. In the practical application of SIFT feature matching, the problem of false matching will occur. This is because there are many feature points in the same feature point, and false matching with similar pixel grey values will occur in the fields around the image. In addition, in the practical application of RANSAC, its main role is to reject error matching. It can randomly select samples to obtain model parameters. All samples can use samples to classify the model. Matching within the allowable error range will be blocked, and pre matching will be screened to guide the acquisition of a certain number of interior points (Chopin et al., 2019). The SIFT matching algorithm applying RANSAC runs as shown in Figure 4. The initial maximum number of interior points is set, the matching is performed using the SIFT feature algorithm, and the projection relationship matrix of the plane is derived by selecting n matching pairs from the matching results. The distance of the remaining matching feature points in the projection matrix is calculated to determine the type of matching points; sampling is performed several times until the maximum number of internal points and the minimum error is reached, at which point the model obtained is the optimal projection transformation matrix.

Figure 2 Running process of SIFT matching algorithm based on RANSAC (see online version for colours)



3.2 Establishment of 3D model of wooden arch bridge based on binocular image reconstruction method

The binocular image reconstruction method is a process in which a computer measures the parallax of the actual target point in two images by observing the same object from two viewpoints, recovers the three-dimensional information of the object and models it based on the information obtained (Cheng et al., 2021). In practice, the camera always undergoes a certain offset when imaging, and the relationship between any point of the image $p(x_0, y_0)$ and the spatial point $P(X_0, Y_0, Z_0)$ is shown in equation (7).

$$Z_S \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & 0 & 0 \\ 0 & f_y & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} = MSP \tag{7}$$

In equation (7), M is the internal parameter matrix of the camera, S is the external parameter matrix, P is the spatial point, and f_x and f_y are the distances of the focal lengths f on the x and y axes. The process of obtaining the matrices M and S is called camera calibration, which is used to improve the accuracy of the 3D model. Cameras have different manufacturing processes, which can cause different degrees of distortion and aberrations in the same shot, resulting in inaccurate parameters. The mathematical expression for the aberration produced by the point p is equation (8).

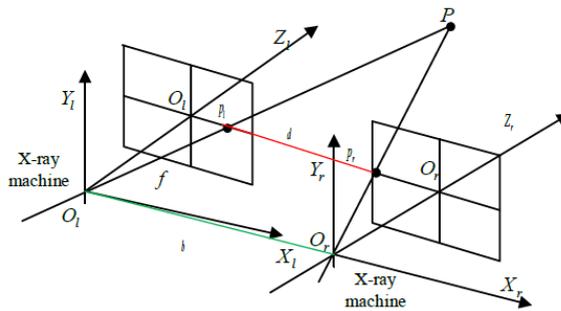
$$\begin{cases} x = x_0 + \Delta x \\ y = y_0 + \Delta y \end{cases} \tag{8}$$

In equation (8), Δx and Δy are the effects of various aberrations on any point of the original image. Among the aberration types, the radial aberration can be corrected by parameter adjustment and the corrected equation is equation (9).

$$\begin{cases} x = x_0 (1 + tr^2) \\ y = y_0 (1 + tr^2) \end{cases} \tag{9}$$

In equation (9), t stands for nonlinear distortion function, $r^2 = x_0^2 + y_0^2$. The relative value of the distortion in x_0 direction and y_0 direction is proportional to the square of the radial radius, and the distortion at the edge of the image is the largest. In 3D model building, binocular calibration requires two cameras to focus on the same object with pixels aligned with each other, as shown schematically in Figure 3.

Figure 3 Dual target schematic diagram (see online version for colours)



There are two cameras, left and right, projecting images of any spatial point P as p_l and p_r , with the distance between images as d and b represents the baseline of the coordinates of the left camera and the right camera. According to the principle of similarity, the depth value of the spatial point P is obtained as shown in equation (10).

$$b_z = \frac{bf}{x_l - x_r} \quad (10)$$

In equation (10), $x_l - x_r$ is the disparity. The expression for the relationship between two imaging points transformed on the coordinate system is shown in equation (11).

$$\begin{cases} p_r = R_r P + T_r \\ p_l = R_l P + T_l \end{cases} \quad (11)$$

In equation (11), R_l and R_r are the rotation matrices of the left and right camera coordinate systems, and T_l and T_r are the translation matrices of the left and right camera coordinate systems. By simplifying equation (11), the relationship between the imaging points p_l and p_r in the coordinate system is given in equation (12).

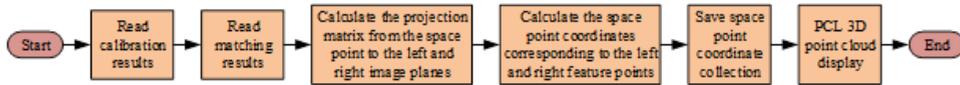
$$p_r = R(p_l - T) \quad (12)$$

In equation (12), $R = R_r R_l^{-1}$ is the rotation matrix between the cameras and $T = T_r - R_r R_l^{-1} T_l$ is the translation matrix between the cameras. The pixel points of the left and right cameras are perfectly aligned and the set relationship between the pixel points and the imaging plane is described by the F matrix to obtain the left and right polar equations as shown in equation (13).

$$\begin{cases} l_r = F p_l \\ l_l = F^T p_r \end{cases} \quad (13)$$

The 3D reconstruction module based on binocular image reconstruction technology mainly calculates the 3D information in the corresponding scene according to the result parameters obtained from camera calibration and the feature points set matched by the stereo matching module, so as to restore the scene in the scene. Combining the contents of the camera calibration module and the stereo matching module, we can get the 3D scene model flow of the wooden arch bridge based on the binocular image reconstruction technology, as shown in Figure 4.

Figure 4 3D scene model flow of wooden arch bridge based on binocular image reconstruction technology (see online version for colours)



4 Performance study of binocular image reconstruction strategy

In order to build the 3D model of the wooden arch bridge structure more accurately, a new binocular image reconstruction technology is proposed to obtain more accurate feature matching of the whole and local details. In order to verify the effect of this technology, the experiment is carried out under the platform of OpenCV. The resolution of the binocular camera in the experiment was $1,024 \times 96$ and the baseline between the cameras was 25 cm. The calibration of the camera was important in the binocular image

reconstruction strategy. The study acquired 40 tessellated images with different angles and the results obtained using the binocular calibration techniques are shown in Table 1.

Table 1 Results obtained by double target determination technique

<i>Parameter</i>		<i>Left camera</i>	<i>Right camera</i>
Internal parameters	f_x	449.6193	451.5329
	f_y	448.9315	452.1034
Distortion parameter	t_1	0.01304	-0.01129
	t_2	-0.07935	-0.02839
	t_3	-0.08160	0.01020
	t_4	0.01097	-0.00936
Rotation matrix	$\begin{bmatrix} 0.9996 & 0.0045 & 0.0197 \\ -0.0048 & 0.9998 & -0.0253 \\ -0.0199 & 0.0257 & 0.9994 \end{bmatrix}$		
Translation vector	(-159.359, -1.298, 2.034)		

Stereo matching is the core point of constructing 3D models, using the Laplace-Sobel algorithm to detect edge feature points and feature stereo matching by the SIFT algorithm. In order to determine the appropriate size of the active window and get the extracted set of edge feature points matched at different sizes, the experiment was unfolded through four groups of images, respectively, in four kinds of wooden-arch corridor bridge structures, group 1 was a photo of Wanan Bridge, group 2 and group 3 used images of Zhejiang corridor bridge structures, group 4 was an image of Xianju Bridge structure, Table 2 shows the average number of edge features extracted from the four groups of photos at different sizes with matching pairs.

Table 2 Average number of feature matching results of four groups of photos under different sizes

<i>Experience group</i>	<i>/</i>	100×100	120×120	140×140	160×160	200×200	240×240	300×300
Group 1	Extraction quantity	2.37	8.33	12.31	18.03	25.17	31.18	40.03
	Matching pair	1.76	6.89	9.57	13.28	18.35	24.32	30.47
Group 2	Extraction quantity	2.53	10.25	15.86	20.59	29.91	38.97	48.92
	Matching pair	1.81	8.96	12.64	15.97	25.49	35.88	42.71
Group 3	Extraction quantity	2.61	9.15	13.19	18.26	27.30	30.12	40.59
	Matching pair	1.82	7.28	9.14	14.61	20.57	25.79	31.81
Group 4	Extraction quantity	2.60	9.34	14.27	19.56	26.89	33.14	42.81
	Matching pair	1.80	7.45	10.33	12.48	22.62	28.17	33.08

From the experimental results in Table 2, we can see that with the image size of 100×100 , the edge feature points obtained by the windows of all four groups are around 2. The system can detect the point when the window moves on that feature point, which will cause a waste of computation time. At 140×140 , the feature points obtained from all four groups of windows are above 10, and the accuracy of the acquisition gradually increases. At a size of 300×300 , the edge feature points are all above 40. Therefore, when the matching window is small, the extracted feature data is relatively sparse, which is not conducive to the extraction of window detail features. When the matching window is large enough, the number of features obtained will increase. However, the number of feature matching pairs is smaller than the number of feature points in the same region, which is because some correct discrete features will also be eliminated after the mis-matching filter. The experimental windows of group 1 are 100×100 , 140×140 , 160×160 and 200×200 , and the obtained stereo matching results of structural features with different windows are shown in Figure 5.

Figure 5 Stereo matching results of structural features of different windows (see online version for colours)

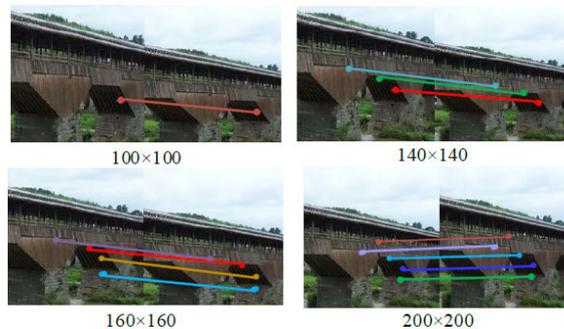


Figure 5 shows the structure matching images of the wooden-arch corridor bridge captured by the system, using median filtering to filter the noise. As can be seen from the figure, the feature matching set is small under the 100×100 window, and as the experimental window size gradually becomes larger, the feature matching results captured in the wooden-arch corridor bridge gradually increase. Setting 160×160 as a large window to extract the overall structural features ensures a clear image. To sum up, when the matching window size is small, the number of features extracted in the window is relatively small, which is not conducive to extracting detailed features in the window. When the window is large, due to the excessive number of features detected, some discrete and correct matching features are eliminated after filtering the wrong match using the improved RANSAC algorithm. Therefore, the size selected for the experiment is 100×100 window is used as a small window to extract local details, and the selected size is 160×160 . The window is used as a large window to extract the overall structural features, so as to enrich the feature points in the scene and avoid too sparse features. The reconstructed 3D scene is not clear enough. To determine the best neighbourhood range for the target window, the best neighbourhood results were searched near the target window in group 1 with pixel value sizes of 10×10 , 10×15 , 10×20 , 15×15 , 20×15 and 20×20 , and the neighbourhood search results are shown in Figure 6.

Figure 6 Neighbourhood search results (see online version for colours)

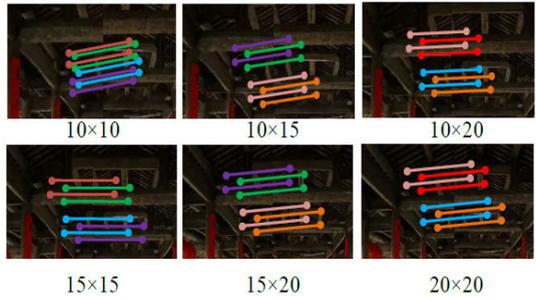
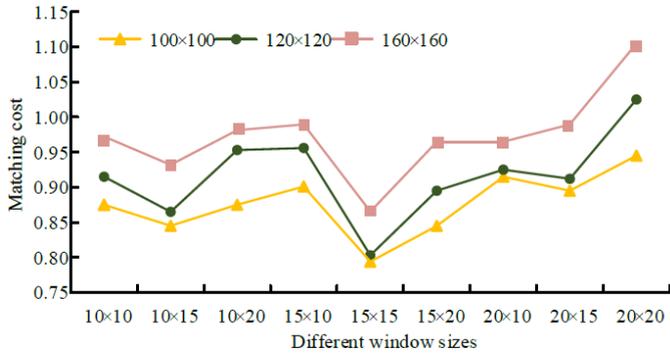


Figure 6 shows the results of searching for the best match for windows in different neighbourhoods. The best match window will be obtained when the neighbourhood range is 15×15 , and the most feature points are searched at this moment. Therefore to verify that this range meets the neighbourhood range requirement, the cost result of searching for the best matching window at window sizes of 100×100 , 120×120 and 160×160 traversing the whole image is shown in Figure 7.

Figure 7 Result diagram of optimal matching window (see online version for colours)



From Figure 7, it can be seen that the images acquired by the binocular camera system have the smallest surrogate value in the optimal range with a window neighbourhood of 15×15 . At the same time, a window neighbourhood of 15×15 has the highest similarity to the source window and the lowest surrogate value. In the 100×100 source window, the generation value is 0.781. Therefore, the matching cost in the 15×15 neighbourhood is the smallest, and 15×15 can be used as the matching window to ensure the quality of the 3D construction of the system image. To illustrate the superiority of the SIFT matching algorithm with the addition of RANSAC, the obtained feature matching results are plotted in Figure 8.

In Figure 8(a), the SIFT algorithm is able to filter out most of the presence of slope anomalies, but there are a large number of column structures in the wooden arch corridor bridge with high feature similarity, and the matching is prone to errors. Figure 8(b) shows the SIFT matching algorithm with the addition of RANSAC, which is able to filter out the mis-matched pairs and the number of mis-matched points in the image is significantly reduced. To further illustrate the matching accuracy of the improved algorithm, the root

mean square geometric error was used for testing, and the experimental results are shown in Table 3.

Figure 8 Comparison of feature matching results, (a) SIFT, (b) RANSAC-SIFT (see online version for colours)

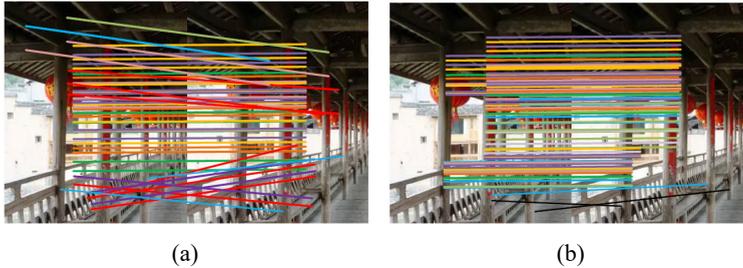


Table 3 Accuracy of RANSAC-SIFT matching algorithm

Experimental image	SIFT		RANSAC-SIFT	
	RMS	Running time	RMS	Running time
Group 1	1.5426	141 ms	1.2854	85 ms
Group 2	1.5712	153 ms	1.3014	90 ms
Group 3	1.4820	187 ms	1.2110	103 ms
Group 4	1.4632	176 ms	1.1951	98 ms

As can be seen from the data in Table 3, the improved RANSAC-SIFT algorithm has excellent performance in multiple sets of image data test results, with higher matching pair accuracy and an average RMS of 1.2482, and better performance in terms of running time compared to the SIFT algorithm, with an average running time of 94 ms. The average RMS of the SIFT algorithm is 1.5148 and the average running time is 164.3 ms. It can be seen that the improved RANSAC-SIFT algorithm has higher accuracy in matching points.

5 Conclusions

The complexity of the wooden arch bridge makes it difficult to build a 3D model. In order to build a more accurate 3D model of the wooden arch bridge, the Laplace Sobel algorithm for edge feature extraction is proposed to accurately extract image features, and RANSAC algorithm is used to improve the matching accuracy. Through accurate feature matching, the 3D model of the wooden arch bridge is built. The experimental results show that as the active window size increases, the number of detected edge feature points gradually increases and the feature stereo matching set increases, but the number of matching pairs is lower than the number of feature points, and the overall structural features are extracted at 160×160 size to ensure the clarity of the image. When the neighbourhood range is 15×15 , the most feature points are searched at this moment and the matching generation value is also the smallest, with a generation value of 0.824, so the matching window is optimal in the 15×15 neighbourhood. The improved RANSAC-SIFT matching algorithm is able to filter the mis-matched pairs, with

significantly fewer mis-matched points and higher accuracy when matching; its average RMS is 1.2482 and the average running time is 94ms, and the relevant data also indicate that the RANSAC-SIFT matching algorithm has better performance. Therefore, the application of this matching algorithm to binocular image reconstruction provides important technical support for the construction of 3D models of wooden arch corridor bridge structures. Although this research has made some achievements, there are still shortcomings. The research only matches the image feature points in the 3D reconstruction. In the future research, we can study the line, surface and other feature matching in the image to further improve the 3D reconstruction effect of the wooden arch bridge structure.

Funding

The research is supported by: Science Foundation for Young Scholars of Chengyi University College, Jimei University (2018); Special fund for first-class discipline construction of Chengyi College, Jimei University (2023): Inheritance and innovation of historic buildings in Fujian and Taiwan (No. C11709).

References

- Ammad, M. and Ramli, A. (2020) 'Construction of 3D-terrain model from contour lines using parameterized B-spline ruled surface', *Pertanika Journal of Science and Technology*, Vol. 28, No. 3, pp.937–950.
- Bailer, R. and Martin, R. (2019) 'The effectiveness of using 3D reconstruction software for surgery to augment surgical education', *The American Journal of Surgery*, Vol. 218, No. 5, pp.1016–1021.
- Cai, H., Chen, Y., Lingling, X.U., Elbaz, K. and Zhang, C. (2020) 'Intelligent building system for 3D construction of complex brick models', *IEEE Access*, Vol. 8, pp.182506–182516.
- Casado, A., Sánchez, M.A.C. and Leon, I. (2021) 'Use of flat interwoven wooden strips in architecture and construction simulation and optimization using 3D digital models', *Sustainability*, Vol. 13, No. 11, pp.6383–6391.
- Chen, Y., Li, Z., Zeng, T., Ning, Y. and Bin, Z. (2019) 'Research and design of 3d reconstruction system based on binocular vision', *International Core Journal of Engineering*, Vol. 5, No. 12, pp.29–35.
- Chopin, A., Chan, S.W. and Guellai, B. (2019) 'Binocular non-stereoscopic cues can deceive clinical tests of stereopsis', *Scientific Reports*, Vol. 9, No. 1, pp.5789–5796.
- Dandapani, S.A., Padmanabhan, P. and Hussaindeen, J.R. (2020) 'Spectrum of binocular vision anomalies in keratoconus subjects', *Optometry and Vision Science*, Vol. 97, No. 6, pp.424–428.
- Guo, X.D., Wang, Z.B., Zhu, W., He, G., Deng, H.B. and Lv, C.X. (2022) 'Research on DSO vision positioning technology based on binocular stereo panoramic vision system', *Defence Technology*, Vol. 18, No. 4, pp.593–604.
- Hauptmann, A., Lucka, F., Betcke, M., Huynh, N., Adler, J., Cox, B., Ourselin, S. and Aeeidge, S., (2018) 'Model-based learning for accelerated, limited-view 3-D photoacoustic tomography,' in *IEEE Transactions on Medical Imaging*, Vol. 37, No. 6, pp.1382–1393.
- Jessell, M., Ogarko, V., Rose, Y.D., Lindsay, M., Joshi, R. and Piechocka, A. (2021) 'Automated geological map deconstruction for 3D model construction using map2loop 1.0 and map2model 1.0', *Geoscientific Model Development*, Vol. 14, No. 8, pp.5063–5092.

- Jin, S., Zhang, Y., Yamazaki, T. and Jiang, Z. (2021) 'Automatic 3D CAD model and 2D drawings generation in construction engineering', *Journal of Physics Conference Series*, Vol. 1827, No. 1, pp.115–122.
- Li, H. and Zhang, B. (2021) 'Application of integrated binocular stereo vision measurement and wireless sensor system in athlete displacement test', *AEJ – Alexandria Engineering Journal*, Vol. 60, No. 5, pp.4325–4335.
- Li, L., Wang, J., Yang, S. and Gong, H. (2021a) 'Binocular stereo vision based illuminance measurement used for intelligent lighting with LEDs', *Optik – International Journal for Light and Electron Optics*, Vol. 237, No. 7, pp.51–56.
- Li, Y., Cheng, Z., Yang, C., Wei, M. and Wen, J. (2021b) 'Application of binocular stereo vision in radioactive source image reconstruction and multimodal imaging fusion', *IEEE Transactions on Nuclear Science*, Vol. 67, No. 11, pp.2454–2462.
- Liu, Y. (2020) 'A full moon in another land: the Moon Bridge in the Japanese garden of the Huntington Library', *Frontier of Architecture Research*, Vol. 9, No. 3, pp.556–567.
- Lu, N., Wu, Y., Feng, L. and Song, J. (2018) 'Deep learning for fall detection: three-dimensional Cnn combined with LSTM on video kinematic data', *IEEE Journal of Biomedical and Health Informatics*, Vol. 23, No. 1, pp.314–323.
- Lyu, F., Goto, Y., Kawanishi, N. and Yan, X. (2020) 'Three-dimensional numerical model for seismic analysis of bridge systems with multiple thin-walled partially concrete-filled steel tubular columns', *Journal of Structural Engineering*, Vol. 146, No. 1, pp.04019164.1–04019164.16.
- Ma, Y., Li, Q., Chu, L., Zhou, Y. and Xu, C. (2021) 'Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision', *Remote Sensing*, Vol. 13, No. 2, pp.230–241.
- Qiu, Z.C. and Wang, T.X. (2019) 'Fuzzy neural network vibration control on a piezoelectric flexible hinged plate using stereo vision detection', *Journal of Intelligent Material Systems and Structures*, Vol. 30, No. 4, pp.556–575.
- Sun, X., Jiang, Y. and Ji, Y. (2019) 'Distance measurement system based on binocular stereo vision', *IOP Conference Series Earth and Environmental Science*, Vol. 252, No. 5, pp.51–60.
- Wang, R., Han, Y., Luo, M.Z., Wang, N.K. and Lu, L.J. (2020) 'Accuracy study of a binocular-stereo-vision-based navigation robot for minimally invasive interventional procedures', *World Journal of Clinical Cases*, Vol. 8, No. 16, pp.3440–3449.
- Wang, Z. (2021) 'Analysis on construction technology of reinforced concrete tied arch bridge', *World Architecture*, Vol. 5, No. 6, pp.67–71.
- Yin, H., Ma, Z., Zhong, M., Wu, K. and Huang, B. (2020) 'SLAM-based self-calibration of a binocular stereo vision rig in real-time', *Sensors*, Vol. 20, No. 3, pp.621–641.