IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 August 2023, accepted 23 October 2023, date of publication 2 November 2023, date of current version 8 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3329580

RESEARCH ARTICLE

3D Reconstruction Cost Function Algorithm Based on Stereo Matching in the Background of Digital Museums

PENG PENG[1] AND JUN HAN[2]

[1]School of Design, Jiangnan University, Wuxi 214122, China
[2]School of Art and Design, Wuhan Institute of Technology, Wuhan 430205, China

Corresponding author: Peng Peng (pengpeng20230224@126.com)

ABSTRACT Stereo matching plays an important role in 3D reconstruction in the context of digital museums. At present, it has problems such as occlusion, weak texture, and discontinuous disparity, which restrict the development of binocular vision. In response to this type of problem, Census transformed algorithms based on mean discrimination and Sobel edge detection were introduced to calculate the cost function. At the same time, the algorithm also incorporated the absolute value method of grayscale difference, making it more adaptable to situations such as discontinuous disparity and weak textures. The results show that the Census transform algorithm, which introduces edge gradients, has the lowest error matching rate on different images, with a minimum value of 25.1%. The classical Census transformation method and the AD Census classical transformation algorithm are 28.3% and 27.4%, respectively. Compared with the other two algorithms, the Census transformation of edge gradient improves the matching performance of the algorithm in the discontinuous area of edge disparity and improves the anti-interference ability of the algorithm. At the same time, the algorithm has the lowest error matching rate in the disparity discontinuous regions of Teddy, Cones, Venus, and Tsukuba images, and the lowest value is only 32.1%. Compared to the classic Census transformation method and the AD Census classical transformation algorithm, the minimum error matching rate has decreased by 13.2% and 4.5%, respectively. In addition, the algorithm has the lowest average effective runtime on all four types of images, with an effective average of 4.6 seconds on Venus images with rich texture features, which is 0.6 seconds lower than the classic Census transform algorithm. The improved Census transform algorithm not only has high matching accuracy but also low time complexity, providing a reliable method reference for modern 3D reconstruction fields.

INDEX TERMS Stereo matching, 3D reconstruction, cost function, edge gradient, Census transformation.

I. INTRODUCTION

Digital museum refers to a new form of museum that utilizes modern information technology to digitize the cultural relics and exhibits of physical museums on a network platform. Through digital museums, people can browse, learn, and study cultural relics online, providing a more convenient and open way for cultural exchange for the general audience. In the digital museum, Iterative reconstruction technology is an important technical means, which can transform the actual 3D objects into digital 3D models, so that the audience can experience the reality through virtual reality technology and other methods. In Iterative reconstruction, stereo matching is a common method, which estimates the depth information of object surface through pixel point matching between two or more images, so as to realize 3D model reconstruction [1], [2]. However, due to the existence of noise, occlusion and illumination changes in the image, stereo matching methods face great challenges. Therefore, designing an efficient and accurate cost function algorithm is of great significance for

The associate editor coordinating the review of this manuscript and approving it for publication was Gongbo Zhou.

solving stereo matching problems [3]. Cost matching is a crucial step in stereo matching algorithms, which belongs to the process of finding homonymous points and reconstructing scenes. However, in the process of scene reconstruction, there are many uncontrollable factors, such as noise, low texture, optical distortion, perspective distortion, etc., which affect the accuracy of stereo matching. The Census cost function based on non parametric transformation is currently widely used in dealing with accuracy issues, but it has the drawback of overly relying on central pixels in the calculation of disparity maps [4]. The research mainly focuses on the problems of stereo matching technology in 3D reconstruction, such as image noise, textureless areas, and edge blur, and proposes an optimized cost function calculation method. This method combines mean discrimination, Census transform method for Sobel edge detection, and absolute value method for grayscale difference to better handle these problems and improve the performance and accuracy of stereo matching. The research content mainly includes four parts. The first part mainly provides an overview of 3D reconstruction technology and SM algorithms. The second part studied the cost function algorithm for 3D reconstruction based on SM. The first section provides a detailed description of the constraints and methods required for 3D reconstruction by SM. The second section focuses on the cost function calculation problem in SM, introducing the Census transform method of mean discrimination and Sobel edge detection, and combining it with the absolute value method of grayscale difference for optimization. The third part verifies the performance and advantages of the improved cost function calculation. The fourth part elaborates on the experimental data and effectiveness of the proposed method, and proposes future improvement directions. The contribution of this study is mainly reflected in the following aspects. Firstly, by applying the method of mean discrimination, the impact of image noise on the matching process was successfully reduced, significantly improving the accuracy of matching. This innovative method provides new solutions for complex image processing tasks. Secondly, the Sobel edge detection operator was used in the study, which can effectively detect edge information in the image and use these edge information as key features for stereo matching. This step not only enhances the robustness of matching, but also provides a more accurate data foundation for subsequent processing. In addition, by utilizing edge information to determine the position of matching points, the method in this study further improves the accuracy and robustness of stereo matching. This method has advantages over region based stereo matching algorithms when dealing with issues such as textureless regions and edge blurring in images, and can better adapt to various complex scenes [5]. Most importantly, the research provides key technical support for iterative reconstruction of digital museums. By applying this technology, the experience and display effect of digital museums have been significantly improved, providing viewers with a richer and more vivid visiting experience. This contribution not only has important value in the academic field, but also has profound impact in practical applications.

II. RELATED WORKS

3D reconstruction refers to the establishment of appropriate computer processing mathematical models for 3D objects. Currently, 3D reconstruction technology is widely utilized in various fields of society. Cui Y's team believed that 3D modeling of indoor environments had crucial effect on interactive visualization and building information modeling. Therefore, the team proposed to use point clouds and trajectories scanned by mobile lasers for 3D reconstruction of indoor environments. This method mainly utilized multi label graph cutting technology to solve the energy optimization function. The results showed that the model reconstructed by this method had better recall and precision than other techniques [6]. Chen et al. designed a 3D gradient echo imaging technique for faster and more efficient high-resolution whole brain imaging. At the same time, this technology utilized an iterative hard threshold algorithm to cut down the cost function. The results indicated that this method effectively improved the accuracy of 3D reconstruction and could achieve fast 3D distortion free high-resolution imaging [7]. Cai and other researchers introduced a stereo network based on multi-level fusion perception feature pyramid to solve the cost volume regularization of consuming memory and dense matching in 3D reconstruction. This method could narrow down the interval of deep search through prior information from the previous level, and used multi-level fusion to establish a feature pyramid. It was validated that this method effectively alleviated the memory consumption and obtained an effective cost representation [8]. Li et al. team believed that the current 3D reconstruction technology did not provide sufficient data accuracy in laser scanning, so they have introduced a laser scanner based on light detection to optimize it. The device needed to first set up several control points to concatenate the laser scanned images. The results indicated that it had good application effects in geometric 3D reconstruction of infrastructure [9]. To better utilize electrical impedance tomography technology for 3D reconstruction of conductivity distribution, Liu et al. introduced a high-resolution algorithm to solve it. This method utilized Bayesian learning based on structural perception to improve the imaging technology. The outcomes proved that this method could effectively improve the accuracy of 3D reconstruction and simplify the computational complexity [10]. Kench and Cooper scholars proposed a SliceGAN architecture based on generative adversarial networks for high-throughput micro structure optimization problems. This network mainly generated 3D imaging data through representative 2D images to reconstruct 3D models. The results indicated that this method had wide applicability and provided a reliable method reference for future high-throughput microstructure optimization [11].

SM is an important branch of computer vision, whose goal is to obtain matching corresponding points from different viewpoint images. Currently, many scholars have conducted extensive research on this topic. Zhang and other researchers introduced a cross form pyramid stereo matching network to solve the disparity regression problem of corrected stereo images. The network regularizes the cost volume through two parallel 3D deconvolution structures with different receptive fields. The results show that this method has achieved significant results on both scene streams and KITTI datasets [12]. Wang H and other researchers proposed a non parallel stereo matching algorithm based on improved dynamic programming. This algorithm can perform stereo matching on different disparity search paths, and find the optimal solution through Dynamic programming. The results indicate that high image processing results have been achieved in various application scenarios and hardware platforms [13]. Liang et al. proposed using serial computation for stereo matching in response to stereo vision problems. Non parallel stereo matching methods mainly obtain the depth information of images through specific algorithms, such as semi global matching or Dynamic programming. The results indicate that this method is suitable for small-scale image processing and offline applications, such as computer vision research and image processing algorithm development [14]. Yan et al. proposed to improve the parallax plane through global and local optimization for the parallax thinning problem in SM. Among them, global optimization used Markov random field to estimate the average parallax of super pixels, and local optimization used Bayesian model to smooth the 3D neighborhood. The results showed that this scheme could improve the accuracy of SM and effectively reduce computational costs [15]. Wang et al. found that there were some problems such as noise in image acquisition when collecting the images of obstacles on the moon surface, so they introduced an improved AD Sense algorithm. This method introduces an improved average window pixel calculation algorithm into the original census algorithm. The results show that this method can effectively detect obstacles in the image [16]. Joung et al. proposed a stereo matching method based on global optimization in order to further improve the accuracy of stereo matching algorithm. This algorithm considers the consistency of each pixel with all other pixels to obtain the best disparity map. Compared to local stereo matching and semi global stereo matching algorithms, global stereo matching algorithms have higher accuracy and robustness, but higher computational complexity [17]. Liu et al. proposed an improved AD Cenus algorithm based on two-stage adaptive optimization and gradient fusion to solve the problem of poor performance of existing traditional local stereo matching methods in ill posed regions. This method calculates the absolute difference cost and census transformation cost of each pixel by weighting, and obtains a disparity map through cost aggregation, disparity selection, and disparity refinement. The results show that this method performs well in textureless and parallax discontinuous regions, and is sufficiently robust to radiation changes and noise [18].

To sum up, researchers at home and abroad have carried out a lot of research on Iterative reconstruction technology and stereo matching methods. Among them, relevant analysis was conducted on the application of parallel and non parallel algorithms in stereo matching. At the same time, the application of local, semi global, and global stereo matching algorithms in different scenarios was introduced. However, these algorithms have poor processing performance and high computational complexity for scenes with weaker textures. Therefore, the Census cost function of edge gradient is introduced into the traditional stereo matching algorithm to optimize the algorithm, so as to better improve the application effect of Iterative reconstruction in modern digital museums.

III. 3D RECONSTRUCTION COST FUNCTION ALGORITHM BASED ON STEREO MATCHING

In the context of digital museums, the application of computer vision technology is becoming increasingly widespread. This chapter first provides a detailed description of the 3D construction methods and conditions for SM. Next, it needs to calculate the cost function in SM. The Census transform of mean discrimination and Sobel edge detection were introduced separately, and combined with the absolute value method of grayscale difference to make the algorithm more adaptable to situations such as discontinuous edge depth information and weak edge images.

A. 3D RECONSTRUCTION BASED ON STEREO MATCHING

SM systems mainly include binocular, tri ocular, and multi ocular SM systems, among which the most widely used is the binocular SM system. Through this technology, museum exhibits can be presented in virtual form on digital platforms, and users can visit exhibitions through devices such as VR glasses or computers. Users can freely choose viewing angles and distances, and can interact with exhibits to provide a more immersive experience. Binocular stereo vision refers to the use of two cameras to simultaneously capture images of an object, but due to differences in the positions captured by the two cameras, there is also a certain degree of parallax between the images they capture [19], [20]. Therefore, the research adopts techniques such as stereo matching and camera calibration to treat the surface of an object as a series of points, and obtains three-dimensional coordinates and constructs a three-dimensional model based on the spatial position relationship between these points and spatial points. The framework structure of binocular stereo vision 3D reconstruction is shown in Figure 1.

The binocular stereo vision model mainly includes a binocular parallel and a binocularnon parallelstructure model, and its main division is based on the placement position of the two cameras. Among them, the two cameras of the binocular parallel structure model are located in the same height plane, and their optical axes are completely parallel, with identical

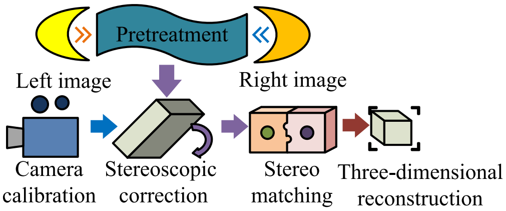


FIGURE 1. The framework structure of 3D reconstruction for binocular stereoscopic vision.

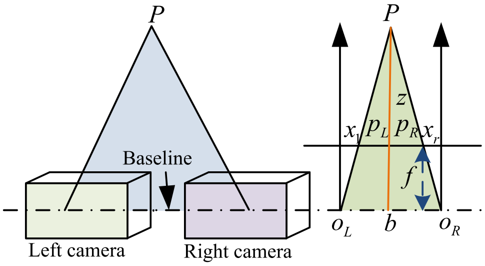


FIGURE 2. Binocular parallel structure model.

parameters [21]. The binocular parallel structure model is shown in Figure 2.

Assuming P(x,y,z) is any point in space, the mathematical representation of its coordinates is shown in equation (1).

$$\begin{cases} \frac{y}{z} = \frac{y_l}{f} \\ \frac{x}{z} = \frac{x_l}{f} \\ \frac{b}{z} = \frac{b-(x_l-x_r)}{z-f} \end{cases} \tag{1}$$

In equation (1), f represents the focal length of two cameras; b represents the baseline, which is the optical center line connecting two cameras; x_l ND x_r represent the horizontal coordinate of the left and the rightcamera shooting points. The calculation of parallax is shown in equation (2).

$$d = x_l - x_r \tag{2}$$

After transformation, various coordinate values of point P can be obtained, and their calculation is shown in equation (3).

$$\begin{cases} x = x_l \cdot \frac{z}{f} \\ y = y_l \cdot \frac{z}{f} \\ z = \frac{bf}{d} \end{cases} \tag{3}$$

From equation (3), the depth value of a spatial point is inversely proportional to the parallax value, that is, the nearer the spatial point is to the binocular camera, the bigger the parallax value corresponding to that point. Therefore, the distance between the spatial point and the camera can be determined by calculating the parallax value, thereby obtaining the 3D coordinates of the spatial point. In addition, the baseline and focal length between cameras are mainly obtained through camera calibration. However, in reality, it is not easy to realize complete parallelism between the optical axes of two cameras, making it difficult to construct a binocular parallel structure model. The binocular non parallel structure model solves this problem well, and it does not have strict requirements for the position relationship of the camera [22]. Therefore, the binocular parallel structure model is used in the actual Iterative reconstruction of the digital museum. The binocular optical axis non parallel model is shown in Figure 3. From where, when one of the cameras observes the spatial point P, the accurate position of P cannot be obtained. But when two cameras are used to observe the P point, the specific orientation of the P point can be obtained.

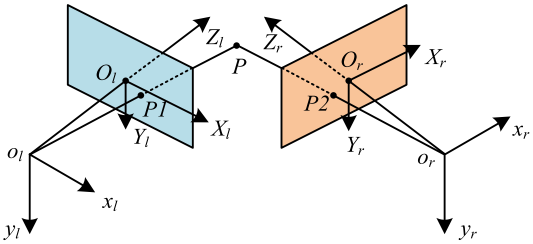


FIGURE 3. Schematic diagram of non parallel model of binocular optical axis.

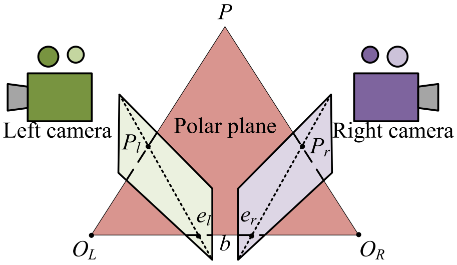


FIGURE 4. Schematic diagram of polar constraint structure.

When using SM algorithms for 3D reconstruction, the SM algorithm plays a role in obtaining depth information. It mainly uses the disparity values of two images to obtain depth indicators, thereby constructing a 3D model. However, during the collecting scene images, it is inevitable to be affected by noise interference, which may result in errors between the collected and the ideal images, leading to matching errors. Therefore, when performing SM, it is necessary to use constraints to reduce the search area to reduce error rates and improve matching efficiency. The commonly used constraints mainly include polar, continuity, disparity range, similarity, uniqueness, and sequential consistency constraints [23], [24]. The study selects polar constraints as the constraint conditions for stereo matching, which can adjust the matching pixel points of the camera to be level with the reference pixel points, thereby effectively improving the accuracy and efficiency of matching. The schematic diagram of polar constraint is shown in Figure 4. Where, P_l and P_r stand for the projection points of spatial points on the left and right cameras; The intersection point between the camera surface and the baseline represents the base point; e_lP_l and e_rP_r represent the polar lines corresponding to two projection points. During the SM, to obtain points on e_lP_l, there is no need to search the entire image, but only search for matching points in the direction of the polar e_rP_r. This transforms the search in the 2D direction into a search in the 1D direction, greatly improving matching efficiency and accuracy.

In the process of Iterative reconstruction using stereo matching algorithm, the cost function needs to be calculated first. The cost function is mainly used to determine the degree of similarity between matching pixels and reference pixels, which is mainly represented by surrogate value. The greater the similarity between matching pixels and reference pixels, the smaller the corresponding cost value. Before performing cost matching, a range of disparity values will be set and the disparity value will be calculated within this range. The study uses a three-dimensional matrix to store the generation value of pixels within the parallax range, where each element corresponds to a pixel in the image. In a three-dimensional matrix, the first dimension represents the position of pixels, the second dimension represents the parallax value, and the third dimension represents the generation value. After completing the calculation of the cost function, proceed with cost aggregation. Cost aggregation mainly optimizes the cost matching stage to enhance the similarity between reference pixels and matching pixels. Cost aggregation overcomes the limitations of cost matching from a local perspective, mainly matching costs from a global perspective, thus reducing the error rate of matching. The study chose Box Filtering to achieve cost aggregation, which has the advantage of being relatively simple in calculation and only needs to calculate the average value around the pixels. At the same time, it can effectively smooth the image, reduce high-frequency noise in the image, and make the image clearer and more natural. In addition, it does not introduce additional sharpening or distortion effects, making it suitable for application scenarios that maintain image details and edges. This method mainly treats the disparity values of each spatial point as equal, and the specific calculation is shown in equation (4).

$$C_d^q = \frac{\sum_q C_d(q)}{N} \tag{4}$$

The cost aggregation optimization calculation of pixel point q along a certain direction is shown in equation (5).

$$\begin{aligned} C_r(q,d) = {} & \min(C_r(q-r,d), C_r(q-r,d\pm 1)) \\ & + T1, \min_k C_r(q-r,k) \\ & + T2, \min_k C_r(q-r,k) + C_l(q,d) \end{aligned} \tag{5}$$

In equation (5), $T1$ and $T2$ represent penalty terms. The total cost aggregation calculation of pixel points along various directions is shown in equation (6).

$$C_{agg}(q,d) = \sum\nolimits_r C_r(q,d) \tag{6}$$

After cost aggregation, the obtained cost values need to be statistically analyzed through disparity calculation. Next, it selects the matching pixel with the smallest replacement value within the disparity range, and the disparity value corresponding to this pixel is that calculated by the algorithm. The disparity calculation method used in the study is the Winner Takes All (WTA) algorithm, which mainly selects the disparity value of the smallest matching pixel point as the final result by counting the cost values within each disparity range. This method is simple, efficient, easy to implement, and suitable for most disparity calculation scenarios, including different types of images with rich textures, sparse textures, and obvious edges. At the same time, this method only needs to compare the surrogate values in the cost volume, without relying on the neighborhood information around the pixels. This makes the implementation of the algorithm relatively simple and does not require additional computational and storage overhead. Parallax optimization is mainly to ensure that the proxy value of matching pixels obtained by disparity calculation is minimized, and it requires refinement of the obtained disparity value to improve the accuracy of disparity value acquisition. The research mainly uses methods such as sub pixel fitting, left and right consistency detection, and disparity filling to optimize disparity maps. This method mainly adjusts the range dynamically when calculating the parallax value based on the content and characteristics of the input image. At the same time, the range of parallax can be adjusted adaptively according to the depth distribution of objects in the image [25]. Among them, subpixel fitting mainly records the neighboring modern values of the minimum generation value, and fits them through a univariate quadratic curve, and then calculates the extreme points of the curve. The abscissa of the extremum point obtained is the sub pixel value of the parallax value. The sub pixel fitting diagram is shown in Figure 5. In Figure 5, if the minimum matching value is 16, the sub pixel value of the visual difference is 16.

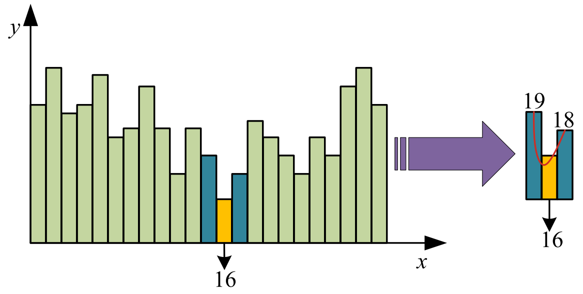


FIGURE 5. Schematic diagram of sub pixel fitting.

B. ANALYSIS OF ALGORITHMS OF ITERATIVE RECONSTRUCTION COST FUNCTION BASED ON STEREO MATCHING

The calculation of the cost function is the most crucial step in SM algorithms. According to different constraint conditions, SM algorithms is composed of global and local SM. Among them, the global matching algorithm first needs to construct a global energy function, then solve the function according to the global optimization theory, and take its minimum value as the optimal disparity value. The definition and calculation of this energy function are shown in equation (7).

$$E(d) = E_{data}(d) + \delta E_{smooth}(d) = \sum_{p,q\in R} T(d_q, d_p) + \sum_{p\in R} S(p, d) \tag{7}$$

In equation (7), $E_{smooth}(d)$ represents the smoothing term, which represents the degree of continuity between the current pixel and adjacent pixels. The data item $E_{data}(d)$ represents the degree of matching between left and right images, which is mainly utilized to record the similarity between pixel points. $S(p, d)$ refers to the matching proxy value;R represents the penalty term, which mainly increases with the difference between two pixels. If the disparity values of two pixels are equal, the penalty term value is 0. Compared with global SM, local SM has higher real-time performance and is suitable for real-time scenes with general accuracy requirements. The local matching algorithm requires the construction of a local cost function, which does not involve smoothness between adjacent pixels and therefore does not have a smoothing term. It also requires constructing a support window and matching based on the information carried by the pixels around the points to be matched. The feature information of these pixels is relatively similar, and the algorithm mainly measures the similarity among pixels through similarity measurement criteria, and the best matching point is the point with the highest similarity. The study uses this method to calculate the cost function. The common similarity measurement functions in local matching algorithms include normalized cross correlation method, sum of squares and absolute valuesof grayscale differences. They mainly describe the similarity between pixels through grayscale information [26]. The calculation of the sum of absolute values of grayscale difference is expressed in equation (8).

$$SAD(x, y, d) = \sum_{(i,j)\in m} |I_L(x+i, y+j) - I_R(x+i+d, y+j)| \tag{8}$$

In equation (8), m is the size of the support window; I_L and I_R represent the left and right images, respectively; and i and j represent the offset of pixels. The calculation of the sum of squares of grayscale differences is shown in equation (9).

$$SSD(x, y, d) = \sum_{(i,j)\in m} [I_L(x+i, y+j) - I_R(x+i+d, y+j)]^2 \tag{9}$$

The calculation of the normalized cross correlation method is shown in equation (10).

$$NCC(x, y, d) = \frac{\sum\limits_{(i,j)\in m} I_L(x+i, y+j) \cdot I_R(x+i+d, y+j)}{\sqrt{\sum\limits_{(i,j)\in m} I_L(x+i, y+j)^2 \cdot I_R(x+i+d, y+j)^2}} \tag{10}$$

Among them, the smaller the matching value obtained by the sum of absolute values and squaresof grayscale differences, the higher the similarity between the pixels to be matched. The closer the calculation result of the normalized cross correlation method is to 1, the higher the similarity between the pixels to be matched. In addition, the Census transform based on non parametric transformation can also measure the similarity of pixels, which can detect local features such as corner information and edge information. The cost function expression is shown in equation (11).

$$S_{census}(x, y, d) = ham\,[I_L(x, y), I_R(x+d, y)] \tag{11}$$

In equation (11), $ham\,[I_L, I_R]$ represents the Hamming distance. The Census transform first constructs a support window and sets the reference pixel as the center pixel of the window. Next, it compares the grayscale values of each pixel in the center reference pixel and the support window. When the grayscale value of the reference pixel is lower, it is recorded as 0, otherwise it is recorded as 1. Finally, it generate a binary bitstream and use it as a Census sequence of central reference pixels. Essentially, Census transformation transforms the gray value of the reference pixel into a binary code stream, and its replacement expression is shown in equation (12).

$$\zeta\,[I(p), I(q)] = \begin{cases} 0, & I(p) \geq I(q) \\ 1, & I(p) < I(q) \end{cases} \tag{12}$$

In equation (12), $I(p)$ and $I(q)$ are the grayscale values of the central reference pixel and the remaining pixels, respectively. The conversion connection expression for binary bitstream sequences is expressed in equation (13).

$$S_{census}(x, y) = \otimes_{i=-l}^{l} \otimes_{j=-r}^{r} \zeta\,[I(x, y) I(x+i, y+j)] \tag{13}$$

In equation (13), $\otimes$ represents bitwise connection, and $S_{census}(x, y)$ refers to the Census sequence code of the central reference pixel. $I(x, y)$ means the grayscale value of the reference pixel, and $I(x+i, y+j)$ indicates the grayscale value of other pixel points. After obtaining the Census sequence codes of all pixels, the matching surrogate value is calculated with the similarity metric of Hamming distance. The calculation of Hamming distance is shown in equation (14).

$$S_{census}(p, d) = Ham\min g\,[S_l(p)]\, S_r(p, d) \tag{14}$$

In equation (14), $S_l(p)$ and $S_r(p, d)$ respectively represent the Census sequence codes of the pixel p in the left and right images, while $S_{census}(p, d)$ infers the matching value of the pixel p. Among them, the smaller the Hamming value, the greater the similarity between pixels. The specific transformation of Census is shown in Figure 6.

The traditional Census transform can improve the matching accuracy of images under the influence of noise, but it has certain drawbacks. Firstly, the selection of the support window has a significant impact on the matching effect. When the window is too big, it can cause points with

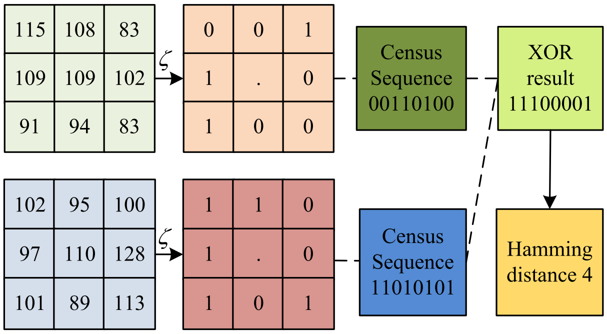


FIGURE 6. The specific transformation process of census.

significant disparity changes to be included, resulting in a decrease in matching accuracy. Secondly, the Census transform has an excessive dependence on the central reference pixel, and sudden changes in the grayscale value of the reference pixel can affect the entire matching effect [27], [28]. In response to the dependency of the Census transform on the central reference pixel, an improved Census transform method with mean discrimination was studied. This method first calculates the average grayscale values of all pixels within the window and uses them as reference values. Next, it needs to work out the absolute value of the difference between the reference and the center pixel value, and compare it with the set threshold. Finally, the grayscale value of the reference pixel is decidedwith the comparison results, and the definition calculation of the comparison is shown in equation (15).

$$I_z(x,y) = \begin{cases} I(x,y), & \left|I(x,y) - \bar{I}(x,y)\right| \leq \delta \\ \bar{I}(x,y), & \left|I(x,y) - \bar{I}(x,y)\right| > \delta \end{cases} \tag{15}$$

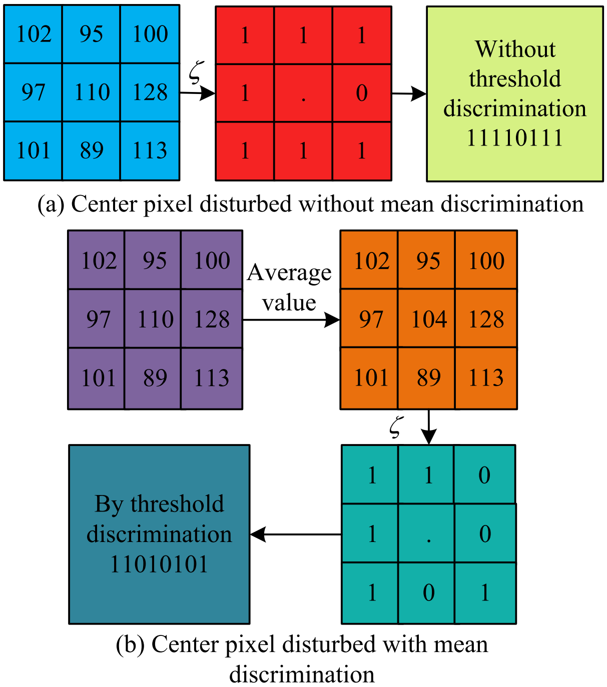
In equation (15), δ represents the set threshold;$I(x, y)$ is the grayscale value of the center pixel in the window;$\bar{I}(x, y)$ indicates the average grayscale value of pixels except for the center point, and also represents the final reference grayscale value. The calculation of $\bar{I}(x, y)$ is presented in equation (16).

$$\bar{I}(x,y) = \frac{\sum I(x,y)}{N-1} \tag{16}$$

In equation (16), N denotes the number of pixels, and after mean discrimination, the transformation relationship of Census is represented as shown in equation (17).

$$\zeta\,[I_Z(p), I(q)] = \begin{cases} 0, & I_Z(p) \leq I(q) \\ 1, & I_Z(p) > I(q) \end{cases} \tag{17}$$

In equation (17), $I_Z(p)$ means the grayscale value of the center reference pixel after mean discrimination. The comparison of the Census transformation results between the center pixel without mean discrimination and the mean discrimination is shown in Figure 7. From Figure 7, when the central reference pixel changes due to interference, the Census sequence code obtained by the traditional Census transformation method also undergoes significant changes. The Census transformation method after mean discrimination is almost unaffected. This indicates that the Census transformation method after mean discrimination has strong adaptability to the environment.



(a) Center pixel disturbed without mean discrimination

(b) Center pixel disturbed with mean discrimination

FIGURE 7. Comparison of census transformation results.

An increase in the Census transform window can improve the matching accuracy to a certain extent, but an excessive window can lead to an increase in the mismatch rate at the edge of the region. To address this issue, a Sobel edge detection operator was proposed to constrain image edges. The implementation step of this method is to first convert the original image into a Grayscale and smooth it, so as to reduce the impact of noise on the edge detection results. Then apply horizontal and vertical Sobel operator templates to the smoothed image, respectively. The horizontal Sobel operator template is used to detect horizontal edges in the image, while the vertical Sobel operator template is used to detect vertical edges. The calculation results of these two templates obtained gradient images in the horizontal and vertical directions of the image, respectively. Then merge the horizontal and vertical gradient images. This can be achieved by calculating the amplitude of two gradient images, that is, calculating the gradient size of each pixel point, as shown in the formula (18).

$$G = sqrt((g_x^2 + g_y^2) \tag{18}$$

In equation (18), g_x, g_y represents the gradient values of the pixel in the horizontal and vertical directions, respectively. Finally, perform threshold processing on the merged gradient image. According to the set threshold, set the pixels below the threshold in the gradient image to 0, and the pixels above the threshold to 255 to obtain a binarized edge image. Among them, the Sobel edge operator is a discrete difference operator used for edge detection, which mainly marks a specific point as an edge point based on the approximate value of the brightness of the pixels around the edge. This operator mainly assigns weights to the distance between neighboring pixels and the current pixel to highlight the edge contour of the image. In addition, when selecting the scale of the support window, the issue of disparity continuity should also be considered. The study introduced the method of Absolute value of grayscale difference (AD), which assigns more weights to pixels with higher feature similarity to the matched pixels, while assigning fewer weights to pixels with lower similarity [29].

IV. ANALYSIS OF 3D RECONSTRUCTION COST FUNCTION ALGORITHM BASED ON STEREO MATCHING

To verify the performance of the Census transform matching algorithm that introduces edge gradients, the study selected Census and AD Census classic transform methods for comparison. The dataset used in the experiment was Middlebury, and Teddy, Cones, Venus, and Tsukuba images were selected as the research subjects. Among them, Teddy images are characterized by a relatively large parallax range, the presence of some textureless areas and tilted planes, and a high degree of occlusion in the scene. Cones images are characterized by the presence of repetitive texture regions and some non planar objects in the scene. There are mainly many object planes in the Venus image scene, and the texture features of planar objects are very rich. The depth information of the scene changes regularly, and there are also slightly inclined planes with weaker textures. The disparity range of Tsukuba images is relatively small, and the view scene information is mainly the information of the front image parallel to the imaging plane. Several objects in the scene have different depth features, and the edge information between objects is relatively complex, especially in the tripod area of the black camera and the long lamp tube part of the orange desk lamp. There are also areas in the image where texture features are not obvious. These four typical sets of images include most of the feature situations of real-world scenes, making them suitable for testing and evaluating the performance of matching algorithms. At the same time, the development platform for the experiment was the Win 10 × 64 system, with Intel (R) Core (TM) i5-9500 as the central processor, 8GB of running memory, and MATLAB R2016a as the experimental environment. In addition, the standard disparity maps among the four images are expressed in Figure 8.

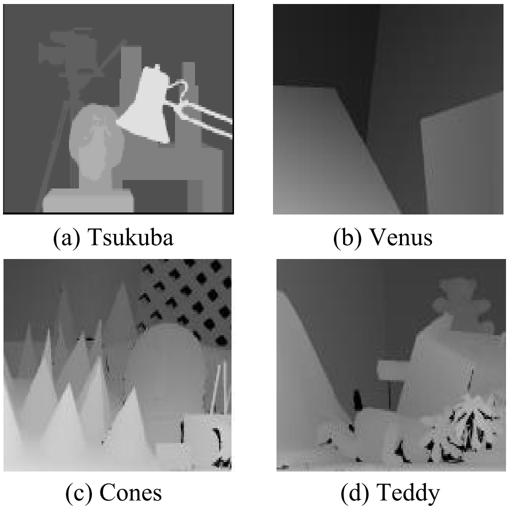

(a) Tsukuba
(b) Venus
(c) Cones
(d) Teddy

FIGURE 8. Standard disparity maps in four types of images.

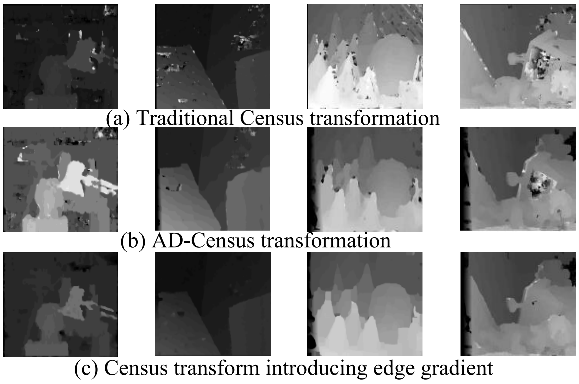
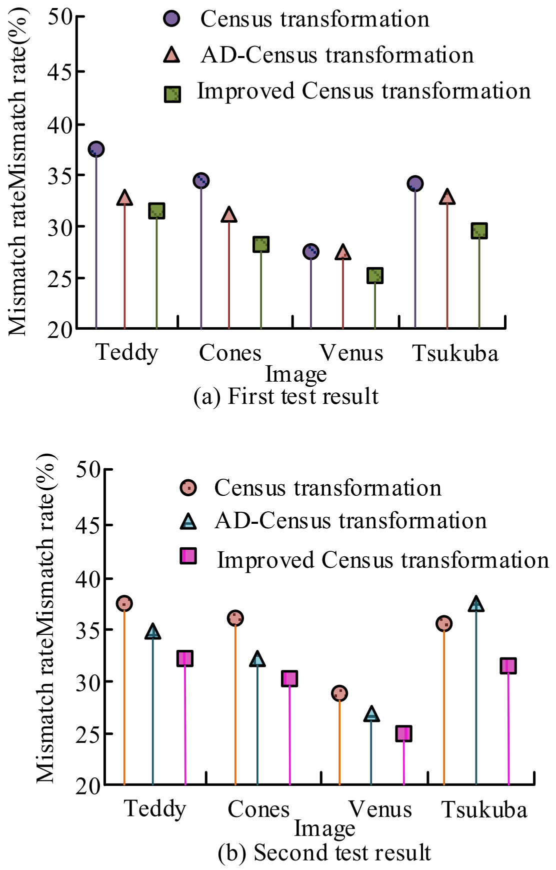

(a) Traditional Census transformation
(b) AD-Census transformation
(c) Census transform introducing edge gradient

FIGURE 9. Parallax maps obtained by different algorithms.

The disparity maps obtained by different algorithms are shown in Figure 9. From Figure 9, the disparity map effect of the Census transform algorithm introducing edge gradients was significantly better than traditional algorithms. The contours of each object in the disparity map of this algorithm were relatively clear. Among them, the Census transform algorithm that introduced edge gradients detects the best sharpening effect of the book contour edge in the Venus image and the desk lamp contour edge in the Tsukuba image, with the highest edge clarity. At the same time, the amount of mismatched points in this algorithm was significantly cut down in contrast to the classic Census transformation method and the AD Census classical transformation. The main reason was that the detection of edge points helps with image resolution, while the Sobel operator can obtain approximate grayscale differences in the horizontal and vertical directions by obtaining gradient information, thereby improving the resolution of image edges.

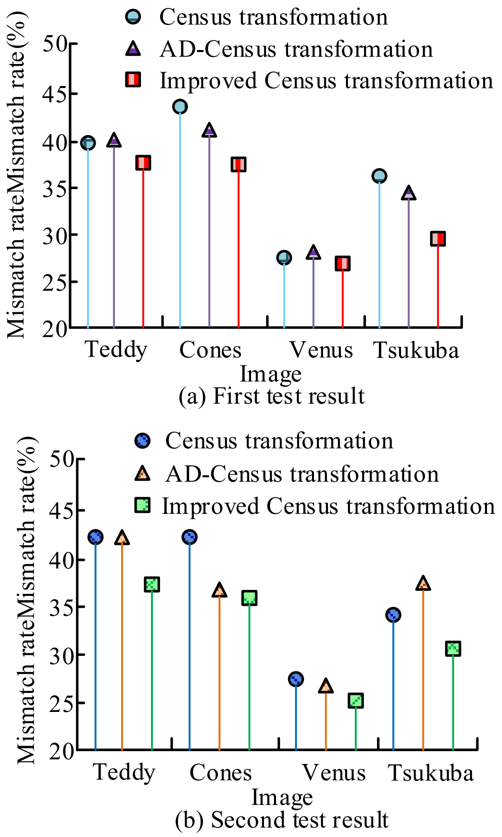
The research continued to validate the matching performance of the Census transform algorithm that introduced edge gradients. Experiments were conducted on non occluded regions, all regions, and disparity discontinuous regions of the image. At the same time, to ensure the effectiveness of the validation results, the experiment was conducted twice in different regions. Among them, the non occluded area referred to the matching area after excluding the mismatched points in the occluded area. The results of the two mismatched rates obtained by different algorithms in the non occluded area are shown in Figure 10. As shown in Figure 10, the Census transform algorithm that introduces edge gradients has the lowest error matching rate on different images, with a minimum value of 25.1%. The minimum error matching rates of the other two algorithms are 28.3% and 27.4%, respectively. Meanwhile, its average mismatch rate in the four



(a) First test result

(b) Second test result

FIGURE 10. Mismatching obtained by different algorithms in non occlusive regions.



(a) First test result

(b) Second test result

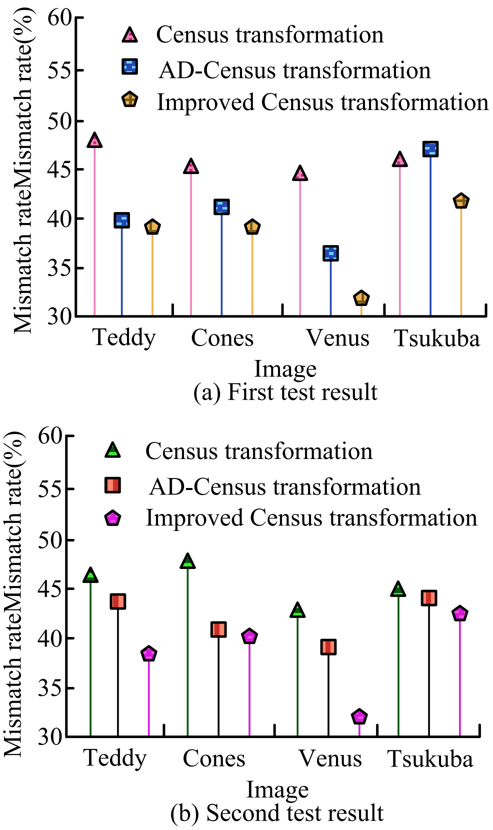
FIGURE 11. Mismatching obtained by different algorithms in all regions.

images is only 28.75%, while the classic Census transform algorithm is as high as 35.37%. The Census transform algorithm that introduced edge gradients had a high matching accuracy in non occluded regions.

The results of the three algorithms tested in all regions of the disparity map are shown in Figure 11. As shown in Figure 11, the Census transform algorithm that introduces edge gradients has the lowest error matching rate in all regions, and the lowest value is only 26.4%. The minimum values of the classic Census transformation method and the AD Census classical transformation algorithm are 29.2% and 28.3%, respectively, which are 2.8% and 1.9% higher than the proposed algorithm in the research. Meanwhile, its average error matching rate in the four images is only 32.5%, while the AD-Census classical transformation algorithm is as high as 37.32%. The Census transform algorithm, which introduces edge gradients, ha the lowest mismatch rate in all regions and the best performance.

The error matching rate results measured by different algorithms in the discontinuous areas of the disparity map are shown in Figure 12. From Figure 12, it can be seen that the Census transform algorithm, which introduces edge gradients, has the lowest error matching rate in the disparity discontinuous regions of Teddy, Cones, Venus, and Tsukuba images, and the lowest value is only 32.1%. The minimum error matching rates of the classic Census transformation method and the AD Census classical transformation algorithm are 45.3% and 36.6%, respectively. Meanwhile, its average mismatch rate in the four images is only 37.9%, while the other two algorithms are as high as 49.51% and 41.47%, respectively. In addition, from the test data of the three test areas, the Census transform algorithm introducing edge gradients had a smaller fluctuation range of mismatch rate, indicating better stability performance. The results indicated that the Census transform algorithm, which introduced edge gradients, had higher matching accuracy and optimal performance in discontinuous areas of disparity maps.

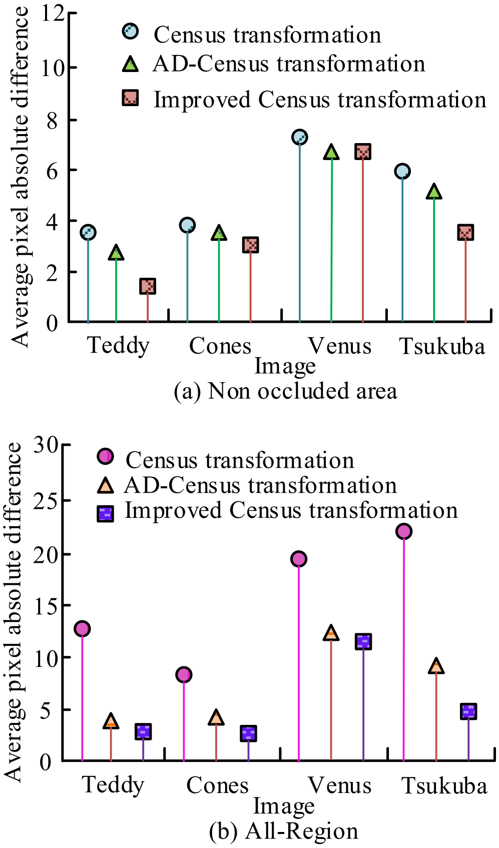
The study continued to validate the performance advantages of the improved algorithm by calculating the absolute difference of average pixels in different regions. The experiments were conducted on the entire and the non occluded region of the disparity map, and the average pixel absolute difference results of different algorithms on the four images are shown in Figure 13. From Figure 13, it can be seen that the Census transform algorithm, which introduces edge gradients, has the lowest average pixel absolute difference between the entire region and the non occluded region in all four images. Among them, the minimum value of this algorithm in non occluded areas is only 1.62, while the other two algorithms are 3.75 and 2.9, respectively. At the same time, the algorithm has a minimum average pixel



(a) First test result

(b) Second test result

FIGURE 12. Mismatching obtained in discontinuous regions of disparity maps.



(a) Non occluded area

(b) All-Region

FIGURE 13. Average pixel absolute difference results for different regions.

absolute difference of only 2.49 in the entire region of the disparity map, while the minimum value of the classic Census transform algorithm is as high as 8.83. The improved Census transform algorithm had high matching accuracy.

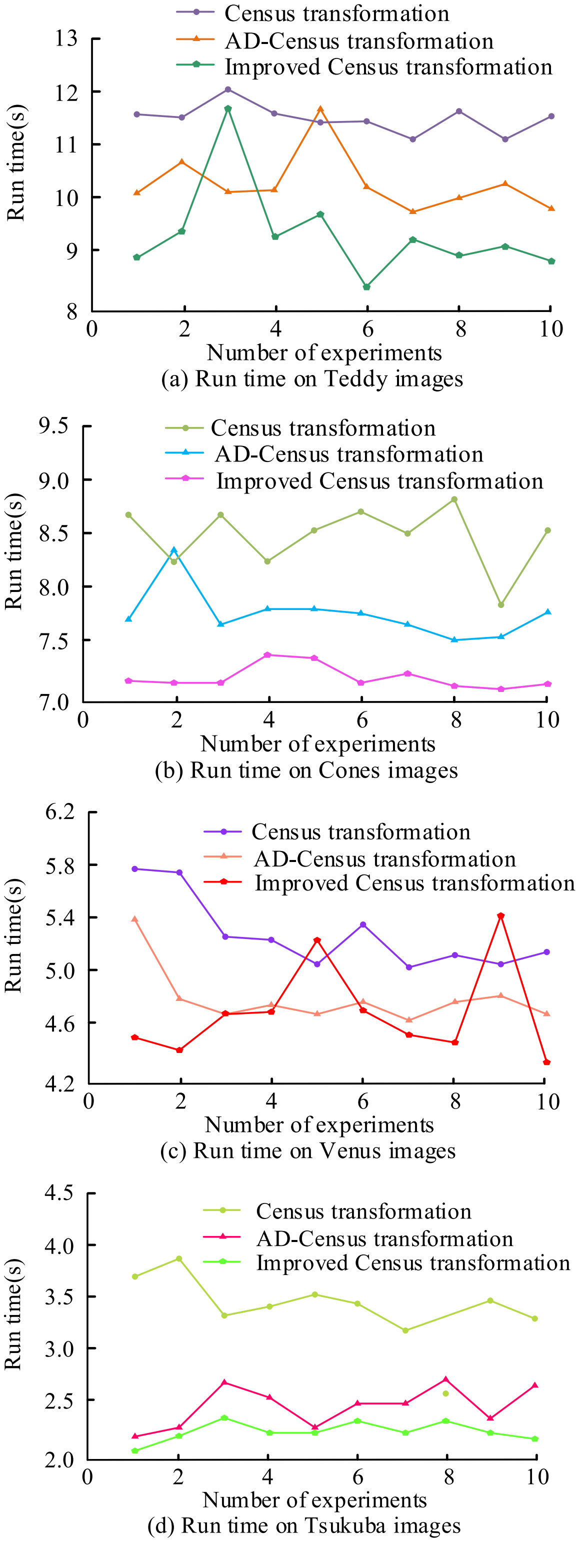
To further validate the performance advantages of the improved algorithm, the study conducted a calculation of the running time. The study also selected Census transform, AD-Census classical transform method, and improved Census transform algorithm with edge gradient for comparison of results. Meanwhile, the experimental images selected were Teddy, Cones, Venus, and Tsukuba images from the Middlebury dataset. To guarantee the experimental data was valid, a total of ten tests were conducted. The runtime results of different algorithms on four images are shown in Figure 14. As shown in Figure 14, the improved Census transform algorithm had the lowest runtime on all four images. Among them, the Census transform algorithm that introduced edge gradients had a minimum running time of only 8.2s on Teddy images, which was 2s and 1.6s lower than the other two algorithms. Meanwhile, the minimum running time on Tsukuba images was only 2.1s. The improved Census transformation algorithm had shorter runtime and better performance.

The experiment continued to process the data from ten measurements, selecting the average of the valid data as the final test result, as shown in Table 1. From Table 1, the Census transform algorithm, which introduced edge gradients, had the lowest average effective runtime on all four images. Among them, the effective average value of this algorithm on Venus images was 4.6s, which was 0.6s lower than the classic Census transform algorithm. Meanwhile, its effective average value on the Cones image was only 7.1s, which was 0.5s less than the AD-Census classical transformation method. The improved Census transform algorithm could not only guarantee the accuracy of matching, but also decreases the complexity of time, with significant performance advantages.

TABLE 1. Average effective running time of different algorithms(s).

Algorithm	Census transformation	AD-Census transformation	Improved Census transformation
Teddy	11.5	10.8	8.3
Cones	8.6	7.6	7.1
Venus	5.2	4.8	4.6
Tsukuba	3.4	2.5	2.2

The research continues to verify the matching accuracy of the Census transform algorithm that introduces edge gradients. The selected digital museum image dataset is the Metropolitan Museum of Art's open access image dataset. This dataset contains high-resolution images of over 40000 artworks collected by the museum, covering various fields from ancient artifacts to modern art. Extract four representative images as research objects, namely Ritterstrasse, Flowers, Smokers, and Farrier. The experiment uses advanced



(a) Run time on Teddy images

(b) Run time on Cones images

(c) Run time on Venus images

(d) Run time on Tsukuba images

FIGURE 14. Running time of different algorithms.

methods based on polar plane images, perspective translation, and traditional depth estimation algorithms to compare their performance. In the method based on polar plane image, polar plane image is taken as input directly, and Convolutional neural network is used to estimate parallax information. The method based on perspective translation mainly increases the Receptive field of the network to better capture the parallax information. The evaluation index is Root-mean-square deviation (MSE), which is mainly used to evaluate the matching accuracy of the matching algorithm. The MSE values of different algorithms are shown in the table 2. As can be seen from the table 2, the Census transform algorithm that introduces edge gradients has the lowest MSE value in all images. Among them, this method has the lowest MSE value in Flowers images, only 0.003, which is 1.506 lower than the method based on polar plane images. Meanwhile, its MSE value in the Ritterstrasse image is 3.025, a decrease of 0.556 compared to the perspective shift based method. This indicates that the proposed method has excellent matching performance in various complex scenarios.

TABLE 2. MSE values for different algorithms.

Algorithm	Ritterstrasse	Flowers	Smokers	Farrier
Polar plane images	3.702	1.509	0.958	0.453
Perspective translation	3.581	1.047	0.869	0.469
Depth estimation	4.251	2.871	1.247	0.872
Census of edge gradient	3.025	0.003	0.813	0.286

V. CONCLUSION

SM, as a key link in binocular vision, has critical effect on 3D reconstruction technology. To address the issue of unclear texture features in SM algorithms, a Census transform algorithm based on mean discrimination and Sobel edge detection was introduced into the calculation of the cost function. At the same time, the method of AD was introduced for optimization. The results showed that the sharpening effect of the book contour edge in the Venus image detected by the improved algorithm was the best compared to the desk lamp contour edge in the Tsukuba image. Moreover, the amount of mismatched points in this algorithm was significantly cut down compared to the classical Census transformation method [30] and the AD Census classical transformation [31]. At the same time, the Census transformation algorithm that introduced edge gradients had a minimum error matching rate of only 26.4% in all regions, which was reduced by 2.8% and 1.9% compared to the classical Census transformation method and the AD Census classical transformation algorithm, respectively. And the minimum value of this algorithm in non occluded areas was only 1.62, which was 2.13 and 1.28 lower than the other two algorithms, respectively. In addition, the Census transform algorithm that introduced edge gradients had a minimum running time of only 8.2s on Teddy images, which was 2s and 1.6s lower than the other two algorithms. This indicated that the algorithm had significant performance advantages and excellent practical application results. However, the test subjects in the study were all taken from the Middlebury platform, where the datasets were all processed images and did not have good representativeness. Therefore, future research can consider using datasets from complex real-world scenarios for testing to further confirm the performance of the algorithm.

REFERENCES

[1] M. S. Hamid, N. A. Manap, R. A. Hamzah, and A. F. Kadmin, "Stereo matching algorithm based on deep learning: A survey," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1663–1673, May 2022.

[2] M. Poggi, S. Kim, F. Tosi, S. Kim, F. Aleotti, D. Min, K. Sohn, and S. Mattoccia, "On the confidence of stereo matching in a deep-learning era: A quantitative evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5293–5313, Sep. 2022.

[3] A. Yang, C. Zhang, Y. Chen, Y. Zhuansun, and H. Liu, "Security and privacy of smart home systems based on the Internet of Things and stereo matching algorithms," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2521–2530, Apr. 2020.

[4] Z. Lu, J. Wang, Z. Li, S. Chen, and F. Wu, "A resource-efficient pipelined architecture for real-time semi-global stereo matching," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 2, pp. 660–673, Feb. 2022.

[5] W. Yin, Y. Hu, S. Feng, L. Huang, Q. Kemao, Q. Chen, and C. Zuo, "Single-shot 3D shape measurement using an end-to-end stereo matching network for speckle projection profilometry," *Opt. Exp.*, vol. 29, no. 9, pp. 13388–13407,2021.

[6] Y. Cui, Q. Li, B. Yang, W. Xiao, C. Chen, and Z. Dong, "Automatic 3-D reconstruction of indoor environment with mobile laser scanning point clouds," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3117–3130, Aug. 2019.

[7] Z. Chen, C. Liao, X. Cao, B. A. Poser, Z. Xu, W. Lo, M. Wen, J. Cho, Q. Tian, Y. Wang, Y. Feng, L. Xia, W. Chen, F. Liu, and B. Bilgic, "3D-EPI blip-up/down acquisition (BUDA) with CAIPI and joint Hankel structured low-rank reconstruction for rapid distortion-free high-resolution T_2^* mapping," *Magn. Reson. Med.*, vol. 89, no. 5, pp. 1961–1974, May 2023.

[8] Y. Cai, L. Li, D. Wang, and X. Liu, "MFNet: Multi-level fusion aware feature pyramid based multi-view stereo network for 3D reconstruction," *Appl. Int.*, vol. 53, no. 4, pp. 4289–4301, Feb. 2023.

[9] Q. Li, F. Wu, and G. Chen, "An efficient, fair, and robust image pricing mechanism for crowdsourced 3D reconstruction," *IEEE Trans. Services Comput.*, vol. 15, no. 1, pp. 498–512, Jan. 2022.

[10] S. Liu, H. Wu, Y. Huang, Y. Yang, and J. Jia, "Accelerated structure-aware sparse Bayesian learning for three-dimensional electrical impedance tomography," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5033–5041, Sep. 2019.

[11] S. Kench and S. J. Cooper, "Generating three-dimensional structures from a two-dimensional slice with generative adversarial network-based dimensionality expansion," *Nature Mach. Intell.*, vol. 3, no. 4, pp. 299–305, Apr. 2021.

[12] Y. Zhang, Y. Chen, X. Bai, S. Yu, K. Yu, Z. Li, and K. Yang, "Adaptive unimodal cost volume filtering for deep stereo matching," in *Proc. AAAI Conf. Artif. Int.*, 2020, vol. 34, no. 5, pp. 12926–12934.

[13] H. Wang, R. Fan, P. Cai, and M. Liu, "PVStereo: Pyramid voting module for end-to-end self-supervised stereo matching," *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 4353–4360, Jul. 2021.

[14] Z. Liang, Y. Guo, Y. Feng, W. Chen, L. Qiao, L. Zhou, J. Zhang, and H. Liu, "Stereo matching using multi-level cost volume and multi-scale feature constancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 300–315, Jan. 2021, doi: 10.1109/TPAMI.2019.2928550.

[15] T. Yan, Y. Gan, Z. Xia, and Q. Zhao, "Segment-based disparity refinement with occlusion handling for stereo matching," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 3885–3897, Aug. 2019.

[16] Y. Wang, M. Gu, Y. Zhu, G. Chen, Z. Xu, and Y. Guo, "Improvement of AD-census algorithm based on stereo vision," *Sensors*, vol. 22, no. 18, p. 6933, Sep. 2022.

[17] S. Joung, S. Kim, K. Park, and K. Sohn, "Unsupervised stereo matching using confidential correspondence consistency," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 5, pp. 2190–2203, May 2020.

[18] H. Liu, H. Zhang, X. Nie, W. He, D. Luo, G. Jiao, and W. Chen, "Stereo matching algorithm based on two-phase adaptive optimization of AD-census and gradient fusion," in *Proc. IEEE RCAR*, Xining, China, Jul. 2021, pp. 726–731.

[19] Y.-Z. Hsieh and S.-S. Lin, "Robotic arm assistance system based on simple stereo matching and Q-learning optimization," *IEEE Sensors J.*, vol. 20, no. 18, pp. 10945–10954, Sep. 2020.

[20] C. Won, J. Ryu, and J. Lim, "End-to-end learning for omnidirectional stereo matching with uncertainty prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 3850–3862, Nov. 2021.

[21] Q. Xie, X. Hu, L. Ren, L. Qi, and Z. Sun, "A binocular vision application in IoT: Realtime trustworthy road condition detection system in passable area," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 973–983, Jan. 2023.

[22] X. Wang, M. Cheng, J. Eaton, C.-J. Hsieh, and S. F. Wu, "Fake node attacks on graph convolutional networks," *J. Comput. Cognit. Eng.*, vol. 1, no. 4, pp. 165–173, Oct. 2022.

[23] M. Mahato, S. Gedam, J. Joglekar, and K. M. Buddhiraju, "Dense stereo matching based on multiobjective fitness function—A genetic algorithm optimization approach for stereo correspondence," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3341–3353, Jun. 2019.

[24] M. U. U. Haq, M. Ashfaque, S. Mathavan, K. Kamal, and A. Ahmed, "Stereo-based 3D reconstruction of potholes by a hybrid, dense matching scheme," *IEEE Sensors J.*, vol. 19, no. 10, pp. 3807–3817, May 2019.

[25] H. Zhang, Y. Yao, K. Xie, C. W. Fu, H. Zhang, and H. Huang, "Continuous aerial path planning for 3D urban scene reconstruction," *ACM Trans. Graph.*, vol. 40, no. 6, pp. 225:1–225:15, 2021.

[26] H. Zheng, L. Yao, and Z. Long, "Reconstruction of 3D images from human activity by a compound reconstruction model," *Cogn. Comput.*, vol. 14, no. 4, pp. 1509–1525, Jul. 2022.

[27] H. Zhang, S. Ma, M. Li, H. Jiang, and J. Li, "Recent reviews on machine vision-based 3D reconstruction," *Recent Patents Mech. Eng.*, vol. 15, no. 1, pp. 12–24, Feb. 2022.

[28] J. Zheng, Q. Yang, N. Makris, K. Huang, J. Liang, C. Ye, X. Yu, M. Tian, T. Ma, T. Mou, W. Guo, R. Kikinis, and Y. Gao, "Three-dimensional digital reconstruction of the cerebellar cortex: Lobule thickness, surface area measurements, and layer architecture," *Cerebellum*, vol. 22, no. 2, pp. 249–260, Mar. 2022.

[29] M. Metzger, Z. Újvári, and G. Gárdonyi, "Criminal application of photogrammetry: Three-dimensional reconstruction of crime scenes, human corpses and objects," *Belügyi Szemle*, vol. 68, no. 11, pp. 27–70,2020.

[30] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.*, vol. 29, pp. 736–746, 2020, doi: 10.1109/TIP.2019.2934572.

[31] H. Qiu, Q. Zheng, G. Memmi, J. Lu, M. Qiu, and B. Thuraisingham, "Deep residual learning-based enhanced JPEG compression in the Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 2124–2133, Mar. 2021, doi: 10.1109/TII.2020.2994743.

PENG PENG was born in Wuhan, Hubei, China, in 1979. She received the master's degree from the Wuhan University of Technology. She is currently pursuing the Ph.D. degree with the School of Design, Jiangnan University. Her research interests include the inheritance and regeneration of design heritage and urban space research.

JUN HAN was born in Jingmen, Hubei, China, in 1978. He received the master's degree from the Wuhan University of Technology, China. He is currently with the School of Art and Design, Wuhan Institute of Technology. His research interests include computer-aided design and industrial design.