IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 February 2023, accepted 20 March 2023, date of publication 23 March 2023, date of current version 29 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3260847

RESEARCH ARTICLE

Automatic Registration of Panoramic Image and Point Cloud Based on the Shape of the Overall Ground Object

BUYUN WANG[1], HONGWEI LI[1], SHAN ZHAO[1], LINQING HE[1], YULU QIN[2], AND XIAOYUE YANG[2]

[1]School of Geoscience and Technology, Zhengzhou University, Zhengzhou 450001, China
[2]School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China

Corresponding author: Shan Zhao (475283764@qq.com)

This work was supported in part by the "Key Technology of Intelligent Robot Space Perception" project of Key Program of National Natural Science Foundation of China under Grant 42130112, and in part by the "Research on Intelligent Identification and Extraction Method of Intelligent Urban Management Department/Event Based on Multi-Modal Data Fusion" project of Science and Technology Tackling Plan Program of Henan Province under Grant 222102320220.

ABSTRACT This paper presents a novel method for registering panoramic images and 3D point clouds using the shape of the overall ground object in the scene as registration primitives. Firstly, a semantic segmentation method is applied to the panoramic image to extract the ground object and remove the sky. Next, the cloth simulation filtering algorithm (CSF) is employed to eliminate the ground points in the 3D point cloud. The remaining 3D ground objects are then projected onto a two-dimensional plane using the imaging model of the panoramic camera to obtain the registration primitives. Finally, we adopt the whale algorithm to perform a coarse-to-fine registration, utilizing overlap degree and mutual information as the similarity measures. The proposed method is evaluated in four different scenes and compared with the other four registration methods. The results demonstrate that the proposed method is accurate and effective, with an average registration error of 11.48 pixels (image resolution is 11000 × 5500 pixels) compared to the EOPs of the system of 101.67 pixels.

INDEX TERMS Point cloud, panoramic image, semantic segmentation, registration.

I. INTRODUCTION

The mobile measurement systems (MMS) is a novel surveying and mapping system that integrates various sensors including LiDAR, panoramic camera, global satellite positioning system (GPS), and inertial navigation system (IMU) [1]. LiDAR is capable of acquiring point cloud of buildings on both sides of the street in the scene to obtain highly accurate position information, while the panoramic camera can capture high-resolution, large-angle panoramic image data of the same scene to obtain rich texture information. These two types of data complement each other in describing the scene, and their combination can provide the best results for 3D scene reconstruction in the digital city, infrastructure maintenance, and engineering planning applications [1], [2], [3], [4].

During the actual measurement process, the GPS signal can be obstructed by buildings or trees, resulting in localization errors [5]. Furthermore, panoramic images captured by multiple fisheye lenses may introduce errors during the stitching process [6]. These issues make it impossible to directly and accurately integrate the point cloud and the panoramic image. To achieve this, data registration is required, which involves calculating a transformation matrix that converts the two sets of data into the same coordinate system and eliminates any geometric inconsistencies. As point clouds and panoramic images are cross-modal data, existing registration methods mostly rely on the initial EOPs provided by GPS/IMU [7],

The associate editor coordinating the review of this manuscript and approving it for publication was Gerardo Di Martino.

which unfortunately results in poor registration accuracy. For the 2D-3D registration problem, scholars have proposed many solutions.

The existing registration methods can be divided into three categories [8]: Feature-based methods, regional statistics-based methods, and multi-view geometry-based registration methods.

The feature-based registration approach involves extracting homologous feature points from both images and point clouds, which are then used as registration primitives to establish a transformation model for matching. Such features typically include points (such as endpoints of road lamps and lanes [6], skyline points [9], [10], or key-points at line intersections [11], lines (such as street light poles [12], building boundary lines [13], roof edge [14], or intersection points of two planes [15]), planes (such as roof planes [16]), or hybrid features [17]. Although feature-based registration can achieve higher accuracy, it remains a challenge to automatically extract homonymous features from point clouds and images due to their significant differences.

The region-based registration method by converting the point cloud into an intensity or distance image, and then comparing the pixel similarity between the point cloud image and the optical image for registration. This comparison can be done using various metrics such as gradient information [18], phase correlation [19], mutual information [5], [20], [21], normalized joint mutual information [22], joint entropy [23], etc. Unlike feature-based registration methods, the region-based approach does not require feature extraction from the data and instead compares the data's correlation in the corresponding regions. This makes it robust to noise and grayscale differences in optical images, but it ignores the spatial location information that corresponds to grayscale features in the image. Moreover, this method is only suitable for urban scenes and not natural scenes [24].

The multi-view geometry registration method employs a two-step approach for precise registration. Initially, the optical image is reconstructed in 3D using either Structure from Motion (SfM) or multi-view stereo (MVS) techniques to generate an image reconstruction point cloud. Subsequently, the 3D-3D registration technique is employed with laser point cloud data, (such as ICP [2], 4PCS [25], etc.) to register the reconstructed image point cloud with the laser point cloud data. However, this method generally requires the system to provide more accurate initial parameters.

In summary, the registration of 3D laser point cloud and optical image can be divided into three parts: extraction of registration primitives, selection of registration model and parameter optimization. The accuracy of the selection of registration primitives is directly related to the accuracy of registration. Due to the discreteness of point cloud and the special imaging model of panoramic image, errors will inevitably occur when extracting registration primitives. How to extract and match the registration primitives robustly and automatically is a challenge for 2D-3D registration.

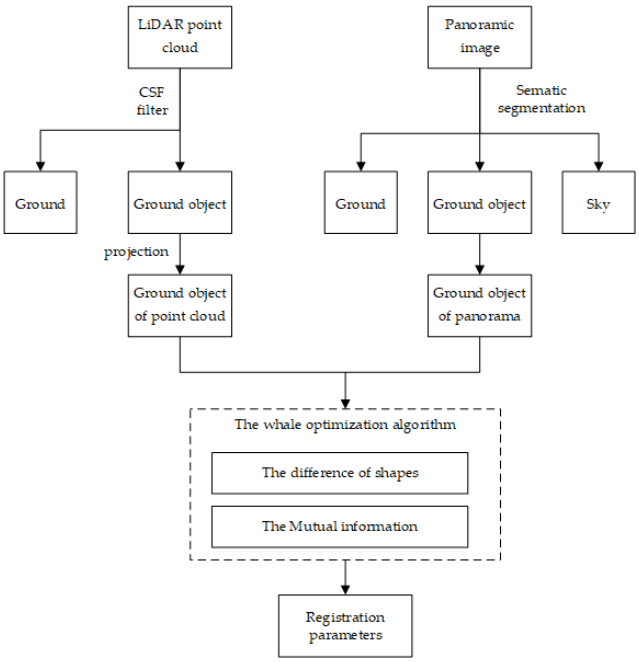

FIGURE 1. Flow chart of the proposed registration method.

Most current researches use traditional feature extraction methods to extract features from point clouds and optical images. In recent years, deep learning has developed rapidly and has played a significant role in the fields of feature extraction, object recognition and semantic segmentation. Since deep learning does not rely on prior knowledge to manually design features and parameters, new and effective feature representations can be quickly learned from training data for new applications, we propose to combine the deep learning with the registration of point cloud and optical image, and automatically extract the registration primitives to improve the automation and accuracy of registration. The overall registration method is shown in Figure 1.

The main contributions of this paper are as follows:

1) we utilize the shape of ground objects in the scene to perform registration of the laser point cloud and panoramic image. This approach solely relies on straightforward semantic segmentation of panoramic images, eliminating the need for extracting geometrical characteristics from point clouds and images.

2) Based on the mutual information method, we incorporate spatial location information into the matching process, which enhances the accuracy and efficiency of matching.

The rest of the paper is organized as follows: Section II describe the principle of the proposed method in detail. Section III conducts the experiment, then compare with other methods and conducts simulation experiments to discuss the characteristics of the method. Finally, Section IV summarizes the article.

II. PROPOSED METHOD

A. GROUND OBJECT SHAPE EXTRACTION FROM THE PANORAMIC IMAGE

Panoramic images often contain complex features that are distorted, posing a challenge for image segmentation. Traditional segmentation methods (such as the threshold and boundary detection techniques) rely on low-level semantic features of images (such as color and shape, etc.), which are insufficient to obtain optimal segmentation results in real-world scenarios. In contrast, deep learning methods are capable of automatically learning complex features and exhibit good generalization. In this study, we employ Deeplabv3+ [26] as the segmentation network to perform semantic segmentation of panoramic images. The network identifies the type of each pixel in the image and categorizes the image into three categories: sky, ground, or ground objects. It is worth noting that buildings, plants, vehicles, and other objects present in the scene all belong to the overall ground object category. Subsequently, we remove the sky and ground parts from the image and convert the remaining portion containing only ground objects into a binary image. Due to limitations in segmentation accuracy, we down-sample the panoramic image to reduce subsequent errors, yielding the shape of the ground objects in the panoramic image. The entire process is illustrated in Figure 2.

B. GROUND OBJECT SHAPE EXTRACTION FROM POINT CLOUD

Initially, statistical filtering is applied to the point cloud to eliminate noise points and outliers. Subsequently, the cloth simulation filtering algorithm (CSF) [27] is used to divide the point cloud into two parts: the ground object and the ground. Ground points adhered to ground objects are removed, and only the point cloud data corresponding to the ground object is preserved.

As point clouds and panoramic images belong to different dimensions, direct comparison between them is not feasible. Therefore, it is necessary to transform them into a common frame of reference. To achieve this, we utilize the imaging model of the panoramic image to transform the point cloud from the three-dimensional space (the LiDAR coordinate system) to a two-dimensional plane (the image coordinate system). Figure 3 illustrates the transformation from the LiDAR coordinate system to the image coordinate system.

Initially, the point cloud is transformed from the world coordinate system to the local coordinate system within the POS system. Subsequently, the point cloud is mapped onto the camera coordinate system by utilizing the initial EOPs of the system, which can be achieved through equation (1):

$$\begin{bmatrix} \bar{X} \\ \bar{Y} \\ \bar{Z} \end{bmatrix} = R\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T \right) \tag{1}$$

$$T = \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix} \tag{2}$$

$$\begin{aligned} R &= R_X R_Y R_Z \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_X & -\sin\theta_X \\ 0 & \sin\theta_X & \cos\theta_X \end{bmatrix} \begin{bmatrix} \cos\theta_Y & 0 & \sin\theta_Y \\ 0 & 1 & 0 \\ -\sin\theta_Y & 0 & \cos\theta_Y \end{bmatrix} \\ &\quad \times \begin{bmatrix} \cos\theta_Z & -\sin\theta_Z & 0 \\ \sin\theta_Z & \cos\theta_Z & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{3}$$

where, the $\left[\bar{X}, \bar{Y}, \bar{Z}\right]$ is the camera coordinate system, $[X, Y, Z]$ is the LiDAR coordinate system, T is the translation vector between the LiDAR coordinate system and the camera coordinate system, and R is the rotation matrix between two coordinate systems, T_X, T_Y, T_Z are the translation vectors in the X, Y, Z directions respectively, θ_X, θ_Y, θ_Z are the rotation angles about the X, Y, Z axes respectively.

Next, Equation (4) is used to convert the point cloud from the camera coordinate system to the spherical coordinate system, with the camera location serving as the center.

$$\begin{cases} \varphi = \sin^{-1} \dfrac{\bar{Z}}{\sqrt{r}} \\ \theta = \tan^{-1} \dfrac{\bar{Y}}{\bar{X}} \\ r = \bar{X}^2 + \bar{Y}^2 + \bar{Z}^2 \end{cases} \tag{4}$$

where, θ and φ represent the angle between the target point and the X-axis and Z-axis in the spherical coordinate system respectively.

Then, equation (5) is utilized to transfer the point cloud from the spherical coordinate system to the image coordinate system, where the panoramic image is located. As a result, the point cloud is converted from three-dimensional space to the corresponding two-dimensional panoramic image plane.

$$\begin{cases} v = \left(\dfrac{1}{2} - \dfrac{\varphi}{\pi} \right) H \\ u = \left(\dfrac{1}{2} - \dfrac{\theta}{2\pi} \right) W \end{cases} \tag{5}$$

where, H and W are the height and width of the panoramic image respectively, u and v are the image coordinates in the image.

As point clouds are composed of discrete points with relatively low resolution, they are susceptible to noise. To mitigate this, we down-sample the point cloud image and apply Gaussian filtering to smooth out the point cloud data. Next, we utilize the boundary tracking algorithm [28] for topological analysis of the point cloud image to automatically detect the boundary of ground objects from the binary image. We fill the closed polygon formed by the contour to obtain the shape of the ground object from the point cloud image. The whole process is shown in Figure 4.

C. REGISTRATION BASED ON GROUND OBJECT SHAPE MATCHING

The overall shape of ground objects extracted from the panoramic image is denoted as ζ^{img}, while the shape of the ground objects extracted from the point cloud data of the

FIGURE 2. Extraction of ground objects from the panoramic image.

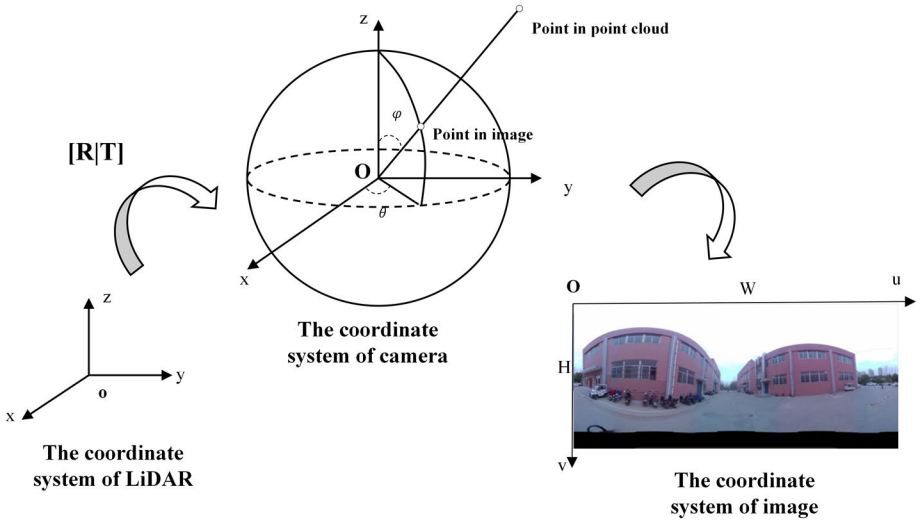

FIGURE 3. The transformation of coordinate systems.

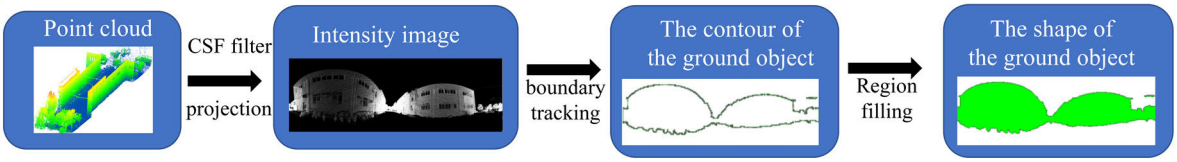

FIGURE 4. Extraction of ground objects from the point cloud.

corresponding region is represented as ζ^{pc}. The objective is to minimize the difference between ζ^{img} and ζ^{pc} by finding the registration parameters between the LiDAR and camera coordinate systems, thus achieving spatial alignment of the two types of data.

1) GROUND OBJECT SHAPE DIFFERENCE

The alignment of the point cloud and the panoramic image is achieved by minimizing the difference area between ζ^{img} and ζ^{pc}. To obtain the shape of the ground objects ζ^{pc} in the point cloud, the point cloud is projected and converted into a binary image using the method in 2.3. On the other hand, the panoramic image is subjected to semantic segmentation, and the pixel values of the ground object are set to 255 while those of the sky and ground are set to 0, thus obtaining the ground object shape ζ^{img}. Subsequently, the difference operation between ζ^{img} and ζ^{pc} is performed, where if the pixel values of ζ^{img} and ζ^{pc} at (u, v) are non-zero, the pixel value at (u, v) is set to 0. On the other hand, if ζ^{img} has a non-zero pixel at (u, v) and the pixel value of ζ^{pc} is 0, the pixel is filled with a value of 255, and vice versa.

Consequently, the dissimilarity between the shapes of the overall ground objects is computed, yielding the difference image S. The alignment degree of the point cloud and panoramic image can be quantified by the ratio ε_1 between the number of non-zero pixels $Pixel_s$ in S and the number of non-zero pixels $Pixel_{\zeta^{pc}}$ in ζ^{pc}. A smaller value of ε_1 indicates a higher degree of matching between the point cloud and panoramic image. The whole process is shown in Figure 5.

$$\varepsilon_1 = \frac{Pixel_s}{Pixel_{\zeta^{pc}}} \times 100\% \tag{6}$$

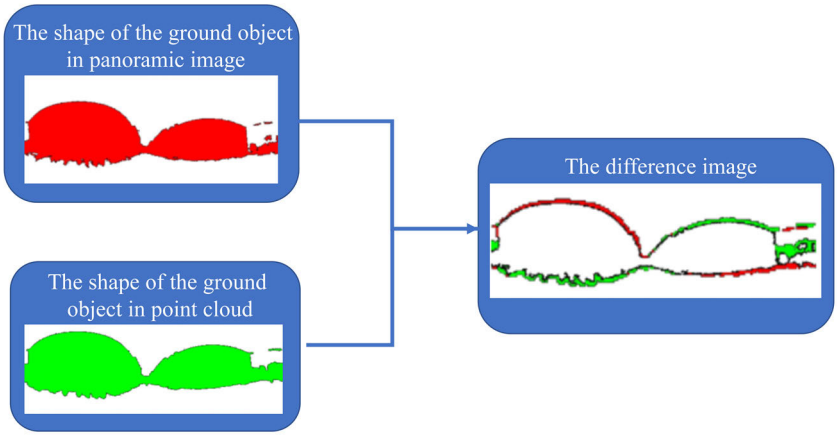

FIGURE 5. Ground object difference.

2) MUTUAL INFORMATION ENTROPY MATCHING OF GROUND OBJECT SHAPE

In the field of information theory, entropy is commonly used to quantify the degree of uncertainty of a random variable. Specifically, larger entropy values correspond to greater levels of uncertainty. The joint entropy of two random variables, as derived from their joint distribution, represents the uncertainty that arises when both variables are observed simultaneously. Meanwhile, mutual information serves as a measure of the extent to which one random variable contains information about another, and it can be used to describe the statistical correlation between two random variables. A higher mutual information value (MI) indicates a stronger correlation between the two variables [20].

We regard the reflection intensity of the LiDAR point and the gray value of the corresponding panoramic image pixel point as two random variables. The part of the ground object in the image is used to replace the whole image to participate in the calculation of mutual information.

For the panoramic image, we perform semantic segmentation on the panoramic image to extract the overall ground objects, which are then converted into grayscale images with values of 0-255 to obtain the ground object representation I^{pano}; for point cloud, we normalize the intensity values of the point cloud data to the same range, and then project it onto a blank image using equation (1-5), where the intensity value of each pixel is taken as the corresponding grayscale value to generate the point cloud feature image I^{pc}.

The probability of I^{pc} and I^{pano} is defined as H^{pc} and H^{pano} respectively, and their joint probability is defined as H^{joint}:

$$H^{pc} = -\sum p_i^{pc} \log\left(p_i^{pc}\right) \tag{7}$$

$$H^{pano} = -\sum p_i^{pano} \log\left(p_i^{pano}\right) \tag{8}$$

$$H^{joint} = -\sum\sum i_i^{\frac{pano}{pc}} \log\left(p_i^{\frac{pano}{pc}}\right) \tag{9}$$

where, p_i^{pc} and p_i^{pano} are respectively the frequency of pixel i in point cloud image and panoramic image. $p_i^{pano/pc}$ is obtained by counting the position of the pixel i in I^{pc} and calculating the frequency of the pixel value in the same position in I^{pano}. Then, we can calculate the mutual information entropy M between the I^{pc} and the I^{pano}:Based on these values, the mutual information entropy M between I^{pc} and I^{pano} can be calculated.

$$M = H^{pc} + H^{pano} - H^{joint} \tag{10}$$

The greater the value of M, the greater the correlation between the panoramic image and the intensity image. Therefore, we can use f to describe the degree of matching between the point cloud and the panoramic image, where a smaller f value indicates a higher degree of matching.

$$f = \varepsilon_1 - \varepsilon_2 \tag{11}$$

where $\varepsilon_2 = \alpha M$, α is the weight coefficient, which is set to 5 after many trials.

3) MATCHING BASED ON THE WHALE OPTIMIZATION ALGORITHM

The primary objective of the cost function is to minimize the shape matching coefficient ε_1 and the shape mutual information entropy ε_2 to correct the translation vector $T(T_X, T_Y, T_Z)$ and the rotation angle $\theta(\theta_X, \theta_Y, \theta_Z)$. However, solving this problem using numerical methods like Newton's method or gradient descent may lead to local optima due to the non-convexity of the optimization problem. To tackle this, we utilize the whale optimization algorithm (WOA) [29] to find a nearly optimal set of orientation parameters.

WOA is a meta-heuristic optimization algorithm that aims to find global optimal solutions. Its fundamental concept is to simulate the hunting behavior of humpback whales, the search range of whales is the global solution space. While hunting, each whale updates its position using one of two behaviors selected at random with a 50% probability:

(1) Searching, which involves moving towards other whales to surround the prey, or (2) Hunting, which entails spewing a spiral bubble-net to encircle and capture the prey. These behaviors are mathematically represented by equations (12) and (13) respectively:

$$x_i^{t+1} = \begin{cases} x_{best}^t - A\left|Cx_{best}^t - x_i^t\right|, & |A| < 1 \\ x_{rand}^t - A\left|Cx_{rand}^t - x_i^t\right|, & |A| \geq 1 \end{cases} \tag{12}$$

Equation (12) describes two movements in the WOA. The first movement involves moving towards the optimal position represented by x_{best}^t, while the second movement entails moving towards a randomly selected position, represented by x_{rand}^t. Here, x_i^t and x_i^{t+1} denote the current and updated positions of the i^{th} whale, respectively. Additionally, C is a random number within the interval [0 2], and A is a random number within the range [−a, a].

The mathematical model simulates the process of a whale searching and approaching its prey by reducing the value of A, which decreases from 2 to 0 with an increase in the number of iterations, resulting in a reduced range of A. When |A>1|, the individual updates its positions based on the positions of randomly selected whales, forcing the whale to deviate from its current prey and search for more suitable prey, Conversely, when |A<1|, the individual moves in the direction of the current optimal value to update its position, narrowing the search range for a more precise search.

$$x_i^{t+1} = \left|x_{best}^t - x_i^t\right| * e^{bl} * \cos(2\pi l) + x_{best}^t \tag{13}$$

where b is a constant 1, l is a random number uniformly distributed in [−1,1], x_{best}^t represents the position of the individual with the best value currently, in this case, the whale group moves along a spiral path.

Based on the WOA, we take $|A| = 1$ as the dividing point, and divide the process of searching for the optimal solution into two stages, corresponding to the rough registration and the fine registration:

Stage 1: During this stage, the value of |A| is randomly generated from the range (1,2) and the search range is set to be around the initial registration parameter. The shape matching coefficient ε_1 is used as the cost function for this stage. Multiple whales are used to form the initial whale group, with each whale calculating its fitness ε_1 at the beginning of the algorithm, and the minimum value in the current group is recorded as the global optimal value x_{best}^t. Subsequently, each individual randomly selects whether to search or hunt based on x_{best}^t, leading to continuous updates of their respective positions and the optimal value of the population x_{best}^t.

Stage 2: During this stage, the value of |A| is randomly generated from the interval (0, a), where a linearly decreases from 1 to 0 with an increase in the number of iterations. Following this, we employ the parameter associated with the optimal value of group x_{best}^t from stage 1 as the search center. Equation (11) serves as the cost function for successive iterative computations.

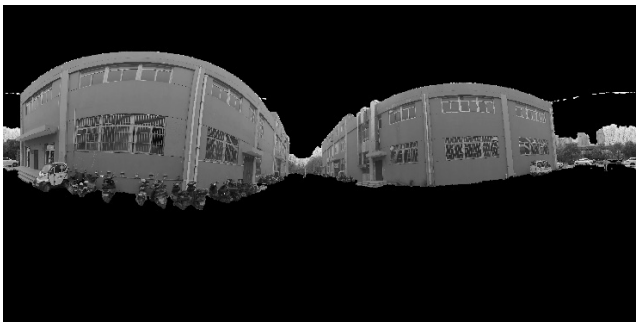

FIGURE 6. Ground object in panoramic image.

FIGURE 7. Ground object in intensity image.

Algorithm 1 Automated Registration By Shape Matching

Input: Point Cloud Shape ζ^{pc}, Panoramic Image Shape ζ^{pano}, Reflectivity, Initial Extrinsic Parameter $\Theta_i(\mathrm{R}|\mathrm{T})$

Output: Estimated Extrinsic Parameter $\Theta(\mathrm{R}|\mathrm{T})$;

```
Initialize: maxiter, x_best, y_best;
for i = 1, 2, ..., (maxiter/2) do
    Calculate The Shape Matching Coefficient ε(ζ^pc, ζ^pano|Θ_i)
    if ε < y_best then
        y_best = ε
        x_best = Θ_i
    end if
end for
for i = (maxiter/2), ..., maxiter do
    Calculate The Shape Matching Coefficient ε(ζ^pc, ζ^pano|Θ_i)
    Calculate Mutual Information MI(ζ^pc, ζ^pano|Θ_i)
    Calculate f(ζ^pc, ζ^pano|Θ_i) = ε - MI*5
    if f < y_best then
        y_best = f
        Estimated Extrinsic Parameter = Θ
    end if
end for
```

FIGURE 8. Automatic registration method based on the shape matching.

Where, stage 1 is to maximize the overlap of the shape of the ground objects in the point cloud image and the panoramic image, which enables a rough alignment of the two; Stage 2 is to maximize mutual information of ground objects based on stage 1, so as to achieve more accurate matching. Ultimately, the two-stage WOA search is utilized to determine a set of orientation parameters $T(T_X, T_Y, T_Z)$ and $\theta(\theta_X, \theta_Y, \theta_Z)$ that enable the best approximate match between point cloud and

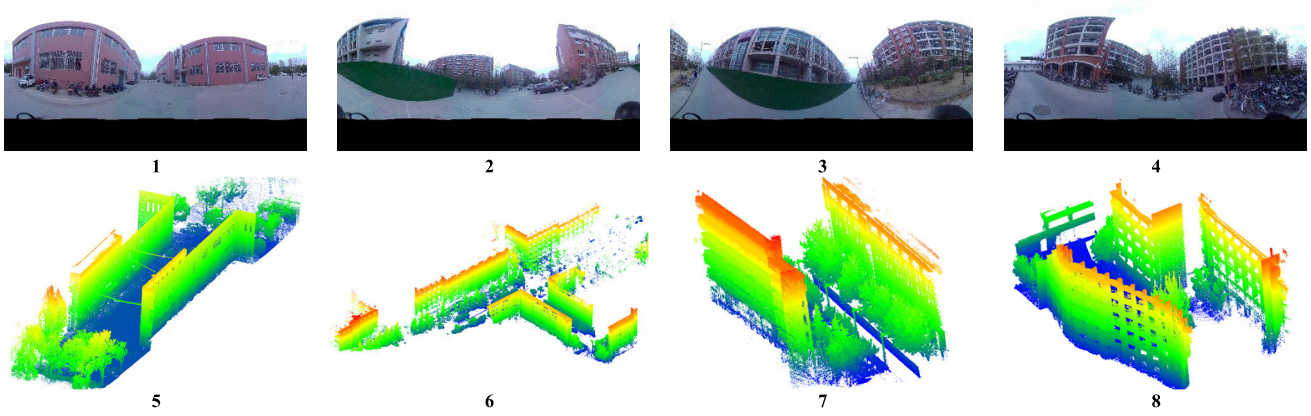

FIGURE 9. Panoramic images and point clouds used in the experiment. The 1-4 are panoramic images and the 5-8 are point clouds in the corresponding areas.

TABLE 1. Automatic registration method based on the shape matching.

	Number of points	Number of panoramas
Scene I	6,789,923	31
Scene II	7,880,755	23
Scene III	5,685,755	27
Scene IV	8,455,287	19

panoramic image objects. The complete method is shown in Figure 8.

III. EXPERIMENTS AND RESULTS

A. DATA PREPARATION

The experimental data was acquired primarily using the GeoSLAM backpack-type mobile measurement device ZEB Discovery, equipped with a Velodyne VLP-16 LiDAR and a panoramic camera consisting of four fisheye lenses. The lidar has a scanning range of 100m and a field of view of 360°x 270°, generating high-density point clouds with an average minimum spacing of 0.03m. The internal parameters of the panoramic camera have been calibrated, and the captured panoramic images have a resolution of 5500 × 11000. The equipment was used to collect 3D point cloud and panoramic images in the main campus of Zhengzhou University, from which we selected data in four different scenarios for the experiment. Figure 9 depicts the panoramic image data and laser point cloud data used in the experiment, and Table 1 provides a detailed description of the dataset.

B. RESULT OF EXPERIMENT

We performed semantic annotation on 100 panoramic images with a high resolution of 5500 × 11000 pixels in our dataset and expanded the dataset through various data augmentation techniques such as rotation, flipping, and cropping. Ultimately, we obtained 5600 images with a lower resolution of 1375 × 655, which we used as our training dataset. To train our model, we utilized the deeplabv3+ network for semantic segmentation on the panoramic images across four different scenes to obtain the overall ground object present in each panoramic image.

By equation (1-5), the overall ground objects of the point cloud are projected as the image, then use the whale algorithm for optimization. The parameters involved in our method include: iteration number, population number, angle error range and distance error range. After several experiments, we set the total number of iterations to 100 and the number of populations to 60, and then took the initial parameters provided by GPS/IMU system as the starting point for search, and set the search range of distance and angle to ±0.1 m and ±2° respectively.

The registration quality can be evaluated visually by projecting the point cloud into the image space using the registration parameters and overlaying the panoramic image. This comparison is illustrated in Figure 9 and Figure 10, which demonstrate the differences between the images before and after the registration process.

Figure 10 and Figure 11 reveal significant discrepancies in the initial state, indicating errors in the initial EOPs of the MMS. The coarse registration procedure achieves rough alignment of the point cloud and panoramic images, albeit with some residual errors that maximize the overlap of the ground object shapes. Subsequently, fine registration enables optimal optimization, yielding point cloud and panoramic image alignment with high accuracy across all four scenes.

To quantitatively evaluate the method, we selected 15 groups of corresponding checkpoints from panoramic images and point clouds. The checkpoints were chosen as prominent corners or inflection points that were evenly distributed throughout the scene, as illustrated in Figure 12. Subsequently, we projected the 3D checkpoints from the point cloud onto the panoramic image, and calculated the pixel offset between them and the image checkpoints as a quantitative measure of the accuracy of the algorithm [6], [8],

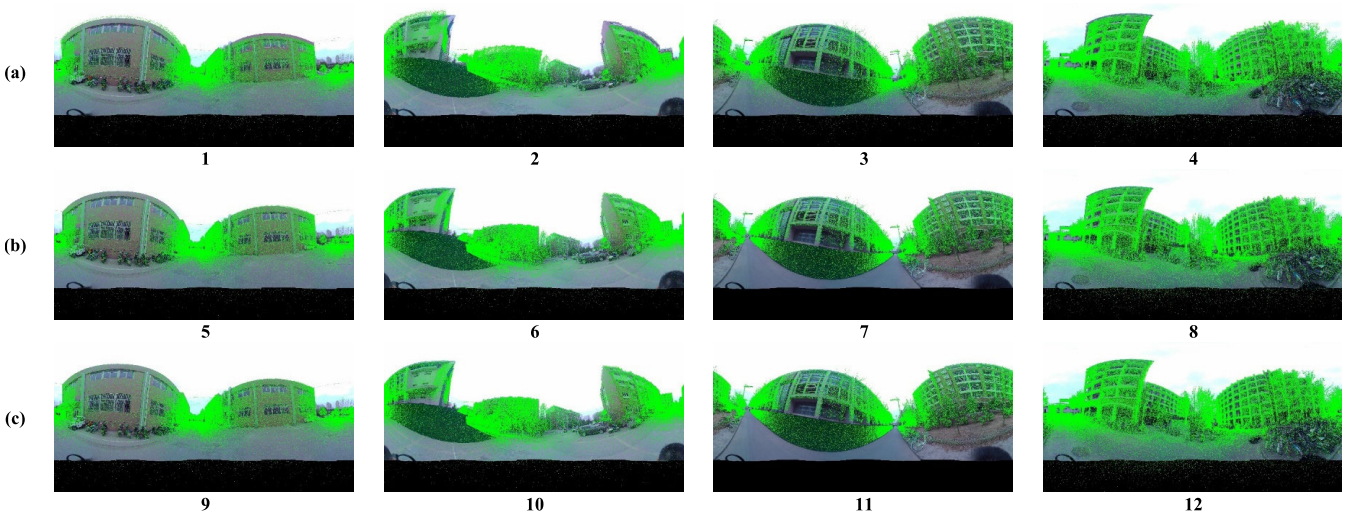

FIGURE 10. Comparison of registration results. (a) Results of original EOPs. (b) Results of the rough registration. (c) Results of the fine registration.

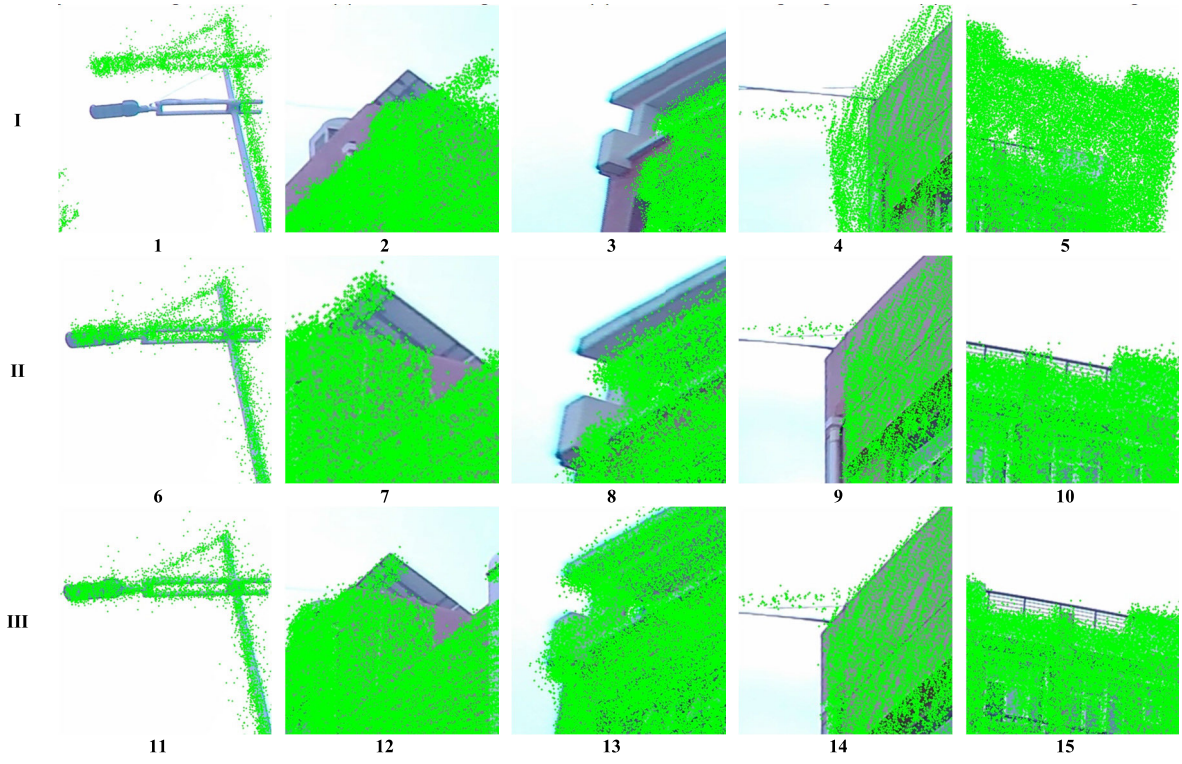

FIGURE 11. Local enlargement of registration results. From top to bottom are (I) the results of the initial EOPs, (II) rough registration and (III) fine registration respectively.

[10], [29]. Figure 13 shows the statistics of registration errors in four different scenes.

As depicted in Figure 13, the effect of coarse-to-fine registration is evident as the error reduces in a progressive manner. Specifically, the average error of the four scenes after registration drops from an initial value of 101.67 pixels to 11.48 pixels, indicating the efficacy of the registration approach in enhancing the accuracy of the system's initial EOPs.

C. COMPARISON WITH OTHER METHOD

To verify the accuracy of the method, we compare our method with the method based on area overlap maximization, the method based on mutual information maximization [5], [20], [21], the skyline-based matching method [10], and the control point based method [30] (referred to as methods I, II, III and IV, respectively). Method I takes the overlap degree of ground object area as the similarity measure to optimize.

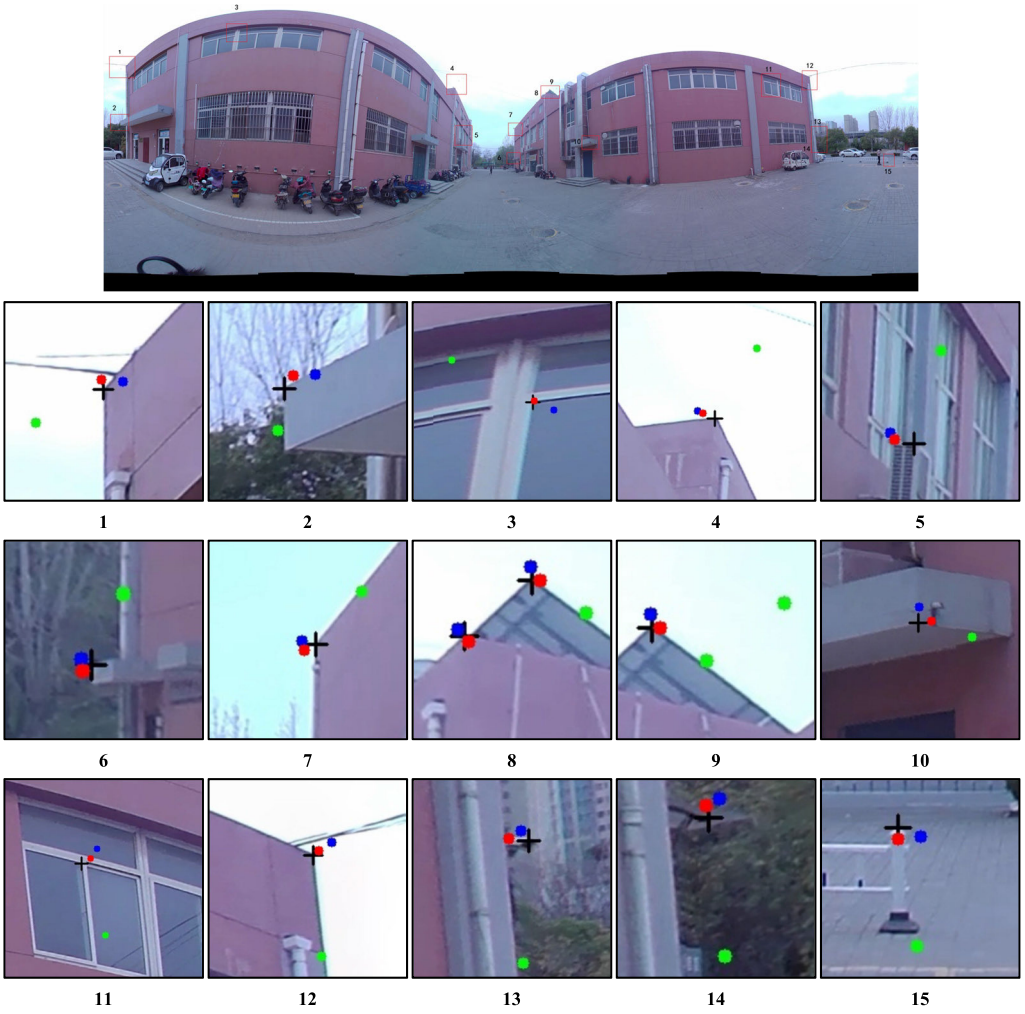

FIGURE 12. Checkpoints in scene I. The black cross mark is the location of the real checkpoint in the panoramic image, the green point is the projection result of the initial EOPs of the system, the blue is the rough registration result, and the red is the fine registration result.

Method II uses the mutual information between the intensity image of the point cloud and the gray scale image of the image as the similarity measure, and then performs the optimization. In Method III, 2D skyline pixels and 3D skyline points were extracted from the image and point cloud respectively, and then 3D points were projected into 2D points by coordinate transformation. Finally, the matching number between 2D skyline points was used as the cost function, and then the optimization was carried out. Method IV is to manually select several pairs of corresponding control points in the image and the Point cloud respectively, and then use these corresponding 3D-2D pairs of points to solve the relative conversion between camera and lidar by the UPnP (Unified Perspective-n-Point) method [30].

Since our method and Method I, II, III are both based on iterative optimization, for the convenience of comparison, we both choose the whale algorithm for optimization. We set the same overall size and maximum number of iterations for comparison experiments. To better compare the two registration methods, we calculated the checkpoint projection errors and the time required for single calculation of the five methods in all scenarios. Table 2 and Table 3 show the registration errors and operational efficiency of each method.

As shown in Table 2, method IV has the highest accuracy among all methods, with an average error of 11.10 pixels. However, the method based on control point requires manual selection of corresponding 2D and 3D control points, which has a low level of automation and is difficult to be effectively applied in real large scenes. The average error of our method is 11.48 pixels, which is better than methods I, II and III, indicating that our method can achieve automatic extraction of registration primitives on the premise of ensuring high registration accuracy.

As for the comparison of computational efficiency, as shown in Table 3, Method III takes the shortest time to calculate. However, Method III relies on skyline in the scene for registration. In some cases, skyline in the image

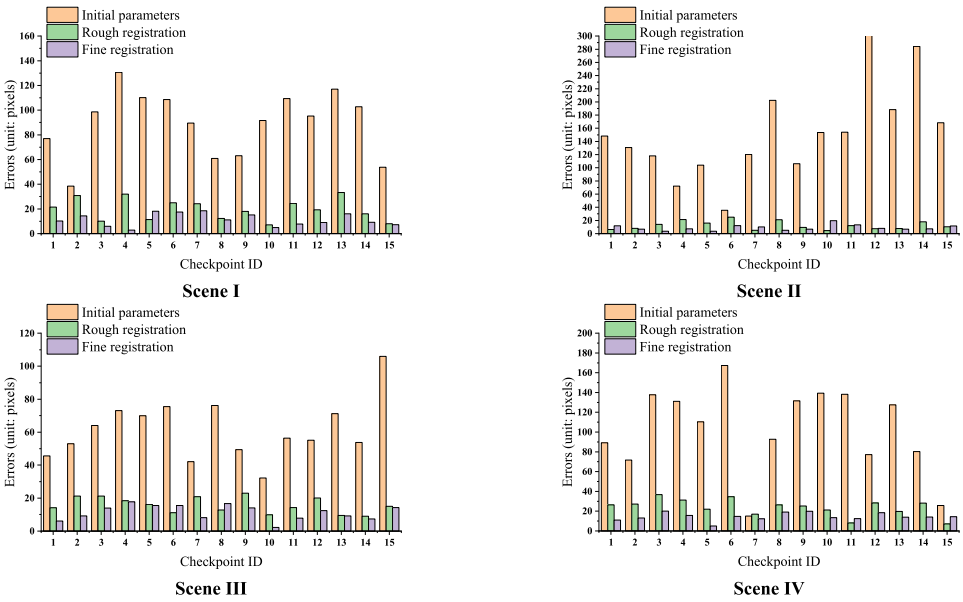

FIGURE 13. The registration error in the registration process of four different scenes. The orange, green and purple columns correspond to the checkpoints error of initial parameters, rough registration and fine registration respectively.

TABLE 2. The registration errors of different methods.

		Our Method	Method I	Method II	Method III	Method IV
Checkpoint projection error	Scene I	11.21	21.26	31.27	38.11	10.10
	Scene II	8.94	15.53	10.47	18.10	6.69
	Scene III	11.32	43.32	19.84	27.39	7.85
	Scene IV	14.46	33.11	25.26	29.53	19.75
	Average	**11.48**	**28.30**	**21.71**	**28.28**	**11.10**

records distant objects, while lidar is difficult to collect distant objects, so skyline cannot be well matched; Method II can achieve relatively good registration accuracy, but it requires all point clouds and the whole panoramic image to participate in the mutual information calculation, which consumes a lot of time. Meanwhile, our method combines the shape of ground objects and mutual information as the constraint condition. Since the background in the data (i.e. the sky and ground pixels in the panoramic image and ground points in the point cloud) are eliminated in advance, the calculation time is significantly faster than Method II, which indicates that our method can still achieve good computational efficiency while maintaining a relatively high precision and degree of automation. It shows that this method is accurate and effective.

D. THE INFLUENCE OF INCOMPLETE DATA

Our method mainly registers point cloud and panoramic image by matching the shape of the overall ground object. However, the semantic segmentation of the image and the limitation of the sensor may lead to the incomplete acquisition of the ground object, which may affect the registration results. Therefore, in order to verify the robustness of our method, we use the method in [8] to add noise points with different radii and intervals into the panoramic image segmentation results, so as to simulate the over-segmentation and under-segmentation of the image. Similarly, we add noise points in the process of point cloud intensity image generation to simulate possible holes and incompleteness in point cloud data.

As shown in Figure 14, we simulated three different levels of noise points and applied them to our method. Table 4 shows the average error of 15 groups of checkpoints under the influence of different degrees of simulated noise points and the deviation between them and the original results. The findings show that the addition of noise has a greater impact on the results of the rough registration stage than on the fine registration stage. This is attributed to the fact that the rough

TABLE 3. The operational efficiency of different methods.

		Our Method	Method I	Method II	Method III
Time of each registration	Scene I	13.43	8.53	23.47	8.33
	Scene II	17.64	12.27	28.67	13.28
	Scene III	18.57	13.13	29.33	8.43
	Scene IV	14.00	9.15	21.75	11.63
	Average	**15.91**	**10.77**	**25.81**	**10.41**

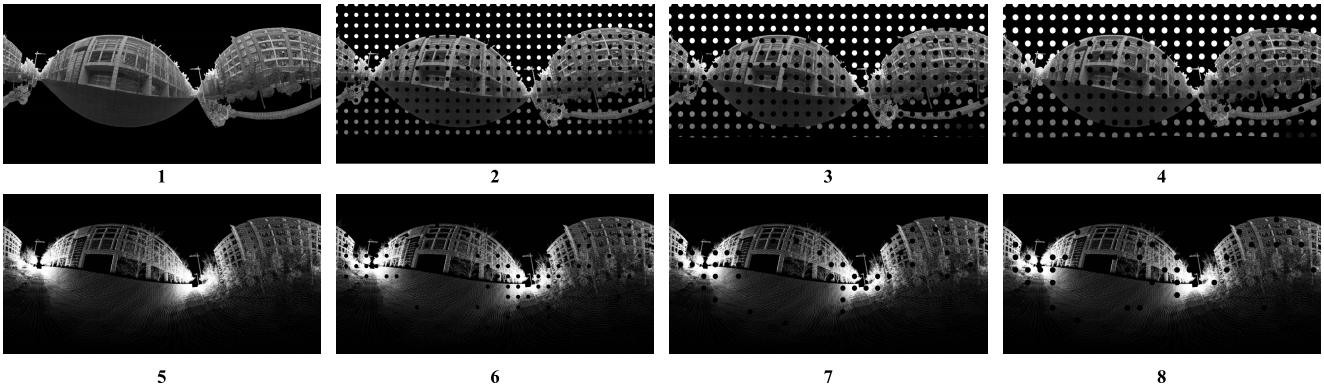

FIGURE 14. Data incomplete simulation. The top line and the bottom line respectively represent the incomplete data of the image and point cloud simulated by adding noise points. (1) and (5) are the initial results. (2) and (6) the radius of noise points is 70 pixels, and the interval is 350 pixels; (3) and (7) the radius of noise points is 90 pixels, and the interval is 400 pixels; (4) and (8) Noise points with a radius of 100 pixels and an interval of 430 pixels.

TABLE 4. The average error under different noise interference. (Unit: pixel).

	Panoramic image with noise				Point cloud with noise			
	Rough registration		Fine registration		Rough registration		Fine registration	
	Average error	Δ	Average error	Δ	Average error	Δ	Average error	Δ
original	15.77		11.32		15.77		11.32	
Noise I	20.59	4.82	14	2.68	24.00	8.23	15.06	3.74
Noise II	34.58	18.81	17.46	6.14	30.89	15.12	18.88	7.56
Noise III	25.78	10.01	17.54	6.22	20.34	4.57	11.55	0.23

* **Δ** is the deviation between the result of adding noise and the original segmentation.

registration stage heavily relies on the area difference of the overall ground object shapes, thus image segmentation errors have a greater impact. Conversely, the fine registration stage leverages the mutual information between the point cloud image and the panoramic image, and the similarity of the background sky region in both images partially offsets the impact of segmentation errors [21]. Therefore, although noise points can influence the cost function calculation, our method can still achieve good registration accuracy. In summary, our proposed method maintains a good registration effect under

various noise point interferences and demonstrates robustness against data incompleteness caused by sensor limitations or false segmentation.

IV. CONCLUSION

This paper proposes a novel automatic registration method to address the registration problem between LiDAR point clouds and panoramic images. Unlike traditional methods that extract geometric features, our method uses the overlap of the overall ground object shapes and the mutual information between the two types of data to estimate the rigid body transformation between the panoramic camera and LiDAR. We conducted experiments in four different scenes and compared our approach to the other four. Our method achieved an average error of 11.48 pixels on panoramic images with a resolution of 11000×5500 pixels, resulting in a 0.73% (error/image diagonal line) improvement in accuracy compared to the initial EOPs. Additionally, simulation experiments were conducted to test the method's robustness against different levels of noise point interference, and the results showed that our method has excellent robustness. Finally, simulation experiments are carried out for different noise point interference, the results show that the method has good robustness.

Despite the promising results of our proposed method, some limitations must be acknowledged. Firstly, as our approach utilizes the overall ground object shape as the registration primitive, it may face difficulties in scenarios where there are fewer objects or smaller targets. Additionally, our method is currently optimized for low-resolution images, and its performance for high-resolution images, such as aerial or terrestrial frame camera images, remains an area for further exploration. Addressing these challenges and expanding our method's capabilities in more complex scenarios will be crucial directions for our future research.

REFERENCES

[1] I. Puente, H. González-Jorge, J. Martínez-Sánchez, and P. Arias, "Review of mobile mapping and surveying technologies," *Measurement*, vol. 46, no. 7, pp. 2127–2145, Aug. 2013, doi: 10.1016/j.measurement.2013.03.006.

[2] B. O. Abayowa, A. Yilmaz, and R. C. Hardie, "Automatic registration of optical aerial imagery to a LiDAR point cloud for generation of city models," *ISPRS J. Photogramm. Remote Sens.*, vol. 106, pp. 68–81, Aug. 2015, doi: 10.1016/j.isprsjprs.2015.05.006.

[3] M. Elhashash, H. Albanwan, and R. Qin, "A review of mobile mapping systems: From sensors to applications," *Sensors*, vol. 22, no. 11, Jun. 2022, Art. no. 4262, doi: 10.3390/s22114262.

[4] M. Soilan, B. Riveiro, J. Martinez-Sanchez, and P. Arias, "Automatic road sign inventory using mobile mapping systems," in *Proc. 23rd Congr. Int. Soc. Photogramm. Remote Sens. (ISPRS)*, vol. 41, Prague, Czech Republic, 2016, pp. 717–723, doi: 10.5194/isprsarchives-XLI-B3-717-2016.

[5] M. Miled, B. Soheilian, E. Habets, and B. Vallet, "Hybrid online mobile laser scanner calibration through image alignment by mutual information," in *Proc. 23rd ISPRS Congr.*, vol. 3, Prague, Czech Republic, 2016, pp. 25–31, doi: 10.5194/isprsannals-III-1-25-2016.

[6] N. Zhu, Y. Jia, and X. Huang, "Semiautomatically register MMS LiDAR points and panoramic image sequence using road lamp and lane," *Photogram. Eng. Remote Sens.*, vol. 85, no. 11, pp. 829–840, Nov. 2019, doi: 10.14358/pers.85.11.829.

[7] N. Zhu, B. Yang, and Y. Jia, "Registration of MMS LiDAR points and panoramic image sequence using relative orientation model," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLIII-B1, pp. 291–298, Aug. 2020.

[8] J. Li, B. Yang, C. Chen, R. Huang, Z. Dong, and W. Xiao, "Automatic registration of panoramic image sequence and mobile laser scanning data using semantic features," *ISPRS J. Photogramm. Remote Sens.*, vol. 136, pp. 41–57, Feb. 2018, doi: 10.1016/j.isprsjprs.2017.12.005.

[9] S. Hofmann, D. Eggert, and C. Brenner, "Skyline matching based camera orientation from images and mobile mapping point clouds," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. II-5, pp. 181–188, May 2014.

[10] N. Zhu, Y. Jia, and S. Ji, "Registration of panoramic/fish-eye image sequence and LiDAR points using skyline features," *Sensors*, vol. 18, no. 5, p. 1651, May 2018, doi: 10.3390/s18051651.

[11] J. W. Zhu, Y. S. Xu, Z. Ye, L. Hoegner, and U. Stilla, "Fusion of urban 3D point clouds with thermal attributes using MLS data and TIR image sequences," *Infr. Phys. Technol.*, vol. 113, Mar. 2021, Art. no. 103622, doi: 10.1016/j.infrared.2020.103622.

[12] T. Cui, S. Ji, J. Shan, J. Gong, and K. Liu, "Line-based registration of panoramic images and LiDAR point clouds for mobile mapping," *Sensors*, vol. 17, no. 12, p. 70, Dec. 2016, doi: 10.3390/s17010070.

[13] A. Taneja, L. Ballan, and M. Pollefeys, "Registration of spherical panoramic images with cadastral 3D models," in *Proc. 2nd Int. Conf. 3D Imag., Modeling, Process., Vis. Transmiss.* Zürich, Switzerland: ETH Zurich, Oct. 2012, pp. 479–486, doi: 10.1109/3dimpvt.2012.45.

[14] S. Peng and L. Zhang, "Automatic registration of optical images with airborne LiDAR point cloud in urban scenes based on line-point similarity invariant and extended collinearity equations," *Sensors*, vol. 19, no. 5, p. 1086, Mar. 2019, doi: 10.3390/s19051086.

[15] C. Yuan, X. Liu, X. Hong, and F. Zhang, "Pixel-level extrinsic self calibration of high resolution LiDAR and camera in targetless environments," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 7517–7524, Oct. 2021, doi: 10.1109/LRA.2021.3098923.

[16] T.-S. Kwak, Y.-I. Kim, K.-Y. Yu, and B.-K. Lee, "Registration of aerial imagery and aerial LiDAR data using centroids of plane roof surfaces as control information," *KSCE J. Civil Eng.*, vol. 10, no. 5, pp. 365–370, Sep. 2006, doi: 10.1007/BF02830090.

[17] M. Eslami and M. Saadatseresht, "A new tie plane-based method for fine registration of imagery and point cloud dataset," *Can. J. Remote Sens.*, vol. 46, no. 3, pp. 295–312, May 2020, doi: 10.1080/07038992.2020.1785282.

[18] Y. Belkhouche, Y. Belkhouche, S. Jackson, K. Namuduri, and B. Buckles, "Automated two-dimensional–three-dimensional registration using intensity gradients for three-dimensional reconstruction," *Proc. SPIE*, vol. 6, no. 1, Apr. 2012, Art. no. 063517, doi: 10.1117/1.Jrs.6.063517.

[19] N. Shorter and T. Kasparis, "Autonomous registration of LiDAR data to single aerial image," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2008, pp. V-216–V-219.

[20] G. Pandey, J. R. Mcbride, S. Savarese, and R. M. Eustice, "Automatic extrinsic calibration of vision and LiDAR by maximizing mutual information," *J. Field Robot.*, vol. 32, no. 5, pp. 696–722, Aug. 2015, doi: 10.1002/rob.21542.

[21] R. Wang, F. P. Ferrie, and J. Macfarlane, "Automatic registration of mobile LiDAR and spherical panoramas," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, p. 8.

[22] E. G. Parmehr, C. S. Fraser, and C. Zhang, "Automatic parameter selection for intensity-based registration of imagery to LiDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7032–7043, Dec. 2016, doi: 10.1109/TGRS.2016.2594294.

[23] A. Mastin, J. Kepner, and J. Fisher, "Automatic registration of LiDAR and optical images of urban scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2639–2646.

[24] R. K. Mishra, "A review of optical imagery and airborne LiDAR data registration methods," *Open Remote Sens. J.*, vol. 5, no. 1, pp. 54–63, Jul. 2012.

[25] M. Corsini, M. Dellepiane, F. Ganovelli, R. Gherardi, A. Fusiello, and R. Scopigno, "Fully automatic registration of image sets on approximate geometry," *Int. J. Comput. Vis.*, vol. 102, nos. 1–3, pp. 91–111, Mar. 2013, doi: 10.1007/s11263-012-0552-5.

[26] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. 15th Eur. Conf. Comput. Vis. (ECCV)*, in Lecture Notes in Computer Science, vol. 11211, Munich, Germany, Sep. 2018, pp. 833–851, doi: 10.1007/978-3-030-01234-2_49.
[27] W. Zhang, J. Qi, P. Wan, H. Wang, D. Xie, X. Wang, and G. Yan, "An easy-to-use airborne LiDAR data filtering method based on cloth simulation," *Remote Sens.*, vol. 8, no. 6, Jun. 2016, Art. no. 501, doi: 10.3390/rs8060501.
[28] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following," *Comput. Vis., Graph., Image Process.*, vol. 29, no. 3, p. 396, Mar. 1985.
[29] W. Zhang, J. Zhao, M. Chen, Y. Chen, K. Yan, L. Li, J. Qi, X. Wang, J. Luo, and Q. Chu, "Registration of optical imagery and LiDAR data using an inherent geometrical constraint," *Opt. Exp.*, vol. 23, no. 6, pp. 7694–7702, Mar. 2015, doi: 10.1364/oe.23.007694.
[30] L. Kneip, H. Li, and Y. Seo, "UPnP: An optimal $O(n)$ solution to the absolute pose problem with universal applicability," in *Proc. 13th Eur. Conf. Comput. Vis. (ECCV)*, in Lecture Notes in Computer Science, vol. 8689, Zürich, Switzerland, Sep. 2014, pp. 127–142.

BUYUN WANG received the B.E. degree from the School of Surveying and Geo-Information, North China University of Water Resources and Electric Power, Zhengzhou, China, in 2020. He is currently pursuing the M.S. degree with Zhengzhou University, Henan, China. His main research interest includes the fusion of point cloud and image.

HONGWEI LI is a Professor with the School of Earth Science and Technology, Zhengzhou University, Henan, China. His main research interests include geospatial data mining, machine vision measurement, and map and spatial cognition.

SHAN ZHAO is a Lecturer with the School of Earth Science and Technology, Zhengzhou University, Henan, China. Her main research interest includes the technology and application of earth information systems.

LINQING HE received the B.E. degree from the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China, in 2019. She is currently pursuing the master's degree with the School of Geo-Science and Technology, Zhengzhou University, Zhengzhou, China. Her research interest includes the research on geographic information service based on microservice.

YULU QIN received the B.E. degree from the Nanjing University of Finance and Economics, Nanjing, China, in 2020. She is currently pursuing the master's degree with the School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, China. Her research interest includes the multi-robot collaborative SLAM.

XIAOYUE YANG received the B.E. degree from the University of Luoyang Normal, Luoyang, China, in 2020. She is currently pursuing the master's degree with the School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, China. Her main research interest includes the mapping and path planning of robot multi-sensor fusion.