Article

Automatic Registration for Panoramic Images and Mobile LiDAR Data Based on Phase Hybrid Geometry Index Features

Genyi Wan¹,²,³, Yong Wang⁴, Tao Wang¹,²,³, Ningning Zhu⁵, Ruizhuo Zhang¹,²,³ and Ruofei Zhong¹,²,³,*

1 Key Laboratory of 3D Information Acquisition and Application, MOE, Capital Normal University, Beijing 100048, China
2 Base of the State Key Laboratory of Urban Environmental Process and Digital Modeling, Capital Normal University, Beijing 100048, China
3 College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China
4 Chinese Academy of Surveying and Mapping (CASM), Beijing 100036, China
5 The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
* Correspondence: zrf@cnu.edu.cn

Abstract: The registration of panoramic images and mobile light detection and ranging (LiDAR) data is quite challenging because different imaging mechanisms and viewing angle differences generate significant geometric and radiation distortions between the two multimodal data sources. To address this problem, we propose a registration method for panoramic images and mobile LiDAR data based on the hybrid geometric structure index feature of phase. We use the initial GPS/IMU to transform the mobile LiDAR data into an intensity map and align the two images to complete registration. Firstly, a novel feature descriptor called a hybrid geometric structure index of phase (HGIFP) is built to capture the structural information of the images. Then, a set of corresponding feature points is obtained from the two images using the constructed feature descriptor combined with a robust false-match elimination algorithm. The average pixel distance of the corresponding feature points is used as the error function. Finally, in order to complete the accurate registration of the mobile LiDAR data and panoramic images and improve computational efficiency, we propose the assumption of local motion invariance of 3D–2D corresponding feature points and minimize the error function through multiple reprojections to achieve the best registration parameters. The experimental results show that the method in this paper can complete the registration of panoramic images and the mobile LiDAR data under a rotation error within 12° and a translation error within 2 m. After registration, the average error of rotation is about 0.15°, and the average error of translation is about 1.27 cm. Moreover, it achieves a registration accuracy of less than 3 pixels in all cases, which outperforms the current five state-of-the-art methods, demonstrating its superior registration performance.

Keywords: mobile LiDAR points; co-registration; panorama images; nonlinear radiometric differences

1. Introduction

In recent years, with the rise of vehicle navigation [1], urban planning [2,3], cultural relics protection [4], and assisted driving [5,6], the demand for 3D laser point clouds has dramatically increased. The accuracy of the 3D laser point cloud is also getting higher and higher. However, although the 3D laser point cloud can accurately record objects’ positions and surface features, it lacks a description of object texture information. Additionally, images have good spatial resolution and can provide rich textures and semantics, but lack a description of the depth information of objects. The fusion processing of the 3D laser point cloud and an image can effectively make up for the lack of a single data source, and this combination is widely used in building extraction [7], 3D reconstructions of large-scale scenes [8,9], point cloud classification [10,11] and dynamic object detection [12–14]. The precise registration of the 3D laser point cloud and the image is the premise of fusion
processing. Only accurate registration can enable all the respective advantages of the two sensors.

The registration between the 3D laser point cloud and the image is necessary to calculate the rigid transformation relationship between the two. Through translation and rotation, the projected image of the 3D laser point cloud can be aligned with the image captured by the camera. Laser scanners are used to acquire 3D point cloud data. Depending on the acquisition method, laser scanning can be divided into terrestrial laser scanning (TLS), mobile laser scanning (MLS) and airborne laser scanning (ALS) [15]. In general, ALS can be used for large-scale building surveys, roads and forests; TLS can be used for more detailed but slower urban surveys in outdoor and indoor environments; and MLS is less accurate than TLS, but has a higher efficiency due to sensors mounted on the same vehicle. A mobile mapping system (MMS), usually equipped with a mobile laser scanner, panoramic camera, inertial measurement unit (IMU) and global positioning system (GPS), can easily obtain the LiDAR point cloud, image and orientation [16]. Theoretically, the LiDAR point cloud and the image obtained by the MMS are precisely registered. However, because pre-calibrated laser scanners and panoramic cameras are likely to have unpredictable shifts in their long-term use, an IMU or a GPS may not be able to obtain accurate poses due to environmental problems. Therefore, even if the panoramic image and the mobile LiDAR data have been georeferenced before fusion [17], it is still difficult to accurately calculate the rigid transformation relationship between the mobile LiDAR data and the panoramic image to complete the accurate registration between the two.

Since LiDAR point clouds are discrete space points, and the image is a continuous two-dimensional optical image, the different data structures make the two kinds of data have significant differences [18]. At the same time, the density of the LiDAR point cloud and the spatial resolution of the image [19], as well as the differences in geometry and radiation between the two types of data, will also affect registration. These situations make the registration of LiDAR point clouds and images more complex and challenging. The registration process of the LiDAR point cloud and images generally includes 2D–2D registration methods [18–23], 3D–2D registration methods [16,17,24–27] and 3D–3D registration methods [28–31]. Most of the existing registration methods extract the same features from the image and the LiDAR point cloud to complete registration, but these methods generally have characteristics that require specific scenes, tedious extraction steps and complicated calculations. The 3D–3D registration methods are not popular because they require 3D image reconstruction first. The 2D–2D registration methods may not match in the finished work because measurements of the similarity of two images are not accurate enough.

The method proposed in this paper belongs to the 2D–2D registration method group. To realize the refined registration of the panoramic image and the mobile LiDAR data, we propose a registration method for the panoramic image and the mobile LiDAR data based on the hybrid geometric structure index feature of phase. First, the mobile LiDAR data are transformed into an intensity map using the initial GPS/IMU data and panoramic image. Then, the feature descriptor of the developed HGIFP combined with a robust false matching point elimination algorithm is used to obtain a set of corresponding feature points and a 2D–3D correspondence. Finally, according to the local motion invariance assumption of the 3D–2D corresponding feature points proposed by us, the registration of the mobile LiDAR data and the panoramic image is completed after several iterations.

The main contributions of this paper are as follows:

1. A registration method for panoramic images and mobile LiDAR data based on the hybrid geometric structure index feature of phase is proposed.
2. A multi-scale and multi-directional feature descriptor called HGIFP is developed to capture the shape and structural features of the image and combined with a robust false matching point elimination algorithm to complete the corresponding feature points extraction work of the multimodal images.
3. The assumption of local motion invariance of 3D–2D corresponding feature points is proposed to solve the registration process of the mobile LiDAR data and panoramic image.

4. The method in this paper can correct the rotation error within $12^\circ$ and the translation error within 2 m, and the average error is within 3 pixels.

2. Related Works

The registration process of the LiDAR point cloud and images has been widely studied and can be divided into the airborne LiDAR point cloud and the mobile LiDAR point cloud according to the different acquisition methods. The registration methods of these two acquisition methods can be divided into the three methods mentioned above: 2D–2D, 3D–2D and 3D–3D. However, the registration method for the airborne Lidar point cloud and images is likely unsuitable for when the mobile LiDAR point cloud and panoramic images are collected simultaneously. Compared with the airborne Lidar point cloud, the mobile LiDAR point cloud has a smaller angle perturbation with a more severe impact [22], which makes registration more complex and challenging. Below we present a detailed review of the registration process for mobile LiDAR data and panoramic images.

3D–2D registration methods. 3D–2D registration methods aim to extract the corresponding feature primitives from panoramic images and the mobile LiDAR point cloud and convert them into linear relations to complete registration. Cui et al. [26] completed registration by extracting line features from the LiDAR point cloud and the panoramic image, respectively, and projecting the line features in the LiDAR point cloud to the panoramic image. This method requires specific features, such as curbs, light poles and urban buildings. Zhu et al. [25] proposed to use the skyline features of panoramic images and the LiDAR point cloud to complete registration. This registration method is more generic than the method of finding city line features. However, this method is not applicable when the skyline cannot be obtained due to dense building scenes. Peng et al. [17] further used point features to constrain based on the single-use of line features and proposed an “infinitesimal feature analysis method” to refine edge lines. Compared with 2D–2D registration methods, 3D–2D registration methods are more dependent on the specified features in the registration scene, which means that when a scene cannot provide the specified features required by the registration method, or the number of features is not enough and the distribution is not uniform enough, the accuracy of the registration method will be greatly reduced.

2D–2D registration methods. 2D–2D registration methods need to transform mobile LiDAR data into an image, so 2D–2D registration methods can also be regarded as the registration between images of different modalities. These methods can generally be divided into region-based [21,22,32,33] and feature-based [16] registration methods. The region-based registration methods measure the similarity of two images by constructing a similarity measure from the space or frequency to realize the registration of the mobile LiDAR data and panoramic image. Commonly used similarity measures generally include mutual information (MI), normalized mutual information (NMI) and the sum of squared differences (SSD). Wang et al. [22] used the method of maximizing mutual information to complete the registration of mobile LiDAR data and panoramic images. Zachary et al. [33] introduced gradient components and normalized mutual information to complete the registration process. Zhao et al. [32] improved the robustness of registration by setting weights for different data sources according to different scenarios when calculating with mutual information. However, region-based approaches often require a lot of computation and are also subject to image noise and nonlinear radiation changes, limiting their application in the registration field. Feature-based methods generally complete registration by searching for corresponding features between two images. For example, Li et al. [16] completed the registration process by finding the corresponding vehicle unit on the panoramic image and the point cloud projection image. Feature-based registration methods have made significant progress with the development of multimodal image registration algorithms. Zhu et al. [18] and Shi et al. [20] used excellent multimodal image registration algorithms on point cloud projection images, completing registration with aerial images. However,
due to the difference between the airborne LiDAR point cloud and the mobile LiDAR point cloud, these methods still need to be further improved to better register the mobile LiDAR point cloud and the panoramic image.

3D–3D registration methods. 3D–3D registration methods use structure from motion (SFM) and multi-view stereo (MVS) to generate point clouds from adjacent images, transforming the registration of LiDAR point clouds and images into the registration between the point clouds [16]. These registration methods require sufficient overlap between the two images taken before and after and require multiple overlapping images to obtain densely connected points to generate reliable point clouds [34]. 3D matching algorithms, such as normal distribution transform (NDT) [35], iterative closest point (ICP) [36] and variants [37,38], are commonly used for registration between point clouds. Zhao et al. [30] aligned a point cloud computed from a video onto the point cloud directly obtained from a 3D sensor. However, accurately performing the automatic initialization of the ICP algorithm is an important issue. Li et al. [39] first utilized the global navigation satellite system (GNSS) and IMU-assisted SFM to obtain an accurate image orientation and correct the error of low-precision POS through a two-step registration method. Moreover, they then transformed the original 2D–3D registration into 3D–3D registration, minimizing the differences between the depth map derived from SFM and the original laser scan to achieve accurate registration. In general, 3D–3D registration methods may generate errors when using images to create point clouds and point clouds to register each other. A sequence of images is also required to complete registration.

3. Methodology

This paper proposes an automatic registration method for panoramic images and mobile LiDAR using GPS/IMU data, which includes four steps: (1) Transform the mobile LiDAR data into an intensity map according to the GPS/IMU; (2) Use a novel structural feature descriptor called HGIFP and a robust mismatch elimination algorithm to obtain corresponding feature points; (3) Iteratively obtain the optimal solution using a particle swarm optimization (PSO) algorithm; (4) Combine the results obtained in step (3) into new POS data, and repeat steps (1)–(3) twice to complete the registration of the panoramic image and the mobile LiDAR data. Figure 1 shows a flowchart of the steps of the proposed method.

3.1. Multimodal Image Corresponding Feature Points Extraction

The extraction of the corresponding feature points is generally realized by relying on the local grayscale characteristics of the feature points. However, for multimodal image data, the grayscale characteristics will change non-linearly or non-functionally. The traditional method of extracting corresponding feature points based on image grayscale features is no longer applicable. In contrast, the geometrical features of the image are more resistant to nonlinear radiation differences. In this paper, a new method called HGIFP is proposed to extract corresponding feature points from multimodal images by using the geometrical features of images, which includes four steps: (1) Construct a weighted moment of local features taking anisotropy into account and use a FAST detector to detect the interest points on the weighted moment map of local features taking anisotropy into account; (2) Use a new structural feature descriptor called HGIFP to obtain panoramic images; (3) For the problem of the viewing angle difference, use the local consensus of matching points to remove outliers to obtain the correct corresponding feature points, and robustly complete the extraction of the corresponding feature points.
3.1. Multimodal Image Corresponding Feature Points Extraction

The extraction of the corresponding feature points is generally realized by relying on the local grayscale characteristics of the feature points. However, for multimodal image data, the grayscale characteristics will change non-linearly or non-functionally. The traditional method of extracting corresponding feature points based on image grayscale features is no longer applicable. In contrast, the geometrical features of the image are more resistant to nonlinear radiation differences. In this paper, a new method called HGIFP is proposed to extract corresponding feature points from multimodal images by using the geometrical features of images, which includes four steps: (1) Construct a weighted moment of local features taking anisotropy into account and use a FAST detector to detect the interest points on the weighted moment map of local features taking anisotropy into account; (2) Use a new structural feature descriptor called HGIFP to obtain panoramic images; (3) For the problem of the viewing angle difference, use the local consensus of matching points to remove outliers to obtain the correct corresponding feature points, and robustly complete the extraction of the corresponding feature points.

3.1.1. Point Feature Detection

There are apparent nonlinear radiation differences (NRD) between multimodal images, which increases the difficulty of feature point detection. As a nonlinear filter, anisotropic filtering can achieve the smooth operation of the image and preserve the edge and texture information of the image. Therefore, a local feature map with good edges and corner features can be obtained by subtracting the to-be-registered and the smoothed images. Meanwhile, phase consistency (PC) has been proven to detect feature points [40], and the edges and corner features of the image can be accurately obtained by using the maximum and minimum moments [41]. Therefore, our operation steps are as follows: (1) Smooth the image using optimized anisotropic filtering [42] to complete the anisotropic image diffusion; (2) Subtract the smoothed image from the to-be-registered image to obtain a local feature map that takes anisotropy into account; (3) Construct a weighted moment Equation and generate an anisotropic weighted moment map to extract feature points. The anisotropic diffusion Equation is:

$$\frac{\partial I}{\partial t} = \text{div} \left[ \frac{1}{1 + |\nabla I|^2/k^2} \cdot (x, y, t) \Delta I \right]$$

where $t$ is the metric value of the time scale; $\text{div}$ is the divergence operator; $1/\left(1 + |\nabla I|^2/k^2\right)$ is the diffusion coefficient; $k$ is the contrast factor; $\nabla$ is the gradient operator; $\Delta$ represents the Laplacian operator; $I$ is the current two-dimensional image; $L$ is the difference image;
\( m \) means the dimension of the image; \( A_i(L^n) \) represents the diffusion coefficient matrix of the image in each dimension; \( L^{n+1} \) is the result after diffusion; and \( \tau \) represents the diffusion time step, \( \tau = t_{n+1} - t_n \). When the diffusion calculation is calculated, \( t_n = 1/2\sigma^2 \) represents the scale, and \( n \) is the number of layers in the scale space.

After completing the anisotropic diffusion of the image, we calculate a local feature map taking anisotropy into account. It is calculated as follows:

\[
I_F(x, y) = I_{pm}(x, y) - G_{\sigma} * I(x, y)
\]

where \( I_F(x, y) \) is the local feature map taking anisotropy into account, \( I(x, y) \) is the original image, \( G_{\sigma} \) is a Gaussian filter, and the result of the anisotropic filtering of the original image is \( I_{pm}(x, y) \).

We use phase congruency to detect the edge features and corner features. Unlike the grey-based edge feature extraction method, this method is not affected by the image’s local light and dark changes. It can include information, such as corners, lines, textures, etc., especially when the edge contrasts of the image are relatively low (Figure 2). The formula of phase congruency is as follows:

\[
PC(x, y) = \frac{\sum c \sum n W_o(x, y) |A_{no}(x, y)\Delta \phi_{no}(x, y) - T|}{\sum c \sum n A_{no}(x, y) + \epsilon}
\]

where \( PC(x, y) \) is the magnitude of phase congruency; \( (x, y) \) is the pixel coordinate of a specific point on the image; \( A_{no}(x, y) \) is the amplitude at \( (x, y) \) at wavelet scale \( n \) and orientation \( o \); and \( \Delta \phi_{no}(x, y) \) is a more sensitive phase deviation. \( W_o(x, y) \) is the weight factor for a frequency spread, \( T \) is a noise threshold and \( \epsilon \) is a small constant to avoid division by zero. \( | \cdot | \) denotes that the enclosed quantity is equal to itself when its value is positive and zero otherwise.

Figure 2. Comparison of phase congruency with Sobel edge detection. (a) Original images; (b) Sobel edge detection; (c) phase congruency.

According to the moment analysis algorithm, the axis corresponding to the minimum moment is called the principal axis, and the principal axis indicates the direction information of the feature; the axis corresponding to the magnitude of the maximum moment generally reflects the distinctiveness of the feature. We construct a weighted moment map taking local anisotropy features into account according to the maximum moment and the minimum moment, and the calculation results are as follows:

\[
M_{\text{max}} = \frac{1}{2} \left( c + a + \sqrt{b^2 + (a - c)^2} \right)
\]

\[
M_{\text{min}} = \frac{1}{2} \left( c + a - \sqrt{b^2 + (a - c)^2} \right)
\]
\[ M_{pm\text{-}pc} = \frac{1}{2}(M_{\text{max}} + M_{\text{min}}) + \frac{k}{2}(M_{\text{max}} - M_{\text{min}}) \]  

where

\[ a = \sum_g (pc(\theta)\cos(\theta))^2 \]  
\[ b = 2\sum_g (pc(\theta)\cos(\theta))(pc(\theta)\sin(\theta)) \]  
\[ c = \sum_g (pc(\theta)\sin(\theta))^2 \]

where the maximum moment map \( M_{\text{max}} \) represents the edge map of the image; the minimum moment map \( M_{\text{min}} \) represents the corner map; \( M_{pm\text{-}pc} \) is a weighted moment diagram taking anisotropy local features into account; \( k \) is the weight coefficient, and its value is between \(-1\) and \(1\); and \( a, b, c \) are three intermediate quantities. Moreover, we use a FAST detector to detect interest points.

3.1.2. Feature Description Based on Mixed Index Map and Orientation of Phase Congruency Index Map

After the feature points are detected from the image, feature descriptors are needed to describe the difference between the feature points. The RIFT [34] has proved that compared with the classical feature descriptor, the feature descriptor of the multimodal image using the geometric structure information of the PC map still cannot describe features well. Thus, the RIFT constructs feature descriptors based on a maximum index map (MIM) to solve this problem. However, this method cannot provide enough corresponding feature points when matching visible light and LiDAR intensity images. We use the visible light and LiDAR intensity images for testing, as shown in Figure 3a. The detection of feature points is performed via the same operation, so the number of the initial feature points of the two methods is the same, and the distribution is consistent. Figure 2b shows that the number of the detected feature points is abundant and evenly distributed. However, in the matching process, the RIFT cannot obtain enough corresponding feature points, the distribution of the matching pairs is uneven and the matching effect is poor, as shown in Figure 3e. The reasons may be as follows. For images with severe NRD, the MIM-based method cannot describe the detail of features well, because each pixel has only six description latitudes, as shown in Figure 3c. However, if the number of phase indices increases by increasing the convolution direction, it will increase the time complexity. Therefore, we propose a method to construct HGIFP feature descriptors based on the mixed maximum index map and orientation of the phase congruency index map (MIM-OPCIM), as shown in Figure 3d; the extraction effect is shown in Figure 3f.

![Figure 3](image-url)

**Figure 3.** Comparison of our method and the RIFT: (a) input image pair; (b) feature detection results; (c) MIMs; (d) MIM-OPCIMs; (e) matching based on MIM; (f) matching based on MIM-OPCIM.
MIM is constructed via the log-Gabor convolution sequence. The convolution sequence was obtained in the PC map calculation stage, so the computational complexity of MIM is very small and does not entail too many additional calculations. The specific construction method of MIM is shown in Figure 4. Given an image \( I(x, y) \), \( M^{c}_n \) and \( M^{o}_n \) represent the even-symmetric and odd-symmetric log-Gabor wavelets at scale \( n \) and orientation \( o \). The two wavelet functions are convolved with the image signal to obtain the response components \( e_{no}(x, y) \) and \( o_{no}(x, y) \). For the obtained response component \( e_{no}(x, y) \) and \( o_{no}(x, y) \), we first calculate the wavelet transform amplitude \( A_{no}(x, y) \) at scale \( n \) and orientation \( o \); then, for orientation \( o \), all magnitudes of the scale \( S \) are summed to obtain the log-Gabor convolutional layer \( A_o(x, y) \).

\[
[e_{no}(x, y), o_{no}(x, y)] = [I(x, y) * M^{c}_n, I(x, y) * M^{o}_n]
\]

\[
A_{no}(x, y) = \sqrt{e_{no}(x, y)^2 + o_{no}(x, y)^2}
\]

\[
A_o(x, y) = \sum_{n=1}^{S} A_{no}(x, y)
\]

![Index map generation](image)

**Figure 4.** Index map generation: (a) original image; (b) log-Gabor convolution image sequence; and (c) index map.

We arrange \( A_o(x, y) \) to obtain the convolution graph \( \{A_o(x, y)\}_1^O \). In the calculation, \( O \) is the number of orientations. For each pixel position of the image, we can obtain an O-dimensional ordered array, and the subscript corresponding to the maximum value on each O-dimensional ordered array is selected to obtain the maximum index map.

The orientation of phase congruency represents the fastest-changing direction of the image’s geometric structure features, which is similar to the gradient direction and is very important for feature description. We collect the horizontal and vertical components of \( o_{no}(\theta) \) at all scales \( S \) and all orientations \( O \) to obtain an orientation of phase congruency map (OPCM). It is calculated as follows:

\[
\text{OPCM} = \arctan \left( \sum_\phi \sum_\theta (o_{no}(\theta) \sin(\phi)), \sum_\phi \sum_\theta (o_{no}(\theta) \cos(\phi)) \right)
\]

Due to the different modes of images, the orientations of phase congruency of each pixel between the two OPCMs are not closely related. It is impossible to use the OPCM to establish a feature description directly. Based on this, we construct the orientation of the phase congruency index map (OPCM) for feature description. We reclassify each pixel of each OPCM image according to the numerical value, and the value of the reclassification is \( N \) to obtain the OPCIM. It is calculated as follows:

\[
\text{OPCIM} = \text{mapminmax}(\text{OPCM}, 1, N)
\]
After obtaining the OPCIM, to better describe the image, we use MIM to constrain the OPCIM and then propose the MIM-OPCIM, calculated as follows:

\[
a = \text{mapminmax}(\text{MIM}, 1, \mu)
\]

\[
\text{MIM - OPCIM} = \text{mapminmax}(a \times \text{OPCIM}, 1, N)
\]

where \(a\) is the constraint coefficient matrix of the OPCIM, \(\mu\) is the size of the termination value for reclassifying MIM from 1, and the size of \(N\) is the same as that in Equation (14).

For each pixel, the method based on the MIM-OPCIM has a better description than that based on the RIFT because the description dimension increases from 6 to \(N\). Therefore, the HGIFP feature descriptor constructed based on the MIM-OPCIM is more reliable than the feature descriptor constructed by the RIFT. After obtaining the MIM-OPCIM, we generate feature descriptors for each feature point in the same way as the RIFT to obtain HGIFP feature descriptors. We build a local image patch of size \(J \times J\) pixels centered on each feature point and assign weights to each pixel using a Gaussian function with a standard deviation of size \(J/2\). For the sub-grid in each local pixel block, the size is set to \(r \times r\), \(r > 6\). This way includes as much grid information as possible. Figure 5 shows the construction process of the HGIFP feature descriptor.

![Figure 5](image.png)

**Figure 5.** The main processing chain of the HGIFP descriptor.

3.1.3. Elimination of Mismatched Points Based on Local Preserving Matching

Obtaining a reliable correspondence from two feature point sets is an essential step in extracting corresponding feature points. For panoramic images and intensity maps, the difference in imaging viewpoints makes it easier to have wrong matches in the matching results. Filter out the correct matching pairs in the initial matching set is particularly critical. To address this problem, we use locality preserving matching (LPM) [43] to filter from the collection of matching points to obtain the correct matching pairs.

The local preserving matching of feature points means that the motion displacements of the two corresponding feature points in the neighborhood should be consistent. So, we can eliminate the wrong matching pairs by calculating the difference between the motion displacement values of each feature point and its neighborhood and the topological relationship.

Assuming that the unknown interior-point set is \(I\), and the initial set is \(S\), then the final matching set \(I^*\) has the following relationship:

\[
I^* = \text{argmin}_{I} C(I; S, \lambda)
\]

The loss function \(C(I; S, \lambda)\) is defined as follows:

\[
C(I; S, \lambda) = \sum_{i \in I} \frac{1}{4} \left( \sum_{j \in S_i \cap N_{\mu}} (d(x_i, x_j) - d(y_i, y_j))^2 + \lambda \sum_{j \in N_{\mu}} (d(x_i, x_j) - d(y_i, y_j))^2 \right)
\]

\[
+ \lambda (N - |I|)
\]

\[
(18)
\]
where \( N_{xi}, N_{yi} \) are the neighborhoods of the feature points \( x_i, y_i \); \( d \) is a certain distance metric, such as Euclidean distance; and \( K \) is the cardinality of the neighborhood points. Here we use a nearest neighbor search; \( 1/2K \) is the normalization of the loss function; \( \lambda \) is the penalty term coefficient, \( \lambda > 0 \); \( N \) is the number of all matching points; and \( |I| \) is the number of correct matching point pairs.

If we use the numerical difference in the local motion displacement of the matching points to measure whether the point is a wrong matching point, this method will fail when there are non-rigid changes in the image. Therefore, we binarize the distance.

\[
d(x_i, x_j) = \begin{cases} 0, & x_j \notin N_{xi} \\ 1, & x_j \in N_{xi} \end{cases} \quad d(y_i, y_j) = \begin{cases} 0, & y_j \notin N_{yi} \\ 1, & y_j \in N_{yi} \end{cases}
\]  

(19)

For Equation (18), a \( 1 \times N \) vector \( P \) can be used to indicate whether each pair of matching points is correct. \( P_i = 1 \) means that the match correctness of the pair of matching points is an inlier and \( P_i = 0 \) means an outlier.

So, we can rewrite Equation (18) by combining \( P \) with Equation (12).

\[
C(P; S, \lambda) = \sum_{i=1}^{N} \frac{P_i}{2K} \left( \sum_{j|x_j \in N_{xi}} d(y_i, y_j) + \sum_{j|y_j \in N_{yi}} d(x_i, x_j) \right) + \lambda \left( N - \sum_{i=1}^{N} P_i \right)
\]  

(20)

In order to better use the local motion consistency of the feature points to screen the error points, we can use the topological relationship of the neighborhood to distinguish the match correctness of correspondence further \((x_i, y_j)\) in the neighborhood \(N_{xi}, N_{yi}\), such as in Figure 6. \( v_i \) represents the displacement vectors corresponding to \((x_i, y_i)\), and \( v_j \) represents the displacement vectors corresponding to \((x_j, y_j)\), more specifically \( v_i = y_i - x_i \) and \( v_j = y_j - x_j \).

![Figure 6. Local neighborhood topology.](image)

The difference between \( v_i \) and \( v_j \) is defined as follows:

\[
s(v_i, v_j) = \min \left\{ \frac{|v_i|}{|v_j|}, \frac{|v_j|}{|v_i|} \right\} \frac{(v_i, v_j)}{|v_i| \cdot |v_j|}
\]  

(21)

where \( s(v_i, v_j) \) represents the topological similarity between two matching point pairs in the neighborhood \((v_i, v_j) \in [-1, 1]\). \((\cdot, \cdot)\) represents the inner product calculation of
vectors. The larger the value of \( s(v_i, v_j) \), the higher the similarity between the two, and binarization is performed according to the similarity threshold:

\[
s(v_i, v_j) = \begin{cases} 
0, & s(v_i, v_j) \geq \tau \\
1, & s(v_i, v_j) < \tau 
\end{cases}
\]

(22)

We combine Equations (21) and (22). Equation (20) can be written as the following minimization problem:

\[
C(P_i; S, \lambda, \tau) = \sum_{i=1}^{N} P_i \left( \sum_{j \in N_{s_i}} d(y_i, y_j) + \sum_{j \in N_{s_i}} s(v_i, v_j) \right) + \lambda \left( N \sum_{i=1}^{N} \sum_{j \in N_{s_i}} 1 \right)
\]

\[
= \sum_{i=1}^{N} P_i (c_i - \lambda) + \lambda N
\]

(23)

where

\[
c_i = \sum_{j \in N_{s_i}} d(y_i, y_j) + \sum_{j \in N_{s_i}} s(v_i, v_j)
\]

(24)

For the above Equation (23), the initially given matching set \( S \) is known, so the positions of all feature points are known, and hence all the cost values \( \{c_i\}_{i=1}^{N} \) can be calculated in advance. That is to say, the only unknown quantity is \( P_i \). If we want to minimize the value of the loss function, we need \( P_i (c_i - \lambda) < 0 \). For the wrong matching pairs \( P_i = 0 \), the calculation result of \( (c_i - \lambda) \) is meaningless. Only when \( P_i = 1 \), \( (c_i - \lambda) < 0 \) can the loss function reduce the value. Thus, we can relate the minimization of the loss function for solving \( P \) to \( \lambda \):

\[
P_i = \begin{cases} 
1, & c_i < \lambda \\
0, & c_i \geq \lambda 
\end{cases}, \quad i = 1, \ldots, N
\]

(25)

Thus, we can see that \( \lambda \) is a threshold for judging whether a pair of matching points is the correct interior point through Equation (25). According to the experiment, we use the above Equations for two iterations to obtain \( P_i \) from the candidate matching set \( S \).

3.2. Panoramic Image and the Mobile LiDAR Data Registration Model Based on Local Motion Invariance in 3D–2D Corresponding Feature Points

After obtaining a reliable set of corresponding feature points in the panorama image and the LiDAR intensity image, we can use the average pixel distance between the sets of corresponding feature points as the error equation. Therefore, the registration problem of panoramic images and mobile LiDAR data can be equivalent to the optimization problem, or minimizing the value of the error equation. However, in the optimization process, if the LiDAR intensity image is regenerated whenever extracting corresponding feature points and calculating the cost function, this will significantly increase the computational complexity and time. It also cannot converge to the optimal camera matrix if we realize registration by minimizing the distance between the projection points and the corresponding pixel points, because the number of extracted corresponding feature points is sparse and unevenly distributed due to the significant difference in angles of view and displacement between the matched images.

Based on the above problems, we propose the following hypothesis: Under a given pose, if the mapping of a point in the point cloud is the corresponding feature point of a pixel on the panoramic image, the mapping of this point should also be the corresponding feature point of the original pixel or within the minimal neighborhood of the corresponding feature point when the pose is changed to other similar viewing angles. This 3D–2D correspondence invariance is called the local motion invariance in the corresponding feature. The specific operation steps are as follows:
1. According to the panoramic image imaging model, there are the following formulas:

\[
\begin{bmatrix}
    u \\
    v \\
    1
\end{bmatrix}
= KFT \begin{bmatrix}
    X_w \\
    Y_w \\
    Z_w
\end{bmatrix}
\] (26)

where \((X_w, Y_w, Z_w)\) is a certain point in the point cloud; \((u, v)\) is the imaging point of the point on the two-dimensional plane; \(K\) is the inner orientation elements of the panoramic camera; \(F\) is the transformation matrix of the camera coordinate to the spherical polar coordinate; and \(T\) is the outer orientation elements of the panoramic camera. From this, we construct the imaging point cloud set \(P_{\text{intensity}}\). Moreover, there is a one-to-one correspondence between \(P_{\text{intensity}}\) and each pixel in the \(I_{\text{intensity}}(x, y)\).

2. After using the initial pose to obtain the LiDAR intensity image \(I_{\text{intensity}}(x, y)\), we extract the corresponding feature points from the panoramic image \(I_{\text{panoramic}}(x, y)\) and \(I_{\text{intensity}}(x, y)\) to obtain the corresponding feature points set \(S_{\text{panoramic}}, S_{\text{intensity}}\). Then, according to the corresponding relationship established in step 1, we obtain the point cloud set \(P_{\text{panoramic-intensity}}\) which corresponds to \(S_{\text{intensity}}\). Finally, we construct the error equation with \(P_{\text{panoramic-intensity}}\) and \(S_{\text{panoramic}}\). We minimize the error equation to complete the registration of the panoramic image and the mobile LiDAR data. The specific Equation is as follows:

\[
T_{\text{error}}^* = \underset{T_{\text{error}}}{\text{argmin}} \sum_{i=1}^{n} || S_i - KFTP_i ||_2^2
\] (27)

where \(K\) is the inner orientation elements of the panoramic camera; \(F\) is the transformation matrix of the camera coordinate to the spherical polar coordinate; \(T\) is the outer orientation elements of the panoramic camera; \(S_i\) is the point in the set \(S_{\text{panoramic}}, \) and \(P_i\) is the point in the point cloud set \(P_{\text{panoramic-intensity}}\) corresponding to \(S_i\).

3. After constructing the error equation, we use the PSO algorithm for iterative optimization. We set the rotation parameter to \(\pm 0.2\) of the input parameter and the translation parameter to \(\pm 2\) of the input parameter. The number of iterations is 200, and the number of particles is 60.

4. A step-by-step approximation method is adopted based on the proposed assumptions to obtain the correct motion pose. After completing step 3, we repeat the process of step 1–step 3 with the obtained camera motion pose as the initial pose, and set the rotation and translation parameter thresholds to 1/2 of the previous process. When the pixel average distance is less than 3 pixels, the iteration terminates. For specific experiments, please refer to Section 4.3.

4. Experiment and Analysis

This section focuses on our experimental verification of the performance of the proposed registration method on panoramic images and the mobile LiDAR data. First, we provide the detailed setup of the experiment, including a description of the experimental dataset and the experiment environment. Secondly, the performance of our proposed algorithm for extracting corresponding feature points based on HGFFP is analyzed. The evaluation is completed by comparing four state-of-the-art methods: SIFT, 3MRS, PSO-SIFT and RIFT. For a fair comparison, we set these four methods’ local description block sizes to be the same. Considering the characteristics of panoramic images, we will evaluate the corresponding feature point extraction algorithm in multimodal images by visual inspection. Then in the discussion of Section 5, the final results of the registration of the mobile LiDAR data and panoramic images are analyzed. We also introduce region-based MI methods for comparison in the discussion, and finally illustrate the robustness and usability of the proposed method by overlaying panoramic images and mobile LiDAR data.
4.1. Setting of Experiments

The experiments were performed using a personal computer (PC) configured with an Inter(R) Core(TM) i7-9750H CPU 2.60 GHz and 16.0 GB RAM. The mobile LiDAR data and panoramic images were collected from Guangzhou Ecological Design Town. In order to facilitate processing, the collected panoramic images were downsampled with a sampling interval of 3 pixels and converted into single-channel grayscale images. The parameters for image matching were set as follows. The log-Gabor filter was set to four scales and six directions \( (0, \frac{\pi}{6}, \frac{2\pi}{6}, \frac{3\pi}{6}, \frac{4\pi}{6}, \frac{5\pi}{6}) \). Three pairs of panoramic images and the corresponding range of the mobile LiDAR data were selected as experimental data. The details are listed in Table 1. These experimental data cover intersections, trees, basketball courts and residential buildings.

Table 1. Detailed description of the test cases.

| Test Case | Data Resource | Size | Location |
|-----------|---------------|------|----------|
| Test 1    | MMS LiDAR points, Panoramic image | 7,802,955 points, 2731 × 1366 pixels | town |
| Test 2    | MMS LiDAR points, Panoramic image | 8,471,353 points, 2731 × 1366 pixels | town |
| Test 3    | MMS LiDAR points, Panoramic image | 9,067,519 points, 2731 × 1366 pixels | town |

Figure 7a shows the first set of experimental data, including residential houses on both sides of the road; Figure 7b shows the second set of experimental data, open intersection data, including high-rise buildings and residential dwellings; and Figure 7c shows the third set of experimental data, including trees, basketball courts and residential houses.

4.2. Parameter Study

The proposed corresponding feature points extraction algorithm in the multimodal image contains four main parameters, namely \( N, \mu, J, \) and \( r \). The parameter \( N \) is the size of the reclassification of the phase congruency orientation. Generally speaking, the larger the \( N \) is, the higher the discrimination of the information in each pixel in the image and the higher the computational complexity. The parameter \( \mu \) is the reclassification size of the MIM. The larger \( \mu \) is, the greater the weight of the MIM in the constructed similarity descriptor. The parameters \( J \) and \( r \) are the sizes of the local image patch and sub-image block used for feature description. If the sizes of the local image patch and sub-image block are too small, the information is insufficient, which does not adequately reflect the distinctiveness of the feature. Conversely, if the sizes of the local image patch and sub-image block are too large, the computational complexity will be higher, and the matching time will be longer. Therefore, the correct parameter design dramatically influences the extraction effect of the corresponding feature points. We use a pair of LiDAR intensity and visible light images for parameter research. This is because the panoramic image is stitched in six directions. In order to accurately assess the extraction effect of the corresponding feature points, we need to calculate the extrinsic camera parameters through the corresponding feature points and then use the calculated extrinsic camera parameters to regenerate the LiDAR intensity image to realize the fusion of the panoramic image and the LiDAR intensity image. In contrast, extracting and verifying corresponding feature points from intensity projection images of mobile LiDAR data and images captured by flat cameras is simpler and more reliable. We only need to use the corresponding feature points to calculate the transformation matrix between the images to complete the evaluation of the image.
registration effect. The LiDAR intensity image and visible light image of the experiment are shown in Figure 8.

Figure 7. Experimental data: (a) Mobile LiDAR data and panoramic image of test 1; (b) Mobile LiDAR data and panoramic image of test 2; (c) Mobile LiDAR data and panoramic image of test 3.
In order to obtain suitable parameters, we design four independent experiments to learn the parameters $N$, $\mu$, $J$ and $r$. Each experiment has only one parameter as a variable, and the other parameters are designed as fixed values. The overall setup details of the experiments are in Table 2. We use the number of correct matches (NCM) and the success ratio (SR) as evaluation metrics for each parameter. The results of the experiments are shown in Tables 3–6. To ensure the fairness of the comparison experiment, the mismatch elimination algorithm used in the parameter study is consistent with that in the RIFT.

Table 2. The detail of parameter settings.

| Experiments | Variable | Fixed Parameters |
|-------------|----------|------------------|
| parameter $N$ | $N = [11, 13, 15, 17, 19]$ | $\mu = 1.3, J = 108, r = 8$ |
| parameter $\mu$ | $\mu = [1.1, 1.3, 1.5, 1.7, 1.9]$ | $N = 15, J = 108, r = 8$ |
| parameter $J$ | $J = [84, 96, 108, 120, 132]$ | $N = 15, \mu = 1.3, r = 8$ |
| parameter $r$ | $r = [6, 8, 10, 12, 14]$ | $N = 15, \mu = 1.3, J = 108$ |

Table 3. The results of parameter $N$.

| Metric | $N, \mu = 1.3, J = 108, r = 8$ |
|--------|--------------------------------|
|        | 11  | 13  | 15  | 17  | 19  |
| NCM    | 64  | 84  | 93  | 43  | 72  |
| SR/%   | 98.4| 100 | 100 | 93.5| 92.3|

Table 4. The results of parameter $\mu$.

| Metric | $\mu, N = 15, J = 108, r = 8$ |
|--------|--------------------------------|
|        | 1.1 | 1.3 | 1.5 | 1.7 | 1.9 |
| NCM    | 65  | 93  | 51  | 52  | 49  |
| SR/%   | 79.2| 100 | 96.2| 94.5| 98  |

Table 5. The results of parameter $J$.

| Metric | $J, N = 15, \mu = 1.3, r = 8$ |
|--------|-------------------------------|
|        | 84  | 96  | 108 | 120 | 132 |
| NCM    | 76  | 80  | 93  | 72  | 51  |
| SR/%   | 100 | 100 | 100 | 100 | 100 |

Figure 8. Experimental data for parameter study. (a) screenshot of video taken by roadside camera; (b) The point cloud intensity image at the location of (a) was taken.
Table 6. The results of parameter $r$.

| Metric | $r$, $N = 15$, $\mu = 1.3$, $J = 108$ |
|--------|--------------------------------------|
|        | 6   | 8   | 10  | 12  | 14  |
| NCM    | 57  | 93  | 92  | 129 | 121 |
| SR/\%  | 100 | 100 | 100 | 100 | 98.3|

The following conclusions can be drawn from the experimental results: (1) The larger the value of $N$, the richer the information of the MIM-OPCIM, so that there will be fewer mismatches and more NCM will be obtained. However, when $N$ is greater than a threshold, due to local geometric distortions, the number of mismatches and the number of NCM will also decrease. As can be seen from Table 3, when $N$ reaches 13 and 15, the SR of the MIM-OPCIM reaches 100%. Considering the number and computational complexity of extracting corresponding feature points for the MIM-OPCIM, we set $N = 15$. (2) It can be seen from Table 4 that the constructed the MIM-OPCIM mismatch decreases with the increase in $\mu$, but when $\mu \geq 1.5$, due to the influence of local geometric distortions, the mismatch begins to occur. At the same time, the number of NCM is also decreasing. Thus, we set $\mu = 1.3$. (3) SR is the highest for all assumed values of parameter $J$. However, considering that our main purpose is to obtain more corresponding feature points, therefore, according to Table 5, we set $J = 108$. (4) The change in the parameter $r$ will impact the NCM because more information is contained in the sub-image block. To balance the time complexity and the number of corresponding feature points extracted, as shown in Table 6, we set $r = 8$.

According to the experimental results and analysis, in the following experiments, these parameters are fixed as $N = 15$, $\mu = 1.3$, $J = 108$ and $r = 8$.

4.3. Matching Performance Analysis and Comparison

In this section, an analysis is performed to verify the matching effect of the proposed method. We use SIFT, 3MRS, PSO-SIFT, and RIFT and our method to extract corresponding feature points on three test data. All test data are subjected to two rotation and translation changes on the correct camera external parameters. The first change randomly adds a rotation error of $12^\circ$ and a translation error of 2 m, and the second change randomly adds a rotation error of $6^\circ$ and a translation error of 1 m (the settings of the first set of parameters are based on the errors encountered in our work, and the settings of the second and third sets of parameters are to check the registration results of our method under different distortions). The solid yellow line indicates all the correctly matched corresponding feature points in the images, and the solid red line indicates that there are incorrectly matched corresponding feature points.

Figures 9 and 10 show the extraction results of corresponding feature points obtained by using five methods on three test data. Figure 9 shows the extraction results of the corresponding feature points of the first test data. Figure 10 shows the number and success rate of the corresponding feature points extracted on three test data. Despite local geometric distortions, NRD, noise and large deformations between images, our proposed method can perform well on all three-test data. In particular, at $12^\circ$ rotations and with 2 m translation errors, the corresponding feature points obtained are 25, 28 and 12, as seen in Figure 10. On three test data, SIFT and PSO-SIFT all fail to match because both SIFT and PSO-SIFT are feature descriptions based on gradient histograms. Although PSO-SIFT redefines the gradient to meet the needs of multimodal image matching, this does not change the gradient’s sensitivity to NRD.
In contrast, the matching effects of the RIFT and 3MRS is significantly better than the SIFT and PSO-SIFT, which use gradients for feature description. The RIFT can better complete the extraction of corresponding feature points in multimodal images when the difference in viewing angle between the two images is not very large. On three test data, the numbers of corresponding feature points obtained by the RIFT under the second set of parameters are 19, 16 and 10, as seen in Figure 10. When there is no viewing angle difference between the two images, 3MRS obtains the largest number of matching points on the three test data, but once there is a viewing angle difference between the images, 3MRS cannot guarantee the accuracy of the obtained corresponding feature points, as shown in Figure 10. These results show that the RIFT has good robustness for matching visible light and intensity LiDAR images but cannot handle significant geometric distortions between
images well. 3MRS uses coarse matching combined with template matching to increase the number of extracted corresponding feature points significantly. Still, it is more sensitive to geometric distortions between images and unsuitable for registration between the mobile LiDAR data and panoramic images.

Figure 9. Comparison in the visualization results of our method, RIFT, 3MRS, PSO-SIFT, and SIFT on first test data. (a) The results of the proposed method; (b) the results of RIFT; (c) the results of 3MRS; (d) the results of PSO-SIFT; and (e) the results of SIFT.

Figure 10. Comparisons on NCM and SR metrics. (a–c) The number of corresponding feature points extracted on three test data by five methods; (d–f) The success rate of corresponding feature points extracted on three test data by five methods.

5. Discussion
5.1. Pixel Error and Runtime Evaluation

We illustrate the superiority of the proposed method by evaluating the pixel root mean square error (RMSE) and running time (RT) of the results using different methods. The GPS/IMU data used in the experiment are shown in Table 7. At the same time, we add the region-based MI method to measure the proposed method's effect more comprehensively. We choose root mean square error (RMSE) and running time (RT) to evaluate the proposed method and the other five methods. RMSE represents the accuracy of the overall match and is calculated by manually selecting 20 checkpoints. The final effects of all methods are shown in Figure 11.

Figure 12a depicts the histograms of RMSE for all methods, and Figure 12b shows RT histograms for all methods. The smaller the RMSE and RT, the better the matching performance. Table 8 reports the matching results of all methods in terms of RMSE and RT. On three test sets, the matching effect of our proposed method is the best, and the errors of the three results are 0.997 pixels, 1.606 pixels and 2.531 pixels. The errors of other methods are more than 50 pixels, and the registration requirements of the mobile LiDAR data and panoramic image cannot be completed. The reasons for the analysis are as follows. Using the feature point matching method to complete the mobile LiDAR data registration and the panoramic image, it is necessary to accurately find the corresponding relationship between the coordinate and pixel points. In the case of a rotation error of 12° and a translation error of 2 m, except for when using the method in this paper, only a few wrong corresponding feature points are obtained. Therefore, when using the PSO algorithm for iterative convergence, all methods except the method in this paper cannot converge to the correct camera extrinsic parameters. In addition, the resolution of the experimental data in this paper is $2731 \times 1366$. The high resolution is also one of the reasons that the pixel offset
The difference of other methods is too large and cannot be registered. Our experiments have also been tested using the histogram of absolute phase consistency gradients (HAPCGs) [44]. The algorithm cannot complete the matching due to the image’s high resolution, which leads to memory overflow during calculations. When the image is processed by multiple down-sampling, the HAPCG still cannot extract the correctly corresponding feature points between the images. To maintain the clarity of the matching image as much as possible and achieve accurate registration, we do not recommend using multiple down-sampling to complete the extraction of corresponding feature points in multimodal images. Due to the large geometric deformation of the LiDAR intensity image and the original image in different poses, the MI method cannot converge to the correct camera extrinsic parameters for the region-based MI method. This is the reason why the MI method has significant pixel errors.

**Table 7.** Experimental data of external orientation parameters: the (Initial value) column represents the initial value of the element, and the (Correct value) column represents the correct value of the element.

| Test | Elements | Correct Value | Initial Value |
|------|----------|---------------|---------------|
|      | X/m      | 472,133.143   | 472,135.143   |
|      | Y/m      | 2,622,689.476 | 2,622,690.876 |
| Test 1 | Z/m     | 70.920        | 72.120        |
|      | Roll/rad | −0.015755     | −0.205755     |
|      | Pitch/rad| −0.020855     | −0.190855     |
|      | Yaw/rad  | −1.565119     | −1.395119     |
|      | X/m      | 471,980.115   | 471,982.115   |
| Test 2 | Y/m     | 2,622,760.311 | 2,622,761.811 |
|      | Z/m      | 65.443        | 64.043        |
|      | Roll/rad | −0.009265     | 0.179265      |
|      | Pitch/rad| 0.028898      | 0.208898      |
|      | Yaw/rad  | 0.807724      | 0.627724      |
|      | X/m      | 472,089.330   | 472,091.401   |
| Test 3 | Y/m     | 2,622,689.692 | 2,622,687.734 |
|      | Z/m      | 69.654        | 70.154        |
|      | Roll/rad | −0.002380     | −0.200804     |
|      | Pitch/rad| 0.045244      | 0.205244      |
|      | Yaw/rad  | 1.597508      | 1.397508      |

In running time (RT), the three times required by our method are 184.4135 s, 187.1741 s and 192.0239 s. Among the comparative methods, the amount of time required by the RIFT is the least, and the three times of the RIFT are 118.803164 s, 100.742531 s and 104.78784 s. Although the RIFT algorithm takes the shortest amount of time among the six methods, the registration of the mobile LiDAR data and the panoramic image cannot be completed by the RIFT. This is because the RIFT cannot obtain correct and evenly distributed corresponding feature points when the viewing angle difference is too large. As shown in Table 8, the matching errors of the RIFT on the three test sets are 93.813 pixels, 51.153 pixels and 111.961 pixels. At the same time, comparing three test sets, we found that the proposed algorithm has the best effect on the first test data and the worst effect on the third test data. By analyzing the three test sets, it can be found that the structure information of the LiDAR intensity image in the first datum is the most obvious, so more uniformly distributed corresponding feature points can be obtained. The local geometric distortions and NRD of the LiDAR intensity image in the third test datum are the most serious, which brings significant challenges to the matching algorithm, so the matching result of the third test datum is the worst.
Figure 11. The final registration results of six methods. (a) The results of the proposed method; (b) the results of RIFT; (c) the results of MI; (d) the results of PSO-SIFT; (e) the results of 3MRS; and (f) the results of SIFT.
Figure 12. Matching performance comparison of different methods in two criteria for experimental images. (a) the histograms of RMSE for all experimental methods, (b) the histograms of RT for all experimental methods.

Table 8. Registration results of the proposed method, RIFT, MI, PSO-SIFT, 3MRS and SIFT.

| Test   | Method         | RMSE (Pixel) | RT (Second) |
|--------|----------------|--------------|-------------|
| Test 1 | Ours           | 0.997        | 184.414     |
|        | RIFT + PSO     | 93.813       | 118.803     |
|        | MI             | 107.721      | 507.088     |
|        | PSO-SIFT + PSO | 188.919      | 255.598     |
|        | 3MRS + PSO     | 91.426       | 267.647     |
|        | SIFT + PSO     | 94.818       | 110.800     |
| Test 2 | Ours           | 1.606        | 187.174     |
|        | RIFT + PSO     | 51.153       | 100.743     |
|        | MI             | 65.89        | 637.487     |
|        | PSO-SIFT + PSO | 85.335       | 217.802     |
|        | 3MRS + PSO     | 53.924       | 283.901     |
|        | SIFT + PSO     | 198.992      | 98.392      |
| Test 3 | Ours           | 2.531        | 192.024     |
|        | RIFT + PSO     | 111.961      | 104.788     |
|        | MI             | 74.089       | 632.097     |
|        | PSO-SIFT + PSO | 93.483       | 291.207     |
|        | 3MRS + PSO     | 176.542      | 256.630     |
|        | SIFT + PSO     | 143.185      | 124.225     |

5.2. Registration Parameters Error Evaluation

In order to better verify the matching effect of the proposed method, Figure 13 shows the results of the overall registration, and Table 9 shows the parameters before and after registration. Although the point cloud projection images of the three test sets have different degrees of local geometric distortions, NRD and noise effects, the three pairs of panoramic images and the mobile LiDAR data are all correctly aligned. According to Table 9, we know that each parameter’s errors after the registration of the first test datum are \(0.016\) m, \(0.010\) m, \(0.008\) m, \(-0.0017\) rad, \(0.002\) rad and \(0.005\) rad. Each parameter’s errors after the second test datum registration are \(0.000\) m, \(0.011\) m, \(0.013\) m, \(-0.002\) rad, \(-0.001\) rad and \(0.005\) rad. After the registration of the third test datum, the errors of each parameter are \(0.007\) m, \(-0.023\) m, \(-0.026\) m, \(-0.002\) rad, \(-0.0005\) rad and \(0.005\) rad. The average translation errors of three test data are all within 3 cm, and the average rotation errors are all within 0.005 rad. We found that the first test datum has the most accurate result of the three test sets, with an average translation error of 0.14 cm and an average rotation error of 0.003 rad. The third test datum has the worst result, with an average translation error of 1.87 cm and an average rotation error of 0.0025 rad. Comparing the point cloud projection
images of the three test sets, we found that the LiDAR intensity image of the first test datum has the best quality and complete structural information. The LiDAR intensity image of the third test datum has the worst quality, lacking the necessary structural information and filled with a lot of noise. Therefore, high-quality mobile LiDAR data are crucial for successfully registering the mobile LiDAR data and panoramic images.

![Figure 13](image1.png)

Figure 13. Effect diagram before and after registration. (a) comparison of the first test data before and after registration; (b) comparison of the second test data before and after registration; (c) comparison of the third test data before and after registration.

Table 9. Optimization of the exterior orientation parameters: the (Initial value) column represents the initial value of the element, the (Correct value) column represents the correct value of the element, and the (Final value) column represents the final optimized value of the element.

| Test | Elements | Correct Value | Initial Value | Final Value |
|------|----------|---------------|---------------|-------------|
| Test 1 | X/m | 472,133.143 | 472,135.143 | 472,133.127 |
| | Y/m | 2,622,689.476 | 2,622,690.876 | 2,622,689.486 |
| | Z/m | 70.920 | 72.120 | 70.928 |
| | Roll/ rad | –0.015755 | –0.205755 | –0.0174915 |
| | Pitch/ rad | –0.020855 | –0.190855 | –0.0187945 |
| | Yaw/ rad | –1.565119 | –1.395119 | –1.5604139 |
| Test 2 | X/m | 471,980.115 | 471,982.115 | 471,980.115 |
| | Y/m | 2,622,760.311 | 2,622,761.811 | 2,622,760.322 |
| | Z/m | 65.443 | 64.043 | 65.456 |
| | Roll/ rad | –0.009265 | 0.179265 | –0.011067 |
| | Pitch/ rad | 0.028898 | 0.208898 | 0.027423 |
| | Yaw/ rad | 0.807724 | 0.627724 | 0.812992 |
| Test 3 | X/m | 472,089.330 | 472,091.401 | 472,089.323 |
| | Y/m | 2,622,689.692 | 2,622,687.734 | 2,622,689.669 |
| | Z/m | 69.654 | 70.154 | 69.628 |
| | Roll/ rad | –0.002380 | –0.200804 | –0.002570 |
| | Pitch/ rad | 0.045244 | 0.205244 | 0.044744 |
| | Yaw/ rad | 1.597508 | 1.397508 | 1.602342 |
6. Conclusions

The registration between panoramic images and mobile LiDAR data is very important for urban applications, such as vehicle navigation, assisted driving, 3D urban change detection, etc. It is difficult to align accurately because pre-calibrated laser scanners and panoramic cameras are likely to have unpredictable shifts in long-term use, and IMU or GPS may not be able to obtain accurate poses due to environmental problems. This paper proposes a registration method for panoramic images and mobile LiDAR data based on the phase’s hybrid geometric structure index feature. First, we use the initial GPS/IMU data and panoramic image to obtain an intensity LiDAR image. Secondly, we obtain the set of corresponding feature points on the panoramic image and the intensity LiDAR image and receive the 2D–3D correspondence. Finally, after three iterative calculations, the registration of the mobile LiDAR data and the panoramic image is completed. The results show that the method can accurately register the panoramic image and the mobile LiDAR data, and the average registration error is less than three pixels.

Compared with other registration methods, the method in this paper does not need to rely on specific environmental scenarios and has strong applicability. By designing a corresponding feature points extraction algorithm based on the hybrid geometric structure index feature of phase, we can obtain the corresponding feature points with a uniform distribution and in an abundant quantity. These well-distributed and abundant corresponding feature points are the keys for us to complete the registration of mobile LiDAR data and panoramic images. At the same time, when analyzing the three test data in this paper, we found that the difference in point cloud quality affected our matching results. The better the image quality of point cloud imaging, the higher the registration accuracy. Low-quality point cloud imaging images will bring greater challenges to the registration algorithm.

In future work, we will consider further reducing the registration time. For sequential panoramic images, we can use the relative motion relationship between sequential panoramic images to reduce the number of registrations between mobile LiDAR data and panoramic images. However, a more robust and fast corresponding feature points extraction algorithm will help us reduce the registration complexity, which would enable the better and faster registration of panoramic images and mobile LiDAR data.

Author Contributions: G.W. carried out the empirical studies and the literature review and drafted the manuscript; Y.W., T.W., R.Z. (Ruizhuo Zhang), N.Z. and R.Z. (Ruofei Zhong) helped to draft and review the manuscript and communicated with the editor of the journal. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Cornelis, N.; Leibe, B.; Cornelis, K.; Van Gool, L. 3D urban scene modeling integrating recognition and reconstruction. Int. J. Comput. Vis. 2008, 78, 121–141. [CrossRef]
2. Armenakis, C.; Gao, Y.; Sohn, G. Co-registration of aerial photogrammetric and LiDAR point clouds in urban environments using automatic plane correspondence. Appl. Geomat. 2013, 5, 155–166. [CrossRef]
3. Megahed, Y.; Yan, W.Y.; Shaker, A. A phase-congruency-based scene abstraction approach for 2d-3d registration of aerial optical and LiDAR images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2020, 14, 964–981. [CrossRef]
4. Hou, M.; Li, S.K.; Jiang, L.; Wu, Y.; Hu, Y.; Yang, S.; Zhang, X. A new method of gold foil damage detection in stone carving relics based on multi-temporal 3D LiDAR point clouds. ISPRS Int. J. Geo-Inf. 2016, 5, 60. [CrossRef]
5. Feraco, S.; Bonfitto, A.; Amati, N.; Tonoli, A. A LiDAR-Based Clustering Technique for Obstacles and Lane Boundaries Detection in Assisted and Autonomous Driving. In Proceedings of the ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, St. Louis, MO, USA, 16–19 August 2020.
6. Takeuchi, E.; Yoshihara, Y.; Yoshiki, N. Blind Area Traffic Prediction Using High Definition Maps and LiDAR for Safe Driving Assist. In Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC), Las Palmas, Spain, 15–18 September 2015; pp. 2311–2316.
Huang, J.; Zhang, X.; Xin, Q.; Sun, Y.; Zhang, P. Automatic building extraction from high-resolution aerial images and LiDAR data using gated residual refinement network. *ISPRS J. Photogramm. Remote Sens.* **2019**, *151*, 91–105. [CrossRef]

Wu, Q.; Yang, H.; Wei, M.; Remil, O.; Wang, B.; Wang, J. Automatic 3D reconstruction of electrical substation scene from LiDAR point cloud. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 57–71. [CrossRef]

Ouyang, Z.; Liu, Y.; Zhang, C.; Niu, J. A cg-based scene reconstruction model using lidar point cloud. In *Proceedings of the 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, Guangzhou, China, 12–15 December 2017; pp. 1107–1114.

Boehm, J.; Becker, S. Automatic Marker-free Registration of Terrestrial Laser Scans using Reflectance Features. In *Proceedings of the 8th Conference Optical 3-D Measurement Techniques*, Zurich, Switzerland, 9–12 July 2007; Volume I, pp. 338–343.

Yang, B.; Dong, Z. A shape-based segmentation method for mobile laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens.* **2013**, *81*, 19–30. [CrossRef]

Chen, X.; Ma, H.; Wan, J.; Li, B.; Xia, T. Multi-view 3D object detection network for autonomous driving. In *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017; Volume 1, p. 3.

Schlosser, J.; Chow, C.K.; Kira, Z. Fusing LiDAR and images for pedestrian detection using convolutional neural networks. In *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, 16–21 May 2016.

Premebida, C.; Carrera, J.; Batista, J.; Nunes, U. Pedestrian Detection Combining RGB and Dense LiDAR Data. In *Proceedings of the International Conference on Intelligent Robots and Systems*, Chicago, IL, USA, 14 September 2014.

Telling, J.; Lyda, A.; Hartzell, P.; Glennie, C. Review of earth science research using terrestrial laser scanning. *Earth Sci. Rev.* **2017**, *169*, 35–68. [CrossRef]

Li, J.; Yang, B.; Chen, C.; Huang, R.; Dong, Z.; Xiao, W. Automatic registration of panoramic image sequence and mobile laser scanning data using semantic features. *ISPRS J. Photogramm. Remote Sens.* **2018**, *136*, 41–57. [CrossRef]

Peng, S.; Ma, H.; Zhang, L. Automatic Registration of Optical Images with Airborne LiDAR Point Cloud in Urban Scenes Based on Line-Point Similarity Invariant and Extended Collinearity Equations. *Sensors* **2019**, *19*, 1086. [CrossRef]

Zhu, B.; Ye, Y.; Zhou, L.; Li, Z.; Yin, G. Robust registration of aerial images and LiDAR data using spatial constraints and Gabor structural features. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 129–147. [CrossRef]

Parmehr, E.G.; Fraser, C.S.; Zhang, C.; Leach, J. Automatic registration of optical imagery with 3D LiDAR data using statistical similarity. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 28–40. [CrossRef]

Shi, W.; Gong, Y.; Yang, M.; Liu, T. Point Cloud Depth Map and Optical Image Registration Based on Improved RIFT Algorithm. In *Proceedings of the 2021 IEEE 13th International Conference on Computer Research and Development (ICCRD)*, Beijing, China, 5–7 January 2021.

Taylor, Z.; Nieto, J. Automatic calibration of lidar and camera images using normalized mutual information. In *Proceedings of the 2013 IEEE Conference on Robotics and Automation (ICRA 2013)*, Karlsruhe, Germany, 6–10 May 2013.

Wang, R.; Ferrie, F.P.; Macfarlane, J. Automatic registration of mobile lidar and spherical panoramas. In *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Providence, RI, USA, 18–20 June 2012; pp. 33–40.

Ye, Y.; Shan, J. A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences. *ISPRS J. Photogramm. Remote Sens.* **2014**, *90*, 83–95. [CrossRef]

Shao, J.; Zhang, W.; Zhu, Y.; Shen, A. Fast registration of terrestrial LiDAR point cloud and sequence images. In *Proceedings of the International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences—ISPRS Archives*, Wuhan, China, 18–22 September 2017; Volume 42, pp. 875–879.

Zhu, N.; Jia, Y.; Ji, S. Registration of Panoramic/Fish-Eye Image Sequence and LiDAR Points Using Skyline Features. *Sensors* **2018**, *18*, 1651. [CrossRef] [PubMed]

Cui, T.; Ji, S.; Shan, J.; Gong, J.; Liu, K. Line-based registration of panoramic images and LiDAR point clouds for mobile mapping. *Sensors* **2017**, *17*, 70. [CrossRef] [PubMed]

Zhu, N.; Jia, Y.; Huang, X. Semiautomatically register MMS LiDAR points and panoramic image sequence using road lamp and lane. *Photogramm. Eng. Remote Sens.* **2019**, *85*, 829–840. [CrossRef]

Kaminsky, R.S.; Snively, N.; Seitz, S.T.; Szeliski, R. Alignment of 3D Point Clouds to Overhead Images. In *Proceedings of the Second IEEE Workshop on Internet Vision*, Miami, FL, USA, 20–25 June 2009.

Zheng, S.; Huang, R.; Zhou, Y. Registration of optical images with LiDAR data and its accuracy assessment. *Photogramm. Eng. Remote Sens.* **2013**, *79*, 731–741. [CrossRef]

Zhao, W.; Nister, D.; Hsu, S. Alignment of continuous video onto 3D point clouds. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1305–1318. [CrossRef]

Abayowa, B.O.; Yilmaz, A.; Hardie, R.C. Automatic registration of optical aerial imagery to a LiDAR point cloud for generation of city models. *ISPRS J. Photogramm.* **2015**, *106*, 68–81. [CrossRef]

Zhao, Y.; Wang, Y.; Tsai, Y. 2D-image to 3D-range registration in urban environments via scene categorisation and combination of similarity measurements. In *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, 16–21 May 2016.
33. Taylor, Z.; Nieto, J.; Johnson, D. Automatic calibration of multimodal sensor systems using a gradient orientation measure. In Proceedings of the IEEE International Conference on Intelligent Robots & Systems (IROS), Tokyo, Japan, 3–7 November 2013; pp. 1293–1300.

34. Yang, B.; Chen, C. Automatic registration of UAV-borne sequent images and LiDAR data. ISPRS J. Photogramm. Remote Sens. 2015, 101, 262–274. [CrossRef]

35. Magnusson, M.; Lilienthal, A.; Duckett, T. Scan registration for autonomous mining vehicles using 3D-NDT. J. Field Robot. 2007, 24, 803–827. [CrossRef]

36. Besl, P.; McKay, N. A Method for Registration of 3-D Shapes. IEEE Trans. Pattern Anal. Mach. Intell. 1992, 14, 239–256. [CrossRef]

37. Bae, K.-H.; Lichti, D.D. A method for automated registration of unorganised point clouds. ISPRS J. Photogramm. Remote Sens. 2008, 63, 36–54. [CrossRef]

38. Gressin, A.; Mallet, C.; Demantké, J.; David, N. Towards 3D lidar point cloud registration improvement using optimal neighborhood knowledge. ISPRS J. Photogramm. Remote Sens. 2013, 79, 240–251. [CrossRef]

39. Li, J.; Yang, B.; Chen, C.; Habib, A. NRLI-UAV: Non-rigid registration of sequential raw laser scans and images for low-cost UAV LiDAR point cloud quality improvement. ISPRS J. Photogramm. Remote Sens. 2019, 158, 123–145. [CrossRef]

40. Li, J.; Hu, Q.; Ai, M. RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform. IEEE Trans. Image Process. 2020, 29, 3296–3310. [CrossRef]

41. Horn, B.; Klaus, B.; Horn, P. Robot Vision; MIT Press: Cambridge, CA, USA, 1986.

42. Weickert, J.; ter Haar Romeny, B.M.; Viergever, M.A. Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Trans. Image Process. 1998, 7, 398–410. [CrossRef]

43. Ma, J.; Zhao, J.; Jiang, J.; Zhou, H.; Guo, X. Locality preserving matching. Int. J. Comput. Vis. 2019, 127, 512–531. [CrossRef]

44. Yao, Y.; Zhang, Y.; Wan, Y.; Liu, X.; Guo, H. Heterologous Images Matching Considering Anisotropic Weighted Moment and Absolute Phase Orientation. Geomat. Inf. Sci. Wuhan Univ. 2021, 46, 1727–1736.