

## Automatic registration of panoramic image sequence and mobile laser scanning data using semantic features

Jianping Li <sup>a</sup>, Bisheng Yang <sup>a,\*</sup>, Chi Chen <sup>a</sup>, Ronggang Huang <sup>b</sup>, Zhen Dong <sup>a</sup>, Wen Xiao <sup>c</sup>

<sup>a</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

<sup>b</sup> Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan, China

<sup>c</sup> School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, UK



### ARTICLE INFO

#### Article history:

Received 3 July 2017

Received in revised form 28 November 2017

Accepted 9 December 2017

#### Keywords:

Panoramic image sequence

Mobile laser scanning data

Semantic features

Registration

### ABSTRACT

Inaccurate exterior orientation parameters (EoPs) between sensors obtained by pre-calibration leads to failure of registration between panoramic image sequence and mobile laser scanning data. To address this challenge, this paper proposes an automatic registration method based on semantic features extracted from panoramic images and point clouds. Firstly, accurate rotation parameters between the panoramic camera and the laser scanner are estimated using GPS and IMU aided structure from motion (SfM). The initial EoPs of panoramic images are obtained at the same time. Secondly, vehicles in panoramic images are extracted by the Faster-RCNN as candidate primitives to be matched with potential corresponding primitives in point clouds according to the initial EoPs. Finally, translation between the panoramic camera and the laser scanner is refined by maximizing the overlapping area of corresponding primitive pairs based on the Particle Swarm Optimization (PSO), resulting in a finer registration between panoramic image sequences and point clouds. Two challenging urban scenes were experimented to assess the proposed method, and the final registration errors of these two scenes were both less than three pixels, which demonstrates a high level of automation, robustness and accuracy.

© 2017 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

## 1. Introduction

A Mobile Mapping System (MMS), usually equipped with laser scanners, panoramic cameras, Inertial Measurement Unit (IMU) and Global Positioning System (GPS) for acquisitions of images, point clouds and orientations, has been widely used for street inventory. The integration of complementary ranging and imaging data such as laser-scanning and electro-optical sensors provides new solutions to map the Earth's surface. The acquired datasets include rich spectrum and geometry information. Image-to-range registration is a critical step for many applications, such as texturing 3D models of large-scale scenes (Abayowa et al., 2015), building extraction (Yang et al., 2013), forest biomass inventory (Pflugmacher et al., 2014), and point cloud classification using the color information (Barnea and Filin, 2013; Yang and Dong, 2013).

Extrinsic calibration of the panoramic camera and the laser scanner is usually performed before data collection to achieve image-to-range alignment. However, there is often a considerable

misalignment between the panoramic images and laser scanning data according to the calibrated extrinsic parameters (Miled et al., 2016; Cui et al., 2017). The main reason for the misalignment is the unforeseen movement of sensors (Levinson and Thrun, 2013). Since the mounting of the sensors will become less stable over a certain period of time, the relative extrinsic parameters should be calibrated manually frequently, which is laborious (Brenner, 2014).

An automated method is proposed in this paper to accurately register panoramic image sequence and mobile laser-scanning point clouds in an urban environment by estimating the transformation parameters between the panoramic camera and laser scanner using parked vehicles as registration primitives. Firstly, the original EoPs of panoramic images are adjusted in a GPS and IMU aided panoramic structure from motion (SfM). The accurate rotation between the panoramic camera and the laser scanner is obtained. Secondly, parked vehicles are extracted from both panoramic images and point clouds. To extract vehicles in panoramic images, the Faster-RCNN (Ren et al., 2015) is used to detect vehicle candidate areas, which are then segmented based on CRFASRNN (Zheng et al., 2015) and refined by image matting (Levin et al., 2008). Next, corresponding areas of the extracted

\* Corresponding author.

E-mail address: [bshyang@whu.edu.cn](mailto:bshyang@whu.edu.cn) (B. Yang).

vehicles are determined in the point clouds according to the initial EoPs of the panoramic images obtained in the initialization step. Then, vehicles in those corresponding areas of the point clouds are extracted. Finally, the relative translation between the panoramic camera and the laser scanner is refined by maximizing the overlapping area of the vehicle primitive pairs based on the Particle Swarm Optimization (PSO). The main contribution of the proposed method is that it refines the initial bundle adjustment based alignment by the Particle Swarm Optimization utilizing semantic features, resulting in an automatic and accurate registration between panoramic image sequence and mobile laser-scanning point clouds.

The remainder of this paper is organized as follows. Related literature is reviewed in Section 2. The proposed method is elaborated in Section 3. In Section 4, the experimental studies are undertaken to evaluate the proposed method, after which discussions are presented and conclusions are drawn at the end.

## 2. Literature review

Extensive studies on the registration of airborne laser-scanning data and aerial frame-images have been reported, including feature-based methods and area-based methods (Parmehr et al., 2014). Typical features used in feature-based methods are points (Palenichka and Zaremba, 2010), lines (Habib et al., 2005; Lee and Yu, 2006; Wang and Neumann, 2009), and planes (Kwak et al., 2006; Yang and Chen, 2015). Area-based approaches are usually relying on maximizing the statistical correlation (e.g., mutual information) or grayscale similarity (Parmehr et al., 2014). However, methods designed for airborne platform registration may not be applicable to the registration between terrestrial platforms, due to the substantial difference in scales, complexities, scanning perspectives, etc. For example, airborne and vehicle-borne laser scanning platforms clearly differ in data capture mode, typical project size, scanning orientation, and spatial resolution. Moreover, compared with images captured by aerial photogrammetry systems, images captured by panoramic cameras mounted on a vehicle are more complicated and challenging for registration due to occlusions caused by moving objects or nearby objects, smaller overlapping rate, and drastic changes in depth of view.

In fact, many methods have been explored to register mobile laser-scanning data and panoramic images, including features-based methods, statistical-based methods, and multi-view-based methods, which are reviewed in detail as follows.

**Features-based methods:** Böhm and Becker (2007) adopted the Scale-Invariant Feature Transform (SIFT) (Lowe, 2004) feature detector to extract the corresponding point features to register optical imagery and range images. W. Moussa et al. (2012) achieved registration of close-range images with TLS data by utilizing the affine SIFT (ASIFT) (Morel and Yu, 2009) feature matching strategy based on point-based environment model. Lines, bounding boxes, and skylines were also adopted by many studies for registration (Christy and Horaud, 1999; Liu and Stamos, 2005; Ramalingam et al., 2009). The key factor for feature-based registration is the strategy to find the correct matching feature pairs from point clouds and images.

**Statistical-based methods:** This kind of method aims to optimize the initial EoPs of images by maximizing the statistical correlation between images and point clouds. Miled et al. (2016) optimized the inaccurate transformation between camera and laser scanner by maximizing the mutual information (MI) between these two kinds of data resulting in an accurate online calibration. An extension of MI, normalized mutual information (NMI) was introduced by Wang et al. (2012) to register panoramic image

and point clouds. Corsini et al. (2013) used a global refinement algorithm based on MI to optimize the color projection of aligned photos on the 3D object. The definition of mutual information can significantly affect registration result. When computing the MI, Yipu et al. (2016) set weights to different data sources according to different scenes, which improved the robustness of registration. However, the correlation between depth information and color information on which MI-based methods are built is not significant in many scenes (Wang et al., 2012).

**Multi-view-based methods:** This kind of method generates image point clouds from adjacent images using SfM and multi-view stereo (MVS), transforming the 2D-3D registration to a 3D-3D registration. The point clouds generated by photogrammetry are influenced by drift, leading to non-rigid transformation between photogrammetry point clouds and laser scanning data.

Generally, the complexity of urban environment scenes poses challenges for robust registration, especially for the registration methods based on points and linear features, which have difficulties in finding correspondences between panoramic images and point clouds. To overcome the difficulties, parked vehicles are proposed as registration features and corresponding vehicle candidates are paired automatically to fulfill the registration. The advantages of using parked vehicles are that they are often found in an urban environment and are relatively easy to be recognized in images and point clouds. Moreover, street-side vehicles in both panoramic images and point clouds are foreground objects, which are rarely occluded by other objects, meaning reliable detection and segmentation can be expected.

## 3. Methodology

This aim is to accurately estimate the transformation parameters, i.e. rotation and translation parameters, between the panoramic camera and the laser scanner. The proposed method consists of three main steps: (1) Accurate rotation parameters estimation by GPS/IMU aided SfM bundle adjustment. (2) Parked vehicle extraction from point clouds and panoramic images. (3) Accurate translation parameters estimation by maximizing the overlapping area of corresponding pairs. The key steps of the proposed method are illustrated in Fig. 1 and elaborated as follows.

### 3.1. Notations and definitions for the registration model of panoramic images and point clouds

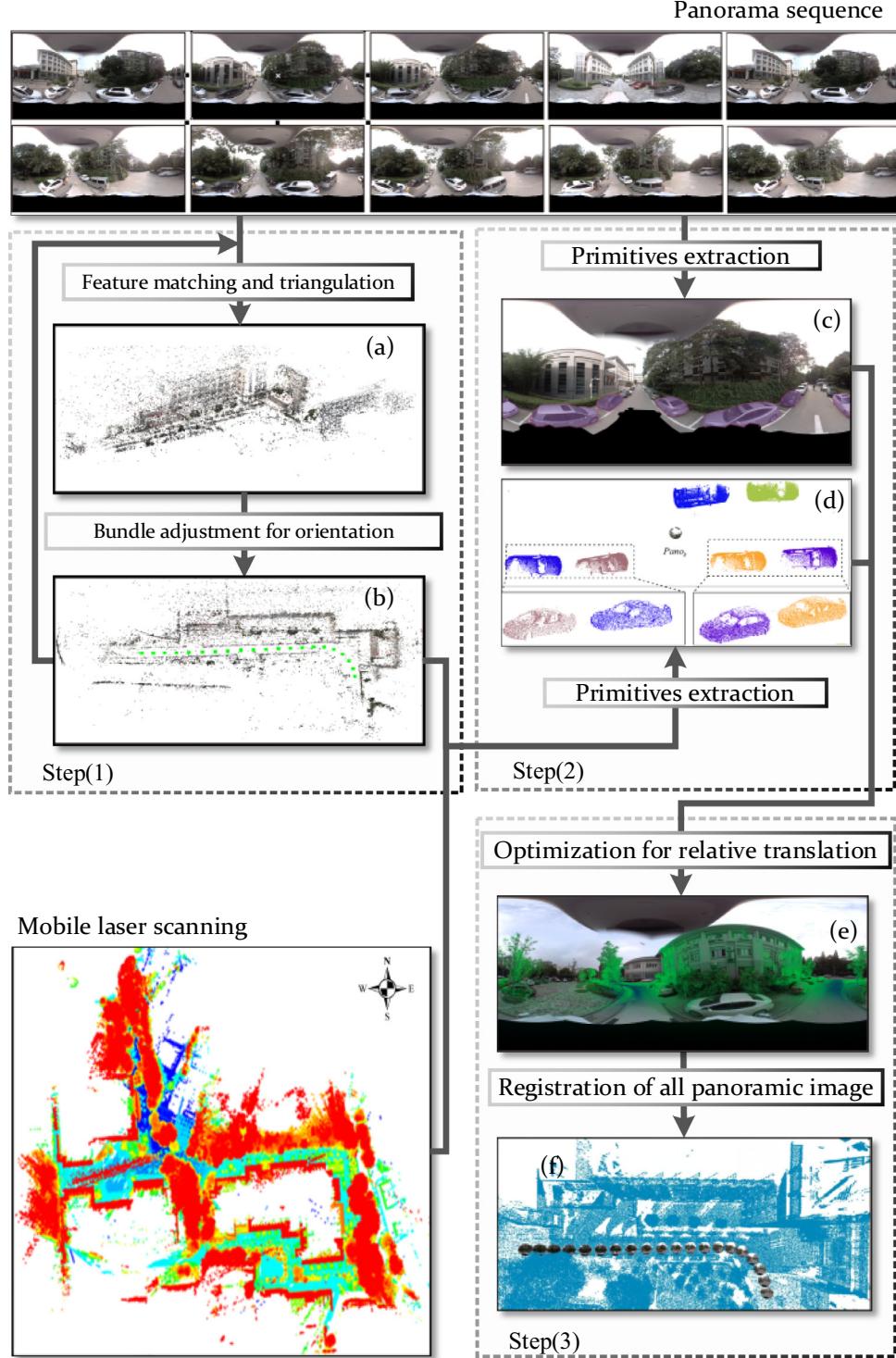
We employ the following notations proposed by Barfoot (2016):  $\mathcal{F}_A$  denotes a reference frame A; a point  $P$  in frame  $\mathcal{F}_A$  is written as a vector  $\mathbf{r}_A^{PA}$ .  $\mathbf{T}_{AB}$  denotes the transformation between  $\mathcal{F}_A$  and  $\mathcal{F}_B$ , and the rotation matrix is represented by  $\mathbf{C}_{AB}$ ; the corresponding quaternion is represented by  $\mathbf{q}_{AB}$ .

The coordinate system of the proposed registration model is shown in Fig. 2.  $\mathcal{F}_W$  denotes the reference frame of the point cloud,  $\mathcal{F}_{C_k}$  denotes the reference frame of the  $k^{\text{th}}$  panoramic image, and  $\mathcal{F}_{\tilde{C}_k}$  denotes the reference frame with the original EoPs of the  $k^{\text{th}}$  panoramic image. The original EoPs of panoramic images are calculated by the pre-calibrated transformation with respect to the laser scanner, and are not accurate due to unforeseen sensor movements.

The EoPs of the  $k^{\text{th}}$  panoramic image is written as  $\mathbf{x}_{C_k}$ :

$$\mathbf{x}_{C_k} := [\mathbf{r}_W^{C_k W^T}, \mathbf{q}_{WC_k}^T]^T \in \mathbb{R}^3 \times SO^3 \quad (1)$$

where  $SO^3$  is the Special Orthogonal Group. The inaccurate EoPs of the  $k^{\text{th}}$  panoramic image is written as  $\mathbf{x}_{\tilde{C}_k}$ :



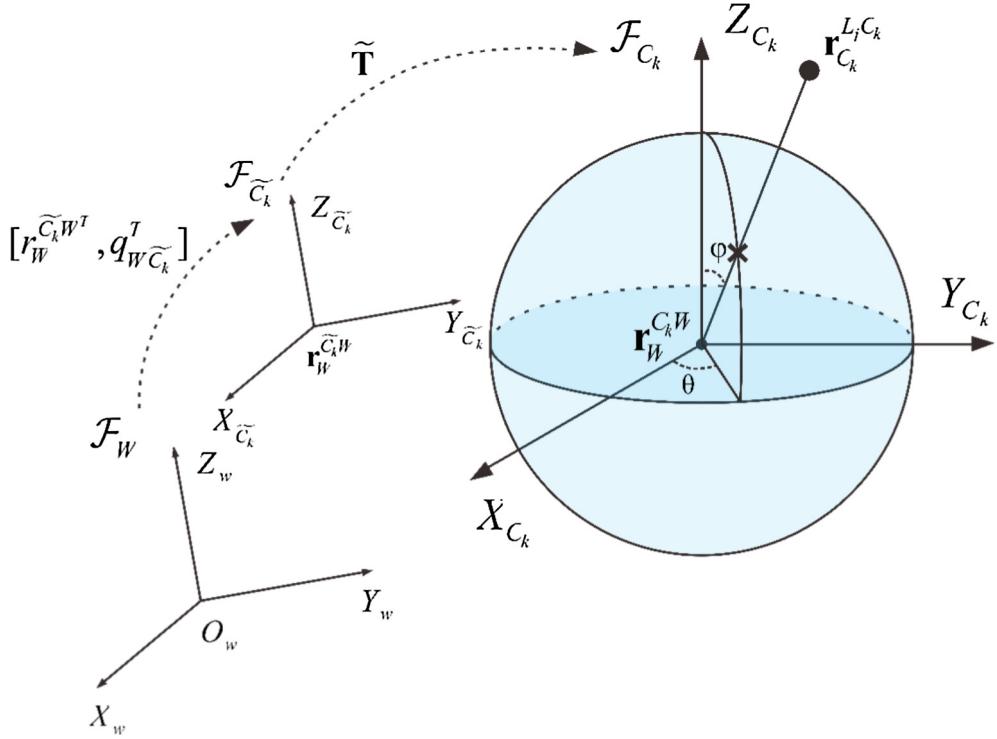
**Fig. 1.** Framework of the registration between laser scanning data and panoramic images. (a) The result of local SfM in which relative EoPs are obtained and feature points are triangulated. (b) The result of absolute orientation by GPS/IMU aided bundle adjustment. (c) Vehicle extraction in a panoramic image. (d) Vehicle primitives in point clouds. (e) The result of reprojected point cloud on a key-frame according to the accurate relative translation. (f) All panoramic spheres rendered in point cloud according to the final accurate transformation.

$$\mathbf{x}_{\tilde{C}_k} := [\tilde{\mathbf{r}}_W^{C_k W^T}, \mathbf{q}_{W \tilde{C}_k}^T]^T \in \mathbb{R}^3 \times SO^3 \quad (2)$$

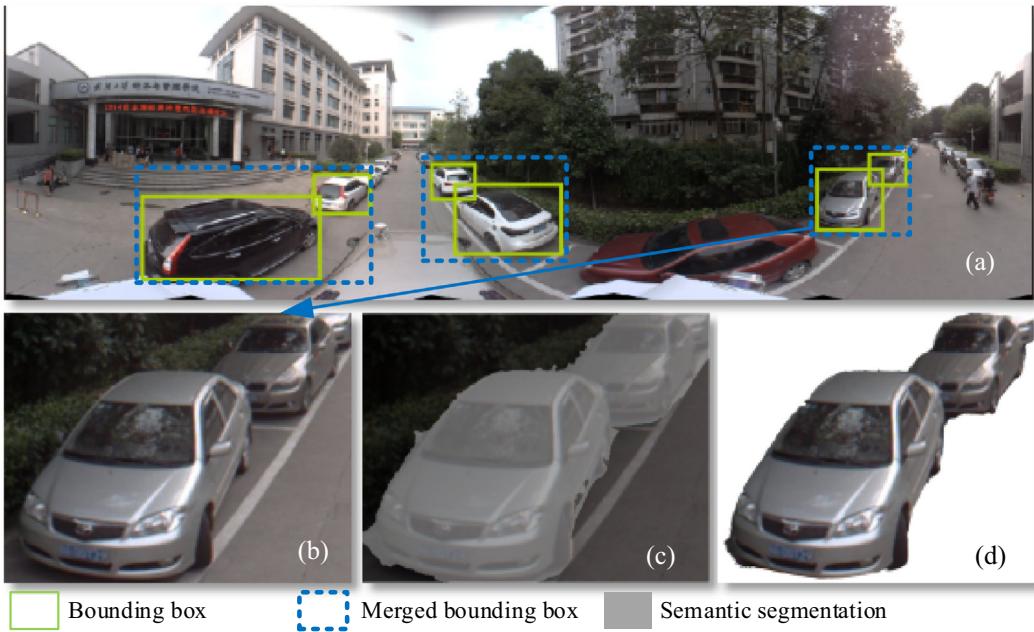
3D feature points are obtained by the SfM, and the  $j^{\text{th}}$  3D feature point is written as  $\mathbf{x}_{l_j}$ :

$$\mathbf{x}_{l_j} := \mathbf{r}_W^{L_j W} \in \mathbb{R}^3 \quad (3)$$

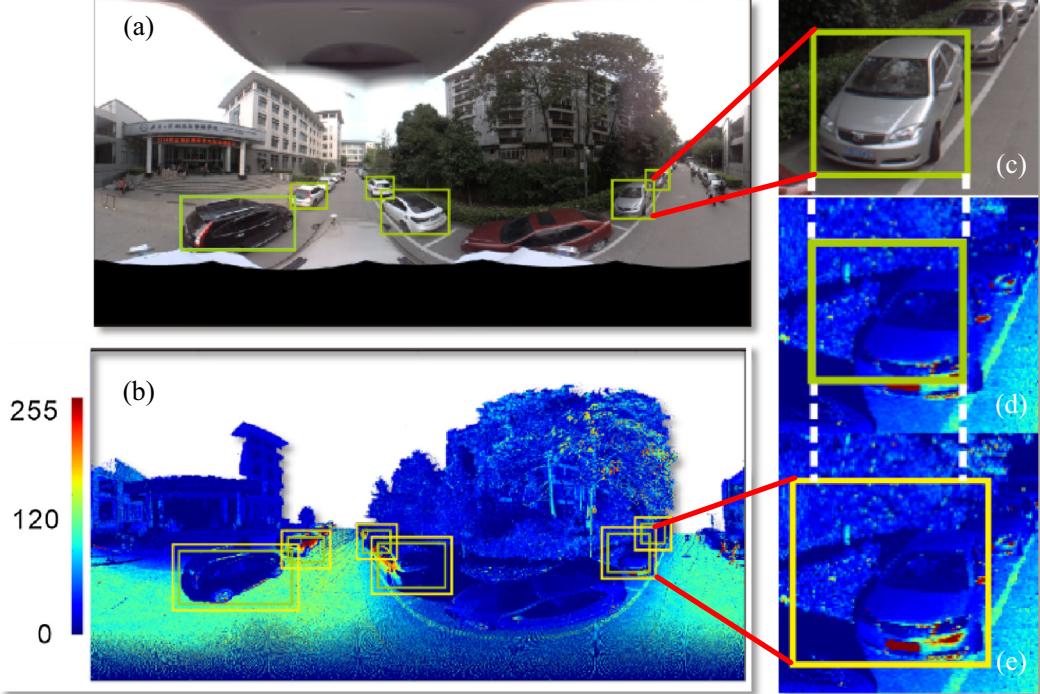
The projection of 3D feature points onto the corresponding image is represented by a spherical coordinate  $(\theta, \varphi)$ . The goal is to estimate the accurate transformation between  $\mathcal{F}_{C_k}$  and



**Fig. 2.** Illustration of the registration model.  $\mathcal{F}_W$  denotes the reference frame of the point cloud.  $\mathcal{F}_{\tilde{C}_k}$  denotes the reference frame with the original imperfect EoPs of the  $k^{\text{th}}$  panoramic image.  $\mathcal{F}_{C_k}$  denotes the accurate reference frame of the  $k^{\text{th}}$  panoramic image.  $\tilde{\mathbf{T}}$  denotes the transformation between  $\mathcal{F}_{\tilde{C}_k}$  and  $\mathcal{F}_{C_k}$ .  $\mathbf{r}_{C_k}^{L_j C_k}$  is the  $j^{\text{th}}$  3D feature point in the reference frame of the  $k^{\text{th}}$  panoramic image, and it can be projected onto the panoramic image using spherical coordinates  $(\theta, \varphi)$ .



**Fig. 3.** Extraction of vehicle primitives in the panoramic image. (a) Vehicle detection using FASTER-RCNN. The green boxes are the detected areas of vehicles. The blue boxes represent the areas of merged overlapping bounding boxes, which are treated as one registration primitive. (b) A merged bounding box. (c) Semantic segmentation result in the merged bounding box. (d) Segment refinement result using the matting technique. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Detection of candidate vehicles from point clouds. (a) Detected vehicle in an image. (b) Projection of points onto the image rendered by intensity. (c) An original vehicle bounding box in the image. (d) The vehicle bounding box on the point cloud projection. (e) Expanded bounding box on the point cloud projection.

$\mathcal{F}_{C_k}$ . During the short time span in the data acquisition process, the relative orientations and translations between the laser scanner and the camera are regarded as unchanged. Hence, the transformation  $T_{C_k C_k}$  is simplified as  $\tilde{T}$ :

$$\tilde{T} := T_{C_1 C_1} \sim = \dots = T_{C_k C_k}, \quad k = 1, \dots, K. \quad (4)$$

$$\tilde{T} = \begin{bmatrix} \tilde{\mathbf{c}} & \tilde{\mathbf{r}} \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (5)$$

where  $\tilde{\mathbf{c}}$  and  $\tilde{\mathbf{r}}$  is the compensatory rotation matrix and compensatory translation vector of the registration model. The transformation equation between  $\mathcal{F}_W$  and  $\mathcal{F}_{C_k}$  is written as:

$$\mathbf{r}_{C_k}^{L_j C_k} = \tilde{\mathbf{c}} \cdot \mathbf{c}_{C_k W} \cdot (\mathbf{r}_W^{L_j W} - \tilde{\mathbf{r}}_W^{C_k W}) + \tilde{\mathbf{r}} \quad (6)$$

The proposed method solves the transformation  $\tilde{T}$  by estimating the compensatory rotation matrix  $\tilde{\mathbf{c}}$  and the compensatory translation vector  $\tilde{\mathbf{r}}$  separately.

### 3.2. Initialization of EoPs by estimating compensatory rotation

In this initialization step, the compensatory rotation matrix  $\tilde{\mathbf{c}}$  of the registration model is estimated based on the GPS and IMU aided SfM. While the 3D location accuracy obtained by GPS and IMU aided SfM is to be improved. Even if the compensatory translation vector  $\tilde{\mathbf{r}}$  is considered in the initialization step, the resulting vector  $\tilde{\mathbf{r}}$  is not accurate enough. In another word, compared with the compensatory rotation matrix  $\tilde{\mathbf{c}}$ , compensatory translation vector  $\tilde{\mathbf{r}}$  (centimeter level) has less effect on the SfM results. In order to speed up the convergence of the GPS and IMU aided bundle adjustment, compensatory translation vector  $\tilde{\mathbf{r}}$  is set to zero vector

in this initialization step and it will be refined in the fine registration step. The GPS and IMU aided panoramic SfM consists of two steps: (1) feature matching and triangulation using a panoramic model (Pagani and Stricker, 2011); (2) GPS and IMU aided bundle adjustment. The two steps above are carried out to obtain the initial EoPs of panoramic images and the compensatory rotation matrix  $\tilde{\mathbf{c}}$ .

The GPS and IMU aided bundle adjustment is performed to reduce the drift of SfM when a new panoramic image is added to the SfM system. The cost function  $J(\mathbf{x})$  for the bundle adjustment consists of three parts: back-projection error term  $\mathbf{e}_r$ , IMU error term  $\mathbf{e}_s$  and GPS error term  $\mathbf{e}_g$ :

$$J(\mathbf{x}) := \underbrace{\sum_{k=1}^K \sum_{j \in \mathcal{J}(k)} \mathbf{e}_r^{j,kT} \mathbf{W}_r^{j,k} \mathbf{e}_r^{j,k}}_{visual} + \underbrace{\sum_{k=1}^K \mathbf{e}_s^{kT} \mathbf{W}_s^k \mathbf{e}_s^k}_{IMU} + \underbrace{\sum_{k=1}^K \mathbf{e}_g^{kT} \mathbf{W}_g^k \mathbf{e}_g^k}_{GPS} \quad (7)$$

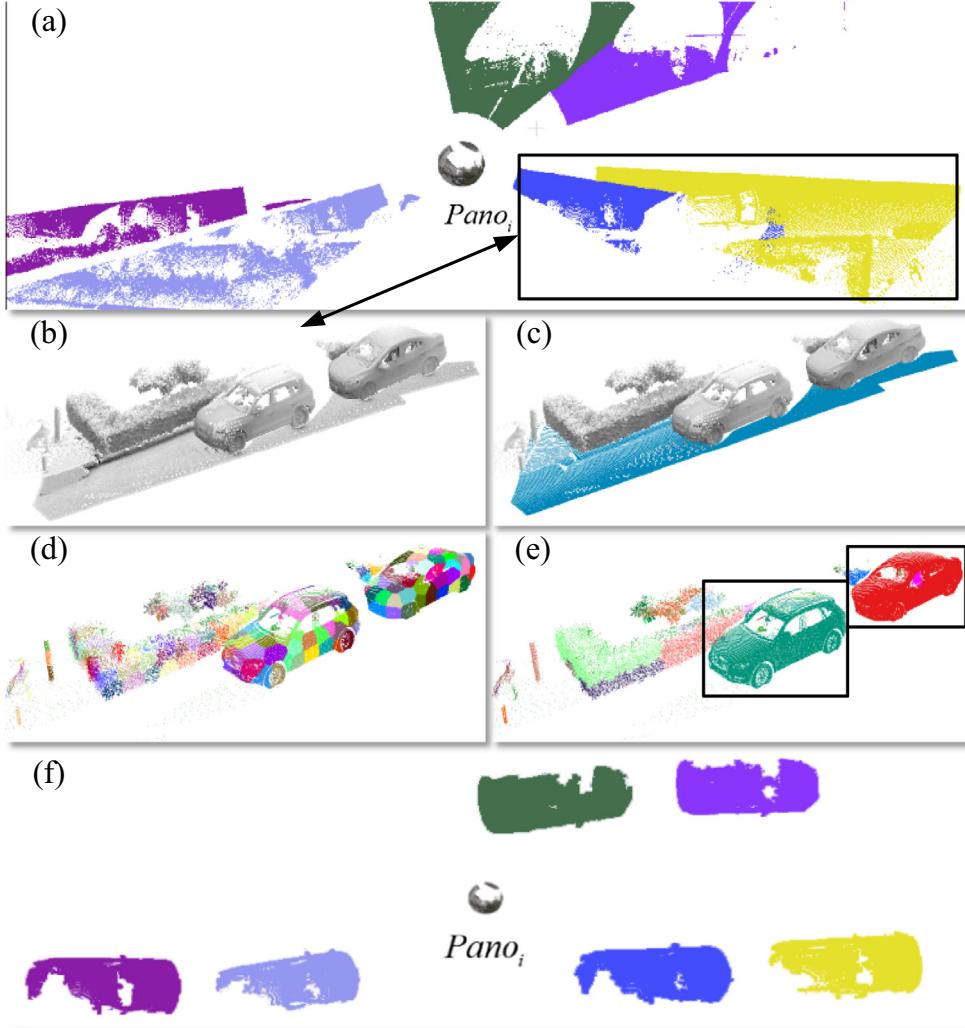
where, if the  $j^{\text{th}}$  3D feature point obtained by the panoramic SfM can be observed in the  $k^{\text{th}}$  panoramic image,  $j$  belongs to set  $\mathcal{J}(k)$ .  $\mathbf{W}_r^{j,k}$ ,  $\mathbf{W}_s^k$  and  $\mathbf{W}_g^k$  are information matrices for the back-projection error, IMU error and GPS error, respectively.

The General Framework for Graph Optimization (G2O) (Kümmerle et al., 2011) is used to optimize the above cost function. The definition of the error equation is described in detail below.

It is assumed that the panoramic camera is calibrated, and there is no distortion in the images. Back-projection error  $\mathbf{e}_r^{j,k}$  is written as:

$$\mathbf{e}_r^{j,k} = \mathbf{f}(\mathbf{z}^{j,k}) - \mathbf{h}\left(\mathbf{c}_{C_k W}(\mathbf{r}_W^{L_j W} - \mathbf{r}_W^{C_k W})\right) \quad (8)$$

where  $\mathbf{h}(\cdot)$  is the projection model of panorama, and it transfers Cartesian coordinates  $(x, y, z)$  to spherical coordinates  $(\theta, \varphi)$  by:



**Fig. 5.** Vehicle extraction in the point cloud. (a) Bird's eye view of vehicle candidates detected in the  $i^{\text{th}}$  panoramic image. Each color shows candidate vehicle points selected in one bounding box. (b) Candidate points in the black box of (a). (c) Ground points elimination. Blue points are ground points and grey points are non-ground points. (d) 3-d voxels of the non-ground points (One color means one super voxel). (e) Segmentation and determination of vehicle segments (One color means one merged segment. The segments in black boxes are the selected vehicle segments). (f) Vehicles extracted from the point cloud in the field of view of the  $i^{\text{th}}$  panoramic image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$r = \sqrt{x^2 + y^2 + z^2}, \quad \theta = \tan^{-1}\left(\frac{y}{x}\right), \quad \varphi = \cos^{-1}\left(\frac{z}{r}\right) \quad (9)$$

$\mathbf{z}^{j,k}$  is the measurement of  $j^{\text{th}}$  feature point on the  $k^{\text{th}}$  panoramic image.  $\mathbf{f}(\cdot)$  transfers the image coordinates  $(x, y)$  of measurement to spherical coordinates  $(\theta, \varphi)$  by:

$$\theta = \pi - \frac{x}{\text{image width}} * 2\pi, \quad \varphi = \frac{y}{\text{image height}} * \pi \quad (10)$$

Error term of IMU is written as:

$$\mathbf{e}_s^k = [\mathbf{q}_{WC_k} \otimes \tilde{\mathbf{q}} \otimes \mathbf{q}_{C_k W}]_{1:3} \quad (11)$$

Error term of GPS is written as:

$$\mathbf{e}_g^k = \mathbf{r}_W^{C_k W} - (\mathbf{C}_{WC_k} \cdot \tilde{\mathbf{c}}^T \cdot \tilde{\mathbf{r}} + \mathbf{r}_W^{C_k W}) \quad (12)$$

In the process of initializing the EoPs, the compensatory translation vector  $\tilde{\mathbf{r}}$  is set to  $0_{3 \times 1}$ , so  $\mathbf{e}_g^k$  can be rewritten as:

$$\mathbf{e}_g^k = \mathbf{r}_W^{C_k W} - \tilde{\mathbf{r}} \quad (13)$$

### 3.3. Extraction of parked vehicle pairs

To detect parked vehicles from panoramic images, vehicle bounding boxes are firstly detected, then segmented to filter the background, followed by a vehicle segment refinement.

The FASTER-RCNN (Ren et al., 2015) was adopted to identify vehicle bounding boxes, described as  $\text{rect}^{\text{origin}} = \{[x, y]^T, w, h\}$ , where  $[x, y]^T$  is the center of the rectangle, and  $w, h$  are the width and height of the rectangle. In the case that bounding boxes overlap, vehicles in these bounding boxes are treated as one registration primitive writing as  $\rho_{\text{image}}$ . A minimum bounding box are obtained and segmented. A semantic segmentation approach, CRFASRNN (Zheng et al., 2015), is utilized to segment vehicles out of the background in the detected bounding boxes. The segments are then refined using the matting technique proposed by Levin et al. (2008). As shown in Fig. 3(a), the green rectangles are bounding boxes of detected vehicle, and the blue dashed rectangles are merged bounding boxes. Fig. 3(b) shows an example of a vehicle candidate area. The segmentation result and refinement result are shown in Fig. 3(c) and (d).

To find corresponding vehicles in the point cloud, points are projected onto the panoramic image using the initial associated EoPs. The points located in the detected bounding boxes from the image are labeled as candidate vehicle points. Due to the inaccurate initial EoPs, points of a vehicle candidate in the point cloud can fall out of the rectangle  $\text{rect}^{\text{origin}} = \{[x, y]^T, w, h\}$ . Thus, the rect-

angle  $\text{rect}^{\text{origin}}$  is expanded to  $\text{rect}^{\text{expanded}} = \{[x, y]^T, w + d, h + d\}$ , where  $d$  is the buffer size, which is set to 45 pixels, as shown in Fig. 4. As an example, Fig. 5(a) shows the result of corresponding vehicle detection in the point cloud by projecting points onto the  $i^{\text{th}}$  panoramic image.

The ground points of the candidate vehicle points are removed using the approach proposed by Hernandez and Marcotegui (2009) before the extraction of vehicle points as shown in Fig. 5(c). Then the method from Yang et al. (2015) is used to extract the vehicle points in non-ground points as follows: Firstly, 3-d voxels of the non-ground points are generated as shown in Fig. 5(d). Secondly, the 3-d voxels are segmented by a graph-based segmentation method (Felzenszwalb and Huttenlocher, 2004), where adjacent segments of planar geometric structure are merged according to the distance between their centroids. The merged segments are shown in Fig. 5(e). Thirdly, the following constraints are imposed to detect vehicle segment from the merged segments:

- (i) the length ( $OL$ ), width ( $OW$ ), and height ( $OH$ ) of the merged segment should meet:

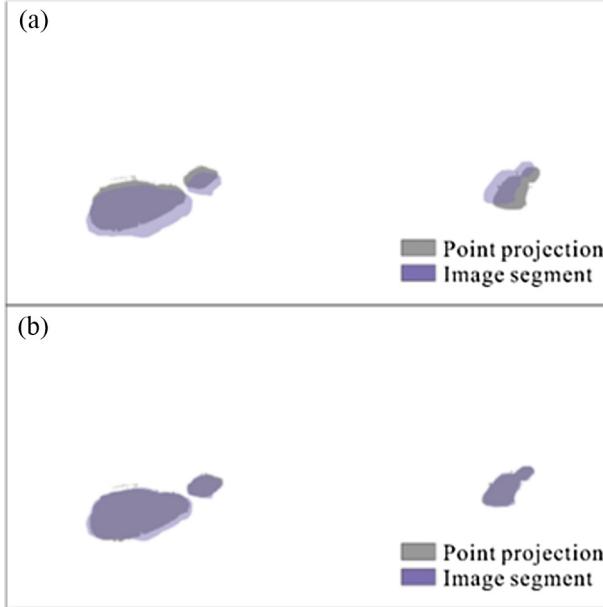
$$\begin{aligned} OL &> T_{l,\min} \&\& OL < T_{l,\max}, \\ OW &> T_{w,\min} \&\& OW < T_{w,\max}, \\ OH &> T_{h,\min} \&\& OH < T_{h,\max}, \end{aligned} \quad (14)$$

where,  $T_{l,\min}, T_{l,\max}, T_{w,\min}, T_{w,\max}, T_{h,\min}, T_{h,\max}$  are all size parameters;

- (ii) at least  $T_N$  planar segments should be included in one merged segment.

As shown in Fig. 5(e), segments in black box are the selected vehicle segment. Fig. 5(f) shows vehicles extracted from a point cloud in the field of view of the  $i^{\text{th}}$  panoramic image.

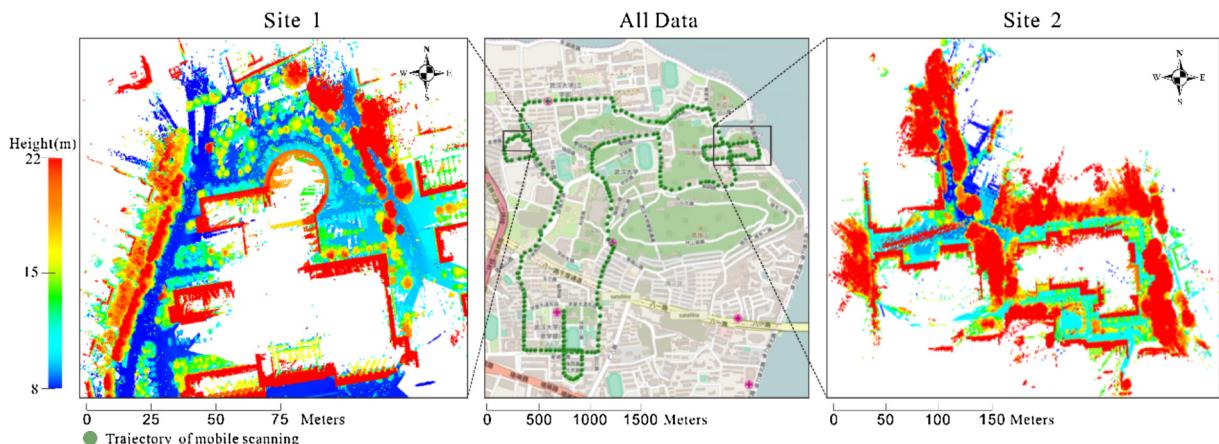
Once the vehicles in the point clouds are detected and segmented, the projected region of each detected vehicle segment



**Fig. 6.** Estimation of  $\bar{r}$  using the PSO. (a) Initial overlapping of primitive pairs. (b) Optimized overlapping of primitive pairs.

**Table 2**  
Data specifications.

	Site 1	Site 2
Number of panorama	80	259
Resolution of panorama (pixel)	$4096 \times 2048$	$4096 \times 2048$
Point density (pts/m <sup>2</sup> )	197	180
Number of points	104,317,056	249,532,656
Length (m)	500	1300



**Fig. 7.** Locations and point clouds of the two study sites.

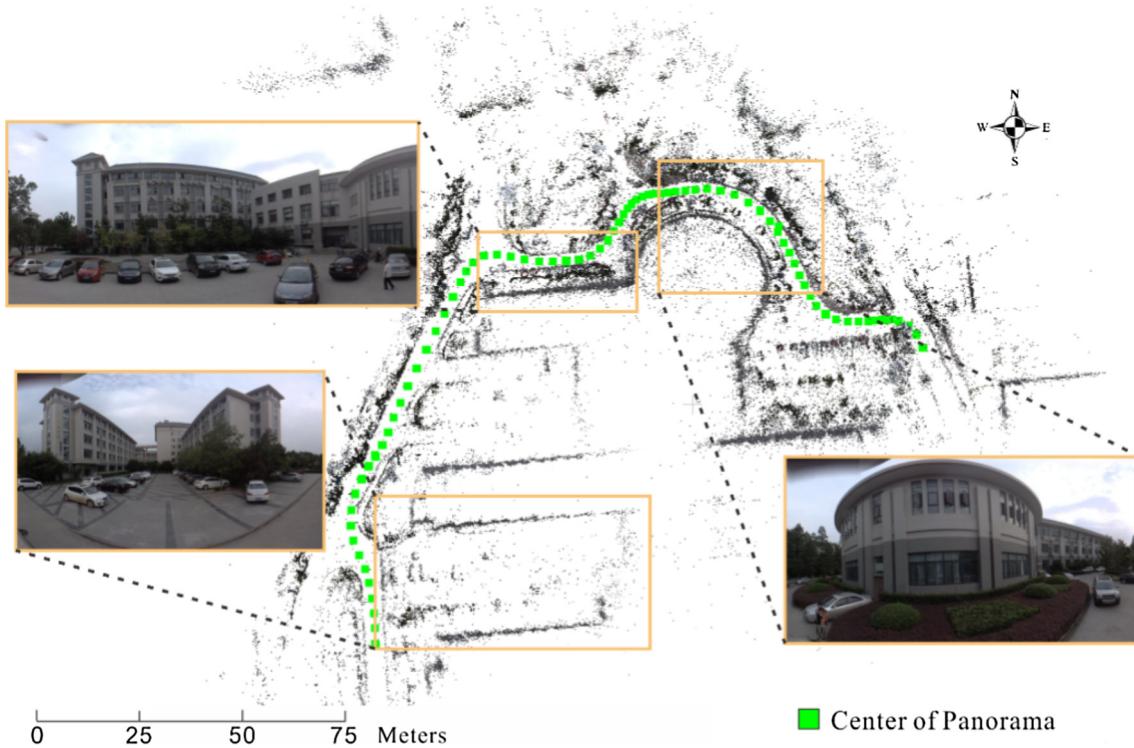


Fig. 8. GPS/IMU aided SfM of Site 1.

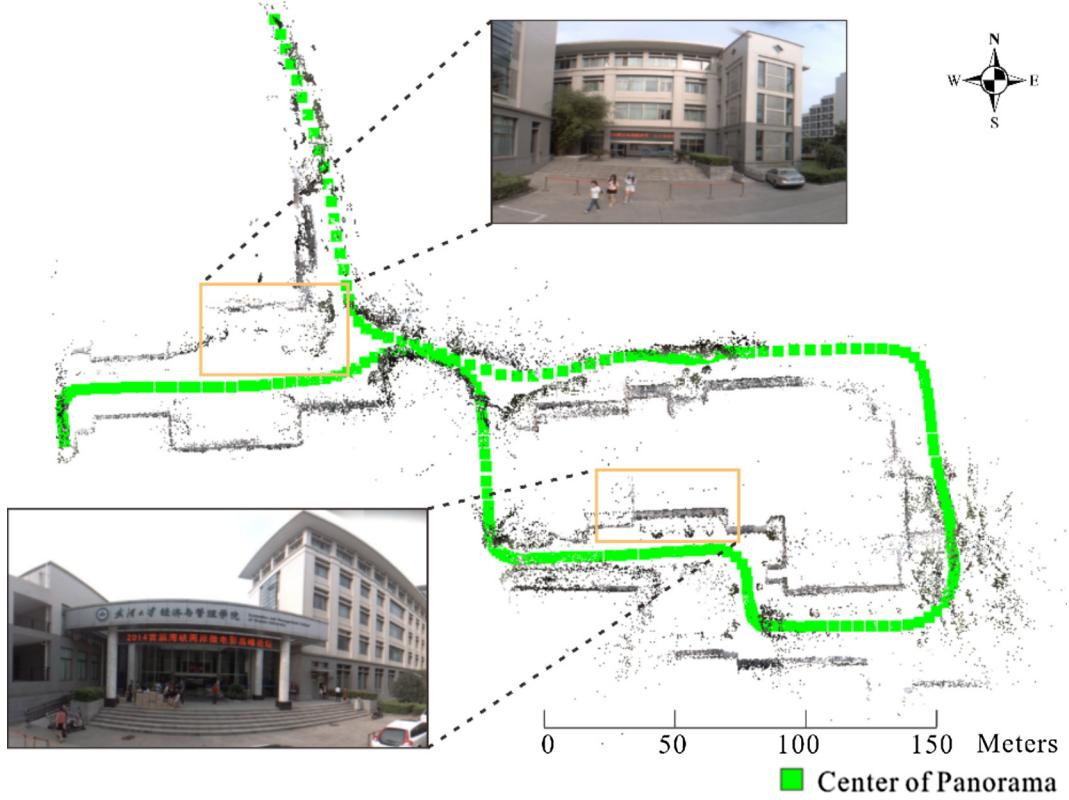


Fig. 9. GPS/IMU aided SfM of Site 2.

on the panoramic image is calculated. Let the area of one image primitive  $\rho_{image}$  be  $S_{image}$  and the projected area of corresponding vehicle segment in the point cloud be  $S_{projected}$ . If a vehicle point segment and an image primitive are one plausible primitive pair, the following constraint must be met:

$$\frac{|S_{projected} - S_{image}|}{S_{image}} < T_{area}, \quad (15)$$

where  $T_{area}$  is a threshold of the overlapping percentage. If the vehicle point segment is not satisfying the overlapping constraint (15), there are two possibilities: (i) there are false detections in the point cloud vehicle detection step, (ii) the detected vehicle is not parked so it will have less or no overlap with the corresponding primitive in the image. In both cases, the detected vehicle will not be used as a registration pair.

All the parameters for vehicle extraction in the point cloud are detailed in Table 1. The vehicle size parameters are set approximately according to the general range of vehicles. As for the number of planar segments ( $T_N$ ) and area threshold ( $T_{area}$ ), they need to be set to relatively strict values to ensure that all the false primitive pairs are removed.

### 3.4. Fine registration by optimizing compensatory translation

The EoPs of panoramic images obtained in the initialization step are further improved by using the primitive pairs extracted in the above extraction step to estimate the accurate compensatory

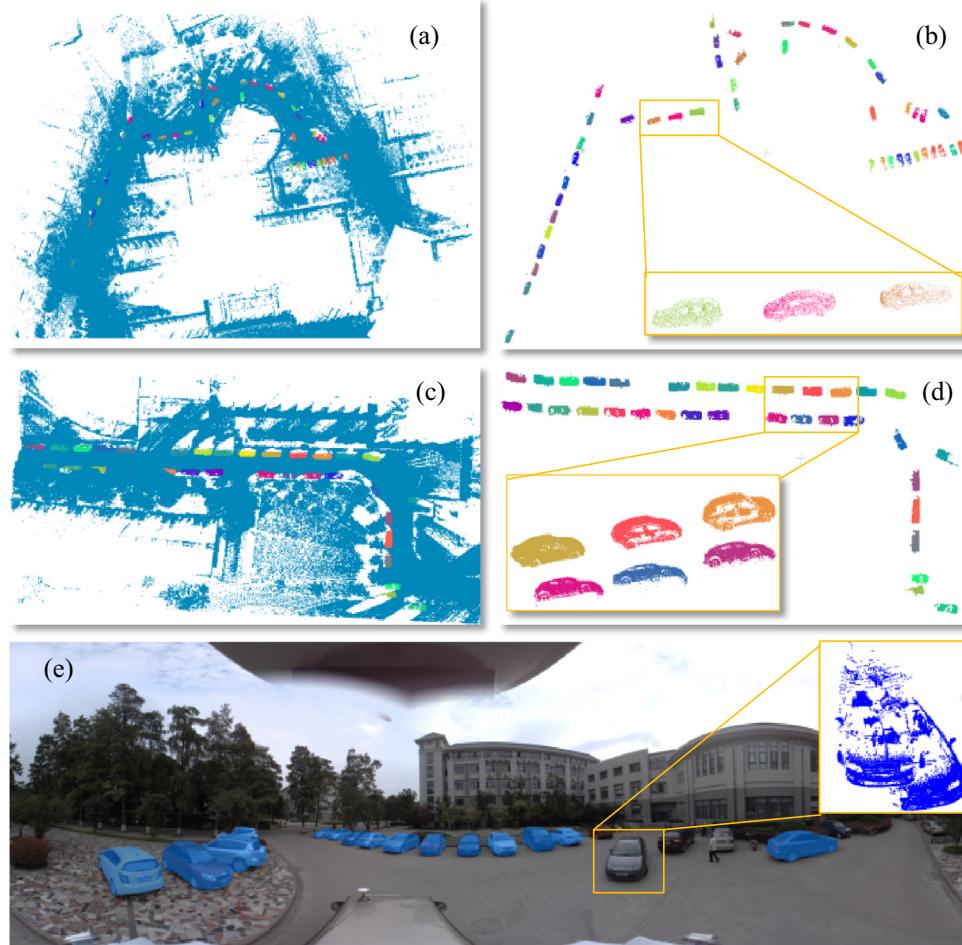
translation  $\tilde{\mathbf{r}}$ . Since not all panoramic images contains vehicles, those in which a vehicle can be detected is defined as key-frame. For each key-frame, a corresponding compensatory translation  $\tilde{\mathbf{r}}$  can be optimized by maximizing the overlapping area of primitive pairs in the key-frame and the point cloud. Then an averaging strategy has been developed to discard wrong translations by calculating the means of correct translation vectors.

Let  $B(k)$  denote a binary image, which has the same image size as the  $k^{th}$  panoramic image. In  $B(k)$ , a pixel value of 1 indicates that the pixel belongs to a vehicle segment, and 0 otherwise. And let  $B(\tilde{\mathbf{r}})$  denote another binary image that shows the projected vehicle segments from the point cloud according to the compensatory translation  $\tilde{\mathbf{r}}$ . As shown in Fig. 6(a), the purple segment represents  $B(k)$ , and the grey segment depicts  $B(\tilde{\mathbf{r}})$ . Ideally, if the  $\tilde{\mathbf{r}}$  is accurate, then  $B(k)$  and  $B(\tilde{\mathbf{r}})$  are equal. Therefore, the goal is to minimize the difference between  $B(k)$  and  $B(\tilde{\mathbf{r}})$ . Then, the cost function to be optimized can be written as follow:

$$J(\tilde{\mathbf{r}}) := \|B(k) - B(\tilde{\mathbf{r}})\|_0 \quad (16)$$

where,  $\|\cdot\|_0$  is the L0-norm, counting the different pixels between  $B(k)$  and  $B(\tilde{\mathbf{r}})$ .

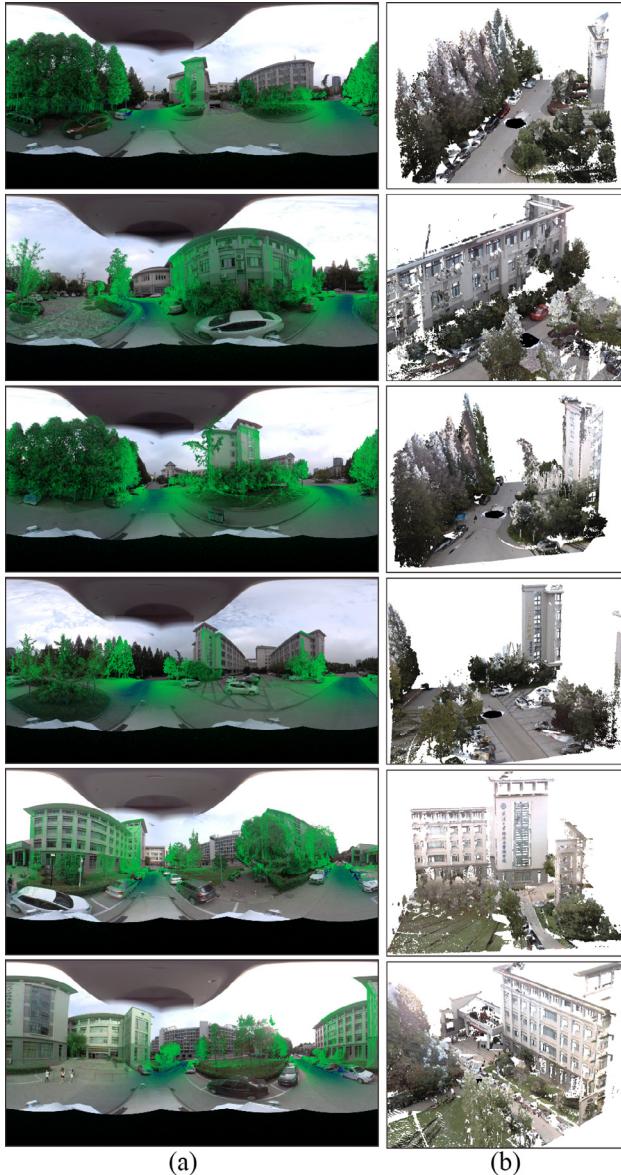
The above nonlinear cost function is a non-convex optimization problem. It is easy to fall into a local optimal solution or causes divergence using numerical solving methods, such as gradient descent. But it can be assumed that the values in vector  $\tilde{\mathbf{r}}$  must be in a small range as the data are approximately aligned already. Based on this hypothesis, we use the genetic sampling algorithm named



**Fig. 10.** Extraction of primitive pairs. (a) and (c) Primitives in point cloud in study area 1. (b) and (d) Details of segmented primitives. (e) Primitives in the panoramic image.

Particle Swarm Optimization (PSO) (Clerc and Kennedy, 2002) to search for the optimal solution of  $\tilde{\mathbf{r}}$ .

The PSO is a bionics algorithm imitating the foraging of a bird community for which each particle memorize their own and other particles' states. Each particle moves in the solution space smoothly influenced by their own experience and the experience of others to obtain the global optimal solution. For the  $i^{th}$  particle,



**Fig. 11.** Examples of the accurate registration results. (a) Points projected onto panoramic images. (b) Colored point clouds. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

it has its current candidate solution  $\tilde{\mathbf{r}}_i^k$ , and its current velocity  $\mathbf{v}_i^k$ . Meanwhile, it stores the best evaluation of the cost function up to the current generation  $k$  in  $\tilde{\mathbf{r}}_i^g$ . All particles are influenced by the best evaluation  $\tilde{\mathbf{r}}_g$  up to the current generation  $k$ . In every generation  $k$ , the velocity of each particle is updated as follows:

$$\mathbf{v}_i^k = \mathbf{v}_i^{k-1} + c_1 r_1 (\tilde{\mathbf{r}}_i^g - \tilde{\mathbf{r}}_i^{k-1}) + c_2 r_2 (\tilde{\mathbf{r}}_g - \tilde{\mathbf{r}}_i^{k-1}) \quad (17)$$

where,  $c_1, c_2$  are the coefficients of individual component and social component, and are set to 3.0 and 1.5 in all our experiments;  $r_1, r_2$  are random variables with uniform distribution between 0 and 1. Its solution is updated according to

$$\tilde{\mathbf{r}}_i^k = \tilde{\mathbf{r}}_i^{k-1} + \Theta \mathbf{v}_i^k \quad (18)$$

$$\Theta = \frac{2}{|2 - c_1 - c_2 - \sqrt{(c_1 + c_2)^2 - 4(c_1 + c_2)}|} \quad (19)$$

The search range of the optimization is restrained by limiting particles moving range in 0.5 m around the initial value. An example of the optimization result is shown in Fig. 6(b).

After the estimation of translation vectors, each key-frame has a corresponding compensatory translation  $\tilde{\mathbf{r}}$ , some of which may be obviously wrong due to distribution of primitives. Inspired by the method proposed by (Brenner et al., 2008), a strategy has been developed to eliminate those incorrect translation vectors. First, the norm of each translation vector,  $d = \sqrt{\tilde{\mathbf{r}}^T \cdot \tilde{\mathbf{r}}}$ , is calculated and clustered using  $k$ -means (Bhatia, 2004). Second, the center of the largest class is obtained using a voting strategy (Fernandes and Oliveira, 2008; Habib et al., 2001), which generates the final accurate compensatory translation vector  $\tilde{\mathbf{r}}$ .

#### 4. Experiments and analysis

More than 10 km mapping data were acquired in the campus of Wuhan University, China, by the Optech LYNX Mobile Mapper. Two study sites were chosen as shown in Fig. 7. Table 2 lists the data specifications in the two studied areas. Both of them are typical urban scenes, containing a large number of vehicles, pedestrians, vegetation and buildings.

##### 4.1. Registration results

###### ● Initialization of the EoPs

Compensatory rotation of the registration model is estimated using GPS/IMU aided panoramic SfM. Figs. 8 and 9 present the result of initialization in the study areas. The green squares are the centers of panoramic images after the GPS/IMU aided bundle adjustment, and the other points are feature points obtained by the SfM. Some corresponding photos of the scenes are shown in the figures too.

###### ● Vehicle primitives extraction

Results of primitive pair extraction are shown in Fig. 10. Fig. 10 (a) shows the primitives in the point cloud in study area 1, and

**Table 3**

Accuracy comparison between the pre-calibration and the proposed methods.

	Original EoPs Error (pixel)			Proposed initialization Error (pixel)			Proposed fine registration Error (pixel)		
	Average	Max	RMSE	Average	Max	RMSE	Average	Max	RMSE
Site1	25.02	35.54	6.51	9.10	18.69	4.45	2.34	3.27	0.68
Site2	25.81	38.29	7.55	9.54	19.49	4.55	2.19	3.51	0.69

vehicle segments are shown in Fig. 10(b), Fig. 10(c) and (d) show primitives in the point cloud in study area 2. Fig. 10(e) shows the primitives in the panoramic image. Notice that there is a moving vehicle, whose corresponding vehicles in the point cloud are shown on the top right of Fig. 10(e) too. According to the proposed selection criterion on the overlapping area of a corresponding primitive pair, it will not be used for registration.

#### • Fine registration

The compensatory translation  $\bar{r}$  is estimated based on maximizing the overlapping area of the primitive pairs. Examples of the registration results are shown in Fig. 11. Using the accurate EoPs of panoramic images obtained by fine registration, the laser points are projected onto the panoramic images as shown in Fig. 11(a). Meanwhile, the colored laser scanning points are shown in Fig. 11(b).

#### 4.2. Registration error analysis

To assess the proposed method, corresponding check points are manually selected from panoramic images and point clouds to compare three registration solutions, namely, the original EoPs from the pre-calibration, those obtained by the proposed initialization, and those by the proposed fine registration. The check points should be easily recognizable in the images as well as in the point clouds, e.g. building corners, vehicle wheel centers. Meanwhile, they should be evenly distributed over the panoramic image. The selected check points in the point cloud are back-projected onto the panoramic image, then the pixel offsets between laser scanning points and corresponding check points on the image are measured as the evaluation of the registration error. The registration evaluation is shown in Table 3. Fig. 12 shows the distribution of the registration errors of the check points at both study sites. Figs. 13 and 14 are two examples of the three registration solutions. By

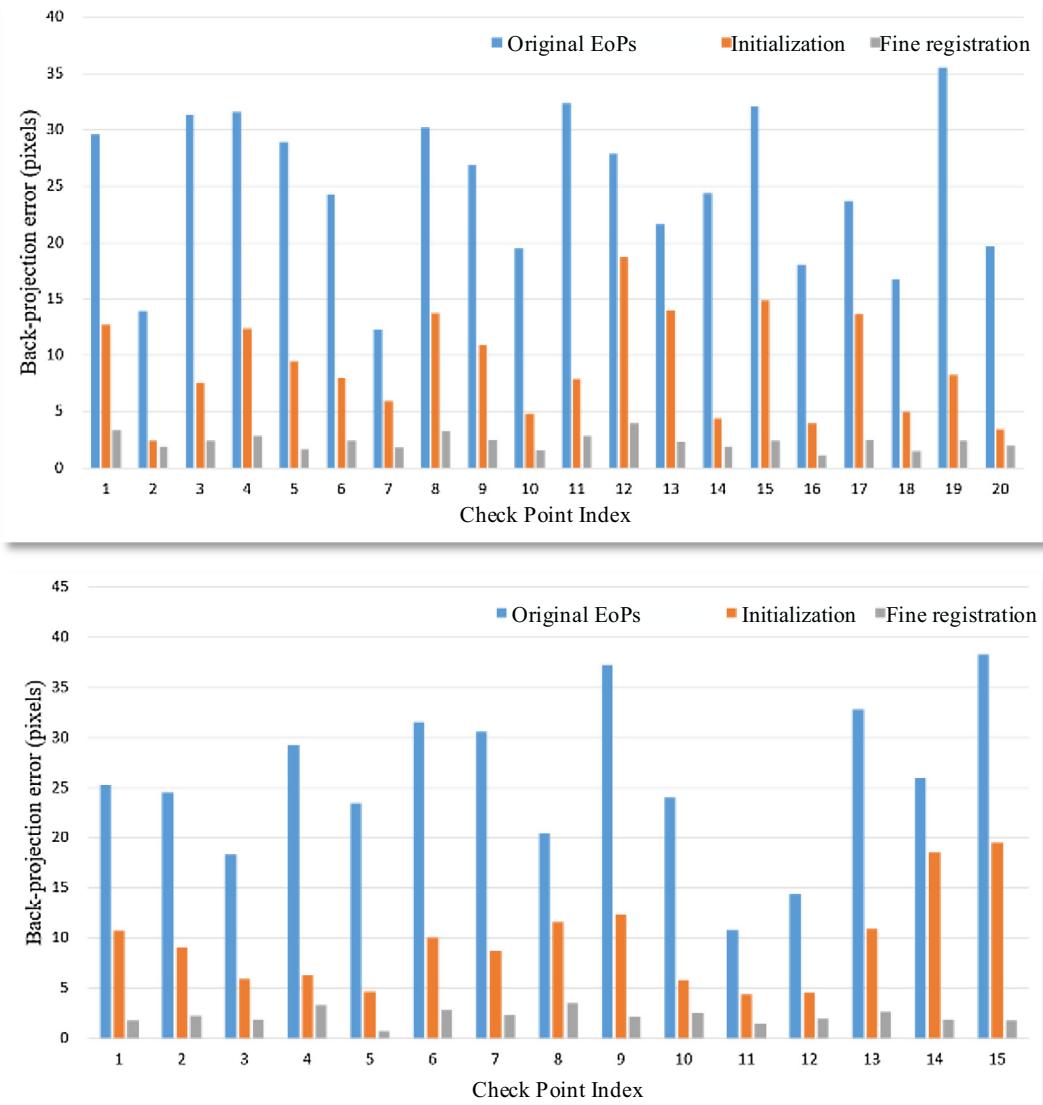
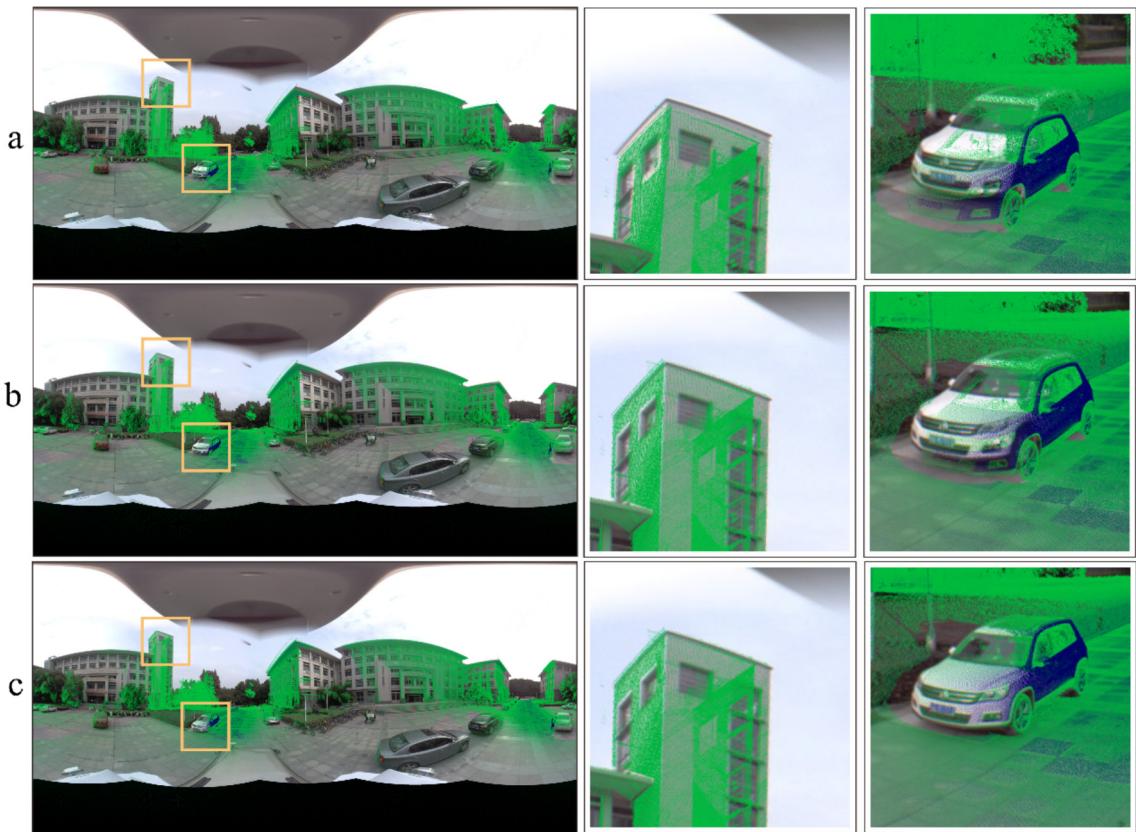


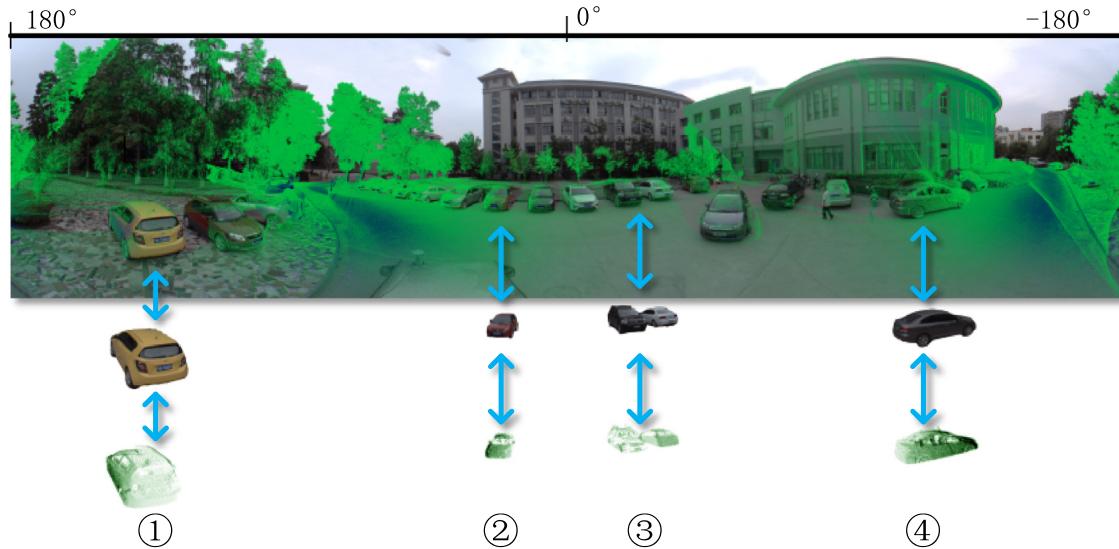
Fig. 12. Check point registration errors in study area 1 (top) and area 2 (bottom).



**Fig. 13.** Registration comparison in study area 1. (a) Results of original EoPs. (b) Results of the proposed initialization. (c) Results of the fine registration.



**Fig. 14.** Registration comparison in study area 2. (a) Results of original EoPs. (b) Results of the proposed initialization. (c) Results of the fine registration.



**Fig. 15.** Groups of primitive pairs.

**Table 4**  
Description of primitive pairs.

Index	Direction (degree)	Distance to the panoramic camera (meter)
1	136	5
2	31	15
3	-18	15
4	-117	10

comparing the back-projection errors, it is evident that the EoPs obtained by the proposed initialization has higher accuracy than the original EoPs. The average pixel offsets are reduced to less than 10 pixels. It means that the proposed initialization method ensures more accurate corresponding vehicle search in the point cloud. It was demonstrated that after the fine registration, the average offset was further reduced to about 3 pixels.

## 5. Discussion

### 5.1. Influence of the distribution and distance of registration primitives

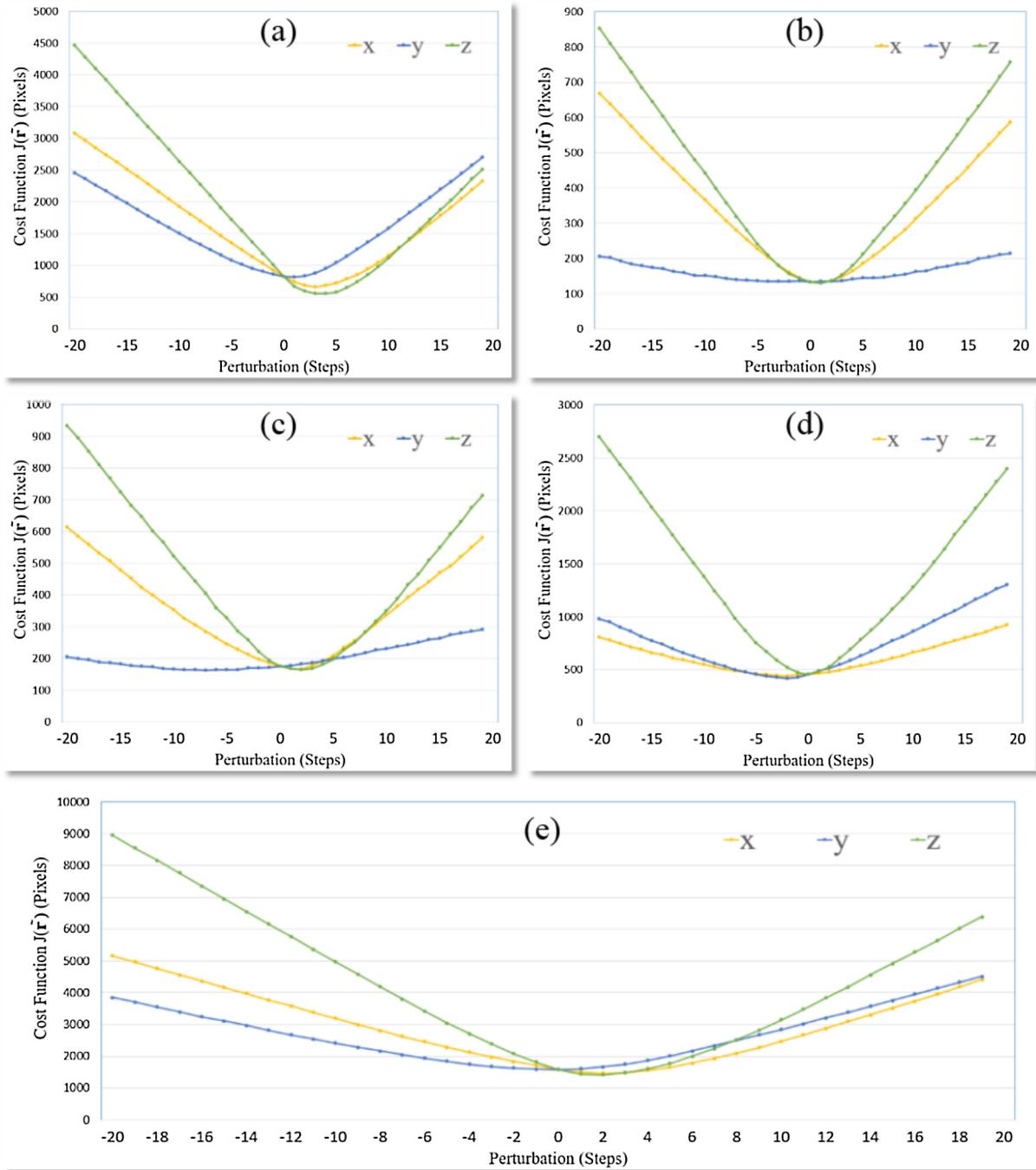
In the proposed fine registration step, the PSO algorithm is utilized to optimize the compensatory translation  $\bar{r}$  by maximizing the overlapping area between registration primitive pairs. In order to assess the effect of the distribution and distance of primitive pairs on the registration accuracy, the perturbation is introduced to the accurate EoPs obtained by manual registration to simulate the cost function under different distributions and distances of primitive pairs. Firstly, a sample panoramic image is manually registered with the point cloud using selected corresponding points, and 4 primitive pairs are selected as shown in Fig. 15. Then, perturbation of translation is added to the  $\bar{r}$  at 0.01 m/step, and the cost functions under different perturbations using different primitive pairs are compared. The different uses of primitives are as follows: (a) using primitive pair 1; (b) using primitive pair 2; (c) using primitive pair 3; (d) using primitive pair 4; (e) using primitive pairs 1–4. The characters of primitive pairs are described in Table 4. Results of the perturbation analysis are shown in Fig. 15.

Comparing the results of (a) and (d), the cost function of (a) has more obvious changes under different perturbations, which indicates that the closer the primitive pair is, the more contribution the pair has to the calculation of the compensatory translation  $\bar{r}$ .

In the experiments of (a) and (d), the effect of the cost function in the y-direction is greater than that in the x-direction. On the contrast, the perturbations in the x-direction contribute more to the cost function than those in the y-direction when comparing (b) and (c). This is caused by the different distributions of primitive pairs. Because the 1st and 4th primitive pairs are located on both sides of the MMS, they are more sensitive to the disturbance in the forward direction (y-direction); whereas the 2nd and 3rd primitive pairs are located in front of the MMS, so they are more sensitive to the side direction (x-direction). The result of (e) demonstrates that if there are primitive pairs evenly distributed on the panoramic image, the PSO algorithm can converge at an optimal point to achieve an accurate registration.

### 5.2. Influence of the quality of the image segmentation

Under-segmentation and over-segmentation are the common problems for the existing image segmentation methods, while it is apparent that the image segmentation quality will affect the registration results. To evaluate the robustness of the proposed method, an experiment is utilized as follow. Figs. 16 and 17(a) illustrate the primitives in the panoramic image extracted by the proposed method, which is the input for the proposed PSO algorithm. This input data is stored as a binary raster with the same size of the panoramic image, in which the values of the pixels belonging to vehicle segments are set to 1 and the value of other pixels are set to 0. Then noise points of different radii and gaps are added to the input binary raster, pixels in the original binary raster occupied by the noise points are set to opposite value to simulate the under-segmentation and over-segmentation of an image. Figs. 16 and 17(b)–(d) illustrate examples of the input raster under different noise levels, where pixels with white color represent the non-vehicle region (pixel value of 0 in the binary raster), the other pixels represent the vehicle region (pixel value of 1 in the binary raster). After the preparation of the input raster, the proposed PSO algorithm is performed using the inputs under different noise levels. Fig. 18 illustrates the average error of checkpoints under different noise levels. The registration results show that the proposed method is robust to the noise of segmentation in images. Although noise points with a gap of 50 pixels and radius of 17 pixels are added, the proposed method can still achieve good registration accuracy. A comprehensive comparison of the average registration error under different conditions has demonstrated that the



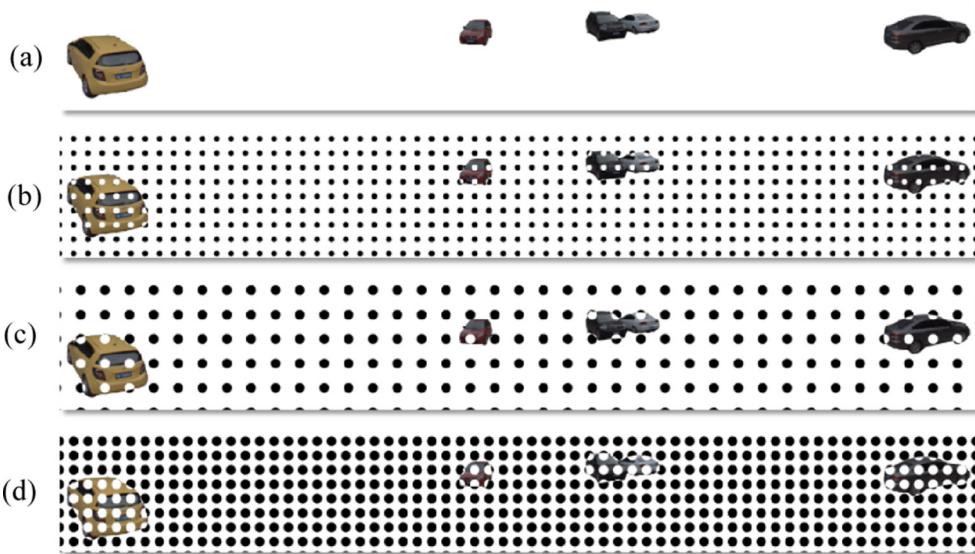
**Fig. 16.** Perturbation analysis. The vertical axis represents the cost function, and the unit is the pixel; the horizontal axis represents the disturbance value, and the unit is the perturbation step. The yellow, blue, and green lines depict the disturbances in the x-direction, y-direction, and z-direction, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

proposed method is robust to the quality of the image segmentation to a certain extent.

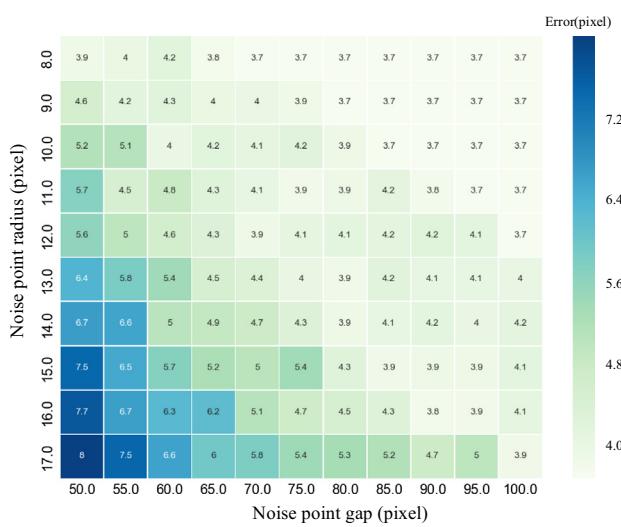
### 5.3. Parameter sensitivity test

The proposed cost function optimization depends on the primitive extraction. So parameters of primitive extraction are can be influential. The vehicle size parameters are set according to the approximate general range of passenger vehicles, hence can be used without tuning. Only two parameters, the number of planar segments ( $T_N$ ) and the area threshold ( $T_{area}$ ) are left for

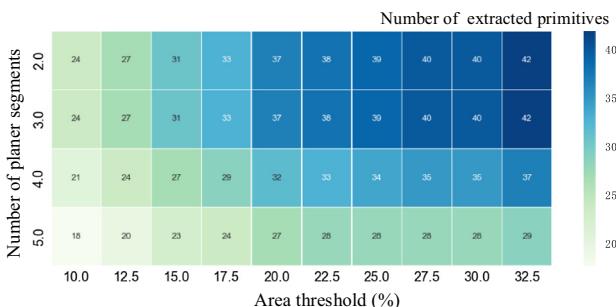
determination. Therefore, we examined the influences of  $T_N$  and  $T_{area}$  by setting different values to test the sensitivity of the proposed extraction method. 50 primitive pairs are randomly selected from the experimental data, which contain five false matches. As a result, the false matches were successfully eliminated using all the tested parameters. The numbers of extracted primitives using different parameters are shown in Fig. 19, and most numbers are greater than 30, which is more than sufficient. Therefore, the proposed method is not sensitive to the parameters and the range of appropriate parameters is large. A high number of extractions is not required by the optimization, so  $T_N$  and  $T_{area}$  are set to



**Fig. 17.** Examples of generated noise points. (a) Primitives without noise points. (b) Noise points with gap of 50 pixels and radius of 10 pixels. (c) Noise points with gap of 85 pixels and radius of 17 pixels. (d) Noise points with gap of 50 pixels and radius of 17 pixels.



**Fig. 18.** Average registration error under different noise levels. The vertical axis represents the radius of noise points; horizontal axis represents the gap between noise points.



**Fig. 19.** The number of extracted primitives using different parameters. The vertical axis represents the threshold of planer segments ( $T_N$ ); horizontal axis represents the area threshold ( $T_{area}$ ).

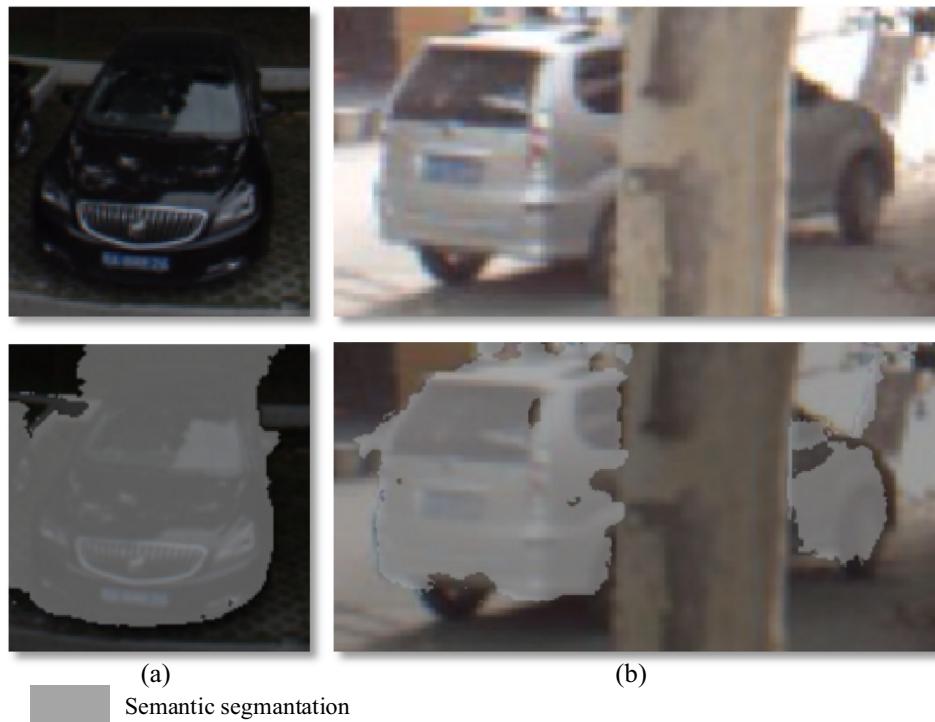
relatively strict values, 3 and 20%, in our experiment to ensure the correctness of the primitive extraction.

#### 5.4. Deficiencies and future work

The automatic registration method proposed in this paper is suitable for urban scenes. As discussed in Section 5.1, the accuracy of the proposed method relies on the extraction of primitive pairs. Primitive extraction may fail due to semantic segmentation errors caused by poor illumination or occlusion as shown in Fig. 20. These ‘false’ primitives will be removed in the process of primitive pair filtering, which will reduce the number of useful primitives. So a reasonable primitive extraction strategy to retain the primitive pairs is worth investigating. Meanwhile, in an urban scene, in addition to parked vehicles, other targets can also be used as registration primitives, such as buildings and traffic lanes. The use of different registration primitives might improve the robustness of registration, so the extraction of other registration primitives will also be explored to potentially improve the registration in the future.

## 6. Conclusion

Accurate registration between panoramic images and mobile laser scanning data is important for many urban mapping applications, such as object extraction, 3D reconstruction, 3D city change detection, etc. Due to unavoidable relative movement between different sensors, it is difficult to have accurate alignment between panoramic images and the point cloud. This paper proposed an automatic and marker-free method for accurate registration between a panoramic image sequence and mobile laser scanning data collected in urban areas. Firstly, the imperfect original EoPs obtained by pre-calibration is used to initialize the panoramic images by estimating the accurate relative orientation between the laser scanner and the panoramic camera using GPS and IMU aided panoramic SfM. Secondly, vehicles are treated as registration primitives, which are extracted from both panoramic images and point clouds. Finally, the relative translation between the laser scanner and the panoramic camera is refined by maximizing the



**Fig. 20.** Examples of failures in primitive extraction. (a) Segmentation error caused by poor illumination. (b) A partially occluded primitive.

overlapping area of corresponding primitive pairs based Particle Swarm Optimization (PSO). Comprehensive experiments have been carried to evaluate the accuracy and robustness of the proposed method. The results demonstrate that the method can achieve an accurate registration with an average registration error fewer than three pixels. The use of other registration primitives in urban scenes will be explored in the future to check whether the registration can be further improved.

## Acknowledgement

This study was jointly supported by NSFC project (No. 41725005, 41531177, 41371431, 41701530), the National Key Technology R&D Program (No. 2014BAL05B07), Public science and technology research funds projects of ocean (No. 2013418025), National Key Research and Development Program of China (No. 2016YFF0103501), Doctoral Fund of Ministry of Education of China (2016M600614), and Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University (No. 2018LSDMIS06).

## References

- Abayowa, B.O., Yilmaz, A., Hardie, R.C., 2015. Automatic registration of optical aerial imagery to a LiDAR point cloud for generation of city models. *ISPRS J. Photogramm. Remote Sens.* **106**, 68–81.
- Böhm, J., Becker, S., 2007. Automatic marker-free registration of terrestrial laser scans using reflectance. In: Proceedings of 8th Conference on Optical 3D Measurement Techniques, Zurich, Switzerland, pp. 338–344.
- Barfoot, T., 2016. State Estimation for Robotics: A Matrix Lie Group Approach. Draft in preparation for publication by Cambridge University Press, Cambridge.
- Barnea, S., Filin, S., 2013. Segmentation of terrestrial laser scanning data using geometry and image information. *ISPRS J. Photogramm. Remote Sens.* **76**, 33–48.
- Bhatia, S.K., 2004. Adaptive K-Means Clustering. The Florida Ai Research Society, pp. 695–699.
- Brenner, C., Dold, C., Ripperda, N., 2008. Coarse orientation of terrestrial laser scans in urban environments. *ISPRS J. Photogramm. Remote Sens.* **63**, 4–18.
- Brenner, S.H.E., 2014. Skyline matching based camera orientation from images and mobile mapping point clouds. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. II-5, pp. 18 II-5, 181–189.
- Christy, S., Horaud, R., 1999. Iterative pose computation from line correspondences. *Comput. Vis. Image Underst.* **73**, 137–144.
- Clerc, M., Kennedy, J., 2002. The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Trans. Evol. Comput.* **6**, 58–73.
- Corsini, M., Dellepiane, M., Ganovelli, F., Gherardi, R., Fusilli, A., Scopigno, R., 2013. Fully automatic registration of image sets on approximate geometry. *Int. J. Comput. Vision* **102**, 91–111.
- Cui, T., Ji, S., Shan, J., Gong, J., Liu, K., 2017. Line-based registration of panoramic images and LiDAR point clouds for mobile mapping. *Sensors* **17** (1), 70.
- Felzenszwalb, P.F., Huttenlocher, D.P., 2004. Efficient graph-based image segmentation. *Int. J. Comput. Vision* **59**, 167–181.
- Fernandes, L.A.F., Oliveira, M.M., 2008. Real-time line detection through an improved Hough transform voting scheme. *Pattern Recogn.* **41**, 299–314.
- Habib, A., Ghanna, M., Morgan, M., Al-Ruzouq, R., 2005. Photogrammetric and LiDAR data registration using linear features. *Photogramm. Eng. Remote Sens.* **71**, 699–707.
- Habib, A.F., Lee, Y.-R., Morgan, M., 2001. Surface matching and change detection using a modified hough transformation for robust parameter estimation. *Photogramm. Rec.* **17**, 303–315.
- Hernandez, J., Marcotegui, B., 2009. Point cloud segmentation towards urban ground modeling. *Joint Urban Remote Sensing Event*, 1–5.
- Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W., 2011. g 2 o: a general framework for graph optimization. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, pp. 3607–3613.
- Kwak, T.-S., Kim, Y.-I., Yu, K.-Y., Lee, B.-K., 2006. Registration of aerial imagery and aerial LiDAR data using centroids of plane roof surfaces as control information. *KSCE J. Civ. Eng.* **10**, 365–370.
- Levinson, J., Thrun, S., 2013. Automatic online calibration of cameras and lasers. *Robotics: Science and Systems*, pp. 24–28.
- Lee, J.-B., Yu, K.-Y., 2006. Co-registration of aerial photos, ALS data and digital maps using linear features. *J. Kor. Soc. Geospat. Inform. Syst.* **14**, 37–44.
- Levin, A., Lischinski, D., Weiss, Y., 2008. A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**, 228–242.
- Liu, L., Stamos, I., 2005. Automatic 3D to 2D registration for the photorealistic rendering of urban scenes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 132, pp. 137–143.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **60**, 91–110.
- Morel, J.-M., Yu, G., 2009. ASIFT: a new framework for fully affine invariant image comparison. *SIAM J. Imag. Sci.* **2**, 438–469.
- Moussa, W., Abdel-Wahab, M., Fritsch, D., 2012. An automatic procedure for combining digital images and laser scanner data. *Int. Arch. Photogr. Rem. Sens. Spat. Inform. Sci.* **39**, B5.
- Miled, M., Soheilian, B., Habets, E., Vallet, B., 2016. Hybrid online mobile laser scanner calibration through image alignment by mutual information. *ISPRS Ann. Photogr. Rem. Sens. Spat. Inform. Sci.* **3**.

- Pagani, A., Stricker, D., 2011. Structure from Motion using full spherical panoramic cameras. In: Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, pp. 375–382.
- Palenichka, R.M., Zaremba, M.B., 2010. Automatic extraction of control points for the registration of optical satellite and LiDAR images. *IEEE Trans. Geosci. Remote Sens.* 48, 2864–2879.
- Parmehr, E.G., Fraser, C.S., Zhang, C., Leach, J., 2014. Automatic registration of optical imagery with 3D LiDAR data using statistical similarity. *ISPRS J. Photogramm. Remote Sens.* 88, 28–40.
- Pflugmacher, D., Cohen, W.B., Kennedy, R.E., Yang, Z., 2014. Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics. *Remote Sens. Environ.* 151, 124–137.
- Ramalingam, S., Bouaziz, S., Sturm, P., Brand, M., 2009. Geolocalization using skylines from omni-images. In: Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, pp. 23–30.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: towards real-time object detection with region proposal networks. *Adv. Neural Inform. Process. Syst.* 91–99.
- Wang, L., Neumann, U., 2009. A robust approach for automatic registration of aerial images with untextured aerial lidar data. In: Computer Vision and Pattern Recognition. CVPR 2009. IEEE Conference on, pp. 2623–2630.
- Wang, R., Ferrie, F.P., Macfarlane, J., 2012. Automatic registration of mobile LiDAR and spherical panoramas. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 33–40.
- Yang, B., Chen, C., 2015. Automatic registration of UAV-borne sequent images and LiDAR data. *ISPRS J. Photogramm. Remote Sens.* 101, 262–274.
- Yang, B., Dong, Z., 2013. A shape-based segmentation method for mobile laser scanning point clouds. *ISPRS J. Photogramm. Remote Sens.* 81, 19–30.
- Yang, B., Dong, Z., Zhao, G., Dai, W., 2015. Hierarchical extraction of urban objects from mobile laser scanning data. *ISPRS J. Photogramm. Remote Sens.* 99, 45–57.
- Yang, B., Xu, W., Dong, Z., 2013. Automated extraction of building outlines from airborne laser scanning point clouds. *IEEE Geosci. Remote Sens. Lett.* 10, 1399–1403.
- Yipu, Z., Yuanfang, W., Yichang, T., 2016. 2D-image to 3D-range registration in urban environments via scene categorization and combination of similarity measurements. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 1866–1872.
- Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H., 2015. Conditional random fields as recurrent neural networks. *Proceedings of the IEEE International Conference on Computer Vision*, 1529–1537.