Automatic registration method for mobile LiDAR data

Ruisheng Wang
Frank P. Ferrie

Automatic registration method for mobile LiDAR data

Ruisheng Wang[a,][*] **and Frank P. Ferrie**[b]
[a]University of Calgary, Department of Geomatics Engineering, Calgary, Alberta T2N 1N4, Canada
[b]McGill University, Department of Electrical and Computer Engineering and the Center for Intelligent Machines, Montreal, Québec H3A 0E9, Canada

Abstract. We present an automatic mutual information (MI) registration method for mobile LiDAR and panoramas collected from a driving vehicle. The suitability of MI for registration of aerial LiDAR and aerial oblique images has been demonstrated under an assumption that minimization of joint entropy (JE) is a sufficient approximation of maximization of MI. We show that this assumption is invalid for the ground-level data. The entropy of a LiDAR image cannot be regarded as approximately constant for small perturbations. Instead of minimizing the JE, we directly maximize MI to estimate corrections of camera poses. Our method automatically registers mobile LiDAR with spherical panoramas over an approximate 4-km drive, and is the first example we are aware of that tests MI registration in a large-scale context. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.OE.54.1.013108]

1 Introduction

Image-to-range registration is a prerequisite for many applications. The registration result is critical not only for texture-mapping three-dimensional (3-D) models of large-scale scenes, but also for applications such as image-based upsampling of range data,[1–4] image-guided range segmentation,[5,6] and 3-D scene modeling.[7] The problem of image-to-range registration involves the alignment of two-dimensional (2-D) images with 2-D projections of 3-D range data, consisting of estimating the relative camera pose with respect to the range sensors.

There has been a considerable amount of research in registering images with range data. Existing methods include keypoint-based matching,[8–10] structural features-based matching,[11–14] and mutual information-based registration.[15] The range data include terrestrial or aerial LiDAR, and the images include vertical or oblique aerial images and ground-level images.

Keypoint-based matching[8,10] is based on the similarity between laser intensity images and corresponding camera images. First, each pixel of the laser intensity image is encoded with its corresponding 3-D coordinate. Then, feature points are extracted by using either SIFT[16] or Förstner operators[17] from both images. A robust matching strategy based on RANSAC[18] and/or epipolar geometry constraint is employed to determine the correspondence pairs for computing the fundamental matrix. Sensor registration is then achieved based on a robust camera spatial resection. Ding et al.[9] registered oblique aerial images with a 3-D model generated from aerial LiDAR data based on 2-D and 3-D corner features in the 2-D images and 3-D LiDAR model. The correspondence between extracted corners was based on a Hough transform and generalized M-estimator sample consensus. The resultant corner matches are used in Lowe's

algorithm[19] to refine camera parameters estimated from a combination of vanishing point computation and GPS/IMU readings. In general, the feature point extraction and robust matching are the key to a successful registration for this type of approach.

Instead of matching points, structural feature-based methods[11–14] match structural features in both 2-D and 3-D space to estimate the relative camera pose. Direct matching single line features is error-prone because of the noise in both LiDAR and image data as well as the robustness of the detection algorithms. High-level structural features are helpful to increase the robustness of both detection and matching. Wang and Neumann[14] registered aerial images with aerial LiDAR based on matching so-called "3 Connected Segments" in which each linear feature contains three segments connected into a chain. They used a two-level RANSAC algorithm to refine the putative feature matches, and estimated camera pose using the method described in Ref. 20. Liu et al.[11–13] extracted so-called "rectangular parallelepiped" features, which are composed of vertical or horizontal 3-D rectangular parallelepipeds in the LiDAR and 2-D rectangles in the images, to estimate camera translation with a hypothesis-and-test scheme. Camera rotation was estimated based on at least two vanishing points. Since vanishing points are required, their methods work well for ground-level data but are not efficient to handle aerial data with a weak perspective effect.

All the above methods are dependent on either the strong presence of parallel lines to infer vanishing points, or availability of feature pair correspondence, which limits their applicability and robustness. A recent method[15] using statistical metrics, such as mutual information (MI),[21] as a similarity measure for registering oblique aerial images and aerial LiDAR does not require any feature extraction process. This method searches for the optimal camera pose through maximizing the MI between camera images and different attributes of LiDAR such as the LiDAR intensity images,

*Address all correspondence to: Ruisheng Wang, E-mail: ruiswang@ucalgary.ca

depth maps, or a combination of both. Instead of using features, the MI method evaluates statistical measures using all the pixels in both images, which avoids the problems of feature extraction and correspondence. Thus, MI registration method holds the potential to be a robust solution.

This paper deals with the problem of the registration between mobile LiDAR and spherical panoramas. The data acquisition platform is designed such that mapping between the LiDAR and spherical camera system is approximately known through mechanical measurements. However, because of vibration induced by motion of the platform and deformations arising from temperature changes, significant registration errors can still occur, requiring estimation of the true mapping parameters.

In this paper, we use a portion of the LiDAR and spherical panoramas (i.e., one-sixth of the entire panorama) for MI computation as we project them onto a plane with a view port of 940×452 pixels through OpenGL rendering. In contrast to Mastin et al.,[15] we explicitly use the MI metric as the similarity measure. In their work, which involved the registration of aerial LiDAR with oblique images, it was assumed that the entropy of the LiDAR images remained approximately constant for small perturbations. Under this assumption, minimizing the joint entropy (JE) is equivalent to maximizing the MI. However, as we show in Sec. 3, this approach does not appear to work for the case of ground-based applications such as those considered in this paper. The statistics of the data are such that the constant entropy assumption breaks down, invalidating the use of JE.

The algorithm presented in this paper is fully automatic and has been designed to run efficiently on a metropolitain LiDAR/spherical image database using different representations of LiDAR from those in Mastin et al. We are not aware of any other examples of MI registration that have been attempted on this scale.

2 Data Acquisition

Data are collected from a mobile mapping system shown in Fig. 1, which is composed of a 360 deg LiDAR sensor (Velodyne HDL-64E), six high-resolution cameras, a Ladybug 3 camera, GPS, inertial measurement unit (IMU), and distance measurement instrument. The Velodyne LIDAR sensor consists of 64 lasers mounted on the upper and lower
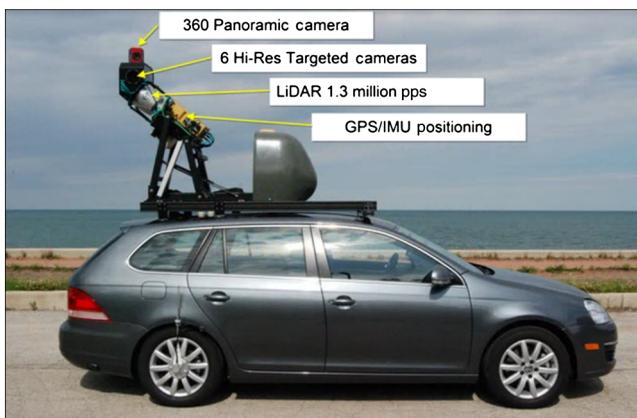

Fig. 1 NAVTEQ TRUE mobile mapping system.

blocks of 32 lasers each and the entire unit spins and generates over 1×10^6 pps. The sensor can spin at rates ranging from 5 to 20 Hz, and the default is 10 Hz with a 905-nm wavelength. The Ladybug 3 covers more than 80% of a full sphere, with six high quality 1600×1200 Sony charge-coupled device sensors, and provides up to 12 MP images at 15 fps. All of these sensors are georeferenced through a GPS and an IMU.

3 Method

We start with the work of Mastin et al.[15] that registers aerial LiDAR with aerial oblique images based on MI. LiDAR intensity images that normally look very similar to gray-scale camera images with, of course, a much lower resolution. This correlation makes MI a suitable measure to evaluate their similarity. In Ref. 15, they define $p(x, y)$ and $l(x, y)$ as the intensity camera image and projected LiDAR features, respectively, on the xy image plane. For a specific camera matrix T, the projected LiDAR features are given by l_T. MI-based registration methods find the optimal camera matrix that maximizes the MI between camera images and projected LiDAR features

$$T_{\mathrm{MI}} = \arg \max_T I(p; l_T). \tag{1}$$

They use a generic camera model, the finite projective camera as described in Ref. 20. Under this camera model, a point in space is mapped to the point on the image plane by

$$P = KR[I - C], \tag{2}$$

where $C = [C_x, C_y, C_z]^T$ is the camera center, I is the identity matrix, and R is the camera rotation matrix. $R = R_x(\gamma)R_y(\beta)R_z(\alpha)$ is given by the three rotation matrices, where α, β, and γ are the Euler angles representing yaw, pitch, and roll. The matrix K is the camera calibration matrix.

In this paper, we show that their assumption is invalid for the ground-based case, as small perturbations in camera pose have larger effect on the LiDAR rendering in the ground-level data. The entropy of a LiDAR image cannot be regarded as approximately constant for small perturbations. This is demonstrated by a perturbation analysis, which shows how the normalized MI and JE vary around the initial registration in terms of altered camera poses as shown in Fig. 2. We select four representative scenes for this test from Fig. 8. Since the correct registration value should be near the initial registration, we set all parameters at their initial values and vary each parameter to view the shape of the cost functions. The range of camera parameter perturbations is ± 2 units, meters for translation and degrees for orientation. The step size for the perturbation analysis is 0.1 units, for a total of 40 measurements for each of the charts shown in Fig. 2. As shown in the figures, the x-axis corresponds to relative displacement and the y-axis corresponds to the normalized value of MI. The traces labeled 1, 2, 3, and 4 in Fig. 2(a)–2(l) correspond to images 1, 4, 5, and 8 in Fig. 8, respectively. The remaining trace, data5, rendered as a solid black line in each chart,
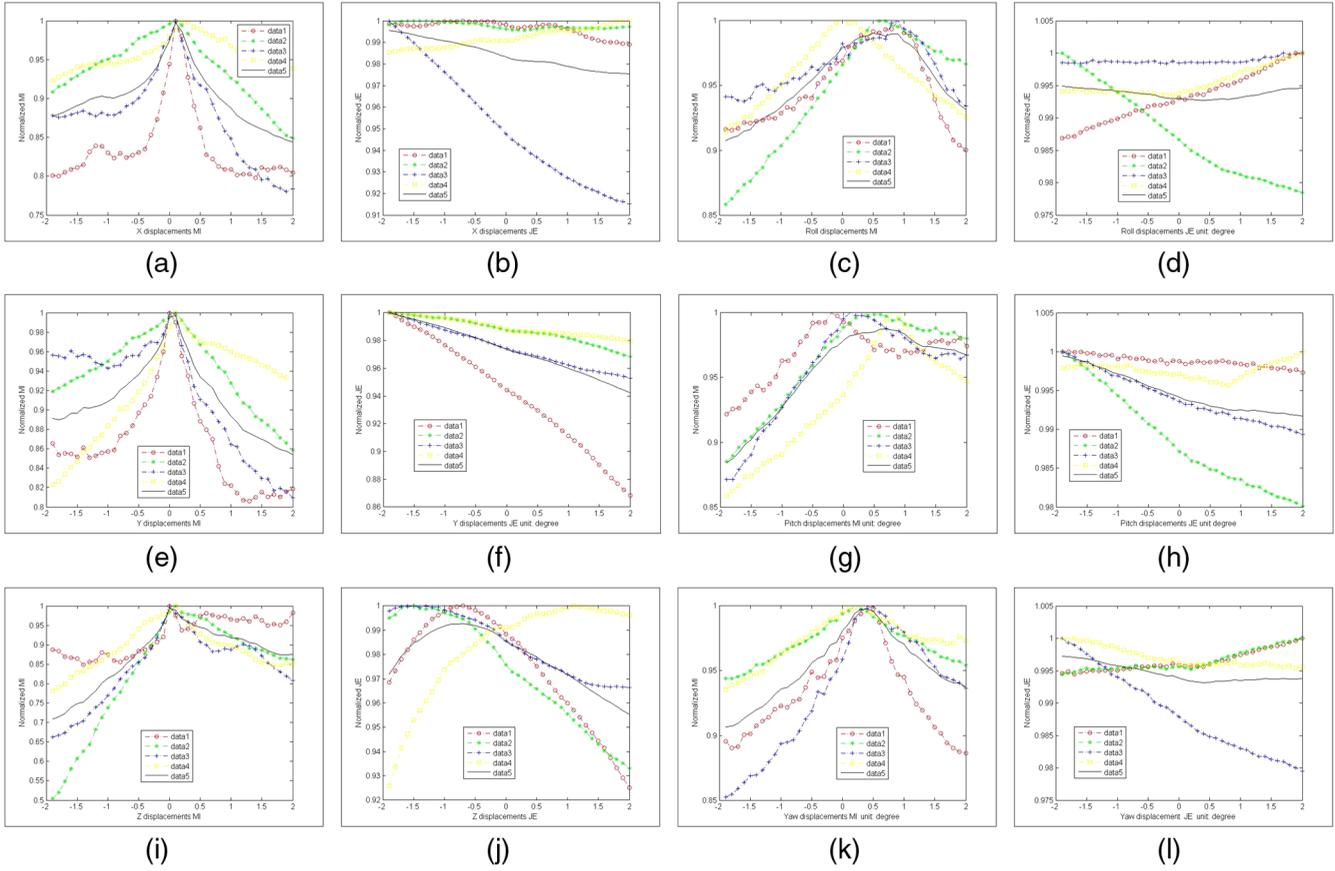
Fig. 2 Probe analysis on camera translation and orientation parameters for evaluation of normalized mutual information (MI) and joint entropy (JE): (a) *X* displacements MI. (b) *X* displacements JE. (c) Roll displacements MI. (d) Roll displacements JE. (e) *Y* displacements MI. (f) *Y* displacements JE. (g) Pitch displacements MI. (h) Pitch displacements JE. (i) *Z* displacements MI. (j) *Z* displacements JE. (k) Yaw displacements MI. (l) Yaw displacements JE.

corresponds to the mean values of the other traces and indicates the general trend. Note that in all cases, the normalized MI yields an unambiguous maximum, whereas the JE measure is ambigous in most cases. Hence, our conclusion is that JE is an inappropriate metric for use in the case of ground-level data.

3.1 Coordinate Framework

Our panoramic images are generated from a Ladybug III system consisting of six Fisheye cameras, and the individual images are then stitched together via α blending. The resulting mosaic is transformed into a spherical panorama via a cylindrical equidistant projection as shown in Fig. 3(a). To generate a linear perspective image, such as the one corresponding to the spherical panorama in Fig. 3(a) and shown in Fig. 3(b), the panorama is mapped onto a sphere and viewed from the center (virtual camera).

Both LiDAR and image data are georeferenced. We first convert the geographic coordinates (latitude and longitude) into Earth-centered, Earth-fixed (ECEF) coordinates, and then they are transformed into local tangent plane (LTP) coordinates. All computations are based on LTP coordinates. Each LiDAR point $p = (x, y, z)^T$ in LTP coordinates is converted into spherical coordinates (θ, φ) by Eq. (3)

$$\theta = \arccos\left(\frac{z}{\sqrt{x^2 + y^2 + z^2}}\right), \qquad \varphi = \arctan 2(y, x), \quad (3)$$

where θ is the inclination ($\theta \in [0, \pi]$), and φ is the azimuth ($\varphi \in (-\pi, \pi]$). Each point's corresponding location in the panoramic image (r, c) is computed by Eq. (4)

$$r = \text{int}\left[\frac{\theta}{\pi} H\right], \qquad c = \text{int}\left[\left(\frac{\varphi}{2\pi} + 0.5\right) W\right], \quad (4)$$

where H and W are the height and width of the panoramic images, respectively.

3.2 Mutual Information Registration

MI methods have been widely used for the multimodal registration problem in the medical imaging domain (e.g., registration of CT and MRI). Recently, they also have been applied to the problem of registering airborne LiDAR data with oblique aerial images.[15] The MI of two random variables X and Y can be defined as

$$I(X; Y) = \int_Y \int_X p(x, y) \log\left[\frac{p(x, y)}{p_1(x)p_2(y)}\right] dx dy, \quad (5)$$
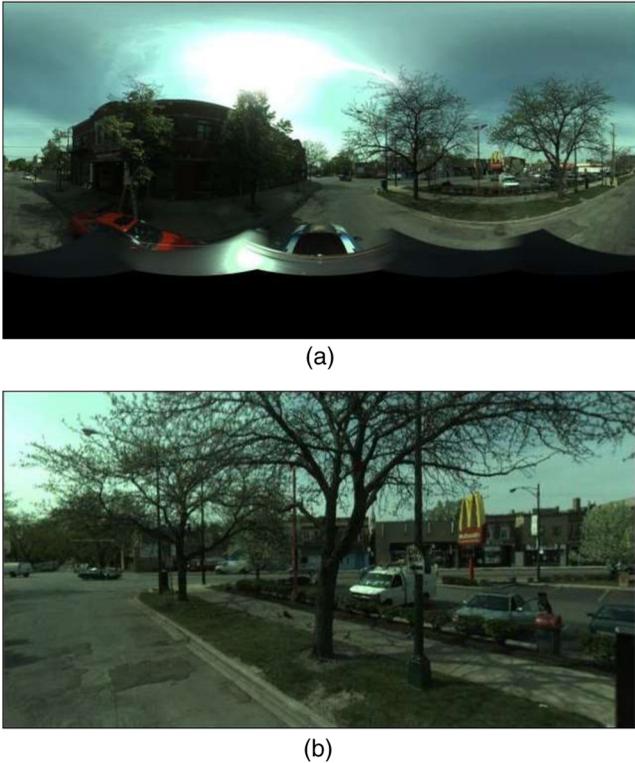
(a)

(b)

Fig. 3 (a) Spherical panorama generated from six spherical images. (b) Linear perspective view corresponding to the spherical panorama.

where $p(x, y)$ is the joint probability density function of X and Y, and $p_1(x)$ and $p_2(y)$ are the marginal probability density functions of X and Y, respectively. The problem here is to estimate the correction of the relative camera pose between the LiDAR sensor and the Ladybug camera. The spherical panorama is chosen as a fixed image, because the camera view point has to stay in the center of the sphere to generate perspective images. Once the camera moves out of the sphere center, the spherical image will be distorted. The LiDAR image is selected as a moving image, where

new LiDAR images are generated at each iteration during the optimization process. Both LiDAR and spherical images are rendered onto a plane from the camera center using OpenGL for the MI evaluation under a pinhole camera model. The perspective camera image is generated by rendering the spherical panorama with a view port of 940×452 pixels. The LiDAR dataset is normally very large. In our experiments, each scene contains 8×10^6 LiDAR points. To make 3-D rendering efficient, we also integrate the OpenGL rendering of the LiDAR features into the registration pipeline to speed up the optimization process.

We use three different representations of the LiDAR data with spherical panoramas for evaluating MI. The first representation of LiDAR is the projected LiDAR points with intensity information [see Fig. 4(b)]. We call it a LiDAR intensity image which looks similar to a gray-scale camera image. We use 256 bins for representing LiDAR intensity images. The second representation is the projected LiDAR points without intensity information [see Fig. 4(c)]. We use two bins for representing the binary images. The third is the depth map of the LiDAR point cloud [see Fig. 4(d)]. The point cloud is rendered with depth intensities with 256 bins, where brighter points indicate a further distance to the camera center. We use gray-scale camera images instead of color images [see Fig. 4(a)] for the MI computation. Note that the bin size in each of the three cases is determined by quantization, e.g., 8-bits for intensity and 1-bit for the binary image corresponding to the LiDAR projection in the image plane. The choice of normalization for depth on the interval [0, 255] was empirically determined. We did not investigate the effect of different quantizations in this study.

We use the Nelder-Mead downhill simplex method[22] to optimize the cost function as it does not require derivatives of the cost function. It is a commonly used nonlinear optimization technique that is well suited to multidimensional, unconstrained minimization as is the case here. As it is applicable only to convex minimization, we must ensure that the initial starting point is in the vicinity of the correct local minimum. In practice, we know the approximate rotation and
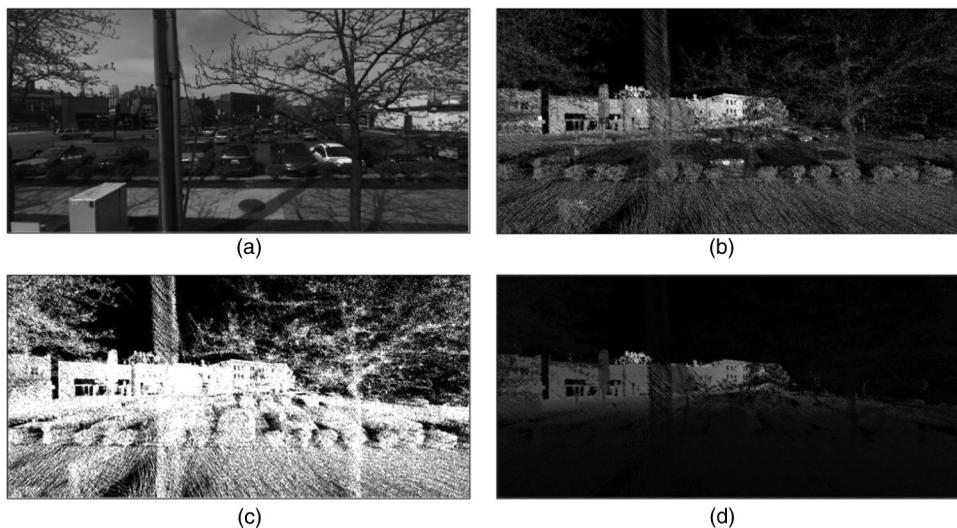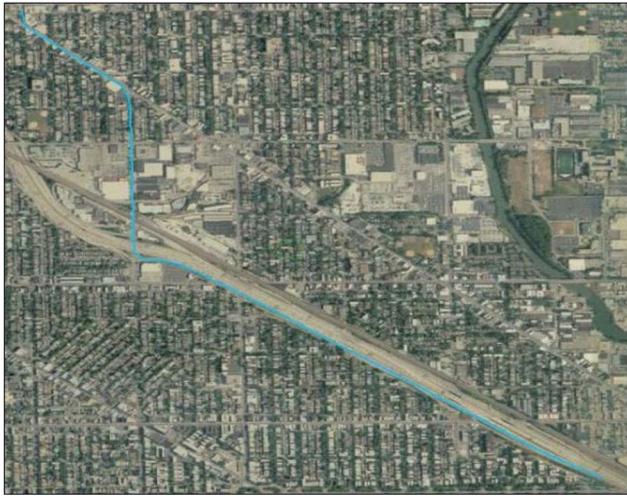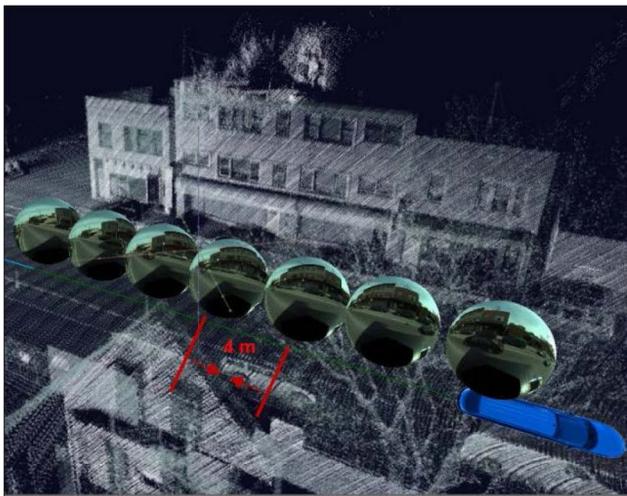

(a)

(b)

(c)

(d)

Fig. 4 Camera image and three representations of LiDAR point clouds in a same scene: (a) camera grayscale image, (b) LiDAR intensity image, (c) LiDAR points, (d) LiDAR depth map.

(a)

(b)

Fig. 5 Test data: (a) test data overview, (b) an illustration of the data.

translation parameters by virtue of measurements of the data acquisition platform, making finding a suitable starting point straightforward.

4 Experiments

For the experiments, we made the algorithm automatically run through an approximate 4-km drive. The driving routine is shown with the light blue line in Fig. 5(a). An illustration of the collected data is shown in Fig. 5(b), where the distance
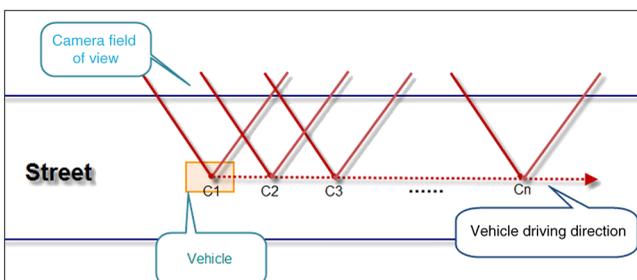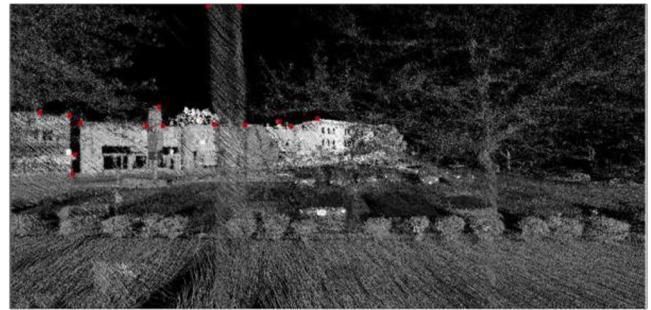

Fig. 6 Camera configuration.

(a)

(b)

Fig. 7 Optical image (a) and LIDAR image (b) with 15 manually selected correspondence points.

between each sperical panorama is around 4 m. The test data were collected in the northwestern suburban of Chicago, Illinois, which include residential, urban streets, and highway scenes. The data are in binary format containing around 4 GB LiDAR data (about 226×10^6 points) and 1 GB panoramic images (814 spherical panoramas). We use the camera views perpendicular to or parallel to the vehicle driving direction to generate perspective images. Figure 6 illustrates the camera views vertical to the vehicle driving direction. The distance between each camera (e.g., C_1, C_2) is around 4 m as the images are taken around every 4 m. The camera views parallel to the driving direction are similar to Fig. 6 except the camera view points to the front.

For our analysis, we selected 10 representative urban scenes shown in Fig. 8 for the evaluation using the three different representations of the LiDAR data described earlier. Images 1 and 2 show a parking lot in a commercial area (possibly shopping mall) from differnet viewpoints. Images 3 and 4 show a urban street in two different scenarios: with or without trees. Images 5 and 6 show two different residential areas. Images 6 and 7 show two different highway scenes. Image 9 shows a scene where trees are major objects, and image 10 shows a scene where houses are major objects. We start with an approximate initial registration that is determined from the data acquisition platform. The initial camera pose corrections are set to zero. The optimization will compute the final camera corrections. The experiments were performed on a laptop PC with a dual core 2.60 GHz processor and 2 GB of RAM. The NVIDIA Quadro NVS 135 M video card was used. The registration algorithms were implemented in C++, and the implementations of MI and amoeba optimization were from insight segmentation and registration toolkit.[23] We adjust the tolerances on the optimizer to define convergence. The tolerance on the six parameters is 0.1 (the
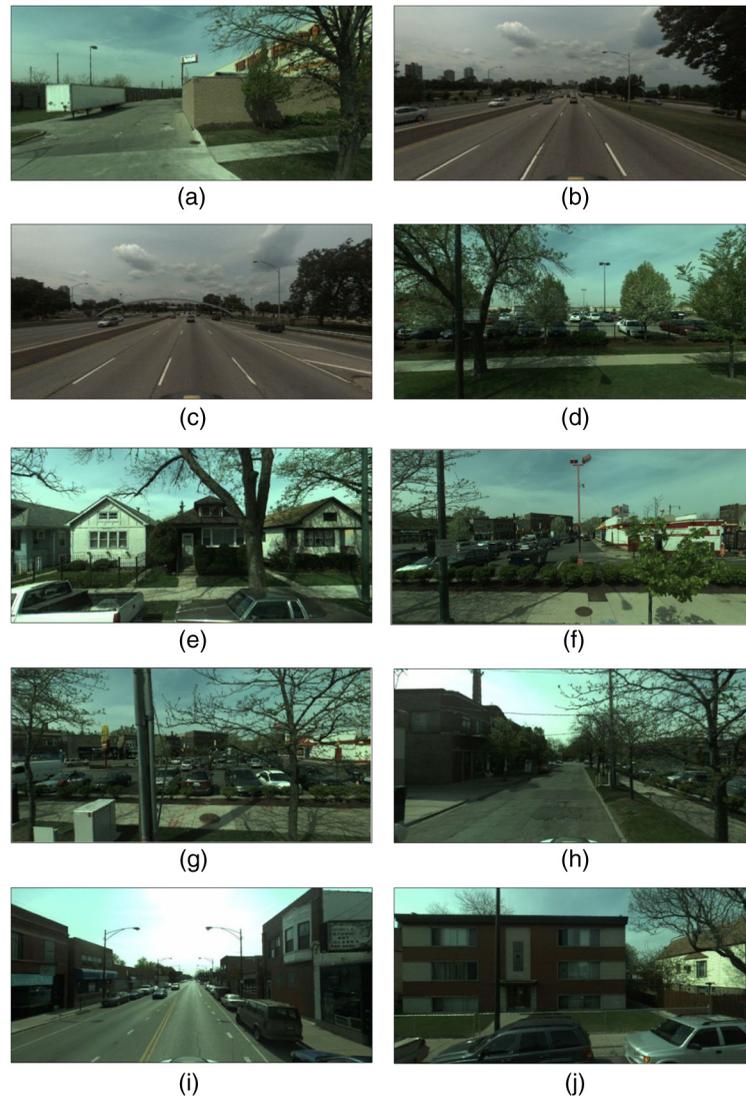
Fig. 8 Images used for experiments: (a) Image 1; (b) Image 2; (c) Image 3; (d) Image 4; (e) Image 5; (f) Image 6; (g) Image 7; (h) Image 8; (i) Image 9; (j) Image 10.

unit for translation parameters is meters and degrees for orientation parameters). We also set the tolerance on the cost function value to define convergence. The metric returns the value of MI, for which we set the tolerance to be 0.001 bits. The initial size of the simplex is set to 1, and the maximum iteration number is set to 200. In our experiments, almost all registrations converged in less than 150 iterations.

4.1 Performance Evaluation

To quantitatively measure the registration results, we compare the registration accuracy in terms of pixel offset between LiDAR and camera images before and after the registration. We manually selected 15 correspondence points in each spherical image and LiDAR intensity image. Figure 7 shows an example of a spherical image and a LIDAR intensity image marked with 15 correspondence points. Figure 9 shows the computed Euclidean distance histogram of the correspondence points for each scene in Fig. 8. In Fig. 9, for instance, Image 1B shows the histogram of the pixel offsets (we compute the

histograms in terms of the offset errors (pixels) among these correspondence points) for the scene indicated by Fig. 8(a) (Image 1) before registration, and Image 1A shows the corresponding pixel offsets after registration. The horizontal axis corresponds to the pixel offsets, and the vertical axis corresponds to the frequency. Image 1B shows that most of the pixel offsets are 2 pixels, and Image 1A shows that most of the pixel offsets are within 1 pixel after MI registration. A similar interpretation applies to the rest of the figures. Figure 10 shows the computed Euclidean distance histograms of the correspondence points for all the 10 images before and after registration. Before the MI registration, most correspondence points have 2 to 3 pixel errors. After the registration, most of the correspondence points are within 1 pixel. The pixel offset histograms using other LiDAR representations are similar.

Table 1 shows the run time for the 10 representative scenes. Testing on the 4-km drive shows that using the LiDAR points without intensity normally runs quickly
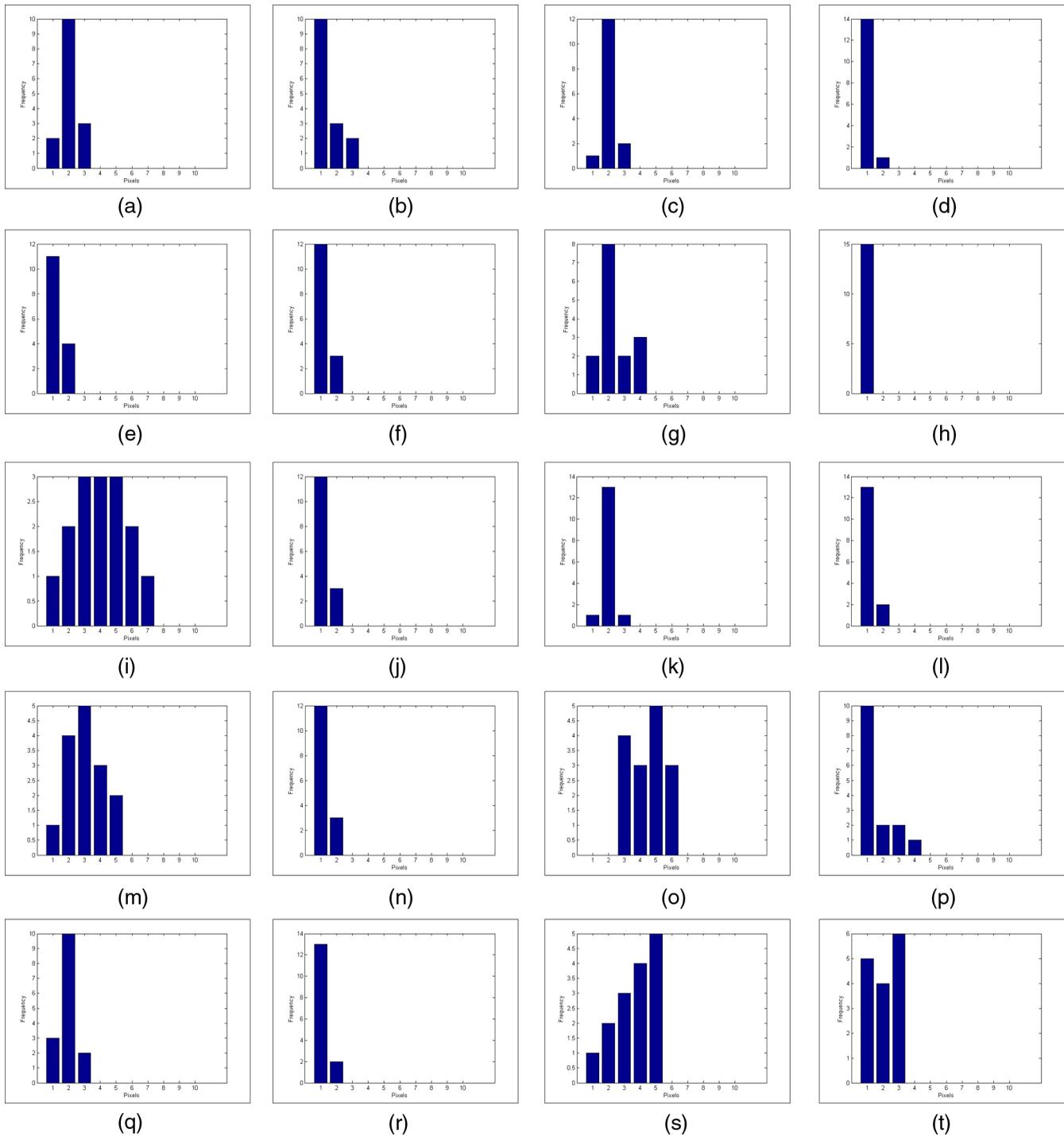
Fig. 9 Registration error analysis, *X*-axis stands for pixels, and *Y*-axis stands for the frequency: (a) Image 1B; (b) Image 1A; (c) Image 2B; (d) Image 2A; (e) Image 3B; (f) Image 3A; (g) Image 4B; (h) Image 4A; (i) Image 5B; (j) Image 5A; (k) Image 6B; (l) Image 6A; (m) Image 7B; (n) Image 7A; (o) Image 8B; (p) Image 8A; (q) Image 9B; (r) Image 9A; (s) Image 10B; (t) Image 10A.

with fewer iterations. Using LiDAR points with intensity normally performs the most robustly followed by using LiDAR points without intensity and using the LiDAR depth maps. We also study the convergence of the optimization using three different measures of MI. Without loss of generality, we choose the data shown in Fig. 4 as an example. Figure 11 shows the sequence of metric values computed as

the optimizer searched the parameter space using these three different representations of the LIDAR data. The measure initially increases overall with the number of iterations. After about 50 iterations, the metric value reaches a steady state without further noticeable convergence.

An example of the registration is shown in Fig. 12. After MI registration, the misalignment is not noticeable.

Table 1 Registration times in minutes for correctly registered images.

MI measure	LiDAR intensity	LiDAR without intensity	LiDAR depth map
Image 1	0.86	0.51	0.93
Image 2	0.93	0.50	1.08
Image 3	1.05	0.63	0.97
Image 4	0.87	0.55	0.88
Image 5	0.70	0.4	1.06
Image 6	0.83	0.38	0.85
Image 7	0.85	0.38	0.65
Image 8	0.73	0.38	0.87
Image 9	0.97	0.50	0.71
Image 10	1.03	0.50	0.75
Mean	0.88	0.47	0.87

4.2 Perturbation Analysis

We plot the normalized MI of Fig. 4(a) in Fig. (13) using the three LiDAR attributes. Figure 13 (red intensity, green point only, and yellow depth map) demonstrates that each curve has a single peak over a subset of the displacement parameters around the initial registration, which demonstrates the effectiveness of the maximization of MI for computing optimal camera corrections.

We also investigated the failure cases. In our experiments, the algorithm works well in feature-rich environments such as residential areas, but often fails in scenes with sparse features or containing moving objects like cars, particularly highway scenes. In our case, the highway scenes mostly fail. The partial overlap between LiDAR point clouds and camera images is another reason. The LiDAR scanner only can reach up to 120 m, while the camera can always have a larger field of view than the LiDAR scanner. Figures 14(a) and 14(b) show one typical failure case in a highway scene. The cars in the camera image [Fig. 14(a)] do not appear in the LiDAR image [Fig. 14(b)]. The LiDAR image only partially covers the camera image; for instance, the trees and buildings in the far distance in the camera image do not appear in the LiDAR image. In Ref. 15, the authors claim that they only use the image pixels with corresponding projected LiDAR points for MI calculation and others are considered background points and discarded. We tried to discard the background points and only use overlapping regions for MI computation, but the results were worse than using entire images. When using entire images, the background such as sky appears similar in both LiDAR and camera images, which largely contributes to the MI score. When using overlapping regions for MI computation, the LiDAR images contain no sky. Therefore, the background is not used in the MI computation, which affects the MI evaluation. Failure cases are also due to overexposed images [Figs. 14(c) and 14(d)],
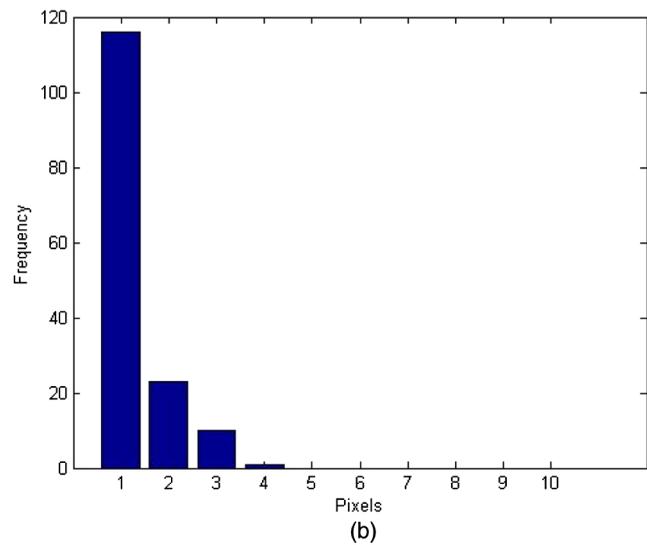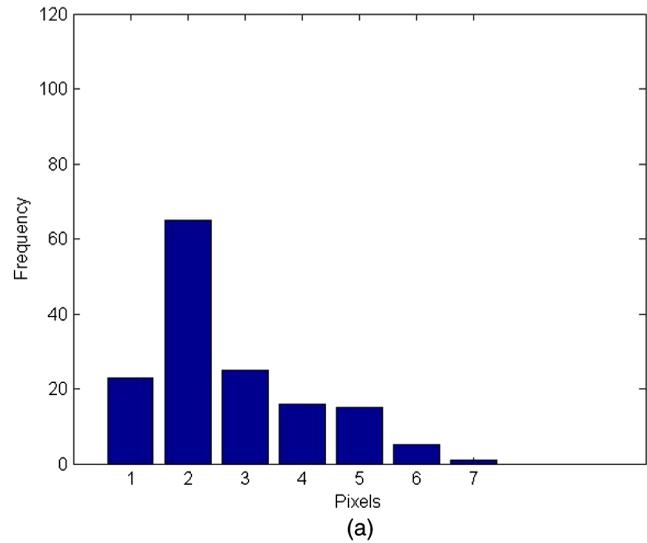

(a)

(b)

Fig. 10 Histogram of Euclidean distance of pixel offset for before registration (a) and after registration (b). The histograms were generated with samples for all 10 test images.
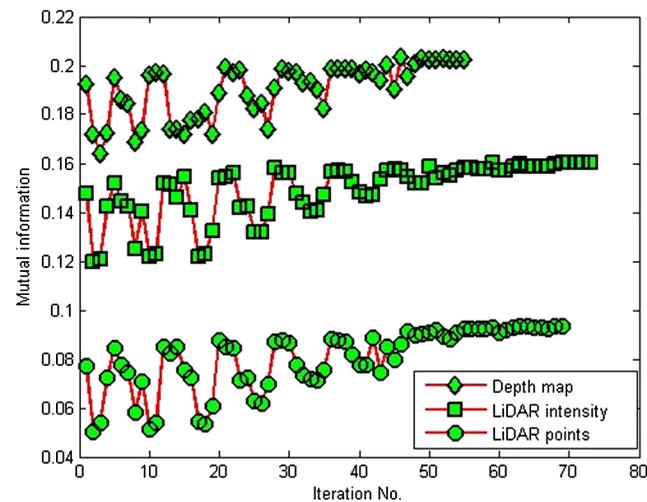

Fig. 11 Mutual information values produced during the registration process.

Fig. 12 An example of the registration results.

particularly in the case where the vehicle drives through/out of a tunnel.

4.3 Camera Pose Corrections

One of our interests is to investigate how camera pose errors change during the data collection. To do so, we manually selected 100 successful registrations (using the registrations from camera views vertical to the vehicle driving direction) by carefully examining the alignment of major features in the registered images, and plotting the camera pose corrections as shown in Fig. 15. Figure 15(a) shows the camera translation corrections, and Fig. 15(b) shows the camera
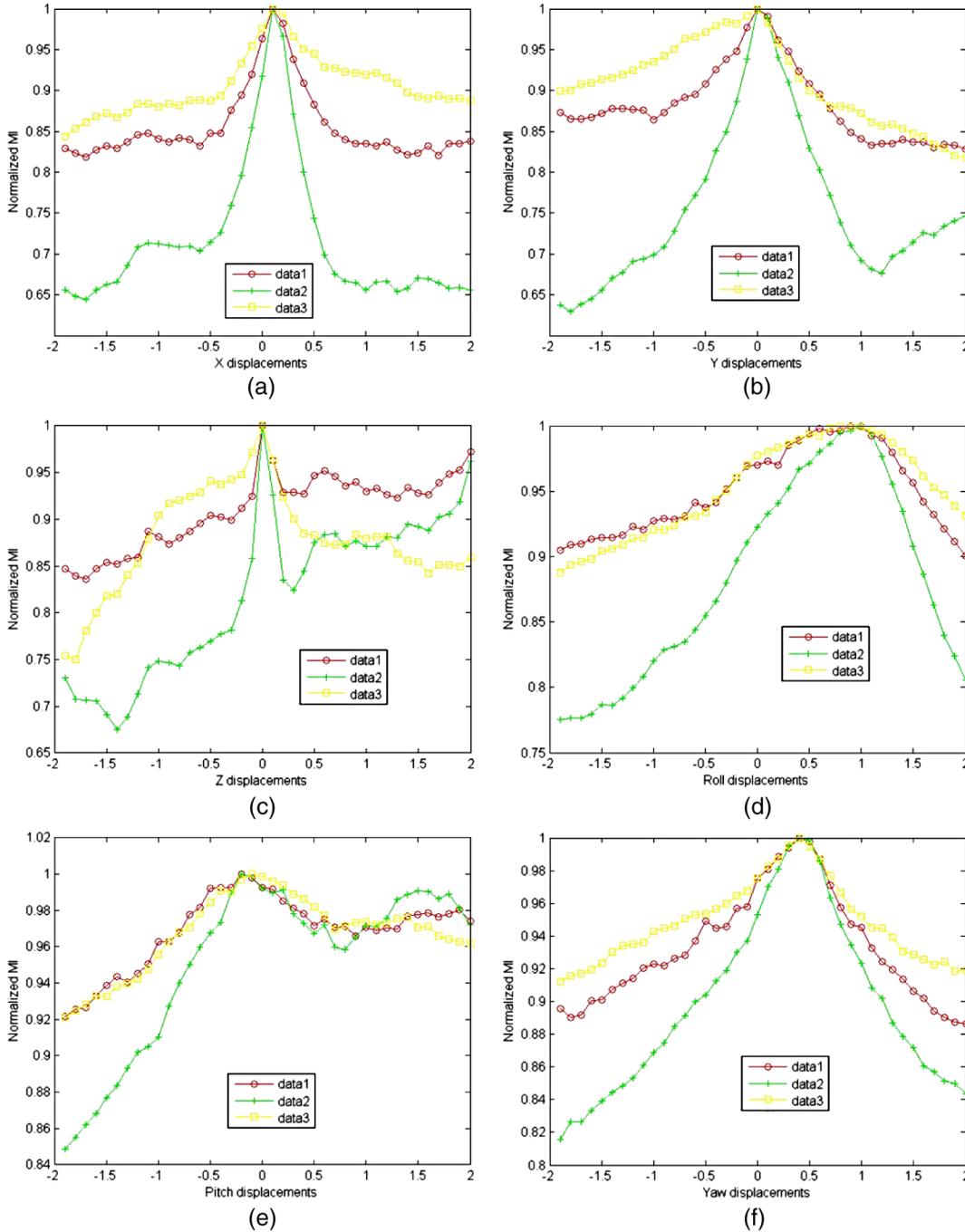

Fig. 13 Plots of normalized mutual information: (a) X displacements. (b) Y displacements. (c) Z displacements. (d) Roll displacements. (e) Pitch displacements. (f) Yaw displacements.
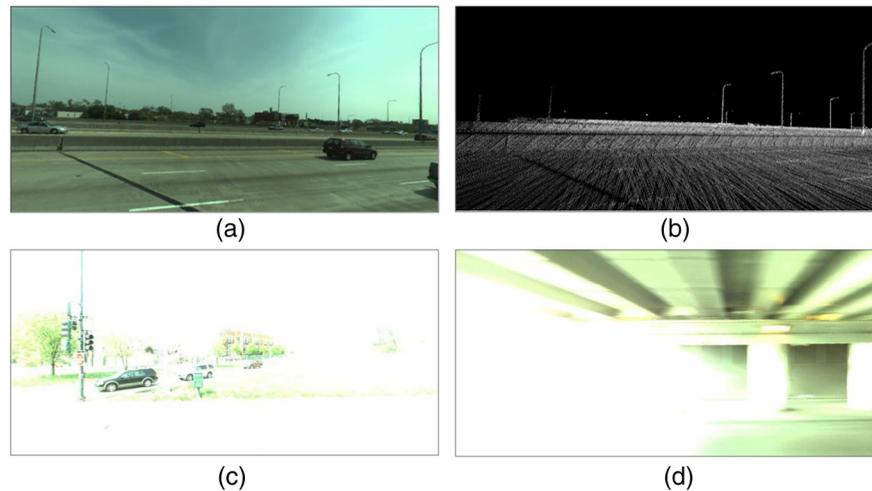
Fig. 14 Failure cases: (a) Ladybug image. (b) LiDAR image. (c) Overexposed Ladybug image I. (d) Overexposed Ladybug image II.
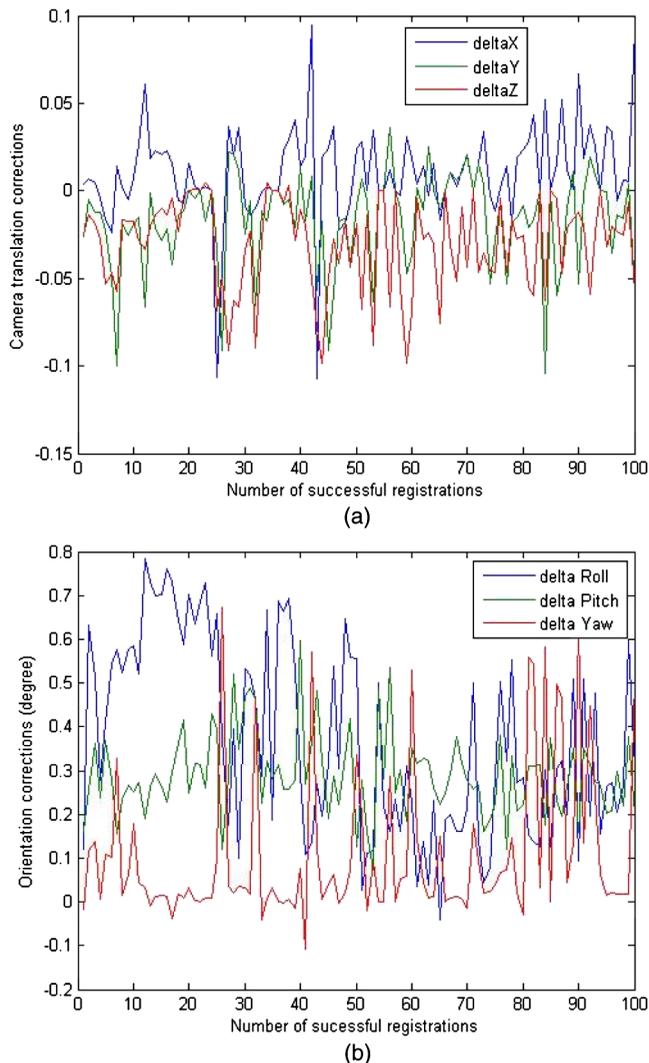

Fig. 15 Plots of camera pose corrections using 100 successful registrations: (a) camera translation corrections, (b) camera orientation corrections.

orientation corrections. Our observation is that almost all of the camera translation corrections are within 0.1 m, while the orientation corrections are within 1 deg.

5 Conclusions and Future Work

In this paper, we have investigated MI registration for ground-level LiDAR and images. The existing method[15] for registering airborne LiDAR with aerial oblique images does not work on the LiDAR and images collected from the mobile mapping system, because the assumption used in Ref. 15 is violated in the case of mobile LiDAR data. Instead of the minimization of the JE, we use the maximization of MI for computing optimal camera corrections. The algorithms work with unstructured LiDAR data and perspective rectified panoramic images generated by rendering a panorama into an image plane using spheric views. We tested the algorithm on various urban scenes using three different representations of LiDAR data with camera images for the MI calculation. Our mutual registration algorithm automatically runs through large-scale mobile LiDAR and panoramic images collected over a metropolitan scale. It is the first example we are aware of that tests MI registration in a large-scale context. With the initiative of urban 3-D modeling from location-based service providers such as Nokia and Google, this work is particularly important for combining ground-level range and visual data for large-scale photorealistic city modeling.

We generated perspective images from spherical images using the view either perpendicular or parallel to the vehicle driving direction. Therefore, we just used one-sixth of the entire spherical image for the MI registration, which does not efficiently use all the available information contained in the 360 deg panoramic images. For future work, one possible approach would be to project the entire LiDAR points along with spherical images onto six cube faces using a quadrilateralized spherical cube mapping[24] or other linear projections. Because the sky and the ground do not provide much useful information, we actually need just four faces for the MI registration. To speed up the computation, a multiresolution approach can be employed by establishing image pyramids on both images. This coarse-to-fine strategy can

improve the performance of the registration algorithm and also increases robustness by eliminating local optima at coarser levels. One of the limitations of the MI metric is that the intensity histograms contain no spatial information. One possible direction is to incorporate spatial context into the metric to improve the robustness of the similarity measure.

Beyond these incremental approaches, there are limits to what can be achieved on a practical basis. However, since the task is 3-D data acquisition, data may be discarded and reacquired as necessary. Thus, future research will also be aimed at automatic detection of the different failure modes so that reacquisition can be automatically initiated.

References

1. J. Diebel and S. Thrun, "An application of Markov random fields to range sensing," in *Proc. Conf. on Neural Information Processing Systems (NIPS), Advances in Neural Information Processing Systems*, Vol. 18, Vancouver, British Columbia (2005).
2. J. Dolson et al., "Upsampling range data in dynamic environments," *IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*, Vol. 0, pp. 1141–1148, San Francisco, California (2010).
3. L. A. Torres-Méndez and G. Dudek, "Reconstruction of 3D models from intensity images and partial depth," in *Proc. 19th National Conf. on Artifical Intelligence, AAAI'04*, pp. 476–481, Advancement of Artificial Intelligence, San Jose, California (2004).
4. Q. Yang et al., "Spatial-depth super resolution for range images," *IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*, Vol. 0, pp. 1–8, Minneapolis, Minnesota (2007).
5. M. Deveau et al., "Strategy for the extraction of 3D architectural objects from laser and image data acquired from the same viewpoint," in *Proc. Int. Arch. of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36, S. El-Hakim, F. Remondino, and L. Gonzo, Eds., Mestre-Venice, Italy (2005).
6. S. Barnea and S. Filin, "Segmentation of terrestrial laser scanning data by integrating range and image content," in *Proc. XXIth ISPRS Congress*, International Society for Photogrammetry and Remote Sensing, Beijing, China (2008).
7. P. Dias et al., "Registration and fusion of intensity and range data for 3D modeling of real world scenes," in *Proc. Int. Conf. on 3D Digital Imaging and Modeling*, pp. 418–425, IEEE, Quebec City, Canada (2003).
8. S. Becker and N. Haala, "Combined feature extraction for facade reconstruction," in *Proc. ISPRS Workshop on Laser Scanning*, ISPRS, Espoo, Finland (2007).
9. M. Ding, K. Lyngbaek, and A. Zakhor, "Automatic registration of aerial imagery with untextured 3D LiDAR models," in *Proc. Computer Vision and Pattern Recognition (CVPR)*, IEEE, Anchorage, Alaska (2008).
10. D. G. Aguilera, P. R. Gonzalvez, and J. G. Lahoz, "An automatic procedure for co-registration of terrestrial laser scanners and digital cameras," *ISPRS J. Photogramm. Remote Sens.* **64**, 308–316 (2009).
11. L. Liu and I. Stamos, "Automatic 3D to 2D registration for the photo-realistic rendering of urban scenes," in *Proc. 2005 IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 02, pp. 137–143, San Diego, California (2005).
12. L. Liu and I. Stamos, "A systematic approach for 2D-image to 3D-range registration in urban environments," in *Proc. Int. Conf. on Computer Vision*, pp. 1–8, IEEE, Rio de Janeiro, Brazil (2007).
13. I. Stamos et al., "Integrating automated range registration with multi-view geometry for the photorealistic modeling of large-scale scenes," *Int. J. Comput. Vis.* **78**, 237–260 (2008).
14. L. Wang and U. Neumann, "A robust approach for automatic registration of aerial images with untextured aerial LiDAR data," *IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition, CVPR 2009*, pp. 2623–2630, IEEE, Miami, Florida (2009).
15. A. Mastin, J. Kepner, and J. Fisher, "Automatic registration of LiDAR and optical images of urban scenes," in *Proc. IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 2639–2646, Miami, Florida (2009).
16. D. G. Lowe, "Distinctive image features from scale-invariant key-points," *Int. J. Comput. Vis.* **60**, 91–110 (2004).
17. W. Förstner and E. Gülch, "A fast operator for detection and precise location of distinct points, corners and centres of circular features," in *Proc. ISPRS Intercommission Workshop Interlaken*, pp. 281–305, Interlaken, Switzerland (1987).
18. M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, **24**(6), pp. 381–395 (1981).
19. D. G. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artif. Intell.* **31**, 355–395 (1987).
20. R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, New York (2004).
21. P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.* **24**, 137–154 (1997).
22. J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comp. J.* **7**, 308–313 (1965).
23. "Insight segmentation and registration toolkit (ITK)," Kitware Inc, http://www.itk.org/ (13 June 2013).
24. F. P. Chan and E. M. O'Neill, "Feasibility study of a quadrilateralized spherical cube earth data base," Computer Sciences Corp., EPRF Tech. Report 2-75, Prepared for the Environmental Prediction Research Facility, Monterey, California (1975).

Ruisheng Wang joined the Department of Geomatics Engineering at the University of Calgary as an assistant professor in 2012. Prior to that, he had worked as an industrial researcher at NOKIA in Chicago since 2008. He holds a PhD degree in electrical and computer engineering from McGill University, an MSc.E. degree in geomatics engineering from the University of New Brunswick, and a BEng degree in photogrammetry and remote sensing from Wuhan University, respectively.

Frank P. Ferrie received his BEng, MEng, and PhD degrees in electrical engineering from McGill University in 1978, 1980, and 1986, respectively. Currently, he is a professor in the Department of Electrical and Computer Engineering at McGill University and codirector of the REPARTI research network in distributed environments. His research interests include computer vision, primarily in the areas of two and three-dimensional shape analysis, active perception, dynamic scene analysis, and machine vision.