A 3D Reconstruction with High Density and Accuracy using Laser Profiler and Camera Fusion System on a Rover

Ryoichi Ishikawa, Menandro Roxas, Yoshihiro Sato, Takeshi Oishi
Institute of Industrial Science, The University of Tokyo
Komaba, Meguro, Tokyo, Japan
{ishikawa, roxas, yoshi, oishi}@cvl.iis.u-tokyo.ac.jp

Takeshi Masuda
National Institute of Advanced
Industrial Science and Technology
Kasumigaseki, Chiyoda, Tokyo, Japan
t.masuda@aist.go.jp

Katsushi Ikeuchi
Microsoft Research Asia
Danling Street, Haidian District,
Beijing, China
katsuike@microsoft.com

Abstract

3D sensing systems mounted on mobile platform are emerging and have been developed for various applications. In this paper, we propose a profiler scanning system mounted on a rover to scan and reconstruct a bas-relief with high density and accuracy. Our hardware system consists of an omnidirectional camera and a 3D laser scanner. Our method selects good projection points for tracking to estimate motion stably and reject mismatches caused by difference between the positions of laser scanner and camera using an error metric based on the distance from omnidirectional camera to scanned point. We demonstrate that our results has better accuracy than comparable approach. In addition to local motion estimation method, we propose global poses refinement method using multi modal 2D-3D registration and our result shows good consistency between reflectance image and 2D RGB image.

1. Introduction

3D digital archiving is a technique that uses 3D scanning systems in creating models of existing real-world objects. It has particularly played an important role in the preservation and restoration of cultural heritage assets. These assets are exposed to the risk of deterioration due to natural and/or man-made causes and preserving their fine shape and structure is of great interest. For example, Fig. 1 (a) shows a bas-relief in a Khmer-style temple which contains a delicately fine detail. Moreover, this relief is carved over a very wide area and it is challenging to produce a high density scan of all its parts.
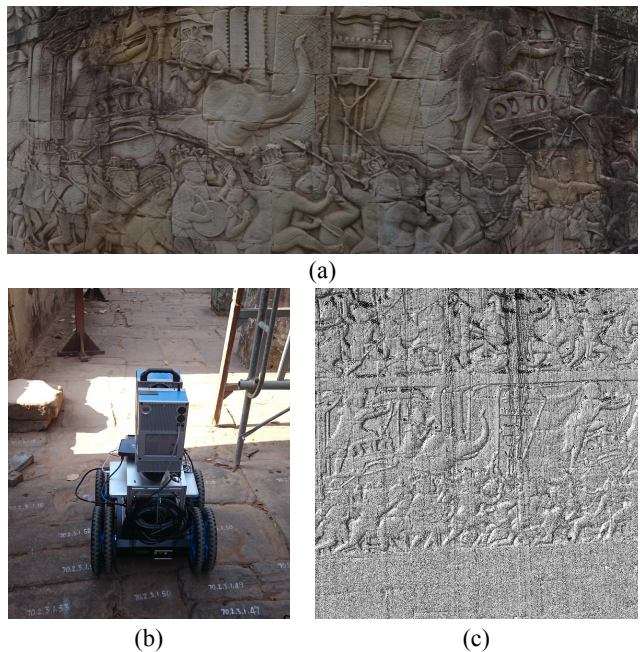

(a)

(b)   (c)

Figure 1. (a) A bas-relief in a Khmer-style temple. (b)Proposed scanning sensor system mounted on a rover. (b)Obtained reconstruction result of relief by our method.

Traditional modeling methods use stationary laser scanner placed in different locations of the scanned object. The obtained data is later aligned and merged. Using these methods on large data is laborious. Alternatively, various scanning systems mounting the laser scanner on a mobile platform can help speed up the scanning process such as LiDAR-only system [26, 31], LiDAR + IMU and GPS sys-

tem [25, 3, 22], LiDAR + camera system [32, 2, 33, 1].
Zheng *et al.* [33] proposed a rail system for scanning re-
lief. However, it is heavy, difficult to transport and has very
limited moving flexibility. To address this problem, we have
developed a rover scanning system (See Fig. 1 (b)) which is
far lighter and less constrained.

Mobile scanning platforms introduce a problem of mo-
tion estimation during merging and alignment. In this pa-
per, we present a sensor motion estimation method by tri-
angulating correspondences in 2D panoramic image and
3D point clouds from calibrated and synchronized omni-
directional camera and laser scanner. We construct 2D-3D
correspondences using projection and point tracking and
only choose pairs where the pixels of projected 3D points
constitutes a high curvature. We then estimate the poses
of the camera frames using triangulation. An error met-
ric based on distance from laser scanner to scanned point
is adopted to reject outlier from 2D-3D mismatches caused
by difference of positions between the laser scanner and the
omni directional camera. Our scanning system can obtain
reconstruction 3D point clouds of relief with good fidelity
as shown in Fig. 1 (c).

In addition to local motion estimation by triangulation,
we also demonstrate a multi-modal 2D-3D registration be-
tween RGB color from camera and reflectance of laser by
deforming estimated motion using mutual information.

This paper makes the following contributions.

- Our rectification method obtains stable construction
  results by directly projecting 3D points onto 2D im-
  ages and only select points on pixels with high curva-
  ture.

- We propose a multi-modal registration for motion esti-
  mation that result in a consistent projected reflectance
  image and 2D RGB image.

- Our method considers mismatches due to the differ-
  ence of positions between the laser scanner and the
  camera and rejects them as outlier by adopting the
  error metric based on distance from omni-directional
  camera to scanned point.

Our hardware system mainly consists of a rover, a 3D
laser scanner and an omnidirectional camera as shown in
Fig. 2. The laser scanner is used in profiler mode, that is,
the scans are round slices that move with the rover. To syn-
chronize the data, a trigger signal is sent from the laser scan-
ner to the camera when scanning starts. The two sensors are
positioned such that there is enough overlapping area on the
obtained data. The motion of laser scanner is then estimated
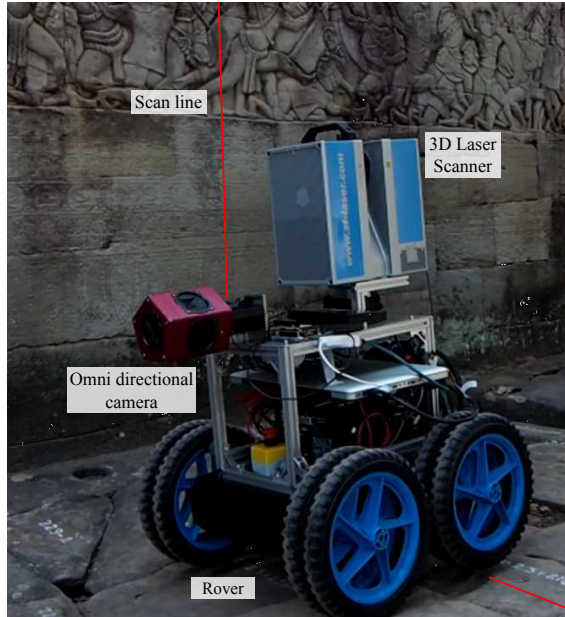using the 2D images and the 3D scans.

Figure 2. Hardware system

2. Related Work

Variety hardware systems are used in researches about
3D reconstruction with mobile platform.

Camera system

Recently, monocular camera systems based on SfM,
such as [7, 12, 20, 6, 13], have shown impressive results.
However, these methods generally do not solve the scale of
the motion without fusion with other sensors or assumptions
about the sensor motion. Alternatively, stereo camera sys-
tems, such as [21, 17], can determine scale of motion which
is provided by the baseline reference between two cameras.
Nevertheless, the accuracy and resolution is still far from
the level of the laser scanners.

RGB-D camera system

Motion estimation using both images and depth data
from RGB-D camera [19, 29, 4, 28] can be conducted eas-
ily and stably without the scale issue. A number of RGB-D
SLAM methods have also been proposed showing promis-
ing mapping results [10, 9, 11]. While the real-time CMOS
RGB-D sensor is good for motion tracking due to low dis-
tortion of depth map, Signal-to-Noise Ratio (SNR) is low
because of the low Light Concentration Ratio (LCR) [18].
Therefore RGB-D camera is not good for outdoor environ-
ments with strong background light and long range scan-
ning.

LiDAR-only system

On the other hands, laser scanning systems concentrate
the available light source power in a smaller region, re-
sulting in a larger SNR [18]. Laser scans can obtain far-
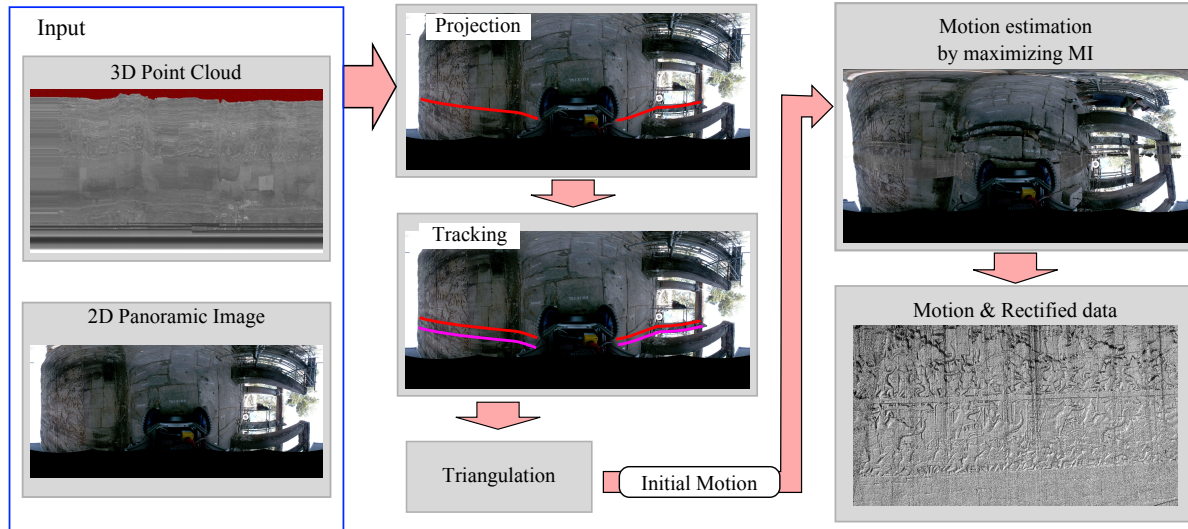ther distance measurements with lower noise than hand-

Figure 3. Method overview. The inputs are the synchronized and calibrated 3D point clouds and 2D image sequence. Initial motion is computed by 3D-2D projection, point tracking and triangulation. Then the motion is optimized using mutual information resulting in a motion and rectified data.

held RGB-D cameras because of large SNR. However it requires long acquisition time, which leads to the distortion of point clouds caused by movement during scanning. State estimation made with LiDARs only [26, 31] is good in terms of low hardware costs. Although LiDAR-only system not required to calibrate or synchronize with any other sensor, it is difficult to obtain accurate estimation results due to its long acquisition time.

LiDAR + IMU, GPS

One way to estimate LiDAR sensor motion is by incorporating other sensors to recover the motion. Incorporating IMU, GPS with LiDAR is simple and direct solution of motion estimation problem [25, 3, 22]. However, the accuracy of them is not enough for 3D digital archive.

LiDAR + Camera fusion system

Motion estimation methods with images or/and additional depth sensor (*e.g.*, 1-axis lidar) are well designed for SLAM system in robotics [32]. A proposed in [22] also combines lidar with RGB camera for indoor/outdoor 3D reconstruction which enhances the robustness of motion estimation. 3D laser scanner is used as profiler mode in scan systems proposed in [33, 2]. Bok *et al.* demonstrate 3D reconstruction of village-level scale heritage in [2] using four cameras and two 2D laser sensors with weak GPS prior. However our work requires millimeter units quality and it is difficult to estimate accurate sensor position using GPS.

3. Local camera pose estimation

We address the motion estimation problem during scanning by using laser scanner and camera fusion system. An overview of our method is shown in Fig. 3. The inputs are

3D point clouds from the laser scanner and 2D panoramic image sequence from the omni directional camera. These two sensors are calibrated and synchronized. The data is taken with the rover moving in the forward direction.

In our motion estimation method, first, 2D-3D correspondences are constructed by projection of 3D points onto 2D images. Then the 2D-3D correspondences are filtered and tracked using KLT tracker [15]. Local camera poses are then triangulated using linear and non-linear optimization.

3.1. Coordinate system

Laser scanner and omni-directional camera are calibrated in advance. Now, let i, j be the index of 2D frame and 3D point, t_i, t_j be a time that 2D frame i is taken and 3D point j is scanned, $\mathbf{X}_j^L$ be 3D coordinates of point j in the laser scanner local coordinates, $\mathbf{x}_i^j$ be the corresponded 2D point on image i of 3D point j. We want to estimate the sensor position $\mathbf{R}_i, \mathbf{T}_i$ at t_i. Translation from laser scanner to camera coordinates is as following. An operation to project $\mathbf{X}_j^L$ onto image is described as following.

$$\mathbf{x}_i^j = Proj(\mathbf{X}_j^L). \tag{1}$$

$\mathbf{X}_j^L$ can be translated to world coordinates X_j^W by the position parameter $\mathbf{R}_j, \mathbf{T}_j$ as:

$$\mathbf{X_j^W} = \mathbf{R}_j \mathbf{X}_j^L + \mathbf{T}_j. \tag{2}$$

$\mathbf{R}_j, \mathbf{T}_j$ are computed through linear interpolation of $\mathbf{R}_i, \mathbf{T}_i$.

3.2. 2D-3D correspondence

To make 2D-3D correspondences, 3D points scanned during $t_i - \Delta t < t_j < t_i + \Delta t$ are projected onto 2D frame i. Using all the projected point results in high computational cost and has the risk to include unstable points for tracking. To prevent this, only the points projected onto pixels which is located on high curvature edge are selected as feature points. Let $\mathbf{X}^L_{(i,j)}$ be local coordinates of 3D point j which is projected onto frame i and selected as a feature point and $\mathbf{X}^W_{(i,j)}$ is global coordinates of $\mathbf{X}^L_{(i,j)}$. Now $\mathbf{X}^L_{(i,j)}$ and $\mathbf{x}^j_i$ are corresponded by projection. Then $\mathbf{x}^j_i$ is tracked on the image sequential by using KLT tracker and we obtain corresponded point of $\mathbf{x}^j_i$ on each frame.

3.3. Triangulation

The triangulation is based on Zheng's method proposed in [33]. Assuming that the poses of camera frame 0 to n are already computed and point j, where $t_i - \Delta t < t_j < t_i + \Delta t$, is aligned to world coordinates $\mathbf{X}^W_j$. The pose of camera $n+1$ $\mathbf{R}_{n+1}, \mathbf{T}_{n+1}$ is initially computed by perspective 3-point algorithm [8] and RANSAC [5] with correspondences between $\mathbf{X}^W_j$ and $\mathbf{x}^j_{i+1}$. Then an energy function is constructed from the correspondences defined as:

$$
\begin{aligned}
\{\mathbf{R_{n+1}}, \mathbf{T}_{n+1}\} & \\
= \quad & \underset{\mathbf{R}, \mathbf{T}}{\arg\min} \sum_{i=1}^{n} \sum_{j} (\|Proj(\mathbf{R}^T(\mathbf{X}^W_{(i,j)} - \mathbf{T})) \\
- \quad & \mathbf{x}^j_{n+1}\|^2) \\
+ \quad & \sum_{i=1}^{n} \sum_{j} (\|Proj(\mathbf{R}^T_i((\mathbf{R}\mathbf{X}^L_{(n+1,j)} + \mathbf{T}) - \mathbf{T}_i)) \\
- \quad & \mathbf{x}^j_i\|^2).
\end{aligned} \tag{3}
$$

Eq. 3 indicates the summation of the reprojection errors of projected 3D-points in the world coordinates onto current frame $n+1$ and projected 3D-points in the local coordinates of current sensor position (at t_{n+1}) onto all of the previous frames (from 0 to n). Figure. 4 shows a schematic diagram of the triangulation method.

3.4. Error metric for outlier rejection

It is inevitable to make a mismatch with 3D-2D projection especially when the center positions of camera and laser scanner are gapped as shown in Fig. 5. This mismatch causes the error accumulation to the red arrow direction in Fig. 5. Moreover, it is difficult to remove this correspondence using reprojection error on the image when the baseline between two sequential camera positions is short because of the small angle θ between the two vectors from camera center to the tracked point and scanned point.

It is difficult to reject the mismatch based on angle even if there is large difference between distance from camera to
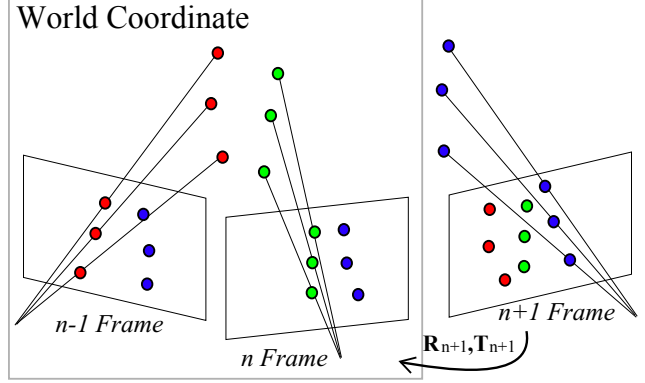

Figure 4. Triangulation. When 0 to n frames are worked out (only $n-1$ and n frames are shown in figure), pose of $n+1$ frame is firstly computed by Perspective 3 Point algorithm [8] using red and green points. Then pose is optimized by non-linear optimization with 2D-3D correspondences of red, green and blue points.
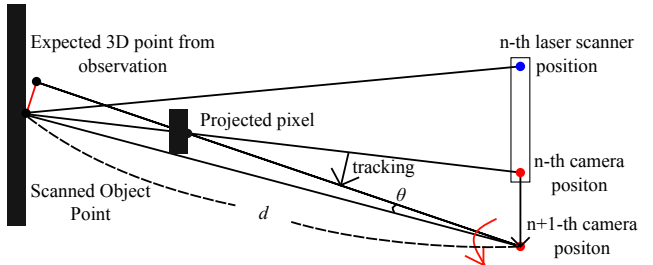

Figure 5. Mismatch of 3D point and 2D image are caused by sensor position gap. Error accumulates in the red arrow direction.

scanned 3D point and depth of corresponded pixel. On the other hand, translated 3D point from n to $n+1$-th camera coordinates must be located on extension line of tracked pixel direction at depth d and observation error in 3D coordinates become large. Therefore, we instead use $d \sin\theta$ (See Fig. 5) as error metric for outlier rejection. This formula indicates an approximate distance of red line in Fig. 5 due to small angle θ.

3.5. Bundle adjustment

Bundle adjustment is implemented to refine motion estimation. An energy function is based on reprojection error when 3D points onto all of the 2D frames. The energy function is defined as:

$$
\begin{aligned}
& \{\mathbf{R}_0 \dots \mathbf{R}_N, \mathbf{T}_0 \dots \mathbf{T}_N\} \\
& = \underset{\mathbf{R}_0 \dots \mathbf{R}_N, \mathbf{T}_0 \dots \mathbf{T}_N}{\arg\min} \sum_i \sum_j \sum_k \\
& \delta_{i,j} \quad \|Proj(\mathbf{R}^T_k((\mathbf{R}_i \mathbf{X}^L_{(i,j)} + \mathbf{T}_i) - \mathbf{T}_k)) - \mathbf{x}^j_k\|^2
\end{aligned} \tag{4}
$$

$$
\delta_{j,k} = \begin{cases} 1 & (\text{if } \mathbf{X}^W_j \text{ is detected on frame k}) \\ 0 & (\text{else}) \end{cases}. \tag{5}
$$

All of camera poses are optimized by minimizing the reprojection errors with Levenberg-Marquard Method.

4. Registration by using MI

Local motion estimation method using 2D-3D correspondence and tracking can obtain good result. However, accumulation of error is inevitable because of the tracking error. To refine estimation results, 2D-3D registration of an entire region between panoramic image and point clouds is introduced.

Although we can extract color information of 3D point clouds using sensor calibration, 3D points to 2D frame projection error is inevitable because of low resolution or fps of omni-directional camera. Therefore, the color of 3D point extracted from 2D frame is not guaranteed. On the other hand, reflectance of laser surely indicates information of surface material of laser footprint. There are no error between scanned 3D point coordinates and reflectance. Therefore, it is valuable to utilize reflectance information for motion estimation. Kurazume, *et al.* conducts multi modal registration using reflectance and RGB image using edges [14].

Our motion estimation method using reflectance is based on multi-modality 2D-3D registration by maximizing mutual information (MI). Multi-modality registration method using MI is proposed by Viola *et al.* [27], and mainly developed in the field of medical science. Pandey, *et al.* conduct sensor calibration between camera and LiDAR using MI [23].

4.1. Motion deformation model

First, we describe the deformation model of the sensor motion line in the optimization phase with MI. Note that $W = (rx, ry, rz, x, y, z)$ indicates 6-DoF parameters. Initial position at time t calculated by method in Sec. 3 is $W(t)$. First, we choose two camera frames at t_1 and t_2 ($t_1 < t_2$). Motion of the section between t_1 and t_2 is then optimized. Let $W(t)$ as the original position at t and $W'(t)$ as the translated position at t. The translated motion $W(t)$ is determined by changing six parameters of final sensor position $W'(t_2)$. Let $P(t)$ as the vector of parameters $(rx, ry, rz)^T$ or $(x, y, z)^T$. The translated motion $P'(t)(t_1 < t < t_2)$ is calculated as:

$$P'(t) = \frac{|P(t) - P(t_1)|}{|P(t_2) - P(t_1)|}(P(t_2) - P'(t_2)) + P(t). \quad (6)$$

Motion of the entire segment between t_1 and t_2 is rescaled by final position $W'(t_2)$ (See Fig. 6). MI is calculated with the deformed motion $W'(t)$.

The entire scan period is divided to this short segments. Although high resolution motion refinement can be proceeded in the case of increasing the number of segments, there could be insufficient information to register with maximizing MI. We determine the number of segments (the
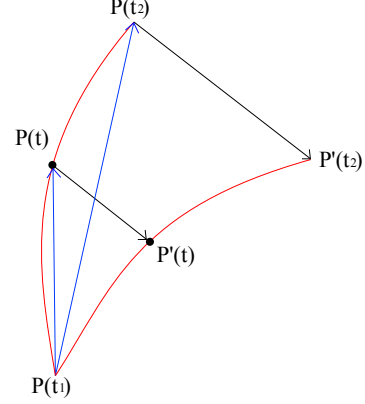

Figure 6. Deformation model of motion. $P(t)$ indicates (x, y, z) or (rx, ry, rz) in 6-DoF parameters. Motion entire the segment between t_1 and t_2 is rescaled by final position $P'(t_2)$

length of the one segment) experimentally by considering this trade-off.

4.2. Mutual information

To optimize motion, MI between gray scale images converted from 2D RGB images and reflectance is used. To compute MI, a technique based on the proposed method by Maes *et al.* [16] using joint histogram is used . The joint image reflectance histogram $\mathbf{H} = \{h_{rf}\}$ with reflectance r of 3D points and pixel value f of gray scale image is calculated by projecting 3D point cloud onto 2D image frame. MI is calculated with $\mathbf{H}$ as following,

$$I = \frac{1}{N} \sum_{f,r} h_{rf} \log \frac{N h_{rf}}{h_f h_r} \quad (7)$$

where $h_f = \sum_r h_{rf}, h_r = \sum_f h_{rf}$ are histogram value of h and r and $N = \sum_{r,f} h_{rf}$.

4.3. Joint histogram construction

Now the problem is how to construct the joint histogram. Let $\mathbf{x}_i^j = Proj(\mathbf{R}^T(\mathbf{X}_j^W - \mathbf{T}))$ be the 2D coordinates of projected 3D point j onto frame i. $q_{(i,j),0}$ is a pixel point where $0 < d_{j,x} < 1, 0 < d_{j,y} < 1$:$(d_{j,x}, d_{j,y})^T = x_i^j - q_{(i,j),0}$, and let $q_{(i,j),1} = (1, 0)^T + q_{(i,j),0}, q_{(i,j),2} = (0, 1)^T + q_{(i,j),0}, q_{(i,j),3} = (1, 1)^T + q_{(i,j),0}$.

The joint histogram $h_r f$ is defined as:

$$h_{rf} = \sum_j \sum_{m=0}^{3} w_{j,m} \cdot \delta(r - r_k, f - f_{k,m}), \quad (8)$$

where $w_{j,0} = (1 - d_{j,x})(1 - d_{j,y}), w_{j,1} = d_{j,x}(1 - d_{j,y}), w_{j,2} = (1 - d_{j,x})d_{j,y}, w_{j,3} = d_{j,x}d_{j,y}$, and δ is discrete unit pulse.

Figure 7. An example of area separation on a panoramic image.

4.4. Area separation

Constructing joint histogram with entire image has the possibility not to capture enough interrelation between reflectance and pixel value because of shade or including excessive kinds of material. To avoid this problem, several rectangle windows the height of which is same as the height of panoramic image are placed in one row on the panoramic image as Fig. 7 shows. In the case of our scanning system, the projected scan line traverses the panoramic image from left to right. Therefore the small windows are located to guarantee that sufficient 3D points for computing MI is projected onto each window. MI is computed in each window and the sum of computed MI values is used for optimization. Finally, the 6-DoF parameters are optimized by maximizing the sum of MI using Powell method.

5. Experimental results

In this paper, we address local motion estimation problem and refinement of estimated motion with multi modal registration and demonstrate validation of our method. We use Z+F imager 5010 [30] as the 3D laser scanner and Ladybug 3 [24] as the omnidirectional camera. The laser scanner works in profiler mode and field of view is $320°$ at one round scan resolution: $10000 points$. Head rotates at $50Hz$. The 2D image resolution from panoramic camera is captured in resolution 5400×2700.

The scanned 3D data and 2D image sequence are stored online in the internal memory of laser scanner and laptop PC connected to camera, respectively. The obtained data is rectified offline with a modern PC: Intel Core i7 @ 3.4G, 64GB memory. The motion estimation program is implemented in C++.

5.1. Motion Estimation

To evaluate the accuracy of our motion estimation method with triangulation, static scans in a same scene are used. An error for evaluation is defined as the distance between rectified 3D point and its closest point. Our method is compared to the method proposed by Zheng *et al.* [33].

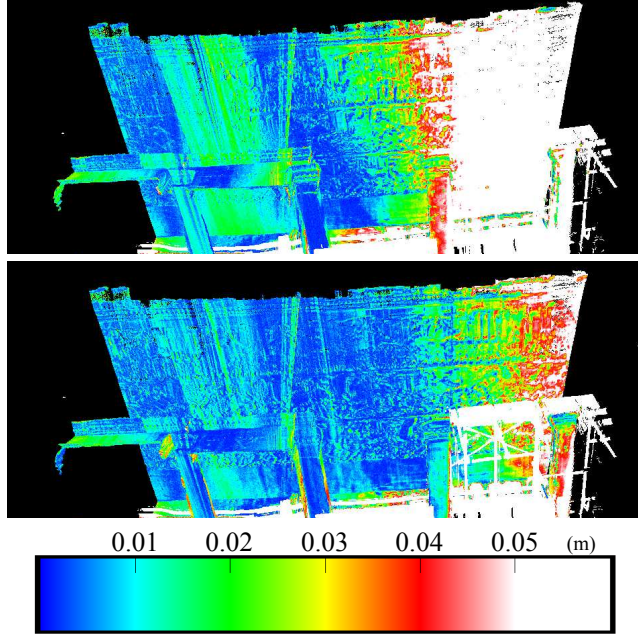Figure. 8 shows the estimated motion, Fig. 9 shows error

Figure 9. Error visualized images and a color bar. From the top: Error visualized image reconstructed by Zheng *et al.*'s method; Error visualized image reconstructed by our method; Color bar indicating error value. Note that scanning scene is the same location but at different time.
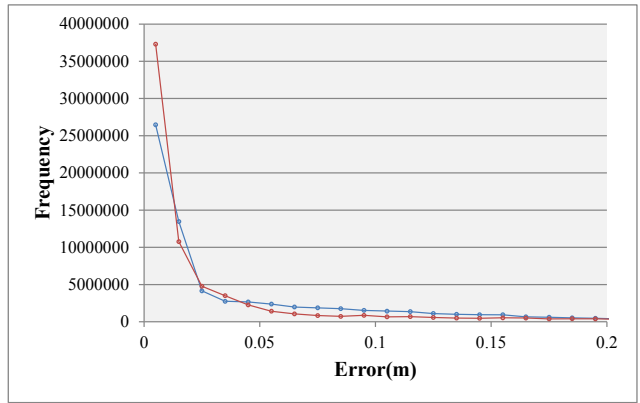

Figure 10. Frequency distribution of error. (blue) Zheng's method [33]. (red) Our method.

visualized image and the frequency distribution of error in each method is shown in Fig. 10. From Fig. 9 and Fig. 10, our results shows larger number of small error points than Zheng's method (small error points are indicated as blue points in Fig. 9). Top image of Fig. 8 shows estimated trajectory of sensor on xy-plane by each method. The blue line in the top image of Fig. 8 is biased in the $+y$ direction. It is likely that the bias is caused by the 2D-3D mismatches at the right side of the sensors (and front side in Fig. 9) due to the complex structure, whereas there is only relief in the left
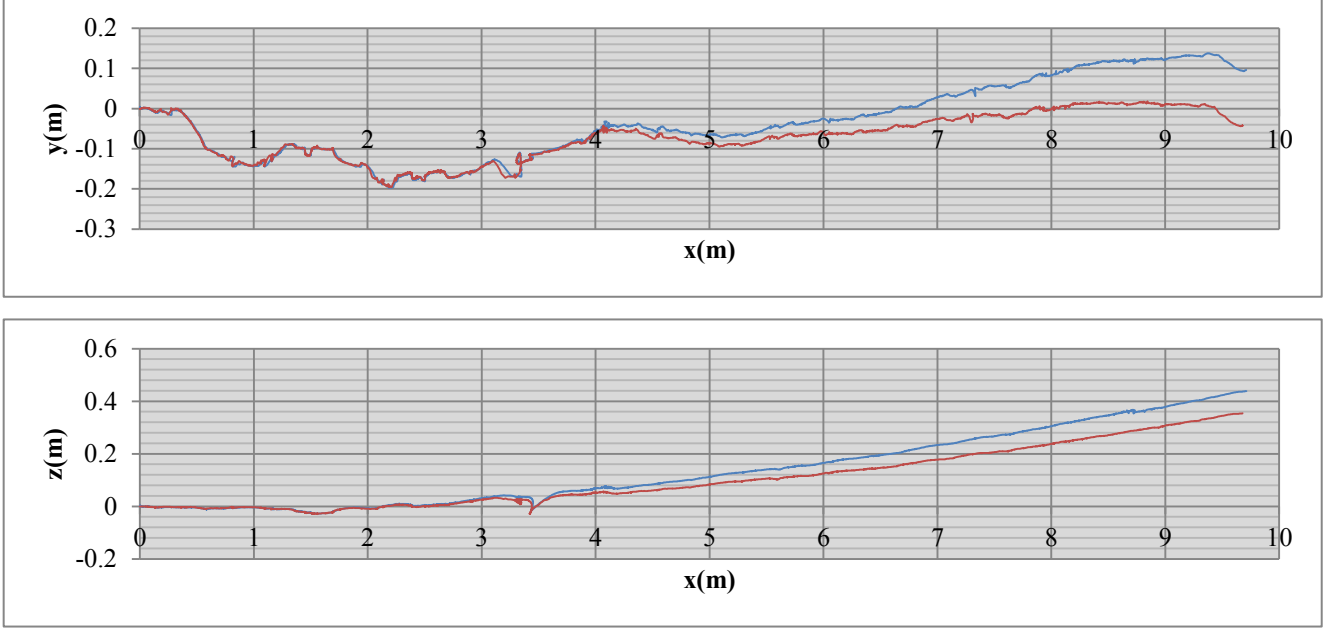
Figure 8. Estimated motion. (top) trajectory on xy-plane, (bottom) trajectory on xz-plane, blue line is sensor trajectory estimated by Zheng's method [33] and red line is estimated by our method.

Figure 12. Occluded 3D points. Several 3D points scanned by laser scanner are not observed by omni-directional camera due to occlusion by the pillar

side. Our method obtain better accuracy results compared to Zheng's method based on these factors without bias on the left side due to the mismatches.

Figure. 11 shows two other reconstructed results in reflectance. Our method can reconstruct 3D model even in the case of curving course. In the right scene in Fig. 11, 2D-3D mismatches is made due to occlusion by the pillar as shown in Fig. 12. Our method can estimate motion without the effect of the mismatches by the outlier rejection using depth based error metric.

5.2. MI registration

To compute MI, we construct joint histogram with 3D points the normalized reflectance of which is 0.05 to 0.55 and gray scale pixels the value of which is 0.1 to 0.7 (Upper and lower limit of gray scale value are 0.0 and 1.0). The numbers of classes in the joint histogram are 30, respectively. Therefore class intervals of normalized reflectance and gray scale value are 1.66×10^{-2} and 2.0×10^{-2}, respectively. We define length of one segment is approximately $14.5 \; sec$. These values are experimentally determined.

Figure. 13 shows the images upon which the 3D points are projected. These pixels are colored by the reflectance value of the 3D points and the projected area is represented as stripes to easily compare the edge of the reflectance image and the 2D image. Figure. 13 shows that the motion is optimized as the edges of the reflectance image converges with the edges of 2D image with good consistency, especially the areas surrounded by red windows.

Although our registration method achieve good consistency on 2D images, this optimization has small effect on reconstructed 3D images. This could be attributed to the error of 2D-3D camera calibration or low resolution motion deformation model.

6. Conclusion

This paper presented 3D reconstruction method in sensor fusion system. Our method is designed for profiler 3D laser scanner and camera fusion system, selects good projection
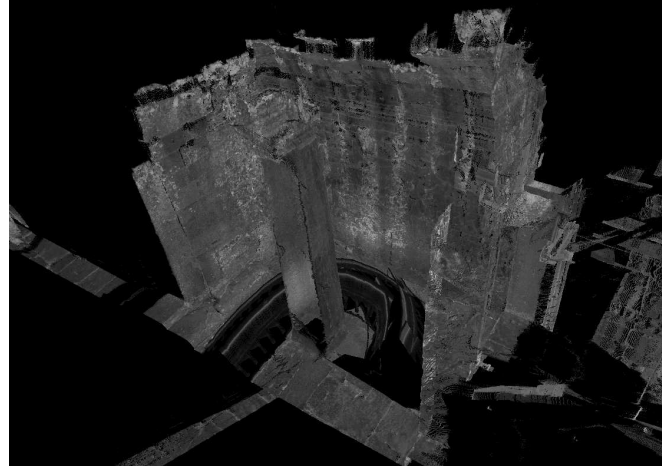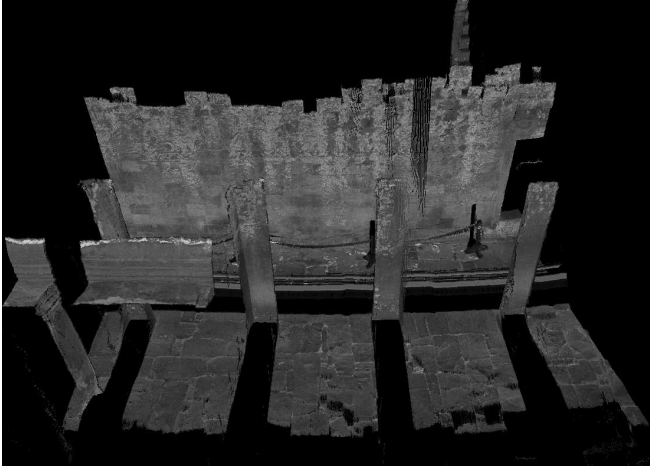
Figure 11. Other results. Reconstructed results are shown as reflectance image.
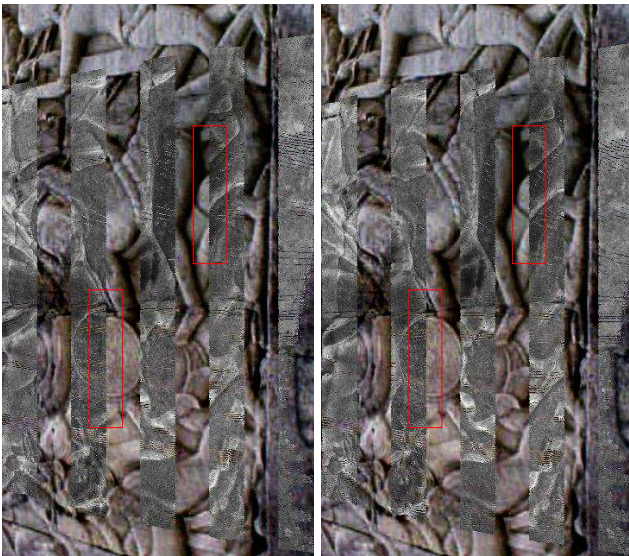

Figure 13. Images that pixels on where 3D points are projected are colored by reflectance. (left) before registration, (right) after registration. Areas surrounded red windows show that reflectance image are well registered by our method.

points for tracking to estimate motion stably and considers the effects of mismatches from the difference of sensor positions between laser scanner and camera. Our scanning system can obtain 3D model of a bas-relief with good fidelity and high density. Although global pose refinement by maximizing MI can be improved, we demonstrate that 3D reflectance image can be registered to 2D RGB image using motion optimization.

Our method will help other laser scanner and camera fusion systems (even those using cheap 1-axis LiDAR). Obtained high density and accuracy 3D data is valuable for preservation, restoration and analysis of cultural heritages. Also it is worthwhile and challenging to recover global optimization using multi modal information for refinement of motion estimation.

Acknowledgment

References

[1] A. Banno, T. Masuda, T. Oishi, and K. Ikeuchi. Flying Laser Range Sensor for Large-Scale Site-Modeling and Its Applications in Bayon Digital Archival Project. *International Journal of Computer Vision*, 78(2-3):207–222, 2008. 2

[2] Y. Bok, Y. Jeong, D. Choi, and I. Kweon. Capturing Village-level Heritages with a Hand-held Camera-Laser Fusion Sensor. *International Journal of Computer Vision*, 94:36–53, 2011. 2, 3

[3] M. Bosse, R. Zlot, and P. Flick. Zebedee: Design of a Spring-Mounted 3-D Range Sensor with Application to Mobile Mapping. *IEEE Transactions on Robotics*, 28:1104–1119, 2012. 2, 3

[4] N. Engelhard, F. Endres, J. Hess, J. Sturm, and W. Burgard. Real-time 3D visual SLAM with a hand-held RGB-D camera. In *RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum*, 2011. 2

[5] M. A. Fischler and R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, 1981. 4

[6] C. Forster, M. Pizzoli, and D. Scaramuzza. SVO: Fast semi-direct monocular visual odometry. In *International Conference on Robotics and Automation*, 2014. 2

[7] Y. Furukawa and J. Ponce. Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010. 2

[8] R. M. Haralick, C.-n. Lee, K. Ottenburg, and M. Nölle. Analysis and solutions of the three point perspective pose estimation problem. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, pages 592–598. IEEE, 1991. 4

[9] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. *International Journal of Robotics Research*, 31(5):647–663, 2012. 2

[10] A. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy. Visual Odometry and Mapping for Autonomous Flight Using an RGB-D Camera. In *International Symposium on Robotics Research*, 2011. 2

[11] C. Kerl, J. Sturm, and D. Cremers. Robust odometry estimation for RGB-D cameras. In *International Conference on Robotics and Automation*, 2013. 2

[12] G. Klein and D. Murray. Parallel Tracking and Mapping for Small AR Workspaces. In *International Symposium on Mixed and Augmented Reality*, 2015. 2

[13] B. Klingner, D. Martin, and J. Roseborough. Street View Motion-from-Structure-from-Motion. In *International Conference on Computer Vision*, 2013. 2

[14] R. Kurazume, K. Nishino, Z. Zhang, and K. Ikeuchi. Simultaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute. In *Asian Conference on Computer Vision*, pages 99–106. Citeseer, 2002. 5

[15] B. D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981. 3

[16] F. Maes, D. Vandermeulen, and P. Suetens. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4):373 – 386, 1999. 5

[17] M. Maimone, Y. Cheng, and L. Matthies. Two years of Visual Odometry on the Mars Exploration Rovers. *Journal of Field Robotics*, 24(2):169–186, 2007. 2

[18] N. Matsuda, O. Cossairt, and M. Gupta. MC3D: Motion Contrast 3D Scanning. In *International Conference on Computational Photography*, 2015. 2

[19] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, A. D. D. Kim, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *International Symposium on Mixed and Augmented Reality*, 2011. 2

[20] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison. DTAM: Dense tracking and mapping in real-time. In *International Conference on Conputer Vision*, 2011. 2

[21] D. Nister, O. Naroditsky, and J. Bergen. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23(1):3–20, 2006. 2

[22] G. Pandey, J. R. McBride, and R. M. Eustice. Ford Campus vision and lidar data set. *International Journal of Robotics Research*, 30(13):1543–1552, 2011. 2, 3

[23] G. Pandey, J. R. McBride, S. Savarese, and R. Eustice. Automatic Targetless Extrinsic Calibration of a 3D Lidar and Camera by Maximizing Mutual Information. In *The Association for Advancement of Artificial Intelligence*, 2012. 5

[24] Point Gray Research. http://www.ptgrey.com/. 6

[25] S. Scherer, J. Rehder, S. Achar, H. Cover, A. Chambers, S. Nuske, and S. Singh. River mapping from a flying robot: state estimation, river detection, and obstacle mapping. *Autonomous Robots*, 32(5):189–214, 2012. 2, 3

[26] C. H. Tong, S. Anderson, H. Dong, and T. Barfoot. Pose Interpolation for Laser-based Visual Odometry. *Journal of Field Robotics*, 31(5):731–757, 2014. 1, 3

[27] P. Viola and W. Wells III. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*, 24(2):137–154, 1997. 5

[28] T. Whelan, H. Johannsson, M. Kaess, J. Leonard, and J. McDonald. Robust real-time visual odometry for dense RGB-D mapping. In *International Conference on Robotics and Automation*, 2013. 2

[29] J. Xiao, A. Owens, and A. Torralba. SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels. In *International Conference on Conputer Vision*, 2013. 2

[30] Zoller + Fröhlich GmbH. http://www.zf-laser.com/. 6

[31] J. Zhang and S. Singh. LOAM: Lidar Odometry and Mapping in Real-time. In *Robotics: Science and Systems Conference*, pages 109–111, 2014. 1, 3

[32] J. Zhang and S. Singh. Visual-lidar Odometry and Mapping: Low-drift, Robust, and Fast. In *International Conference on Robotics and Automation*, 2014. 2, 3

[33] B. Zheng, T. Oishi, and K. Ikeuchi. Rail Sensor: A Mobile Lidar System for 3D Archiving the Bas-reliefs in Angkor Wat. *IPSJ Tran. on Computer Vision and Applications*, 7:59–63, 2015. 2, 3, 4, 6, 7