IPSJ Transactions on Computer Vision and Applications Vol.7 59–63 (July 2015)

[DOI: 10.2197/ipsjtcva.7.59]

Express Paper

Rail Sensor: A Mobile Lidar System for 3D Archiving the Bas-reliefs in Angkor Wat

Bo Zheng[1,a)] Takeshi Oishi[1,b)] Katsushi Ikeuchi[1,c)]

Received: March 13, 2015, Accepted: April 20, 2015, Released: July 27, 2015

Abstract: This paper presents a mobile Lidar system for efficiently and accurately capturing the 3D shape of the Bas-reliefs in Angkor Wat. The sensor system consists of two main components: 1) a panoramic camera and 2) a 2D 360-degree laser line scanner, which moves slowly on the rails parallel to the reliefs. In this paper, we first propose a new but simple method to accurately calibrate the panoramic camera to the 2D laser scan lines. Then the sensor motion can be estimated from the sensor-fused system using the 2D/3D features tracking method. Furthermore, to reduce the drifting error of sensor motion we adopt bundle adjustment to globally optimize and smooth the moving trajectories. In experiments, we demonstrate that our moving Lidar system achieves substantially better performance for accuracy and efficiency in comparison to the traditional stop-and-go methods.

Keywords: 3D reconstruction, mobile Lidar, laser scan, heritage preservation

1. Introduction

Angkor Wat, as one of the largest religious monuments in the world, was built in the early 12th century and located in north of the modern town, Siem Reap, Cambodia. Angkor Wat owns the huge-scale galleries of bas-reliefs in the long corridors which record various scenarios of the mythologies of Hindu, as an example shown in **Fig. 1** (a). These bas-reliefs never fail to fascinate the tourists, but unfortunately they face the problem of deterioration due to natural or man-made breaks. Figure 1 (b) shows the examples of the damage caused by weathering erosion (left) and bullets in civil war (right). 3D digitally preserving them becomes to be very urgent. Also the captured 3D data may play an important role in the areas such as restoration, archaeological/historic analysis and VR tours.

Nowadays, the technique of acquiring the accurate geometric information from the reliefs using 3D laser scanners has been developed. For example, the relief range images can be scanned from different views in a stop-and-go manner and then aligned into same coordinate to make the entire 3D model, such as the method proposed in Ref. [6]. However, the stop-and-go manner is time-consuming for scanning such a long relief gallery (more than 500 meters). Supposing 1 meter can be captured in a day using Vivid9 scanner with 200 shots, the whole scanning project might take more than one year.

To this end, we develop a mobile scanning system called rail sensor (see Fig. 1 (b)) for fast and accurately capturing the 3D range data from the bas-relief carvings. Different from traditional laser scanners, our system only consists of two hardware devices: 1) a laser-line-scan Lidar and 2) a panoramic camera, which are mounted onto a platform moving on rails. While laser

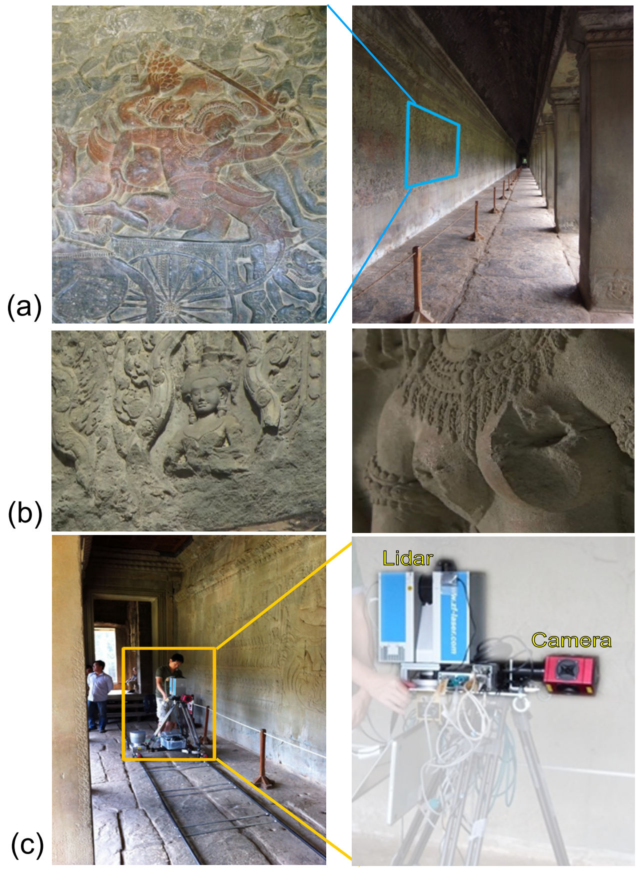


Fig. 1 Proposed sensor system for capturing Bas-reliefs in Angkor Wat. (a) An example of Bas-relief (left) located in the corridor (right) of the first gallery in Angkor Wat. (b) Damages caused by weathering erosion (left) and bullets in civil war (right). (c) Our sensor system moving parallel to the relief on a rail (left) and its close-up (right).

1 Institute of Industrial Science, The University of Tokyo, Meguro, Tokyo 153–8505, Japan
a) zheng@cvl.iis.u-tokyo.ac.jp
b) oishi@cvl.iis.u-tokyo.ac.jp
c) ki@cvl.iis.u-tokyo.ac.jp

scanner works in profiling mode capturing 2D line structures, the panoramic camera captures the panoramic videos used for motion estimation and thus reconstructing 2D lines to 3D shapes.

Our method includes the following contributions: 1) a new but simple method is proposed to accurately calibrate the panoramic camera and the laser scanner; 2) the sensor motion can be estimated robustly from the sensor-fused 2D/3D features tracking method, which helps to reconstruct the structures from the 2D scan lines accurately; 3) furthermore, to reduce the drifting error in sensor motion we adopt bundle adjustment to globally optimize and smooth the moving trajectories. In experiments, we demonstrate that our moving Lidar system achieves substantially better performance for accuracy and efficiency in comparison to traditional stop-and-go methods.

2. Related Work

The study of high-accuracy 3D geometry reconstruction can be traced back to related studies in following three streams:

Range-sensor-only system The stop-and-go scanning manner is suitable for the static laser scanner, e.g., Konica Minolta Vivid 9 adopted in Ref. [6]. However, it is too time-costuming for scanning large-scale reliefs to match our purpose. The high speed structure-light based sensor, such as Ref. [12], works well for motion tracking but has limitation on short scanning range and the weakness of outdoor scanning. One of most related works to ours is a laser-line based moving platform proposed by Ono, S. et al. [9], but however it has to assume the sensor movement in a straight line, which might not reasonable for the long-distance scanning.

Camera-only system One of the most important works is the multi-stereo based 3D reconstruction using only RBG cameras. Furukawa and Ponce [5] proposed an accurate reconstruction that generates and propagates a semi-dense set of patches. This method has shown impressive results but the precision and resolution is still far from the level of the laser scanners and the absolute scale of the data is not easy to obtain. Another notable work is proposed in Ref. [7] which adopted panoramic image sensors for street view scanning but it focuses the problem on city-scale 3D map reconstruction.

Range sensor+Camera system The recent work, such as Refs. [11], [13], using Kinect sensor or combing lidar and RGB camera for indoor/outdoor 3D reconstruction which enhance the robustness for sensor motion estimation, but however these techniques cannot reach the precision required for relief scanning. The most related work to our method are the studies that combine the range sensor and camera (e.g., Refs. [1], [4]) for enhancing the mobile laser scan. Different from these methods, our goal is to build a new sensor fusion system in which a panoramic camera and a high-accuracy lidar are adopted. The depth error of the lidar is around 100 μm, and it can capture 0.5 M points per second. These good properties benefit the scanning accuracy and resolution. In addition, the sensor fusion techniques based on combining third-party positioning sensors such IMU or GPS seem still leaving the gap to reach the precision required.

3. Method

3.1 Sensor System

In our sensor system, we adopted a laser scanner: Z+F Imager 5010c and a panoramic camera: pointgrey Ladybug 3 which are fixed onto a board as shown in Fig. 1 and synchronized by pulse-per-second triggers. Although Z+F Imager 5010c usually works in panoramic scanning mode (the head rotates in 360 degrees), in our system we set it in profiling mode (the head is fixed without rotation). Therefore, as shown in **Fig. 2**, the laser scanner always sends a 2D scan plane orthogonal to the moving directions. The 2D scan lines will be aligned to the world coordinate system after the sensor motion is estimated.

3.2 Calibrating Laser Line to Panoramic Camera

To calibrate the 2D laser scan line to the panoramic camera, there are three local coordinate systems needed to be clarified (as shown in **Fig. 3** (a)): O_l, O_s and O_c corresponding to laser scan line, lidar sensor and camera respectively.

Since the lidar sensor working in profiling mode, O_l looks like a 2D coordinate system. That is, it profiles only in y_l and z_l directions, x_l coordinate is always "0" plus a very small value fed back from the intrinsic calibration of the sensor. O_s is defined in the panoramic scanning mode which starts at the 0 degrees of the horizontal rotation motor ends at 180 degrees. O_c is defined according to the panoramic camera orientation.

Suppose a static scene can be captured by 1) a line laser scan, 2) a panoramic laser scan and 3) panoramic camera, the calibration can be viewed as how to find the relative pose and position between these scene data.

Calibration model. Let x_a denote a point x in coordinate system O_a and let ${}_b\top_a$ denote a transform from O_a to O_b. Thus, $x_b = {}_b\top_a \cdot x_a$. We may write the calibration model as a linear transform ${}_c\top_l$ from the point on scan line "l" to the panoramic image "c":

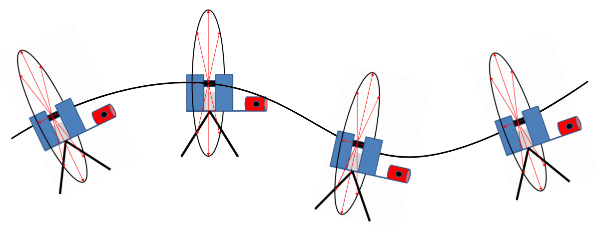


Fig. 2 The proposed moving platform: a lidar sensor (blue) combined with a panoramic camera (red) moves along the trajectory which is alway orthogonal to the planes of scan lines.

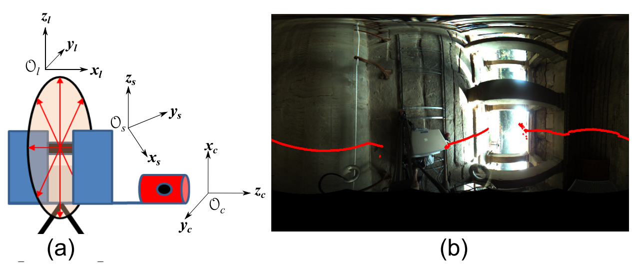


Fig. 3 Sensor calibration. (a) Three local coordinate systems correspond to each senor. (b) After calibration, the scan line (red) is projected onto the panoramic image at the corresponding location.

$$x_c = {}_c\top_l \cdot x_l, \tag{1}$$

where the scan line-to-image transform can be decomposed into spherical projection model and z-axis rotation model:

$${}_c\top_l = {}_c\top_s \cdot {}_s\top_l. \tag{2}$$

The transformation between O_l and O_s, ${}_s\top_l$, is 1 DOF. That is, suppose panoramic scan consists of the scan lines at different degrees around z-axis, the profiling line scan should belong to one of the scan lines in panoramic scan mode at a certain degree. ${}_s\top_l$ can be viewed as z-axis rotation matrix, and the degree can be calculated out by minimizing the 2D ICP problem over all degrees.

The transformation between O_s and O_c, ${}_c\top_s$, is 6 DOF including the rotation and translation. To work out ${}_c\top_s$, we adopt the method proposed in Ref. [2], which first manually selects correspondences from both panoramic image and panoramic laser scan, and then minimize the spherical project error (project 3D points onto a sphere using spherical coordinates) over all 6 parameters using the non-linear optimization such as Levenberg-Marquardt algorithm. Figure 3 (b) shows an example that, after calibration described above, the laser line can be correctly reprojected onto the panoramic image.

3.3 Sensor Motion Estimation

Given the calibrated line scanner and panoramic camera, we propose to build a motion estimation algorithm by using the fused depth and RGB information. However this sensor-fusion system does not like the RGBD camera system such as Kinect Fusion [10], which captures low-resolution depth image in real time and thus is able to directly apply the ICP algorithm [3] for 3D reconstruction. In our case, since only one or few scan lines can be captured at one frame according to shutter speed, the depth information of a line is not enough for estimating the sensor motion.

Triangulation. Fortunately, we can design the triangulation algorithm by fusing the 2D-3D correspondences captured from the sensors. Beyond the traditional SfM method that 3D points are calculated from 2D features, our method can directly obtain the 3D correspondences due to the sensors already well calibrated.

To robustly estimate the relative pose and position between two consecutive panoramic frames, our method consists of four main steps:

- The 2D correspondence (track of image points) along the image sequence is required to geometrically relate consecutive frames using KLT tracker [8] and robust feature descriptor [14]. As shown in **Fig. 4**, the 2D points x^{2D}_{nk} and $x^{2D}_{n+1,k}$ in the n-th and $n+1$-th frames are correspondences.
- Having the relative pose between laser and camera of the calibration, as shown in Fig. 4, the 3D-2D correspondences (in red points) can be extracted by projecting laser line (in black line) on to the frame and searching the nearest tracked feature points to the scan line on the image, e.g., the 2D point x^{2D}_{nk} and 3D point $x^{3D}_{n,k}$ are correspondences.
- Then the 2D-3D correspondences to the next frame can be figured out, i.e., $x^{3D}_{n,k}$ and $x^{2D}_{n+1,k}$ are correspondences.
- The relative relation between the two frames is worked out by minimizing the projection errors between the 2D points projected from the 3D points in previous frame and the 2D tracking points in current frame:

$$\{R_n, T_n\} = \arg\min_{\{R,T\}} \sum_{k}^{\#point} \| \Pr_{n+1}([R|T]x^{3D}_{nk}) - x^{2D}_{n+1,k} \|_2,$$

where function $\Pr_{n+1}(\cdot)$ denotes spherical projection. We apply the Levenberg-Marquardt method for solving the non-linear optimization problem.

Bundle adjustment. In this paper, we use the bundle adjustment (BA) technique similarly adopted in the SfM method which globally optimizes the sensor motion and thus improves the drift-

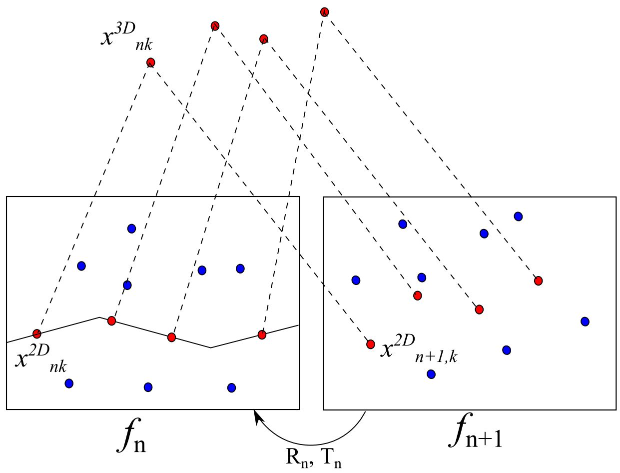
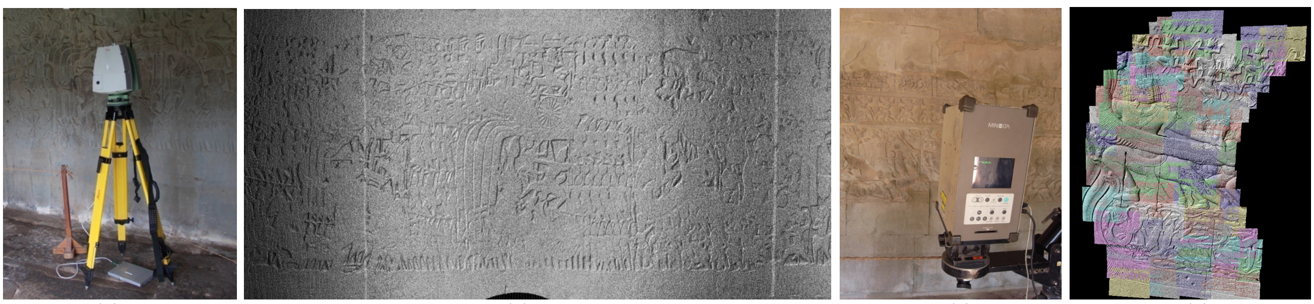


Fig. 4 Triangulation between 2 two continuous frames f_n and f_{n+1}: the points on each frame denote the 2D features obtained by KLT tracker. Red points in f_n denote the features overlapped with scan line (black) and thus the corresponding 3D points can be found out. In f_{n+1}, since the red points can be tracked from f_n, it is easy to calculate rotation and translation by minimizing the projection error between the 3D points and 2D points in f_{n+1}.



(a) (b) (c) (d)

Fig. 5 Qualitative comparison to different type of range scanners: Leica C10 scan station in (a) and Vivid 9 in (c), and their scan results are shown in (b) and (d) respectively.

ing error caused by the accumulation of local triangulation errors. Therefore, given the initial guess, the result of local triangulation, the BA can be formulated as

$$\{R_n, T_n\} = \arg\min_{\{R_i,T_i\}} \sum_{k}^{\#point} \sum_{i}^{\#frame} \sum_{j}^{\#frame} \sigma \parallel \mathrm{Pr}_j([R_i|T_i]x_{ik}^{3D}) - x_{jk}^{2D} \parallel_2,$$

where $\sigma = \begin{cases} 1, \text{if } x_{kj}^{2D} \text{is scanned by } x_{kj}^{3D} \\ 0, \text{otherwise} \end{cases}$. In practice, there is no need to carry out BA through the entire image sequence, according to the move speed and frame rate we do BA for each 16 frames.

4. Experimental Results

Since there seems no very similar sensing system proposed for relief scanning, we first qualitatively compare our result to the traditional sensor scanning in the stop-and-go manner for the bas-relief reconstruction. We adopt two different range scanners: 1) Leica C10 station and 2) Vivid 9 which are at same precision level (around 100 μm) to Z+F imager 5010C adopted in our system. As the scan result shown in **Fig. 5** (b), we can see that, for the omni-directional scanner Leica C10, it is difficult to get a uniform result in density. For Vivid 9, there are 96 scans shown in different colors in Fig. 5 (d) and thus this method is very deficient, not only because of the long time consumed for scan but also a lot of manual work necessary for aligning the scans.

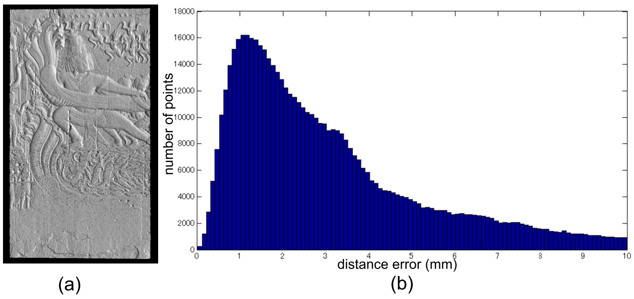


Fig. 6 Distance error comparison. (a) Result obtained by our method. (b) Distance error histogram comparing our result in (a) to Vivid9's result in Fig. 5 (d).

Second we quantitatively compare our result with Vivid9 scanner in accuracy. As shown in **Fig. 6**, we first align the 3D models of our result in (a) to Vivid9's result in Fig. 5. Then we calculate the nearest-point distance between the aligned two models, and plot the point number-vs-distance histogram shown in (b). The result denotes that most points are distributed with the distance error from 1 mm to 5 mm.

Finally, the scan example results of the Bas-relief in Angkor Wat are shown in **Fig. 7**. There are many details that can be preserved in the data.

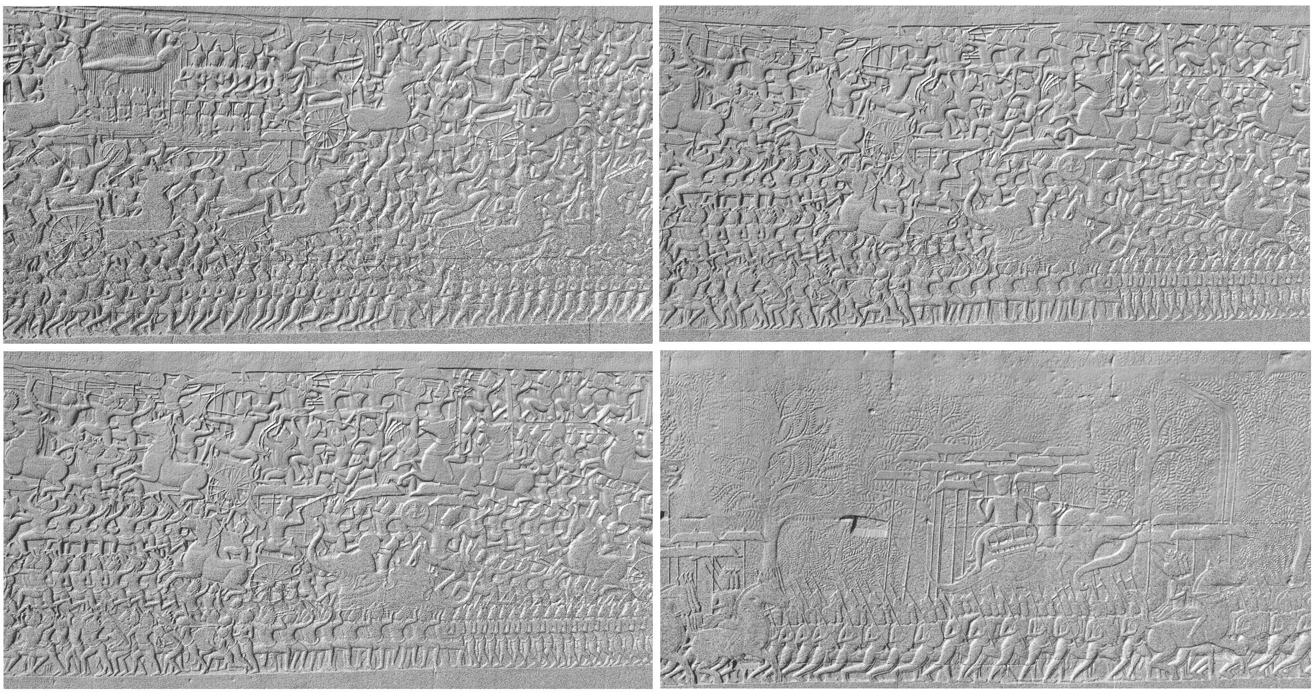


Fig. 7 Scan examples of the Bas-relief in Angkor Wat.

5. Conclusion and Future Work

This paper presents an efficient method for scanning the large-scale bas-reliefs in Angkor Wat. Without losing much accuracy, (compared to Vivid9 sensor, the error should be within 5 mm), our sensor fast scans the bas-reliefs and finished the whole scanning task of Bas-relief in the first gallery in 10 days. This work should not be limited in relief scanning but also has potential for capturing various large-scale and long targets, such as a street, in future.

Acknowledgments We thank to Dr. XiangQi Huang, Dr. Min Lu, Dr. ZhiPeng Wang, Dr. Masataka Kagesawa, Dr. Yasuhide Okamoto, Dr. Shintaro Ono, and Ryoichi Ishikawa in Computer Vision Lab, the University of Tokyo, for various helps on data capture, data processing, sensor setting up and many valuable suggestions for the sensor development. This work was partly supported by JSPS KAKENHI Grant Number 25257303.

References

[1] Banno, A., Masuda, T., Oishi, T. and Ikeuchi, K.: Flying Laser Range Sensor for Large-Scale Site-Modeling and Its Applications in Bayon

Digital Archival Project, *IJCV*, Vol.78, No.2-3, pp.207–222 (2008).

[2] Banno, A. and Ikeuchi, K.: Omnidirectional texturing based on robust 3D registration through Euclidean reconstruction from two spherical images, *CVIU*, Vol.114, No.4, pp.491–499 (2010).

[3] Besl, P.J. and McKay, N.D.: A Method for Registration of 3-D Shapes, *TPAMI*, Vol.14, No.2, pp.239–256 (1992).

[4] Bok, Y., Jeong, Y., Choi, D.-G. and Kweon, I.: Capturing Village-level Heritages with a Hand-held Camera-Laser Fusion Sensor, *IJCV*, Vol.94, No.1, pp.36–53 (2011).

[5] Furukawa, Y. and Ponce, J.: Accurate, Dense, and Robust Multiview Stereopsis, *TPAMI*, Vol.32, No.8, pp.1362–1376 (2010).

[6] Ikeuchi, K. and Miyazaki, D.: *Digitally Archiving Cultural Objects*, Springer-Verlag (2007).

[7] Klingner, B., Martin, D. and Roseborough, J.: Street View Motion-from-Structure-from-Motion, *ICCV*, pp.953–960 (2013).

[8] Lucas, B.D. and Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision, *IJCAI1981*, pp.674–679 (1981).

[9] Matsui, K., Ono, S. and Ikeuchi, K.: The Climbing Sensor: 3-D Modeling of a Narrow and Vertically Stalky Space by Using Spatio-temporal Range Image, *IROS*, pp.3997–4002 (2005).

[10] Newcombe, R., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A., Kohli, P., Shotton, J., Hodges, S. and Fitzgibbon, A.: KinectFusion: Real-Time Dense Surface Mapping and Tracking, *ISMAR*, pp.127–136 (2011).

[11] Pandey, G., McBride, J.R. and Eustice, R.M.: Ford Campus vision and lidar data set, *IJRR*, Vol.30, No.13, pp.1543–1552 (2011).

[12] Sagawa, R., Sakashita, K., Kasuya, N., Kawasaki, H., Furukawa, R. and Yagi, Y.: Grid-Based Active Stereo with Single-Colored Wave Pattern for Dense One-shot 3D Scan, *3DIMPVT*, pp.49–56 (2012).

[13] Xiao, J., Owens, A. and Torralba, A.: SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels, *ICCV*, pp.1625–1632 (2013).

[14] Zheng, B., Sun, Y., Jun, T. and Ikeuchi, K.: A Feature Descriptor by Difference of Polynomials, *IPSJ Tran. on Computer Vision and Applications*, Vol.5, pp.80–84 (2013).

(Communicated by *Akihiko Torii*)

