SEMI-AUTOMATIC CORRELATION FOR INTEGRATING DATA PRODUCED BY LASER SCANNING AND SFM

Kazuki Fujisato, Hiroshige Dan, and Yoshihiro Yasumuro
Faculty of Environmental and Urban Engineering, Kansai University, Japan

ABSTRACT: *The applicability of three-dimensional (3D) laser scanners, which are capable of capturing the surface shapes of objects as "point cloud" sets, has recently been expanded to include examining, re-designing, and preserving existing constructions, as well as collecting on-site information for building information modeling (BIM). However, one of the difficulties involved when collecting complete scans of outdoor constructions is avoiding occlusions. This means, in order to cover the entire surface of a construction, it is normally necessary to scan it from multiple viewpoints. On the other hand, structure from motion (SFM) is a powerful image-based modeling technique that can be used to recover camera parameters, pose estimates, and create sparse 3D scene geometry from image sequences. Utilizing the mobility of unmanned aerial vehicles (UAVs) equipped with high-resolution cameras, it is possible to compensate for unscanned regions in an outdoor target site and combine the obtained information with multi-view stereo (MVS) process data in order to produce dense surface meshes from the SFM output. In this research, we propose a method for correlating laser scanner point cloud data and SFM data in order to integrate them. Herein, we employ SFM data obtained using UAV-mounted digital camera imagery, as well as an images taken by the laser scanner. Usually, laser scanners are equipped and calibrated with color digital cameras in order to capture color information corresponding to scanned points. This feature allows the scanner viewpoint to be integrated into the 3D geometry reconstructed by SFM. More specifically, the 3D geometry produced by the SFM and the laser-scanned data from scanner position viewpoints are overlaid upon each other, after which, selecting the proper number of points from the scanner view via random sample consensus (RANSAC) permits the proper corresponding points to be found and fit in place automatically. In this study, we conducted an experiment using a UAV (DJI Phantom II) and a laser scanner (Riegl LMS-Z420i) equipped with a high-resolution camera, and found that the resultant 3D data set consisting of complementarily scanner point cloud data and SFM process data covered the entire surface shape of the scene.*

KEYWORDS: *Laser scanner, Structure from motion, Point cloud, 3D integration*

1. INTRODUCTION

Laser range scanners are three-dimensional (3D) surface imaging systems that can be used to produce consistent and accurate assessments of the vast spatial conditions required by various kinds of construction applications. These include investigations into construction process management (Shih et al. 2004, 2006), monitoring as-built infrastructures (Miller et al. 2008), and so forth. Since most of these applications require timely spatial

information delivery, collecting the needed information within a limited amount of time is critical for numerous field applications.

Furthermore, as developments of both long-range and vehicle-mounted laser scanners advance, such precise active 3D recording has also come to be effectively used in post-disaster reconstruction work (Watson et al. 2011). On the other hand, due to adverse field conditions at disaster sites, physical access and scanner setup is often limited and difficult. In such cases, scanning the entire shape of a scene is often labor intensive, or may even be impossible because multiple scanning from different viewpoints is often necessary to complement the limited or line-of-sight visibility available from each viewpoint. In such cases, using an unmanned aerial vehicle (UAV), often called a drone, equipped with a digital camera provides an effective alternate strategy for meeting these challenges, particularly since drone flight performance levels and digital cameras image resolution levels have advanced remarkably in recent years.

Additionally, in the computer vision field, 3D reconstruction techniques that use enhanced structure from motion (SFM) techniques can now be easily carried out in combination with those downsized UAVs with small-sized camera. Recently, SFM based on large unorganized photo collections has been successfully used to facilitate 3D reconstruction work. For example, VisualSFM (Wu 2012) achieves $O(n)$ computation cost in practice for n images, while the most commonly known cost for incremental SFM was $O(n^4)$ (Snavely 2006). Since the use of very large photo collections is now acceptable even in chaotic post-disaster situations, small UAVs can examine damaged areas from various distances and viewpoints and gather useful images and/or video footage, not only for reviewing the recorded imagery but also for reconstructing the pre-disaster 3D geometry of the site.

Such 3D information can then be used to facilitate emergency disaster control measures and disaster recovery efforts. However, unlike laser scanner data, SFM data are based on a passive type of 3D imaging measurement

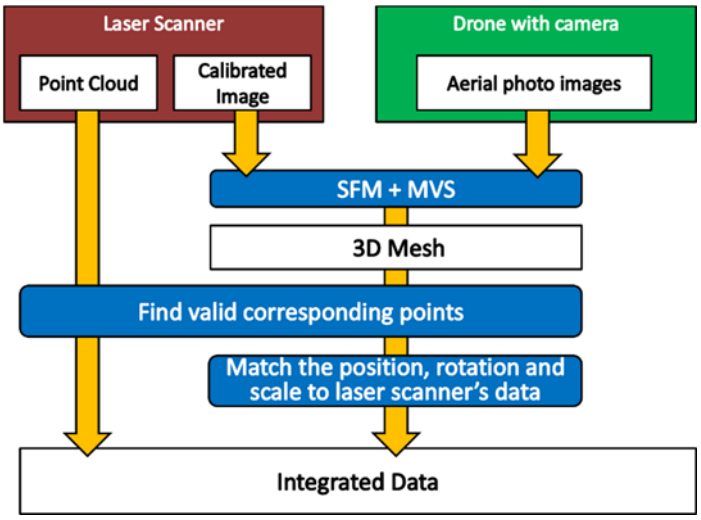

Fig. 1: Diagram depicting the proposed method. A Photo taken by the laser scanner's calibrated camera is combined with photos taken by a UAV for SFM. The generated 3D mesh contains the laser scanner camera viewpoint, which allows valid corresponding points between two data sets to be found easily.

that does not guarantee the absolute scale in nature, and instead depends on the distribution of the natural feature points in the appearance of a scene, which cannot be controlled. As a result, the reconstruction sometimes contains only sparse data points and poor natural features may produce unexpected artifacts, distortions, and noise.

2. METHOD

Conventional methods for registering multiple point cloud data sets are commonly based in the iterative closest point (ICP) algorithm (Besl et al. 1992) and are available for 3D data and designing post-processing in numerous software applications. This algorithm performs difference minimization between point cloud pairs comprising the reference target and the source to be transformed in order to best match the reference. Essentially, the algorithm solves unconstrained optimization problems whose results are sensitive to the initial state, thereby facilitating the iteration start needed to find correspondences between point sets and transform the source points until the best fit is obtained for all points.

When applying simple ICP-based registration to our problem, the initial state is determined by selecting points or mesh vertices as initial correspondences, or by placing the source data set close to the reference data set. This setup is basically manual work, and must take into consideration the acquisition process and characteristic differences between the 3D data sets. More specifically, one is from a laser scanner whose points are aligned on the scan lines, while the other is from an SFM whose points are totally unorganized and may contain huge distortions, noise, and sparse regions. Without this consideration, registration does not work well, even when manual setup is used. This paper addresses this issue by focusing on the reconstructed camera view by SFM.

As shown in Fig. 1, we present a method that can be used to integrate point cloud data from a laser scanner with mesh data generated from SFM. The key idea here is to utilize the photo image taken by the color camera incorporated into many of the laser scanners currently in production. Such cameras are intended to produce pre-calibrated colored point cloud data that allow the 3D coordinates of the measured point data obtained by the laser ranging sensor to be associated with the captured color pixels. This calibration is done by determining the intrinsic camera parameters used to project 3D coordinates onto a two-dimensional (2D) image plane and using extrinsic camera parameters to perform the scaling, translation, and rotation needed transform the 3D coordinates between the laser scanner and the camera image sensor data.

SFM reconstructs 3D geometry by identifying corresponding natural feature points in multiple photo images based on their shooting positions and orientations. More specifically, in the SFM process, photo imagery taken by the camera incorporated into the laser scanner is applied to the SFM input image collection, which consists of numerous images taken by a UAV equipped with a digital camera. From that point, the shooting positions and orientations of both the laser scanner and UAV cameras can be reconstructed in an identical 3D coordinate system determined by SFM.

Since the scanner camera can view both the reconstructed SFM output 3D scene and the point cloud scanned by the laser scanner on an identical image plane, re-projecting the 3D points reconstructed by SFM onto the

scanner's camera image permits the pixel position corresponding to 3D points in the scanner coordinates to be identified. Although SFM reconstructs only sparse 3D points, based on which the further multi-view stereo (MVS) process is able to generate dense vertices and meshes, using MVS output facilitates the identification of numerous corresponding pairs of 3D coordinates between the SFM and the laser scanner.

Moussa et al. (2014) proposed a registration method for non-overlapping laser scanned data sets using SFM reconstruction that assumes the reconstructed camera parameters produced by SFM are accurate and useful for the determination of 3D-to-3D correspondences when registering non-overlapping laser scanned data. However, in practice, the SFM reconstruction precision depends significantly on the characteristics of the input image collection. Accordingly, we only use the SFM result for finding candidate correspondences of point pairs between the scanned point cloud and the reconstructed points produced by SFM. Re-projecting both 3D point sets on an identical image plane enables 3D-to-3D pair correspondences for different point set types to be determined in a stable manner.

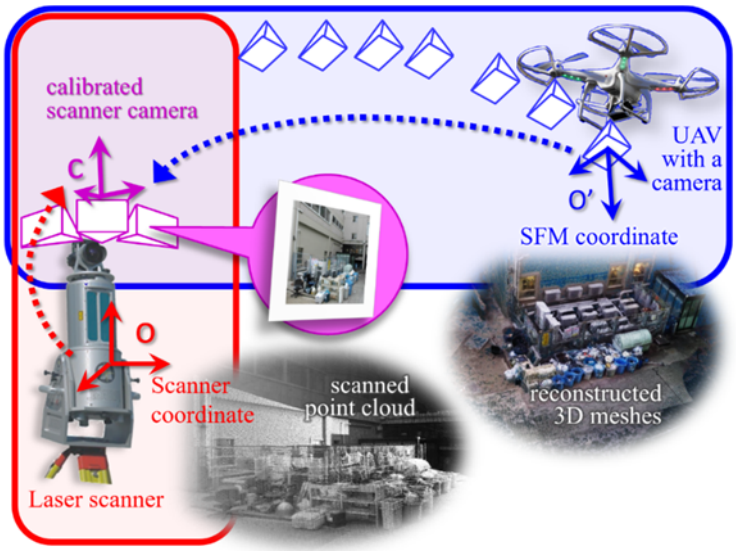

Fig. 2: Schematic concept of the proposed method. Pre-calibrated scanner camera imagery can be used as a medium for transforming SFM and laser scanner coordinate data by incorporating the imagery into the SFM process.

3. Implementation

3.1 Target scene

Assuming that real-world target sites would often have adverse field conditions and complicated structures with numerous visual occlusions, we chose a 20-30-meter-square junkyard on the campus of Kansai University as a test site for our proposed method. In our experiments, we used an LMS-Z420i (Riegl Inc.) laser scanner, and a Phantom2 (DJI Inc.) UAV equipped with a GoPro HERO4 (Woodman Labs Inc.) digital camera. We began by scanning the site from two viewpoints with the laser scanner placed on the ground, and then flew the UAV at a

comparatively low height of about 20 meters at the highest in a manner that ensured the aerial view would cover the entire site.

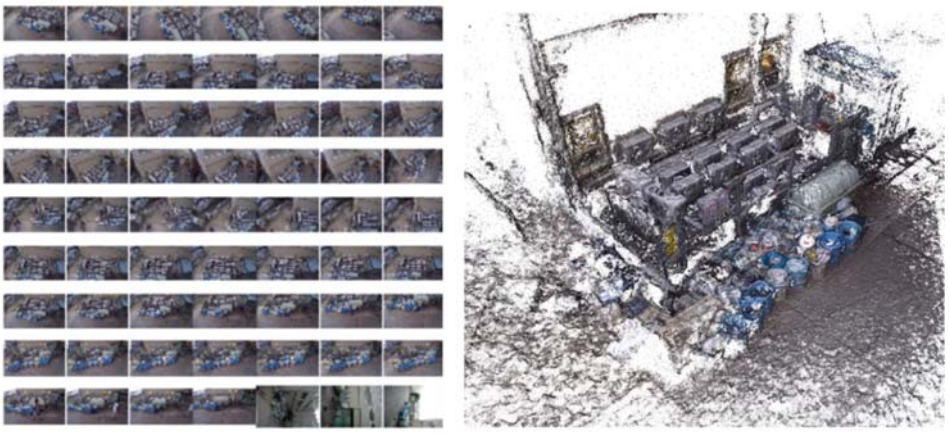

Fig. 3: Portion of the photo collection, including photos taken by both the UAV and laser scanner cameras (left), and the result of the SFM process (right).

3.2 Applying SFM

A total of 133 images were selected for the SFM process. This image collection includes those taken by the scanner camera and captured frames from 4 minutes of video footage taken by the UAV camera. The SFM was produced using VisualSFM, and a further MVS process for creating dense mesh data was performed by CMP-MVS (Jancosek et al. 2012). From these results, 200 million vertices were generated. Moussa addresses the registration of non-overlapping laser scanned data sets using SFM reconstruction (Moussa et al. 2014).

3.3 Selecting Corresponding Points

The scanner camera image plane is associated with the 3D coordinates of the scanned point cloud by using pre-calibrated intrinsic and extrinsic camera parameters. The same image is used in the SFM process, and its estimated viewpoint and orientation results can be used as the extrinsic camera parameters in the SFM coordinate system. From these results, as shown in Fig. 4, corresponding points can be specified in both the scanner point cloud and the SFM reconstructed meshes for each pixel on the image plane. At this point, because the points or reconstructed mesh vertices are unorganized, and thus not aligned on the image plain pixel grid, we use mouse picking to select interest points on the mesh surface produced by SFM.

Mouse picking calculates the ray trajectory from the camera viewpoint toward the mesh surface via pixel selected by mouse, based on the camera parameters. Next, the intersection between the ray and the closest surface from the camera viewpoint is found. Since the intersection point on the mesh polygon surface will be picked up even if the ray does not hit any vertices, it is supposed that point sampling with linear interpolation has

been achieved. The laser scanned point counterpart is then associated with the mouse-picked point.

Fig. 4: Point cloud data view (left) and reconstructed meshes view produced by a SFM and MVS process from an identical viewpoint obtained via the scanner camera. Corresponding points can be found at the same pixel position on the image.

3.4 Registration

Since scan data have precise dimensions and organized points, we set the scan data as the reference and the reconstructed data produced by SFM as the source data. Registration is accomplished by finding a scaling, translation, and rotation transform that can match the source data to the reference data. By mouse picking *n* times, candidate corresponding points set in the source data are given as $\mathrm{X} = \{x_i\}_{i=1}^{n}$, and in the reference data as $\mathrm{Y} = \{y_i\}_{i=1}^{n} \subset R^3$. Using unknown rotation and translation parameters $(R, t) \in \mathrm{SE}(3)$ and scaling parameter s gives the following relation:

$$y_i = sRx_i + t.$$

In order to estimate the best parameters $\mathrm{R}, \mathrm{t}, \mathrm{s}$ from given matrices $\mathrm{X}, \mathrm{Y} \in R^{3\times n}$, we minimize the objective function

$$\min_{(R,t)\in SE(3)} \sum_{i=1}^{n} \|y_i - (sRx_i + t)\|^2,$$

where x_i, y_i are aligned columns of X, Y, respectively. The optimization can then be expressed as follows:

$$\min_{(R,t)\in SE(3)} \sum_{i=1}^{5} \|Y_i - (sRX_i + tI^T)\|_F^2,$$

where $I = (1, 1, \cdots, 1)^T \in R^n$ and $\|\cdot\|_F$ denotes the Frobenius norm, which is defined $\|A\|_F = \sum_{ij} {a_{ij}}^2$ for a matrix $\mathrm{A} = (a_{ij})$. If we let the centroids (mass centers) of X, Y be $\bar{x} = XI/n, \bar{y} = YI/n$, then following holds:

$$\bar{y} = sR\bar{x} + \mathrm{t}$$

Using ${x'}_i = x_i - \bar{x}$, ${y'}_i = y_i - \bar{y}$, the translation t can be canceled and the unknown R is the optimal rotation matrix for following problem:

$$\min_{R \in SE(3)} \sum_{i=1}^{5} \|Y' - sRX'\|_F^2$$

where $\mathrm{X}' = ({x'}_1, {x'}_2, \cdots, {x'}_5), \mathrm{Y}' = ({y'}_1, {y'}_2, \cdots, {y'}_5)$. Using the relation $\|A\|_F = \mathrm{tr}(A^T A)$, the objective function can be written as follows (Arun 1987, Schönemann 1996):

$$\|Y' - sRX'\|_F^2$$

$$= tr\big((Y' - sRX')^T (Y' - sRX')\big)$$
$$= \mathrm{tr}({Y'}^T Y') + s^2 tr({X'}^T X) - 2str({Y'}^T RX')$$

Since the first two terms are not related to minimization, the problem can be redefined as the following maximization:

$$\max_R tr({Y'}^T RX').$$

Let the singular value decomposition of $X'{Y'}^T \in R^{3\times 3}$ be $X'{Y'}^T = U\Sigma V^T$, where Σ is a non-diagonal matrix and $\mathrm{U}, \mathrm{V} \in O(3)$. Then the objective function can be rewritten as follows:

$$tr\big({Y'}^T RX'\big) = tr(RX'V) = tr(RU\Sigma V^T) = tr(V^T RU\Sigma)$$

$\leq \mathrm{tr}(\Sigma)$.

Eventually, when $V^T RU = I_3$, the objective function is maximized and the rotation is given by $\hat{R} = VU^T$. Then the scaling parameter s is given by the flowing equation:

$$\mathrm{s} = \frac{tr({Y'}^T X'\bar{R})}{tr({X'}^T X')}.$$

After s and R are obtained, the translation can be calculated as $\mathrm{t} = \bar{y} - sR\bar{x}$.

In our experiment, we implemented a mouse-picking user interface to select ten points (pixels) on the scanner image view without performing any other data operations. Of these, three points are excluded as outliers by random sample consensus (RANSAC). In Fig. 5, the red dots are the selected points and the blue dots are the outliers. The registration result produced by integrating scanned points and SFM reconstructed meshes is shown in Fig. 6. The residual error of the registration was 0.149 m. When compared to the laser scanner root-mean-square (RMS), the residual is large due to noise in the meshes produced by SFM, as show in Fig. 7.

This is especially prevalent on the ground surface, where less natural features are found in the image even though many roughly shaped artifacts are visible on the smooth surfaces. Despite this result, we found the correspondences works stably under the same view guarantee.

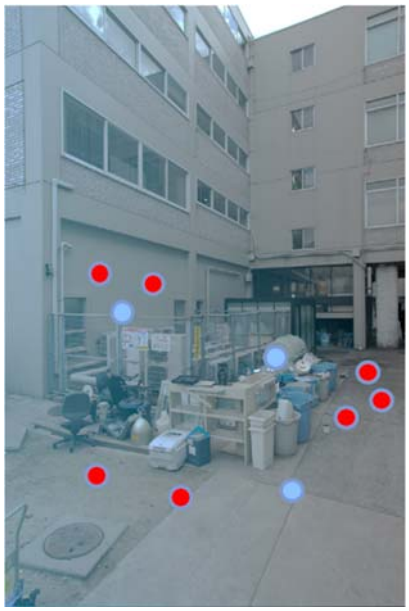

Fig. 5: 3D data view from the scanner image. Both the scanned point cloud and the SFM/MVS reconstructed meshes can be selected from the same pixel section, which permits correspondences to be easily made via mouse picking.

Fig. 6: Registration result. The white dots show the point cloud produced by the scanner and the colored dots are the SFM/MVS reconstruction.

4. CONCLUSIONS

This paper proposed a method for correlating the point cloud data obtained from a laser scanner and UAV camera imagery in order to integrate them together via SFM. More specifically, using a laser scanner equipped with a calibrated color digital camera, the scanner viewpoint is integrated into the 3D geometry reconstructed from UAV collected images via SFM. Then, the 3D geometry of both the SFM and the laser-scanned data are viewed from the scanner position viewpoints, which allows them to be overlaid upon each other. In the next step, selecting the proper number of points on the scanner view by using the RANSAC algorithm facilitates finding the proper corresponding points in order to fit them automatically. Our case study showed effective results based on actual data collected by laser scanner and a UAV-mounted camera that were within the SFM result data quality error rate. Our next step will focus on a fully automatic registration scheme instead of user mouse picking operations.

Fig. 7: Artifacts in the reconstructed SFM and MVS meshes. Reducing the number of image features may cause more unevenly shaped artifacts on originally smooth surfaces.

ACKNOWLEDGMENT

This work was supported by Japan Society for Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (15H02983, 15K02977, 16H05000).

REFERENCES

Shih, N. and Huang, S. (2006). 3D Scan Information Management System for Construction Management, *Journal of Construction Engineering and Management,* 132 (2), 134-142.

Shih, N. and Wang, P. (2004). Point-Cloud-Based Comparison between Construction Schedule and As-Built Progress: Long-Range Three-Dimensional Laser Scanner's Approach, *Journal of Architectural Engineering,* 10(3), 98-102.

Miller P. E., Mills J. P., Barr S. L., Lim M., Barber D., Parkin G. and ClarkeB., Glendinning S. and Hall J. (2008), Ter-restrial laser scanning for assessing the risk of slope in-stability along transport corridors, XXIst ISPRS Con-gress:Commission V, WG 3, Beijing, pp.495-500.

Watson, C., Chen, S., Bian, H., and Hauser, E. (2011). LiDAR Scan for Blasting Impact Evaluation on a Culvert Structure, *Journal of Performance of Constructed Facilities,* doi: 10.1061/(ASCE) CF.1943-5509.0000318.

Wu, C. (2013). Towards linear-time incremental structure from motion, *International Conference on 3DTV*, pp. 127-134.

Wassim Moussa and Dieter Fritsch, (2014), Automatic Registration of Non-overlapping Laser Scans Based on a Combination of Generated Images from Laser Data and Digital Images in One Bundle, EuroMed 2014, LNCS 8740, pp. 1-10.

Snavely, N., Seitz, S. M., and Szeliski, R. D. (2006). Photo tourism: Exploring photo collections in 3D, *ACM Transactions on Graphics (SIGGRAPH Proceedings)*, 25 (3), 835-846.

Jancosek M. et al., (2011), Multi-View Recon-struction Preserving Weakly-Supported Surfaces, CVPR 2011, pp.3121-3128.

P. J. Besl and N. D. McKay, (1992), A Method for Registration of 3-D Shapes, IEEE Trans. Pattern Anal. Machine Intell, vol. 14, NO. 2, pp. 239-256.

K. S. Arun, T. S. Huang, and S. D. Blostein. (1987), Least-squares fitting of two 3-D point sets. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 9, No. 5, pp. 698–700.

Peter H. Sch**ö**nemann, (1996), A generalized solution of the orthogonal procrustes problem. Psychometrika, Vol. 31, No. 1, pp. 1–10.