# STEREO IMAGE POINT CLOUD AND LIDAR POINT CLOUD FUSION FOR THE 3D STREET MAPPING

**Yuan Yang**, Ph.D. Student
**Zoltan Koppanyi**, Post-Doctoral Researcher
**Charles K Toth**, Research Professor
SPIN Lab
The University of Ohio State
Columbus, OH43210
yang.2695@osu.edu
koppanyi.1@osu.edu
toth.2@osu.edu

## ABSTRACT

Combining active and passive imaging sensors enables creating a more detailed 3D model of the real world. Then, these 3D data can be used for various applications, such as city mapping, indoor navigation, autonomous vehicles, etc. Typically, LiDAR and camera as imaging sensors are installed on these systems. Both of these sensors have advantages and drawbacks. Thus, LiDAR sensor directly provides relatively accurate 3D point cloud, but LiDAR point cloud barely contains the surface textures and details, such as traffic signs and alpha numeric information on facades. As opposed to LiDAR, deriving 3D point cloud from images require more computational resources, and in many cases, the accuracy and point density might be lower due to poor visual or light conditions. This paper investigates a workflow which utilizes factor graph SLAM, dense 3D reconstruction and ICP to efficiently generate the LiDAR and camera point clouds, and then, co-register in a navigation frame to provide a consistent and more detailed reconstruction of the environment. The workflow consists of three processing steps. First, we use factor graph SLAM, GPS/INS odometry and 6DOF scan matching to register the LiDAR point cloud. Then, the stereo images are processed by stereo-scan dense 3D reconstruction technique to generate dense point cloud. Finally, ICP method is used to co-register LiDAR and photogrammetric point clouds into one frame. The proposed method is tested with the KITTI dataset. The results show that data fusion of two point clouds can improve the quality of the 3D model.

**KEYWORDS:** LiDAR point cloud, stereo image 3D reconstruction, factor graph, ICP, data fusion

## INTRODUCTION

Over the recent years, the need for 3D city models are increasing. Mobile mapping systems allow for collecting huge datasets around the city to create these models. Therefore, local and national authorities invest to buy mobile mapping platforms, such as UAV (Unmanned Aerial Vehicle), MMS (Mobile Mapping System). One of the main challenge is to extract information relevant to the city model from the data sources captured

by the platform's sensors. The combination of the various data streams allows us to improve the information extraction. LiDAR (Light Detection and Ranging) and stereo camera systems are widely used sensor configuration for mobile mapping systems. Regarding algorithms, Simultaneous Localization and Mapping (SLAM), bundle adjustment (BA), and dense point matching are commonly used to process the data and generate point cloud from LiDAR and images. However, there are critical application-specific implementation problems with each technique. For example, mapping using data from a laser scanner can be difficult because the problem requires the correction of the motion distortion in the Lidar cloud. In addition, LiDAR point cloud can barely capture the surface textures and details (street signs and alpha numeric information on facades) of 3D model. Similarly, the quality of the image-based point cloud might be low due to the lack of tie points (features) and the impact of varying acquisition conditions, such as sun angle, shadows.

To support building 3D maps, this paper proposes a workflow to process LiDAR and image point cloud datasets and co-register them in a common frame to create a more detailed point cloud. First, we apply factor graph SLAM and stereo-scan dense 3D reconstruction to produce LiDAR point cloud and image-based point clouds. Next, we use ICP method to fuse Lidar point cloud and stereo image point cloud into one mapping frame.

The rest of the paper is organized as follows. The related work will be briefly presented in "Related Works" section. In the next section, we will summarize the workflow for building 3D map including algorithms and methods for processing the LiDAR and stereo image datasets. In section of "Results", the comparison and statistics of the results are presented. Finally, the paper ends with conclusion.

## RELATED WORKS

The data fusion of LiDAR and photogrammetric datasets has been wildly used across various applications, such as forest management, city planning, catastrophe control, heritage recording, robotic utilities, etc. Researchers from many disciplines including the remote sensing and computer vision community have shown interest in solving relevant problems with fusing LiDAR and photogrammetric data (Basgall 2013). Among of these comprehensive researches, there are two main types of data fusion approaches.

First, one dataset supports the improvement of the other dataset during the point cloud registration. For instance, in a research, where the authors improved LiDAR point cloud registration acquired by MMS in GPS-denied area (Gajdamowicz et al. 2007), the feature coordinates extracted from images were used to correct LiDAR point cloud. Another example was shown in an application of heritage protection (Nex. F 2010). In this research, LiDAR data was used to aid the edge detection and matching between images for cultural heritage surveys. All these approaches in this category focus on exploiting the merit of image with the characteristics of LiDAR data. In the second approach, the both sensors are used for generating point clouds, and then, these clouds are co-registered. Note that this approach is in the focus of this study. This approach is typically used for building DTM by using airborne LiDAR scanning data and aerial imagery (Basgall 2013; Schenk et al. 2003). In the field of heritage preservation, merging LiDAR and image-based point clouds is also used for creating 3D model of objects, such as monument, statues, or frontage (Muñumer et al. 2015).

## PROPOSED APPROACH

This paper presents a workflow to improve the details of mobile mapping point clouds by merging LiDAR and image-based point cloud with the ability to efficiently process a large dataset, such as long sequence of images. As shows in Figure 1, the proposed method consists of three main steps. In the first step, LiDAR scans is registered by using factor graph SLAM with GICP (Generalized Iterative Closest Point) scan matching and GPS/INS poses as constraints. In the second step, stereo images are processed with the dense 3D reconstruction technique to derive dense point clouds. Finally, as the last step of the workflow, the ICP is used for merging the two point clouds into one common frame. The method is tested with a 100 seconds length of KITTI benchmark data set (Geiger et al. 2012). The coordinate system is the sensor local frame, which specifically is the coordinate system of the left camera.

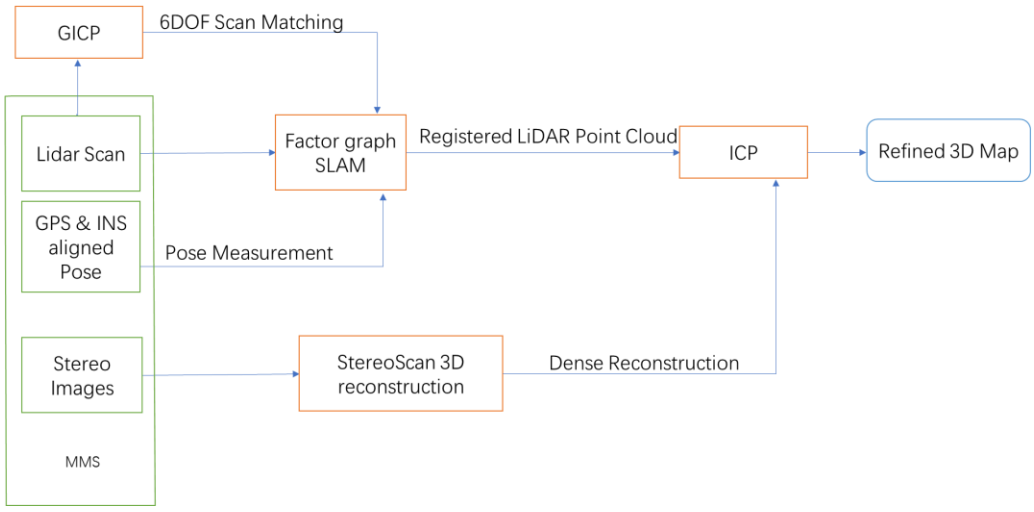



**Figure 1**. Overview of workflow

### Factor graph SLAM for LiDAR point cloud

In this section, we present the LiDAR point cloud registration method applied in this study. The goal is to produce a metrically accurate point cloud. In this paper, we apply the factor graphs to derive the posesof the platform, including position, and attitudes (Dellaert 2012). The structure of factor graph used in this paper is presented in Figure. 2. This factor graph contains poses $(x_n)$ as nodes, i.e. the unknown variables of the graph, and the constrains are poses from the KITTI GPS/INS integrated solution $(o_m)$ and the transformation parameters between the consecutive point clouds obtained by an ICP $(c_i)$ algorithm. Here, we use GICP as ICP implementation. The nonlinear optimization problem of the factor graph is solved with Levenberg-Marquardt optimizer.

The consecutive LiDAR point clouds can be transformed into one mapping frame using the poses obtained from the factor graph; the results can be seen in Figure 3a. We use an SOR (Statistical Outlier Removal) filter to remove noise presented in the point cloud, see Figure 3b. The filtered point cloud contains 17,563,260 points. For the qualitative analysis, the points of the cloud is projected onto the corresponding

image plane, see Figure 3c.

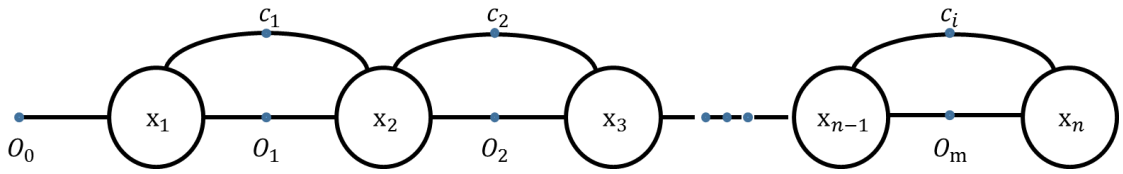



**Figure 2.** Factor graph for estimating the platform poses.

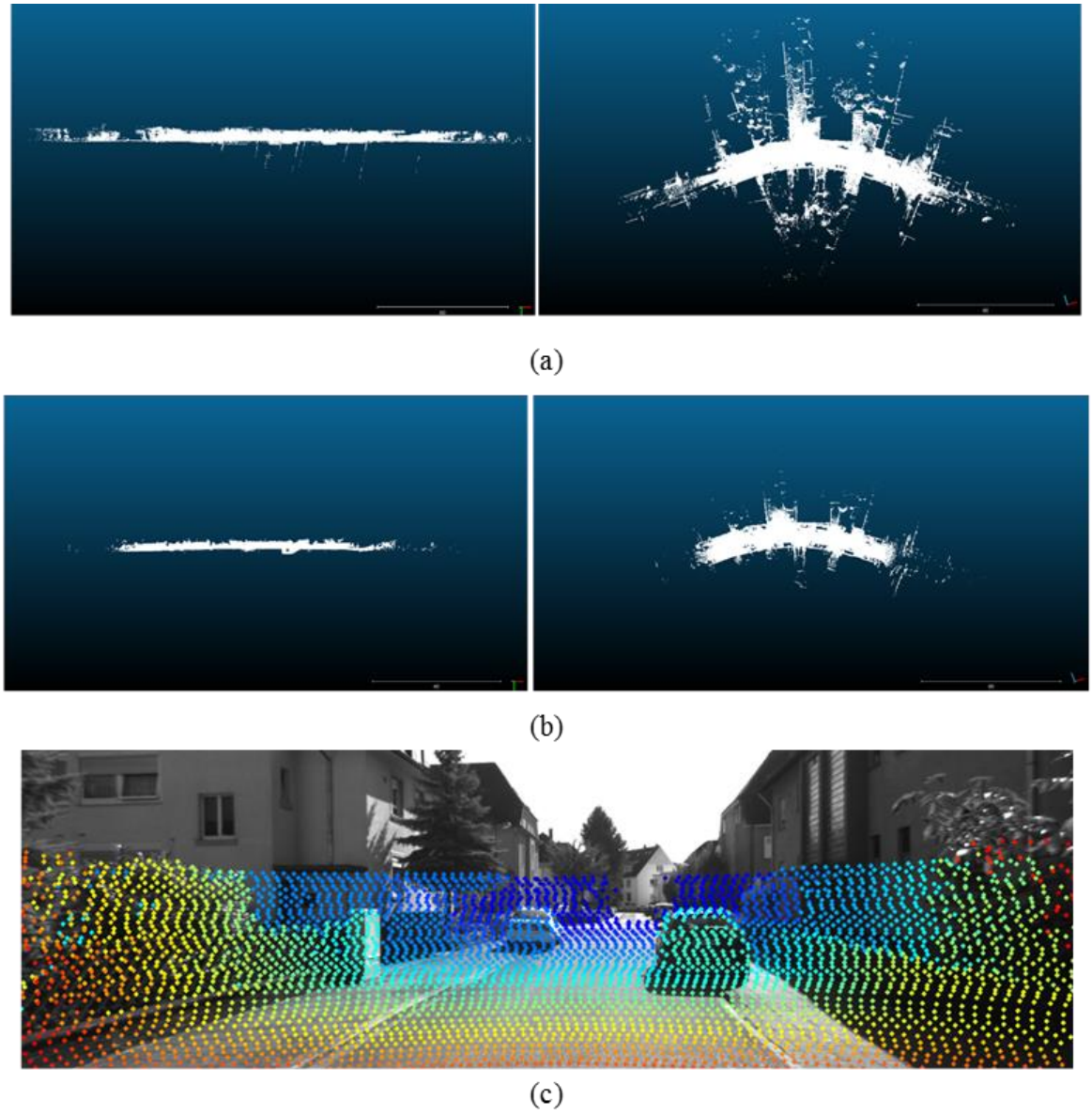
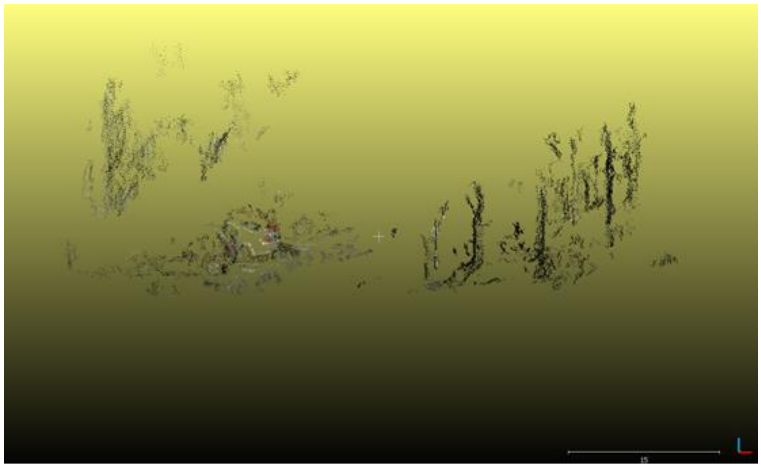



**Figure 3.** Results of the LiDAR registration. In (a), we show the registered point cloud before filtering. The results after filtering is shown in (b). Sample projection of 3D point into image is shown in the (c).

## Photogrammetric 3D Reconstruction

The sequence of stereo gray images with size of 1241×376 is used as inputs for the scene reconstruction. The stereo images are simultaneously acquired with the LiDAR scans. The SFM techniques are widely applied in 3D reconstruction application. However, such techniques have considerably computational burden problem and need several redundant viewpoints for the detailed reconstruction. Our dataset was collected by an MMS, and thus, objects can only be observed for short time periods and from one direction resulting in small numbers of features that are essential for SFM. Overall, this reduces the quality of the point clouds derived from images. An example is shown in Figure 4. One sample image from the scene is presented in Figure 4b. The resulting reconstruction from these images are shown in Figure 4a. Note that the derived point cloud is very sparse. Clearly, this results cannot be used for mapping the environment. In addition, the computation was run on a MacBook Air with a 2.2 GHz Core i7-5650U, 8 GB ram and Intel HD Graphics 6000, took 1.5 hours for 32 images, just for illustrating the computation burden of this process.



(a)

(b)

**Figure 4.** 3D reconstruction by SFM: (a) The point cloud. (b) Pictures of the scene for reconstruction.

To build dense point cloud from stereo image sequence, we use StereoScan 3D reconstruction technique in this paper, for more details, see Geiger et al. 2011. The approach is based on the Libviso2 (Library for Visual Odometry 2) library, LIBELAS (Efficient large-scale stereo matching) library, see Geiger et al. 2010, and STEREOMAPPER. The proposed pipeline consists of four steps:

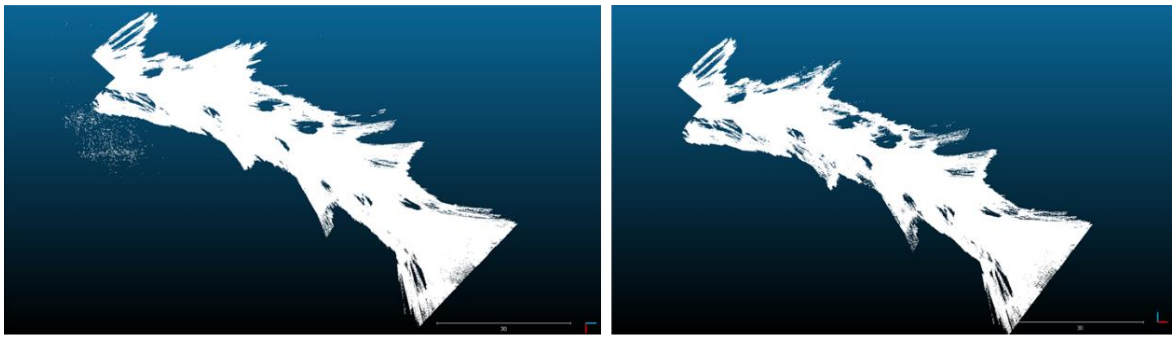
1) **Sparse feature extraction and matching**. In the first step, the features as blob or corners are extracted from images, then match the feature descriptors.
2) **Egomotion estimation.** The motion parameter including translation and rotation of moving platform are estimated. The egomotion estimation is achieved by matching features in a 'circular' way; between the left and right images as well as between the consecutive frames.
3) **Dense stereo matching.** Once, the relative positions between the stereo cameras are known, then, the 3D object space coordinates can be derived for all 2D image points.
4) **3D reconstruction**. In Step 3, one-one point clouds are derived epoch-by-epoch. In the last step, the point clouds obtained from the stereo image matching are merged into one mapping frame. To increase computational efficiency, a filter is used for reducing the amount of points and noise.

After processing stereo images through the proposed pipeline and applying SOR filter, the dense reconstruction contains 17,288,623 points, see Figure 5.

## DATA FUSION WITH ICP METHODS

The last step in our workflow is to match the LiDAR and image point clouds with ICP. ICP iteratively tries to transform one point cloud to the other one with minimizing the distance between corresponding closest point pairs. The objective function can be formulated as

$$ {}^{i+1}_{i}T = \arg\min_{T}(\sum_{k=1}^{K}\|Tp_k - q_k\|_2), \quad (1) $$

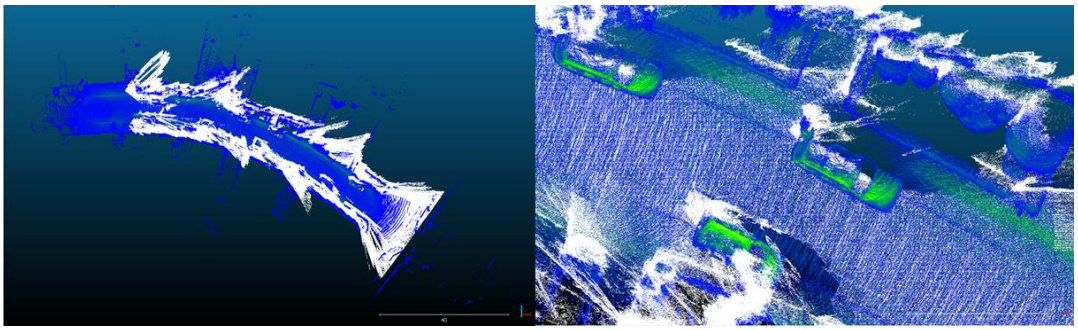


(a) (b)

**Figure 5.** (a) shows the 3D reconstruction via StereoScan 3D reconstruction method. (b) shows the point cloud after SOR filter.

where $K$ is the number of matched points, P= $[p_1 \dots p_n]$, Q= $[q_1 \dots q_m]$ are the two point clouds expressed in homogenous coordinates. $T$ is a 4-by-4 homogenous transformation matrix. Note that the points in P and Q are ordered, in that way, $p_k$ and $q_k$ are the closest point pairs. Comparing to other ICP methods, in GICP, a probabilistic model is applied to this minimization process (Segal et al. 2009).

The main advantages of ICP methods are the simplicity (Segal et al. 2009). However, the initial position of two point clouds can highly influence the matching quality. To minimize this influence, before the ICP process, we transform the two point clouds into one common frame by using the sensor-to-sensor calibration parameters offered by KITTI. The co-registered point clouds and a zoom in detail are shown in Figure 6, in which the white points are from the image-based point cloud and the blue/green points are the LiDAR points. The co-registration is performed with GICP and ICP. These two methods are compared in the "Results" section.



(a) (b)

**Figure 6.** The 3D view of point clouds before ICP process is shown in (a). (b) shows the scene in details.

## RESULTS

### Compare ICP and GICP

In this section, we firstly compare and analyze the results of two ICP methods. Figure 7 illustrates the the registered point clouds produced by ICP and GICP with the same view as the Figure 6. Note that there is an offset between the two point clouds highlighted by the red marked area in Figure 7d, such as the edges of walls and vehicles. On the other hand, in the same areas of Figure 7 (c), the matching result is better than in (d). Thus ICP outperforms GICP in terms of matching accuracy. In Figure 8, we also compared LiDAR odometry and visual odometry trajectories before and after data fusion process. In Table 1, we present the values of RMS between LiDAR and Visual trajectories components (X, Z) from not fused, ICP and GICP results. The quantitative analysis shows that ICP can matching two point clouds with higher accurate in both X and Z direction than the performance of GICP.

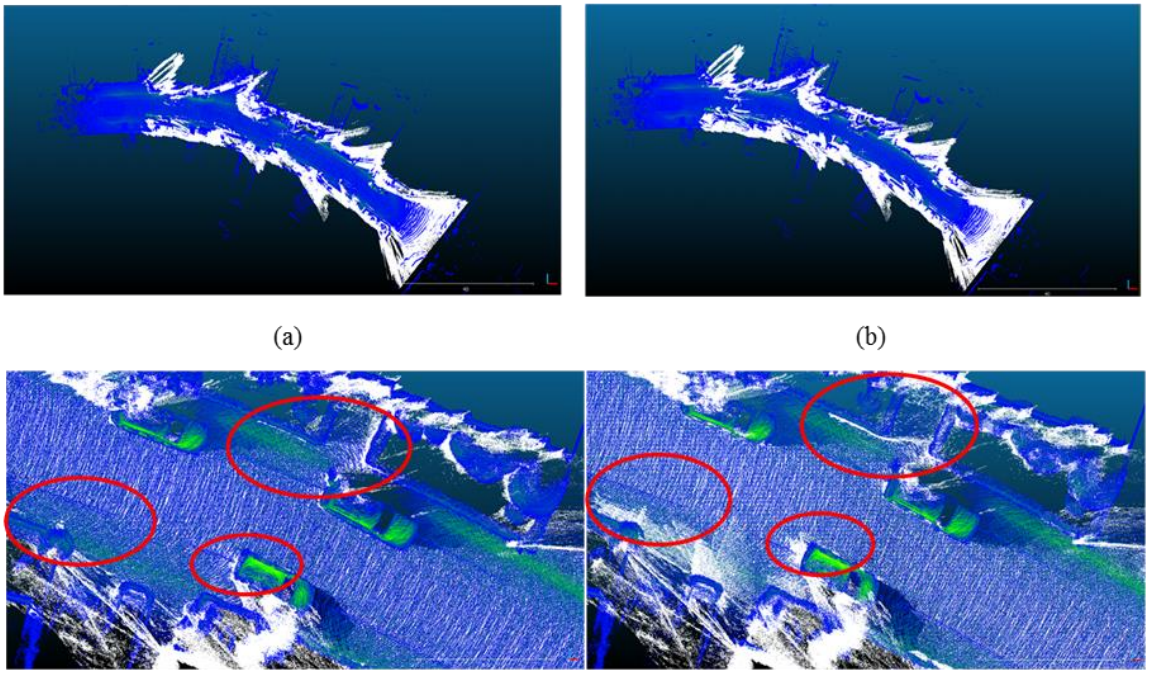



**Figure 7.** Fusion of both point clouds. (a) and (c) are the results derived from ICP tool of CloudCompare. (b) and (d) are the results derived from GICP. The red marks indicate some obvious differences of matching.

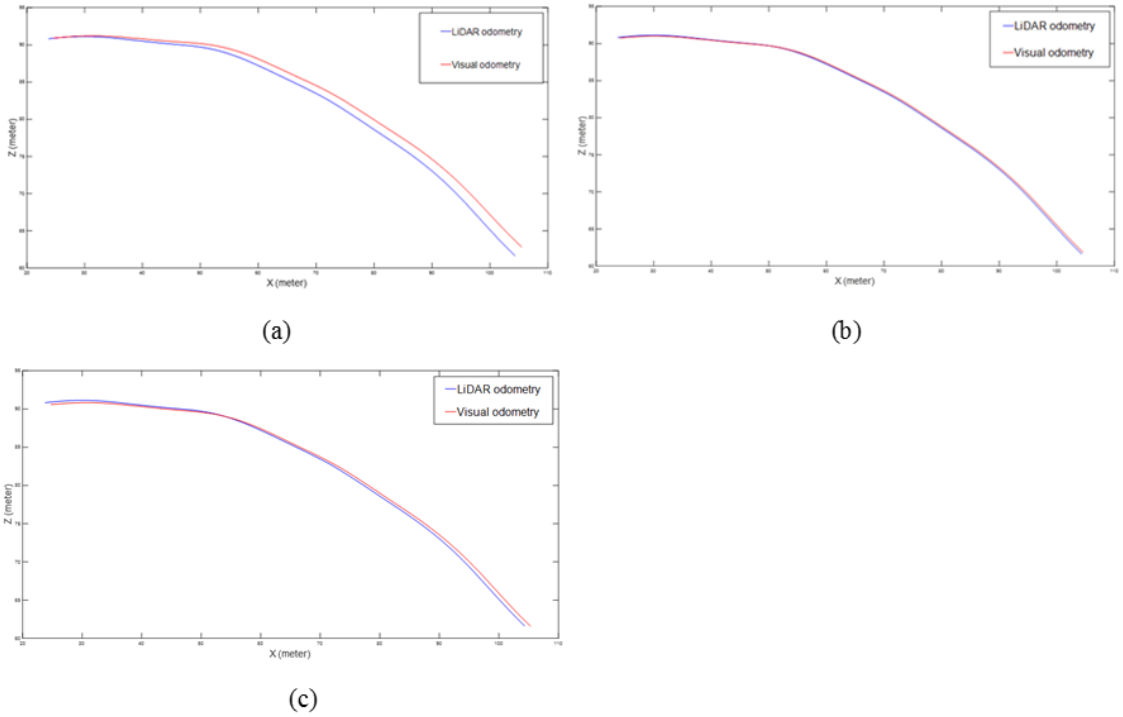



**Figure 8.** LiDAR (blue) and visual (red) odometry results: (a) before fusion; (b) after registration with ICP in CloudCompare; (c) after registration with GICP

**Table 1. Comparison of RMS Values Between LiDAR and Visual Odometrys from Not Fused, ICP and GICP Results**

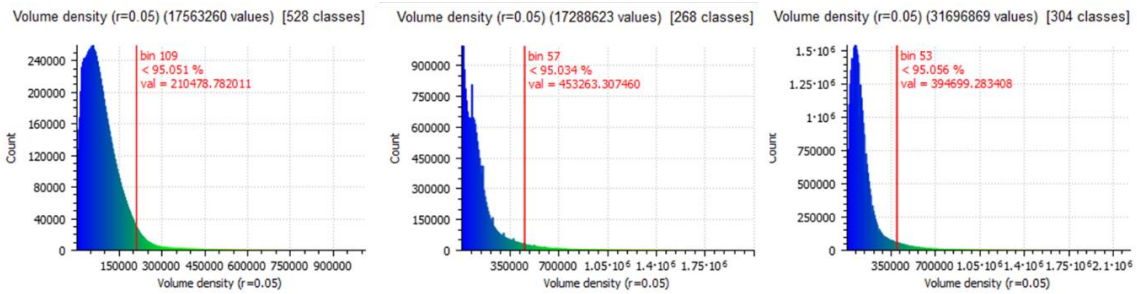
| Odometry RMS | Not fused data | Fusion with ICP | Fusion with GICP |
|---|---|---|---|
| RMS (X) | 1.1223 m | 0.1231 m | 1.1270 m |
| RMS (Z) | 0.4653 m | 0.0117 m | 0.0408 m |

## Investigation on the improvement of the point cloud

We compare the size and density of the LiDAR, image and combined point cloud to assess the improvement of point cloud. The results are shown in Table 2 and Figure 9. In Figure 9, a threshold of 95% of density histogram was depicted by red line. Note that this 95% limit is 200000 in Figure 9a, and 420000 for Figure 9c indicating that the density of original LiDAR point cloud is significantly increased by integrating with image-based point cloud. An example which areas are covered by the two types of point clouds can be seen in Figure 10. Here, we show a pole that can be seen at the left side of picture. Note that not all part of the pole is captured in the LiDAR point cloud, see Figure 10b. In Figure 10c, the pole is complete after merging with the 3D reconstruction. The similar situation is shown in Figure 11. The missing areas around cars can be seen in Figure 11a, which is caused by the vehicles are covered the camera's field of view. This area can be filled up by point cloud obtained from LiDAR, see Figure 11b.

**Table 2. Sizes for LiDAR, image-based and combined point clouds**

| **Category of Point Cloud** | **Amount of Points** |
|---|---|
| LiDAR | 17563260 |
| Image-based | 17288623 |
| ICP Combined | 34851883 |




(a) (b) (c)

**Figure 9.** Density histograms of LiDAR point cloud (a), 3D reconstruction (b) and fusion of both point clouds (c).

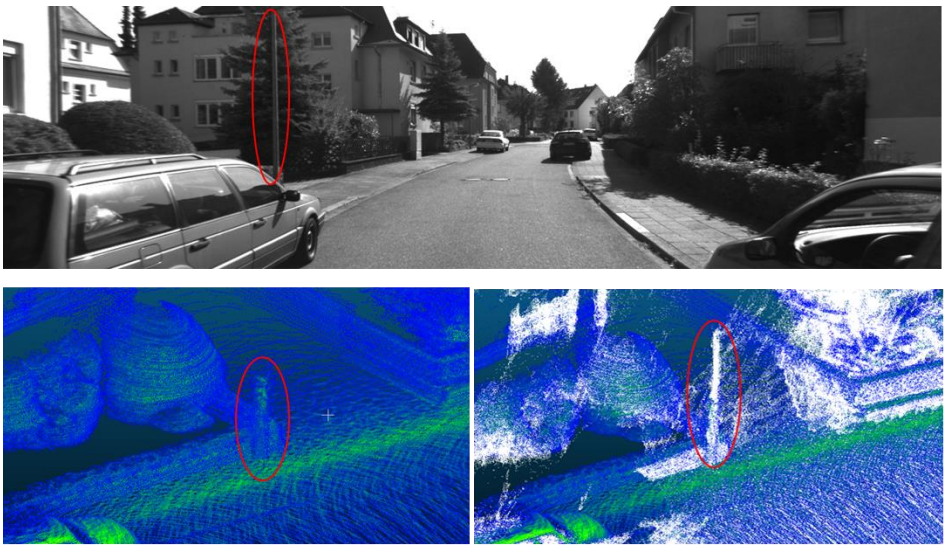
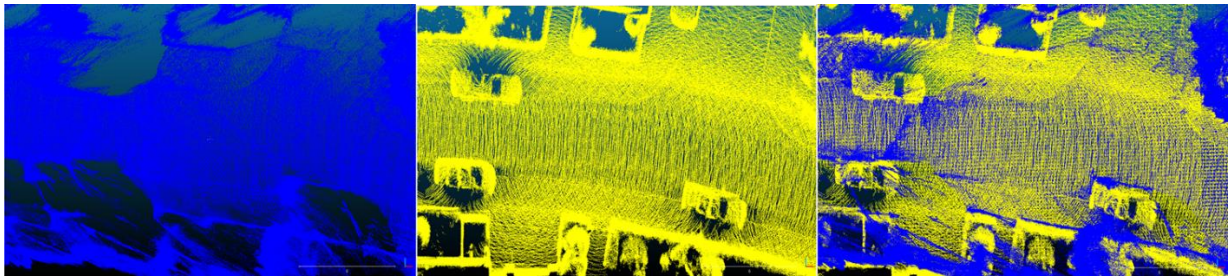


**Figure 10.** Example of image-based point cloud supplement the missing part in LiDAR point cloud



(a) (b) (c)

**Figure 11.** Example of two point clouds supplement each other: (a) is 3D reconstruction, (b) is LiDAR point cloud, (c) is fusion of both point clouds.

## CONCLUSION

In this study, we proposed a workflow for 3D street mapping using data collected from multi sensor system. This workflow utilizes factor graph SLAM to register the consecutive LiDAR scans, a stereo dense 3D reconstruction to generate dense point cloud, and finally, ICP for merging the two point clouds into one mapping frame. The workflow is analyzed on a 100 s long section of the KITTI dataset. This dataset includes LiDAR scans, stereo images and GPS/INS navigation data. In addition, the performance of the two ICP methods is compared in terms of quality of the integrated point cloud. The results show that gaps in the LiDAR point cloud can be filled up with image-based point cloud after the data fusion process. Furthermore, test of ICP methods shows that the ICP in CloudCompare can achieve a better matching accuracy than the GICP on the investigated dataset.

## REFERENCES


Basgall, P.L., 2013. *Lidar point cloud and stereo image point cloud fusion* (Doctoral dissertation, Monterey, California: Naval Postgraduate School).

Biljecki, F., Ledoux, H., Stoter, J. and Vosselman, G., 2016. The variants of an LOD of a 3D building model and their influence on spatial analyses. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, pp.42-54.

Dellaert, F., 2012. *Factor graphs and GTSAM: A hands-on introduction*. Georgia Institute of Technology.

Gajdamowicz, K., Öhman, D. and Horemuz, M., 2007, May. Mapping and 3D Modelling of Urban Environment Based On LiDAR GPS IMU and Image Data. In *International Symposium on Mobile Mapping Technology*.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? the kitti vision benchmark suite, In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference, pp. 3354-3361.

Geiger, A., Roser, M., Urtasun, R., 2010. Efficient large-scale stereo matching, In: Asian conference on computer vision, Anonymous Springer, pp. 25-38.

Geiger, A., Ziegler, J., Stiller, C., 2011. Stereoscan: Dense 3d reconstruction in real-time, In: Intelligent Vehicles Symposium (IV), 2011 IEEE, Anonymous Ieee, pp. 963-968.

Gröger, G., Plümer, L., 2012. CityGML–Interoperable semantic 3D city models. ISPRS Journal of Photogrammetry and Remote Sensing 71 12-33.

Herrero-Huerta, M., Felipe-García, B., Belmar-Lizarán, S., Hernández-López, D., Rodríguez-Gonzálvez, P., González-Aguilera, D., 2016. Dense Canopy Height Model from a low-cost photogrammetric platform and LiDAR data. Trees 30 (4), 1287-1301.

Muñumer, E. and Lerma, J.L., 2015, September. Fusion of 3D data from different image-based and range-based sources for efficient heritage recording. In *Digital Heritage, 2015* (Vol. 1, pp. 83-86). IEEE.

Nex, F., Rinaudo, F., 2010. Photogrammetric and LiDAR integration for the cultural heritage metric surveys. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38* (Part 5), 490-495.

Pomerleau, F., Colas, F. and Siegwart, R., 2015. A review of point cloud registration algorithms for mobile robotics. *Foundations and Trends® in Robotics*, *4*(1), pp.1-104.

Schenk, T., Csathó, B., 2002. Fusion of LIDAR data and aerial imagery for a more complete surface description. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences 34* (3/A), 310-317.

Schenk, T., Seo, S. and Csatho, B., 2001. Accuracy study of airborne laser scanning data with photogrammetry. *International Archives of Photogrammetry and Remote Sensing*, *34*(part 3), p.W4.

Segal, A., Haehnel, D. and Thrun, S., 2009, June. Generalized-ICP. In *Robotics: science and systems* (Vol. 2, No. 4).