2022 EC³
2022 European Conference on Computing in Construction
Ixia, Rhodes, Greece
July 24-26, 2022

PRECISE AND DENSE AI-BASED MOBILE 3D RECONSTRUCTION OF INDOOR SCENES BY CAMERA-LIDAR FUSION AND ODOMETRY

Maciej Trzeciak[1] and Ioannis Brilakis[1]
[1]Department of Engineering, University of Cambridge, United Kingdom

Abstract

We propose a mobile 3D reconstruction method for improving the precision and density of point clouds. It is suitable for hand-held scanners comprised of a colour camera and a lidar. We fuse time-synchronized and spatially registered images and lidar sweeps using deep learning techniques into dense scans, which are then used for progressive reconstruction in an odometry-like manner. We build a prototypic scanner and test our method in an indoor case-study. The results show that our pipeline outperforms reconstructions by other devices and methods, yielding relatively denser and detail-preserving point clouds with a 67% reduction in noise of reconstructed planar surfaces.

Introduction

Construction professionals do not follow hard-and-fast rules while collecting geometric data of objects. On one extreme, a plumber might use a tape to measure the length of a pipe against the design specifications. On the other hand, an inspector will probably use a static laser scanner to document the as-is state of a whole asset while performing quality control. Part of the reason behind the diversity of rules is a lack of general guidelines. The method used by a professional will be rather intuitive and depend on the requirements of the task at hand. There are many technologies currently on the market which can reconstruct infrastructure scenes, filling the needs of some use cases that exist between the two mentioned extremes.

However, professional scanning tools become more and more mobile in contrast to well established static/terrestrial laser scanners such as FARO (2016). Mobile scanners bring the hope of increased productivity unlike their terrestrial counterparts, for which the effort spent on acquiring data is substantial, followed by even more substantial post-processing (BIM task group, 2013; Kalyan et al., 2016). Mobile scanners such as those produced by GeoSLAM (n.d.) or Smith (n.d.) enable faster scanning and eliminate the need for time-consuming single-place data capturing, followed by painstaking registrations of individual scans.

Nevertheless, the mobility of these devices comes at a cost. The top 2 desirable attributes of mobile scanners are point cloud resolution/density and point cloud accuracy according to one of the biggest recent surveys in the AECO (Architecture, Engineering, Construction, Operation) industry (NavVis et al., 2021). The same survey further shows that the key barriers to further market penetration of mobile scanners are the lack of accuracy and reliability of obtained point clouds.

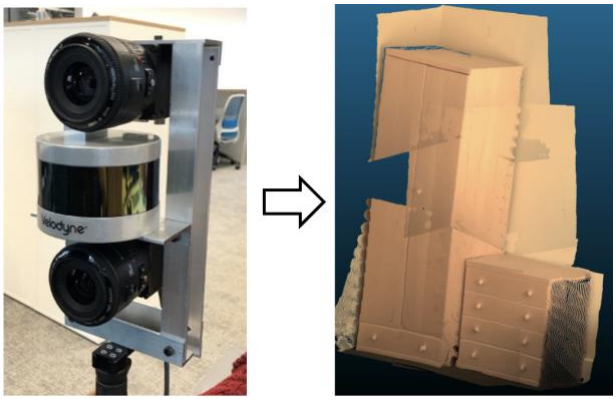

Figure 1: Our prototypic scanner (left); an indoor scene reconstructed using the proposed method (right).

We adopt a geodetic concept of accuracy presented by USIBD (2016) where it is split into 'Precision' and 'Correctness'. We define 'Precision' as the standard deviation (noise) of part of a point cloud (for example, the spread of points on a planar wall) while 'Correctness' is the absolute position of points against their true value. For simplicity, the former can be thought of as a local attribute of a point cloud related to the random noise on measurements of a sensor which produced it while the latter corresponds to the systematic error of a scanning device such as a drift that mobile scanners suffer from.

Another point cloud attribute closely related to reliability and accuracy is resolution. According to USIBD (2016), the resolution of point clouds coming from laser scanners should be such that the points are not far away from each other compared to the diameter of the beam of a laser scanner. For example, some details on a wall would be missed if spaces between points are too large. On the other hand, decreasing the distances between points to a value smaller than the diameter of the beam results in redundancy in data. However, problems with too low resolution are dominant in practice, especially when the distance to the scanning objects increases.

Due to the above reasons, we focus on improving the precision and resolution of point clouds in this paper. We propose a novel mobile 3D reconstruction method where the key distinguisher and novelty compared to all the other

mobile state-of-practice (SOP) and state-of-the-art (SOTA) scanning solutions are twofold: (1) we first considerably densify sparse lidar scans by fusing them with high-resolution colour images; and (2), we use these densified scans to reconstruct a scene progressively. For the 1st part, our system utilizes recent advances in spatial Artificial Intelligence (AI) where colour images are fused with lidar depth measurements in a deep learning-based task named *image-guided depth completion*. This task aims at increasing the resolution of the relatively spare lidar scans. For the 2nd part, we utilize an ICP-like self-localization method to stitch the densified scans with each other in near real-time which results in the progressive creation of the scanned scene. This is otherwise known as odometry, which falls within a broader category of robotics/computer vision called SLAM (Simultaneous Localization and Mapping) (Cadena et al., 2016).

We build a prototypic hand-held scanner consisting of a suite of a camera and a lidar (Figure 1 on the left), implement our method as a software package for the scanner and test the whole in an indoor case study (Figure 1 on the right) to verify our method. The results show that our proposed method outperforms the current indoor scanners, such as Google Tango (Marder-Eppstein, 2016), or LOAM-based reconstruction methods (Zhang and Singh, 2014), with around 67%-increase in the precision of reconstructing flat surfaces. Also, our point cloud more reliably represents the reconstructed object by preserving more of its details.

Before we proceed to the specifics of our method and the case study, however, the subsequent section reviews the SOTA mobile reconstruction methods and the related image-guided depth completion algorithms.

Related work

State of the art in mobile mapping

The performance of general mapping methods is often compared on the KITTI odometry benchmark (Geiger et al., 2012). KITTI ranks the methods according to their average translation error between the estimated trajectory and the ground truth. According to this ranking, visual-lidar methods (Zhang and Singh, 2015) and lidar-only ones (Zhang and Singh, 2014) are the best with the former having a minimal advantage. A recent deep learning-based approach (Yang et al., 2020) has over 60% worse performance compared to the first method, and a monocular visual odometry technique (Buczko and Willert, 2017) has an error of more than twice as high as the baseline. Further subsections will focus on the caveats of these methods.

In visual odometry, a set of features such as two-dimensional intensity discontinuities (commonly called corners or key-points (Cipolla, 2017)) has been used to hand-craft the dominating feature-based approaches. Noisy corners are back-projected into such a 3D point cloud and sensor poses that maximise the probability of obtaining these corners (Engel et al., 2016). Recent approaches include (Mur-Artal and Tardos, 2017; Tang et al., 2019).

Lidar odometry methods minimise the geometric error of subsequent lidar range measurements, named further sweeps or scans. Subsequent sweeps are matched to a progressively built 3D model consisting of previously registered scans. The methods vary across geometric features extracted from scans and the way they are sampled. Some recent works include (Deschaud, 2018; Shan and Englot, 2018; Zhang and Singh, 2014).

Ultimately, there are mixed-commodity mapping approaches combining visual and lidar odometries. They bring a challenge not observed in the previous 2 paragraphs: how to fuse data coming from various domains together, for example, sparse range measurements of a lidar and dense planar images. Some recent approaches designed to address this problem are (Graeter et al., 2018; Zhang and Singh, 2018).

Image-guided depth completion techniques

Described in the previous paragraph camera-lidar fusion can be viewed from a different perspective, namely, how to use a substantially denser camera to densify sparse lidar measurements. Depth densification methods belong to more exotic tasks in computer vision dominated recently by deep learning. Depth densification can be further split by what the specific input to their algorithms is. It is of particular interest to investigate methods which take both images and lidar scans as input because they might synthesize with the camera-lidar odometry methods described in the previous paragraph. Such methods often fall within the term *image-guided depth completion* (IGDC).

IGDC uses substantially denser colour images to densify sparse lidar measurements (Ma et al., 2019). An algorithm predicts how to complete the gaps between sparse lidar measurements guided by images from the camera. More specifically, neural networks regress the depth of each pixel in the depth map. The difference between the sparse lidar depth maps and dense counterparts can be seen in Figure 3 c) and e), or d) and j). Such methods usually compare their performance on 2 datasets: (1) indoor NYU Depth Dataset by Silberman et al. (2012) and (2) outdoor KITTI Depth Completion dataset by Uhrig et al. (2017), both of which provide registered pairs of sparse depth maps and images along with the corresponding dense depth maps as ground truth. Since our paper deals with indoor scenes, only the performance on the first dataset is of interest. According to Imran et al. (2021), the top 3 indoor IGDC methods are those by Imran et al., (2021), Park et al. (2020) and Xiong et al. (2020).

The limitation of IGDC methods is that they only reconstruct dense scenes from single points of view. Also, due to their reliance on end-to-end deep learning

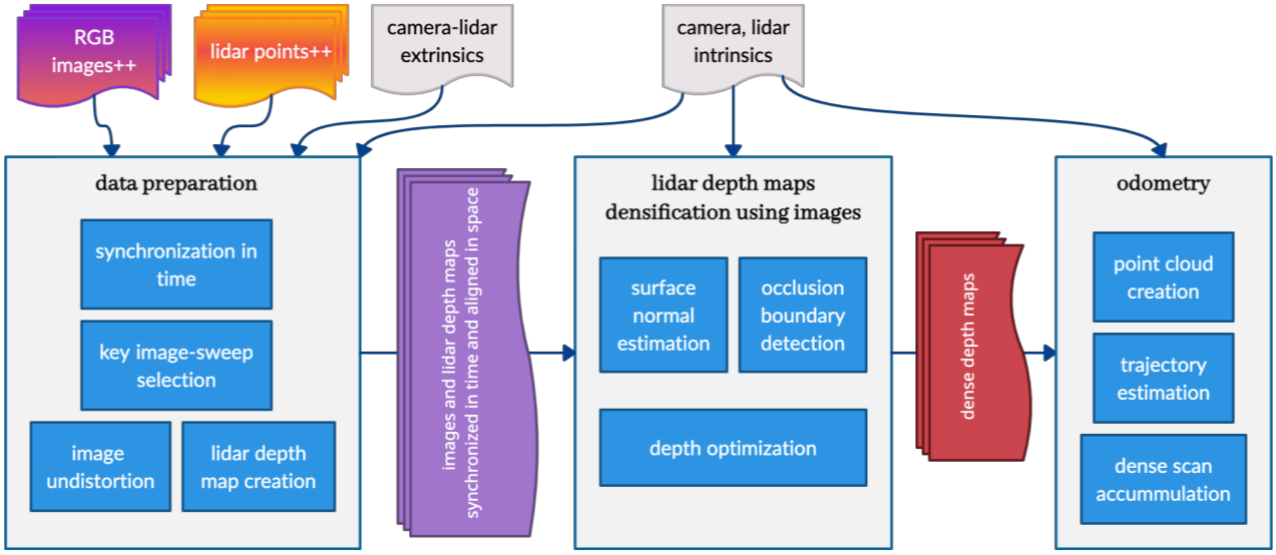

Figure 2: High-level overview of the proposed method (read from top left to right).

techniques, they are very much sensor dependent. This means that they cannot be used "out-of-the-box" with custom cameras and lidars.

Objectives of this paper

The mentioned odometry methods focus on better estimation of trajectory which might then result in improving the correctness of the overall point cloud. The IGDC techniques, on the other hand, try to improve the precision and density of point clouds taken from single points of view. The goal of this paper is to combine these two methods in the hope of improving the precision and density of progressively built scans in an odometric manner.

System overview

Our system aims at achieving precise and dense scans from a mobile scanner by processing the collected sequences of time-synchronized and spatially registered pairs of RGB images and lidar scans. Unlike the other SOP and SOTA odometry solutions, our method: (1) first considerably densifies sparse lidar scans using RGB images; and (2) uses these densified scans to reconstruct a scene progressively. For the 1st part, we utilize deep learning-based image-guided depth completion which uses images to intelligently increase the resolution of the lidar scans. For the 2nd part, we utilize an ICP-like self-localization method to stitch the densified scans to each other which results in the progressive building of the transversed scene. The combination of these two is the key novelty and distinguisher from all the other methods. A high-level description of our solution is presented in Figure 2. It consists of 3 major steps, each of which is described in detail in the following subsections. Figure 3 shows partial outputs of the key processes.

Data preparation

The input (images and lidar sweeps) to the following densification method must be aligned in space and synchronized in time so that the corresponding images and lidar sweeps represent the same scene from the same point of view captured at nearly the same time.

The camera and lidar need to be synchronized in time because they are being moved while scanning, and the sensors send images and lidar sweeps at different frequencies. For example, our camera shoots images at a frequency of around 50 Hz while lidar sends its scans at a frequency of around 10 Hz. These frequencies also slightly fluctuate in time, hence a need to select such an image and a lidar sweep that are maximally consistent with each other at a given moment.

We achieve that by a standard synchronisation policy developed by ROS (n.d.) that selects pairs of images and sweeps such that their difference in timestamps is minimal and neither of them was synchronised before. This will ensure that they represent the same part of the scene and their uniqueness among the other pairs. The timestamped streams of images and lidar sweeps are symbolized by the "++" signs at the top left part of Figure 2.

There is a lot of redundancy in such streams of data given the synchronised signals will be of frequency close to that of the lidar (10 Hz) and a person using the device will be moving with it at around walking speed. Hence, a necessity to discard the already "seen" part of a scene, and, at the same time, to capture images and sweeps that partially overlap with already reconstructed parts of the scene so that they can be registered. Our system is constantly monitoring if the user has moved or rotated the scanner by certain thresholds based on an odometry method described in the following sections. For example, the system picks a new key pair of an image and a scan

when the user holding the device walks 1 meter in any direction or rotates it by 20 degrees. Such key pairs can be seen in Figure 3 in the first 2 rows.

Image undistortion follows a procedure described by OpenCV (n.d.). It corrects images in such a way that all edges that are straight in reality, are also straight in the images. To use this method, however, the camera must be intrinsically calibrated.

Intrinsic calibration of a sensor aims at estimating its internal parameters so that its output is useable and meaningful. We assume that our camera follows a pinhole camera model which includes its intrinsic matrix and lens distortion coefficients. These parameters are estimated using a standard calibration procedure described by OpenCV (n.d.). The camera calibration is also needed to project lidar points onto images which is the last process in our data preparation step.

Corresponding images and lidar scans need to be overlaid/registered onto each other so that the input to the subsequent densification process is consistent in space. We achieve this by projecting lidar points onto image space which creates sparse depth maps with pixels containing the distances from the origin of the camera to the given 3D points. Camera calibration described in the previous paragraph provides a way to project 3D points onto the image space. However, this projection happens in the camera coordinate system which is different from the lidar's coordinate system. We estimate a Euclidean transformation (rigid-body translation and rotation) between the sensors using a SOTA camera-lidar extrinsic calibration method described by (Huang and Grizzle, 2020) to provide a mapping between these two coordinate systems. On a side note, we obtain 3D points of the lidar in Cartesian coordinates based on raw sensor measurements following the manufacturer's manual and incorporating the default lidar intrinsic parameters. Thanks to this, the stream of sweeps coming from the lidar is spatially correct up to the precision level declared by the manufacturer. Also, we assume that the motion distortion caused by holding the scanner at a walking speed while scanning is negligible.

Lidar depth maps densification using images

The time-synchronized and spatially registered key images and sparse lidar depth maps (Figure 3 a, c and b, d) are the input to the first main step in our pipeline – the densification method whose presence is the key distinguisher from all the other SOP and SOTA mobile scanning methods.

We have decided to integrate the image-guided depth completion algorithm by Zhang and Funkhouser (2018) because the way it was designed and trained allows us to use this method with our camera and lidar without significant retraining. Also, we considered such factors as open-source availability and the hardware the method was trained on in our selection process. The algorithm first

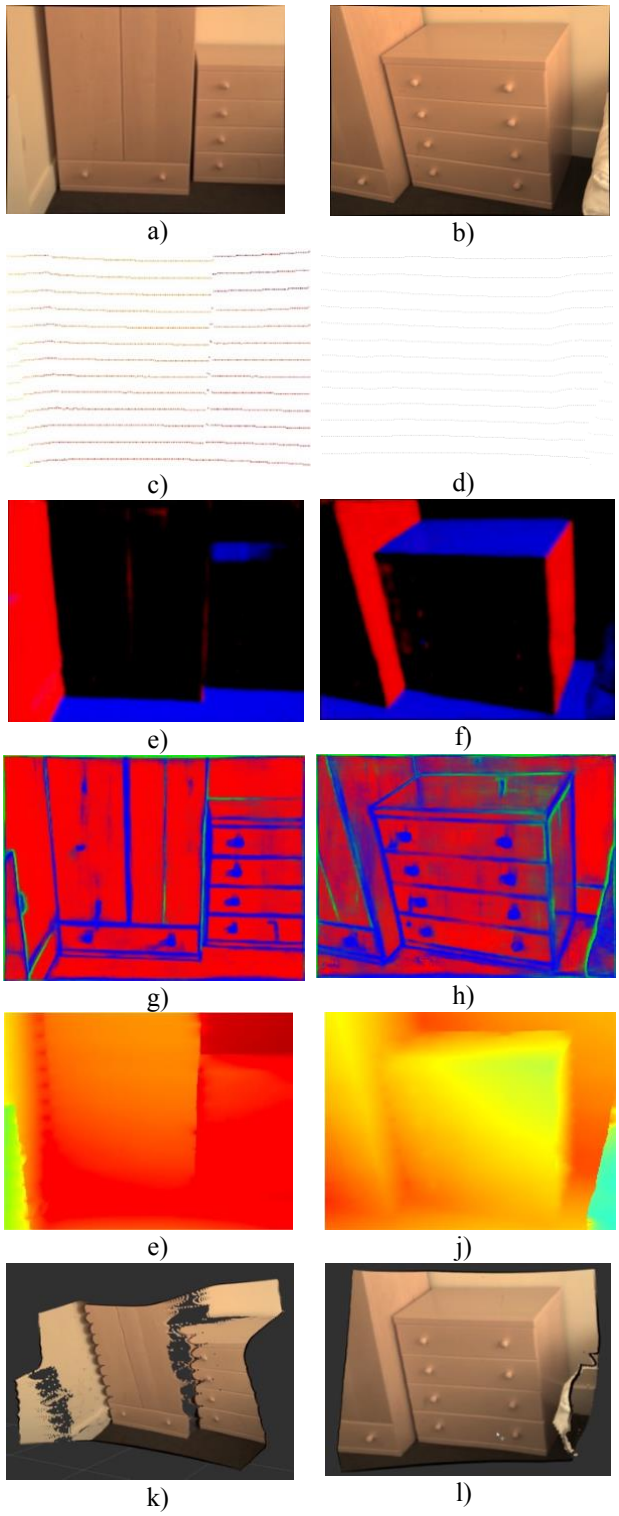

Figure 3: Partial outputs of reconstruction processes outlined in Figure 2. Pairs a) c) and b) d) show time-synchronized undistorted camera images and sparse lidar depth maps; e) and f) portray predicted normal vectors of surfaces; g) and h) represent predicted occlusion boundaries; e) and j) portray regressed dense depth maps; k) and l) show the back-projected depth maps in the form of point clouds

predicts surface normals as well as occlusion boundaries using deep neural networks (the processes in the middle

part of Figure 2). For example, the former is colour-coded in Figure 3 e) and f), with black normals pointing towards the camera and those pointing upwards marked in blue. The latter are shown in Figure 3 g) and h) and are split into 3 classes: i) a boundary (marked in red); ii) an occlusion boundary as depth discontinuity (colour-coded as green); and iii) surface normal discontinuity (blue). Next, the surface normals are weakened at those places where occlusion boundaries have been inferred, and, these two information maps are used to compute a dense depth map (Figure 3 e) and f)) using a standard optimization method (bottom middle of Figure 2) as described by Zhang and Funkhouser (2018).

Odometry

We back-project the individual pixels from the predicted dense depth maps and associated colour images to RGB-XYZ points in 3D space in the camera-centred coordinate system using the camera intrinsic matrix, thus obtaining a colour point cloud (the top process in odometry in Figure 2 with examples in Figure 3 k) and l)). We do that using standard trigonometric relationships between the camera-centred spherical coordinate system and its Cartesian equivalent.

All the processes up to this point described reconstruction relative to the camera's internal reference frame which moves as the device does. However, the final 3D model should be in a static frame, named here *scene* (as shown in the bottom-left part of Figure 4a). The process of the sensor-to-scene translation is achieved via odometry which "glues" the single camera-centred dense point clouds.

The trajectory of the device is estimated in 2 steps: we first compute simple sweep-to-sweep odometry which is then refined with dense-scan-to-dense-scene odometry. For the former, a lidar-only ICP-based method is used. Its result can be seen in Figure 4a where the trajectory (marked in purple) was not estimated good enough to allow for proper registration of the two densified scans. Therefore, this trajectory is corrected by finely registering the dense point clouds to the already accumulated partial scene as shown in Figure 4b.

Finally, the progressive reconstruction can take place by accumulating the scene-referenced point clouds as presented in Figure 4c. This last step closes the proposed framework.

Results

This section provides results on the performance of the proposed system in an indoor environment. We reconstruct a cupboard with a chest of drawers using the proposed method and compare it to the reconstruction from a Google Tango.

We reconstructed the scene around 10 times using different settings which Google Tango offers. The most precise reconstruction was selected for comparison here with our method. It must be noticed, however, that the Google scanner had difficulties reconstructing the scene. The reason for this might be the fact that the scene has very few distinct features with the walls and the pieces of furniture blending and being quite homogenous. We hypothesize that the underlying technology behind Google Tango – visual odometry – struggles in cases of featureless areas. This fact is also pointed out by Engel et al., (2016).

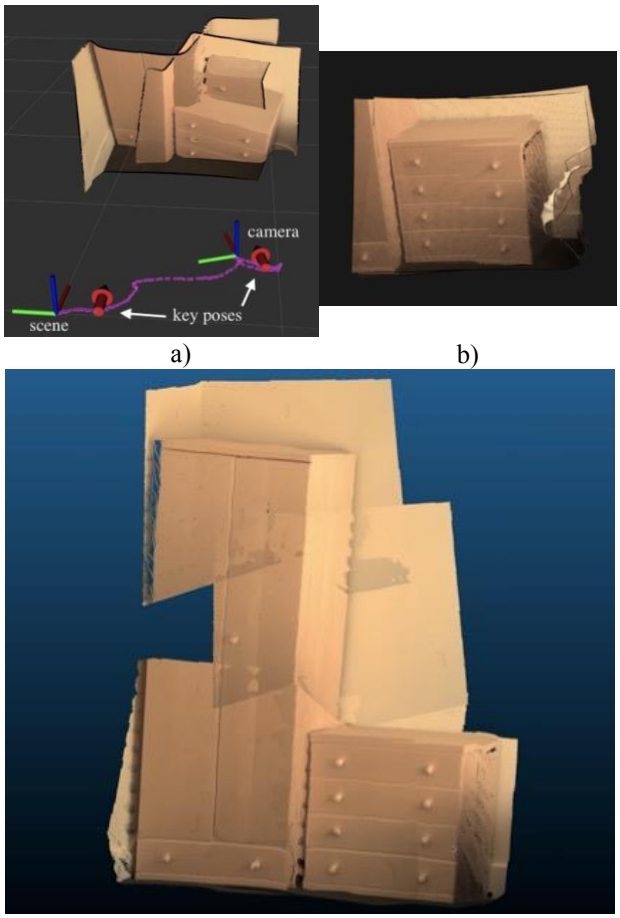

Figure 4: Odometric reconstruction of an indoor scene using our method.

Qualitative

The qualitative comparison between our reconstruction and that of Google Tango can be seen in Figure 5. The green and red circles indicate regions that are of particular interest for the comparison. Judging by images a) and b) in Figure 5, our reconstruction (on the left) seems more precise than that one on the right. Moreover, such details as the handles of the lockers are more realistically reconstructed. Also, the reconstructed surface of the front door in the cupboard is indeed flatter than that on the right. Besides, the side of the drawer is more uniformly and more densly mapped than that of Google Tango which contains voids. Moving on to images c) and d) in Figure 5, the gap between the cupboard and the drawer seems better pronounced in our scan on the left than in that on

the right. In reality, there was around a 2cm gap between these 2 pieces of furniture which is more realistically reconstructed by our method. Pictures e) and f) show that our point cloud is crispier when it comes to reconstructing the handles of the drawer.

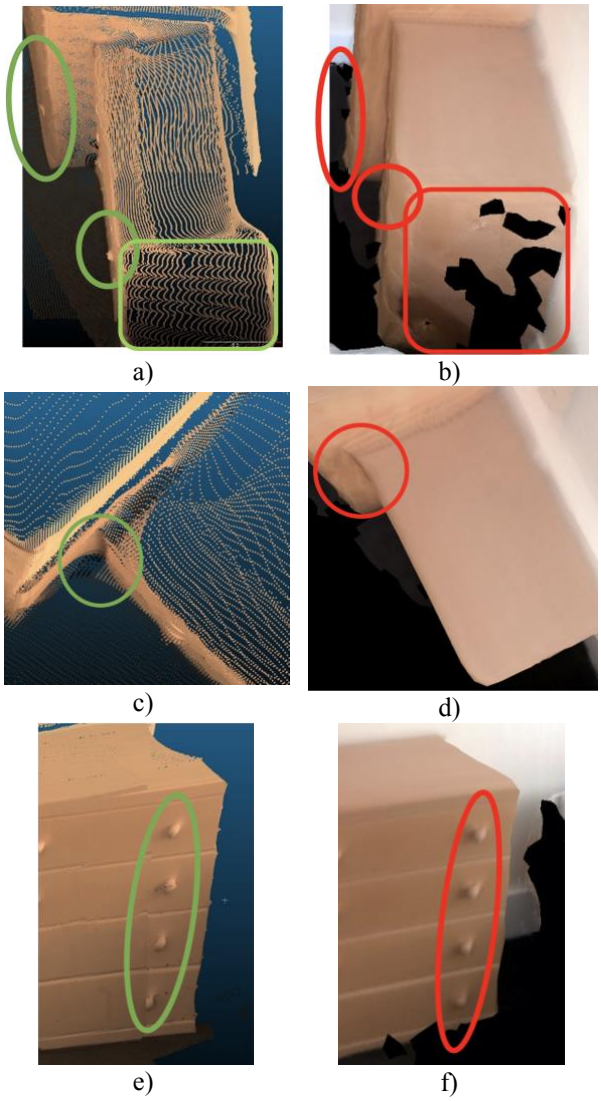

Figure 5: Qualitative comparison of reconstructions by our method (on the left) and by Google Tango (on the right).

On a side note, the final precision of the scene-referenced scans can be influenced by 2 factors: i) precision of our integrated image-guided depth completion method; and ii) the precision of registering the neighbouring dense scans in an odometry-like manner. In our experiment, we measure the combined impact of both. The attentive reader might also notice small undesirable artefacts in the point cloud produced by our method. For example picture k) in Figure 3 shows that some edges contain small indentations (particular on the cupboard) and smarming effects (on the radiator). The former are likely caused by inaccuracies in the camera-lidar calibration, subsequent projection of lidar scans onto image space, and finally, the densification method which is confused between not precisely overlaid image-realted feature maps and lidar depth maps. For the former, individual pixels in images cannot be unambiguously assigned to either the edge of an object or its background due to the nature of the image formation process in the camera. In other words, the boundaries will be blurred after zooming in to edges even when the image seems sharp.

Quantitative

We compare the precision in reconstructing flat surfaces to qualitatively evaluate our method. Specifically, a part of a flat surface has been extracted from the whole reconstructed scene as shown in Figure 6, and its precision was compared to the corresponding patch from the point cloud obtained by the LOAM odometry method which is de facto the best SOTA lidar-only 3D mapping method according to the KITTI odometry benchmark.

We simply fit planes to the extracted planar patches of the point clouds and compute the Root Mean Squared Errors (RMSE) after fitting the planes. We execute this computation in Cloud Compare and the results are presented in Table 1. We multiply the RMSE by 2 in the table to relate it to 2 standard deviations, which in turn, cover 95% of the distances between the points in the extracted patches and the fit planes assuming that the distribution of the points is Gaussian.

Quantitative results seem to confirm that our method yields more precise reconstruction overall. The noise level on a flat surface is decreased from 29.8 to 9.8 mm which results in a 67% reduction compared to LOAM. We could not compute the noise level of the flat surface in the scans by Google Tango though because the feature allowing to download the scans did not seem to work properly.

Discussion & conclusion

In this paper, we provided a novel 3D reconstruction pipeline for hand-held scanners comprising an RGB camera and a lidar. Our contribution lies in the combination of camera-lidar fusion and odometry, where the resulting dense and precise scans are used for progressive mapping in an odometry-like manner. To the best of our knowledge, the aforementioned combination makes for the first such framework in the world.

We built a prototypic scanner and tested our method in an indoor case study where 2 pieces of furniture were reconstructed. Our indoor experiment on reconstructing a cupboard with a chest of drawers shows that our pipeline has the potential to work well as it outperforms the current SOP and even SOTA lidar-based mobile methods. The results show that the proposed pipeline can outperform reconstructions by Google Tango and LOAM. The obtained scans are relatively denser and more precise with a 67% reduction in noise of reconstructed planar surfaces. Moreover, our method allowed us to capture more details

of the scene such as drawer handles and gaps between the pieces of the furniture.

Our method can also improve the estimation of the trajectory of the mobile device if that is the goal. The simple sweep-to-sweep ICP-based lidar odometry revealed that the vertical movements are not estimated even close to the true trajectory. This is because the lidar fixed horizontally to the device and held horizontally by

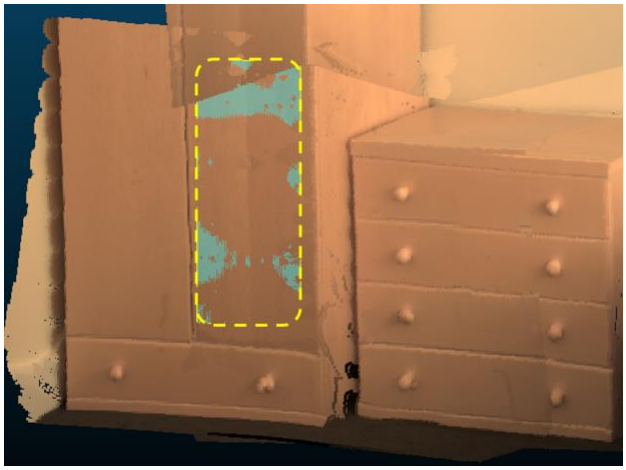

Figure 6: We measure the precision in reconstructing flat surfaces in the place marked with the yellow dashed rectangle. A fit plane there is marked in blue translucent color.

the user does not measure any changes in the vertical direction. Also, we noticed that the drift grows significantly and it makes it hard to reliably reconstruct a scene using only this simple odometric technique. Therefore, the proposed refinement step is justified. It uses the dense scans coming from the camera-lidar fusion which can be way easier registered with each other, which in turn, yields a more reliable trajectory of the mapping device.

Table 1: Comparison of precision in reconstructing flat surfaces marked in yellow in Figure 6 (smaller value is better).

Reconstruction method	2 RMSE [mm]
LOAM	29.8
Our method	9.8

From the end-user perspective, it might seem quite natural that a 3D mapping device should provide a colour point cloud. However, the current state-of-practice mapping devices do not do so by default. A nice "side-effect" of our camera-lidar fusion is that the resulting scans have colour information by default which is directly transferred from the camera.

When it comes to a broader impact of the proposed method, it can potentially unlock certain use-cases popular in the AECO industry, not accessible so far for hand-held scanners. According to band C in the accuracy band table by RICS (2014), scans of infrastructure facilities might require relative accuracy up to 10mm. Our method seems to meet this requirement, unlike other techniques.

The work done here will be a basis for further research in outdoor case studies where larger objects such as buildings or infrastructure facilities will be reconstructed.

Acknowledgements

The research leading to this paper received funding from BP, GeoSLAM, Laing O'Rourke, Topcon and Trimble. We would like to thank these companies for making our research possible. We also thank Hughes Hall College for allowing access to their grounds. We gratefully acknowledge the collaboration of all the industrial partners. Any opinions, findings, conclusions or recommendations expressed in this material are ours and do not necessarily reflect the views of the companies mentioned above.

References

BIM task group, 2013. Client guide to 3d scanning and data capturing.

Buczko, M., Willert, V., 2017. Monocular Outlier Detection for Visual Odometry, in: 2017 IEEE Intelligent Vehicles Symposium (IV). Presented at the 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 739–745. https://doi.org/10.1109/IVS.2017.7995805

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J., 2016. Past, Present, and Future of Simultaneous Localization And Mapping: Towards the Robust-Perception Age. IEEE Trans. Robot. 32, 1309–1332. https://doi.org/10.1109/TRO.2016.2624754

Cipolla, R., 2017. Computer Vision, Handout 2: Feature Extraction and Description.

Deschaud, J.-E., 2018. IMLS-SLAM: scan-to-model matching based on 3D data. ArXiv180208633 Cs.

Engel, J., Koltun, V., Cremers, D., 2016. Direct Sparse Odometry. ArXiv160702565 Cs.

FARO, 2016. Technical Specification Sheet for the Focus3D X 30/130/330 and X 130/330 HDR [WWW Document]. FARO® Knowl. Base. URL https://knowledge.faro.com/Hardware/3D_Scanners/Focus/Technical_Specification_Sheet_for_the_Focus3D_X_30-130-330_and_X_130-330_HDR (accessed 1.19.21).

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074

GeoSLAM, n.d. GeoSLAM ZEB Horzion product specification [WWW Document]. GeoSLAM. URL https://geoslam.com/solutions/zeb-horizon/ (accessed 1.19.21).

Graeter, J., Wilczynski, A., Lauer, M., 2018. LIMO: Lidar-Monocular Visual Odometry.

Huang, J.-K., Grizzle, J.W., 2020. Improvements to Target-Based 3D LiDAR to Camera Calibration. IEEE Access 8, 134101–134110. https://doi.org/10.1109/ACCESS.2020.3010734

Imran, S., Liu, X., Morris, D., 2021. Depth Completion with Twin Surface Extrapolation at Occlusion Boundaries. ArXiv210402253 Cs.

Kalyan, T.S., Zadeh, P.A., Staub-French, S., Froese, T.M., 2016. Construction Quality Assessment Using 3D as-built Models Generated with Project Tango. Procedia Eng., ICSDEC 2016 – Integrating Data Science, Construction and Sustainability 145, 1416–1423. https://doi.org/10.1016/j.proeng.2016.04.178

Ma, F., Cavalheiro, G.V., Karaman, S., 2019. Self-Supervised Sparse-to-Dense: Self-Supervised Depth Completion from LiDAR and Monocular Camera, in: 2019 International Conference on Robotics and Automation (ICRA). Presented at the 2019 International Conference on Robotics and Automation (ICRA), pp. 3288–3295. https://doi.org/10.1109/ICRA.2019.8793637

Marder-Eppstein, E., 2016. Project Tango, in: ACM SIGGRAPH 2016 Real-Time Live!, SIGGRAPH '16. Association for Computing Machinery, New York, NY, USA, p. 25. https://doi.org/10.1145/2933540.2933550

Mur-Artal, R., Tardos, J.D., 2017. ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras. IEEE Trans. Robot. 33, 1255–1262. https://doi.org/10.1109/TRO.2017.2705103

NavVis, BIM+, Geo Week News, Lidar News, GIM International, 2021. State of Mobile Mapping: Survey.

OpenCV [WWW Document], n.d. . OpenCV. URL https://opencv.org/ (accessed 12.8.21).

Park, J., Joo, K., Hu, Z., Liu, C.-K., Kweon, I.S., 2020. Non-Local Spatial Propagation Network for Depth Completion. ArXiv200710042 Cs.

RICS, 2014. Measured surveys of land, buildings and utilities. RICS guidance note, global. 3rd edition.

ROS, n.d. message_filters / ApproximateTime [WWW Document]. URL http://wiki.ros.org/message_filters/ApproximateTime

Shan, T., Englot, B., 2018. LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain. 2018 IEEERSJ Int. Conf. Intell. Robots Syst. IROS. https://doi.org/10.1109/IROS.2018.8594299

Silberman, N., Hoiem, D., Kohli, P., Fergus, R., 2012. Indoor Segmentation and Support Inference from RGBD Images. Presented at the ECCV.

Smith, S., n.d. Stencil 2 for Rapid, Long-Range Mobile Mapping. KAARTA. URL https://www.kaarta.com/products/stencil-2-for-rapid-long-range-mobile-mapping/ (accessed 1.19.21).

Tang, J., Ambrus, R., Guizilini, V., Pillai, S., Kim, H., Gaidon, A., 2019. Self-Supervised 3D Keypoint Learning for Ego-motion Estimation. ArXiv191203426 Cs.

Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T., Geiger, A., 2017. Sparsity Invariant CNNs, in: 2017 International Conference on 3D Vision (3DV). Presented at the 2017 International Conference on 3D Vision (3DV), IEEE, Qingdao, pp. 11–20. https://doi.org/10.1109/3DV.2017.00012

USIBD, 2016. USIBD Level of Accuracy (LOA) Specification Guide.

Xiong, X., Xiong, H., Xian, K., Zhao, C., Cao, Z., Li, X., 2020. Sparse-to-Dense Depth Completion Revisited: Sampling Strategy and Graph Construction, in: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (Eds.), Computer Vision – ECCV 2020, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 682–699. https://doi.org/10.1007/978-3-030-58589-1_41

Yang, N., von Stumberg, L., Wang, R., Cremers, D., 2020. D3VO: Deep Depth, Deep Pose and Deep Uncertainty for Monocular Visual Odometry. ArXiv200301060 Cs.

Zhang, J., Singh, S., 2018. Laser-visual-inertial odometry and mapping with high robustness and low drift. J Field Robot. https://doi.org/10.1002/ROB.21809

Zhang, J., Singh, S., 2015. Visual-lidar odometry and mapping: low-drift, robust, and fast, in: 2015 IEEE International Conference on Robotics and Automation (ICRA). Presented at the 2015 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Seattle, WA, USA, pp. 2174–2181. https://doi.org/10.1109/ICRA.2015.7139486

Zhang, J., Singh, S., 2014. LOAM: Lidar Odometry and Mapping in Real-time, in: Robotics: Science and Systems. https://doi.org/10.15607/RSS.2014.X.007

Zhang, Y., Funkhouser, T., 2018. Deep Depth Completion of a Single RGB-D Image. ArXiv180309326 Cs.