Online Reconstruction of Indoor Scenes with Local Manhattan Frame Growing

Mahdi Yazdanpour, Guoliang Fan, and Weihua Sheng
School of Electrical and Computer Engineering
Oklahoma State University, Stillwater, OK 74078, USA
{mahdiy,guoliang.fan,weihua.sheng}@okstate.edu

Abstract

We propose an efficient approach for robust reconstruction of indoor scenes by taking advantage of the geometric relation between consecutive Manhattan keyframes and local pose refinement to improve the accuracy and fidelity of the reconstructed models. At the core of our framework, we have a Local Manhattan Frame Growing system, which finds the principal directions of the scene and aligns point clouds with the dominant plane, and a Local Pose Optimization, which refines the pose estimation for a specific range of frames. During the reconstruction process, we use Manhattan keyframes for a planar pre-alignment to provide a robust initialization for the final surface registration. All Manhattan keyframes are integrated using a frame-to-model scheme to create local models based on the refined camera poses. The final dense model is reconstructed by adopting a geometric registration between local segments and integrating them into a global frame. The experimental results show the effectiveness of our approach to reduce the cumulative registration error and overall geometric drift.

1. Introduction

Generating high-fidelity dense 3D models of indoor scenes from RGB-D streams has become one of the most challenging research topics in the robotics and computer vision communities. In general, researchers in the robotics community deploy Simultaneous Localization and Mapping (SLAM) framework for sparse 3D mapping using visual and depth information, and their ambition is ultimately to minimize the trajectory error [14, 23, 24, 17, 26] with less concern about the pertinent details and the quality of the generated 3D models. In contrast, the main intention of the researchers in the computer vision community is dense modeling and enhancing the quality and fidelity of the reconstructed models [22, 29, 19, 4]. One of the foremost approaches for volumetric reconstruction, which plays a significant role in the evolution of volumetric modeling is KinectFusion [20]. This approach demonstrated a real-time dense mapping and tracking system for volumetric reconstruction of small scenes. Many methods, both online [22, 23, 8, 17, 7] or offline [4, 29, 30], have been developed for 3D scene reconstruction by promoting the main concepts exposed by KinectFusion. Kinfu Large Scale [15], voxel hashing [19], and scalable surface reconstruction [2] use depth sequences to generate 3D models. Contrastingly, many methods involve both RGB and depth information to enhance the robustness of reconstructed models, including DVO SLAM [17], RGB-D mapping [14], fragment registration [4], and online indoor reconstruction system [22].

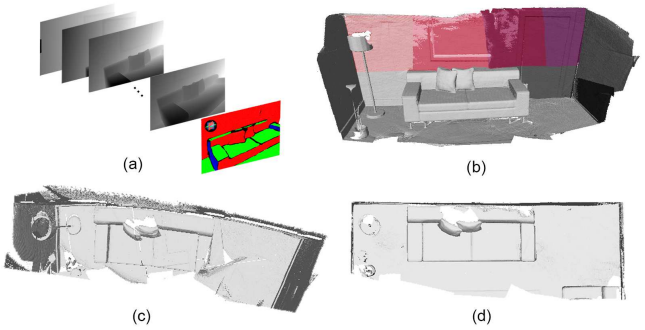

Figure 1. (a) A depth sequence with a Manhattan frame (MF), (b) Local Manhattan Frame Growing (LMFG) via planar alignment, (c) The local reconstruction without using MF, and (d) The local construction result via LMFG.

In this paper, we develop an online framework for 3D indoor scene reconstruction by incorporating only depth information to expedite the reconstruction process. We enhance surface registration by utilizing the geometric similarity cross keyframes by a Local Manhattan frame growing (LMFG) strategy that involves local pose optimization, as shown in Fig. 1). The proposed LMFG approach integrates all local segments into a unified global model and creates a consistent volumetric model. This pipeline reduces the cumulative registration error effectively and reconstructs a drift-free high-quality dense model without using visual features and explicit loop closure detection, which is also comparable with offline reconstruction methods.

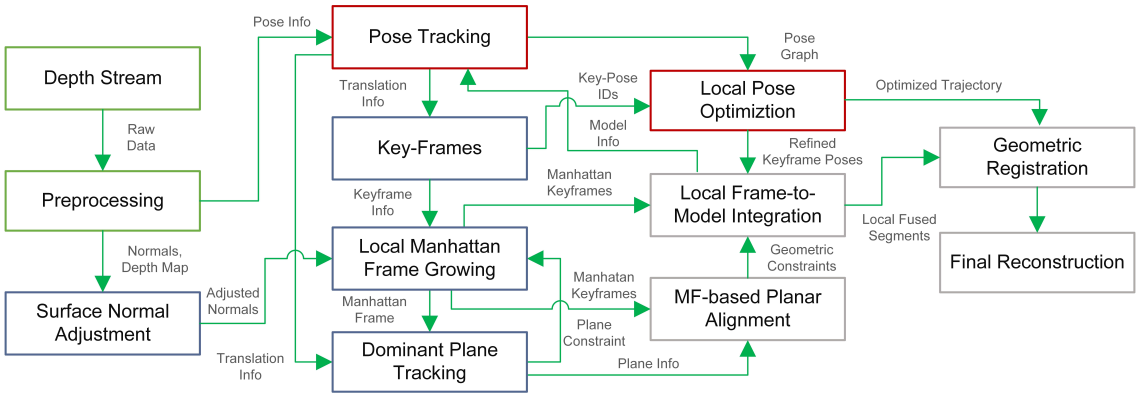

Figure 2. Overview of our framework

2. Related Work

The geometric drift in extended scale scene reconstructions is not entirely avoidable due to the accumulation of the registration error between frames and pose estimation inaccuracy. In large scale environments with complex trajectories, this odometry deflection is likely prone to error. Most of previous works on 3D mapping and dense modeling require a global consistent alignment between all frames and an accurate pose estimation, which is acquirable by a global pose graph optimization, to create drift-free 3D reconstructions. The prevalent solution to the pose optimization problem is using visual feature matching in conjunction with pose graph optimization [18] to refine the pose estimations and handle loop closures, and/or bundle adjustment [21] to minimize the reprojection error. The online methods usually work in near real-time, but they cannot provide an impressive performance in terms of quality and precision. On the other hand, the offline reconstruction methods produce more realistic dense models with superb quality and accuracy, but they may need more computational time for offline processing to access all input frames and to optimize across the entire pose trajectory, which makes them to be practically unusable for the real-time applications.

At the heart of our framework, we use a local Manhattan frame growing system alongside a local pose optimization method to generate dense 3D reconstructions. The Manhattan world (MW) assumption [6] has been used in many 3D scene reconstruction applications [27, 25, 28]. Our algorithm is different from [27, 25], where the Manhattan world assumption is used to create a non-volumetric model using both depth and color data and from [28], where a Manhattan frame (MF) estimation method is adopted to find all Manhattan keyframes to generate a volumetric model and needs a pose optimization method to refine pose estimation errors. Our method is also different from Kintinuous [24], DVO SLAM [17], SUN3D SfM [26], and RGB-D mapping [14], which involve visual odometry and use frame-to-frame tracking and matching system in their pipelines.

3. Proposed Method

We first review the MF-based volumetric reconstruction. Then we propose the new LMFM strategy, which is used to initialize planar alignment across keyframes. Afterwards, we present a local pose optimization method and a frame-to-model registration scheme to generate local fused models, followed by a geometric registration to integrate all segments into a global dense reconstruction.

3.1. Overview

The pipeline of the overall system is shown in Fig. 2. The key idea is benefiting from the geometric information of the scene to enhance registration accuracy by using local Manhattan frame growing and local odometry optimization. To drive more reliable surface registration between depth frames, we have proposed a local optimization for surface geometry, which is used to create regional fused models. After the normal computation, we utilize a surface normal adjustment to refine the surface orientation, which produces a highly persistent distribution of the point normals and facilitates the process of Manhattan frame estimation. The next step is local Manhattan frame growing by expanding the first estimated Manhattan frame along the dominant plane valid in the current scene. This extending strategy is faster than regular MF estimation method used by [28] for volumetric reconstruction and helps to use the local geometry in the reconstruction process. Instead of creating pose graph for whole scene and using a global pose optimization, we optimize the local odometry for each region, which reduces the overall geometric drift error. The estimated Manhattan keyframes by LMFG are used for the planar pre-alignments, which provide a reliable initialization for the final surface registration. The last step is integrating all local models into a global volumetric representation according to the refined camera trajectory using a frame-to-model scheme. We present all major steps in detail below.

3.2. Local Manhattan Frame Growing

The geometric depiction of indoor scenes using MW assumption has different employments in vision application like 3D modeling, semantic segmentation, and scene understanding. This assumption states that all objects in a scene are aligned with one of three orthogonal directions. Determining the MF of the scene based on this assumption simplifies the representation of the objects and planes and provides the accurate and reliable geometric properties, which can be used in the reconstruction process. The Manhattan frame estimation methods can be classified into two groups. The methods in the first group are RGB-based, which rely on extracting lines, edges, and orthogonal vanishing points from RGB images [9, 5]. The approaches that belong to the second group are RGB-D-based, which use the 3D perspective information like surface normals computed from point clouds or depth frames [10, 16].

In our work, we proposed a local Manhattan frame growing framework based on the MF estimation method released by [10] to extend the first estimated MF along each dominant surface plane in the scene. The main idea is to find the best 3D rotation matrix to transform original normals and align them with at least one of the main coordinate directions in the scene. In an ideal Manhattan scene with Manhattan elements, which are mostly aligned to the principal directions, the estimation of MF is not very challenging. However, due to noise, depth discontinuities, and error in depth measurement or normals computation, finding an ideal MF can be an arduous process. In our pipeline, we proposed a surface normal adjustment to flip the normal vectors towards the sensor center, which results in a more consistent normal distribution and facilitates the MF estimation process.

The main intention of our proposed local Manhattan frame growing system is to find the MF of a scene and extend it to the next frames along a most visible surface plane with the same idea of converting normal vectors into the sparsest set of directions as follows:

$$LMFG = \min_{R,X} \quad \frac{1}{2}||(X_{i-1} - (R(N_i) - X_i))||_F^2 + \lambda||X_i||_{1,1} \ , \tag{1}$$

where $N_i \in \mathbb{R}^{3\times m}$ is the matrix of the original surface normals, $R \in SO(3)$ is the rotation matrix to rotate the surface normals to be aligned to principal directions, X_i is the matrix of the rotated normal vectors after applying R to the set of surface normal vectors, and X_{i-1} is a sparse matrix of the previous rotated normals. The second term in the objective function is used as a regularizer, which is the sum of the ℓ_1 norms of the columns in X_i matrix and helps to avoid the overfitting problem and to achieve the higher sparseness. $||X||_{p,q}$ shows the $\ell_{p,q}$ matrix norm of X, and λ is the trade-off for sparsity. We estimate the MF of the first frame and find a desirable rotation matrix R to apply to the original surface normals N and aligned them with dominant coordinate. This calculated rotation matrix will be used as an initial guess for the normals of the next coming frame N_i and transform them to update a new rotated normals X_i, which should be continuance of X_{i-1}, where the camera translation is along the dominant direction. We track the dominant plane and continue updating R and X_i in an iterative manner to find the best rotation matrix and minimize the distance between the rotated surface normals of the incoming frame and the rotated normals of the current Manhattan frame. The local minimum of this non-convex optimization is attainable via alternating optimization, where the solution for variables R and X_i is updated, while the other one is kept fixed.

3.3. Local Pose Optimization

The importance of the pose graph optimization and map correction for minimizing the overall odometry drift in large scale scene reconstruction has been recognized in many SLAM systems and dense modeling approaches. Performing a global pose optimization for all frames clearly increases the computational cost, since a full pose graph has to be optimized and loop closure constraints should be tracked incessantly. In this work, we therefore employ a local pose graph optimization, which provides more local accuracy and consistency to optimize the keyframe poses. The frame-to-model tracking system returns locally more accurate pose estimates than frame-to-frame trackers. The local pose graph is created for Manhattan key-frames and for each new incoming key-frame, a new node will be added to the graph and a new relative transformation will be computed. These relative transformations between Manhattan key-frames provide odometric constraints. We incrementally optimize the pose estimations through minimizing the following general cost function [12, 18]:

$$E(x) = \sum_{i,j} e_{ij}^T \Omega_{ij} e_{ij}, \tag{2}$$

where e_{ij} is a function that computes the difference between the real measurement and the expected measurement of the pose and Ω_{ij} represents uncertainty, which is the inverse square root of the covariance matrix of the measurement. The error function is defined as:

$$e_{ij}(x) = f(x_i, x_j) = \begin{bmatrix} R_i^T(p_j - p_i) - \hat{p}_{ij} \\ \psi_j - \psi_i - \hat{\psi}_{ij} \end{bmatrix}, \tag{3}$$

where $x_i = [p_i^T, \psi_i]^T$ shows the camera poses, $p_i \in \mathbb{R}^2$ and $\psi_i \in [-\pi, \pi)$ are the position and orientation of the i^{th} pose respectively, and R_i denotes the rotation matrix of ψ_i. The relative position and orientation measurements are computed by $R_i^T(p_j - p_i)$ and $(\psi_j - \psi_i)$. The maximum likelihood can be obtained by minimizing the residual errors

as follows:

$$(p^*, \psi^*) = \underset{p,\psi}{\operatorname{argmin}} \sum_{i,j} ||R_i^T(p_j - p_i) - \hat{p}_{ij}||^2 + ||\psi_j - \psi_i - \hat{\psi}_{ij}||^2, \tag{4}$$

By solving the cost function, the pose drift will be corrected by minimizing the error between the real observation and the predicted observation of the poses. In our implementation, we employ `Ceres Solver` [1] for optimizing the local camera trajectories and to minimize sequential constraints between Manhattan key-frames. In addition, we have constructed a whole-scene pose graph and optimized it globally to use with MF reconstruction method [28] (MFR with PGO) in order to generate dense models of the same indoor scenes.

3.4. Local Manhattan Frame based Integration

The final reconstructed models by both visual odometry approaches with a frame-to-frame tracking and matching system, and dense modeling reconstruction methods with a frame-to-model tracking and registration system are inherently prone to accumulate the drift error. In our work, we use a frame-to-model scheme for local scene reconstruction using Manhattan keyframes relying on the local geometry of the scene. This geometric odometry helps for estimating the local camera trajectory and recognizing the dominant plane in the scene by monitoring the frames translations. Using Manhattan keyframes provides reliable geometric constraints to reduce the overall registration error and generate a more accurate model with minimal drift.

We assign pose IDs to all Manhattan keyframes and use these indices to retrieve the refined pose estimations from the local pose optimization step. These refined poses will be used in the following sequential registration. We initially use a geometric registration to effectively align dominant planes in two successive keyframes f_i^k and f_j^k. For the Manhattan frame based plane-to-plane alignment, the metric distance between the point sets on two planar surfaces is minimized by solving:

$$T_{ij} = \underset{T_{ij}}{\operatorname{argmin}} \sum_{i,j} ||T_{ij}(p_i^k) - p_j^k||^2, \tag{5}$$

where P_i and P_j are dominant planes located in two Manhattan keyframes f_i^k and f_j^k, $p_i^k \in P_i$ and $p_j^k \in P_j$ are two set of points on the dominant planes, and T_{ij} is the transformation matrix that minimizes the distance between two planes. This planar alignment reduces the computational complexity, provides a reliable and robust initialization for the surface registration, and expedites the reconstruction process. After obtaining the definitive pose estimations for key-frames, we perform a frame-to-model registration scheme similar to [20] to integrate Manhattan keyframes to the TSDF (Truncated Signed Distance Function) model. The TSDF model is represented in GPU as an array of voxels. Suppose p is the location of each voxel that contains two values, the signed distance TSDF value $\upsilon(\mathrm{p})$ and the voxel weight $w(\mathrm{p})$. To integrate the i^{th} incoming Manhattan keyframe to the reconstructed model, the value of the voxel is updated by a weighted running average as follows:

$$\upsilon_i(\mathrm{p}) = \frac{\upsilon_{i-1}(\mathrm{p}) w_{i-1}(\mathrm{p}) + \upsilon_i(\mathrm{p}) w_i(\mathrm{p})}{w_{i-1}(\mathrm{p}) + w_i(\mathrm{p})}, \tag{6}$$

and

$$w_i(\mathrm{p}) = min(w_{i-1}(\mathrm{p}) + w_i(\mathrm{p}), w_{max}), \tag{7}$$

where $w_i(\mathrm{p})$ is the weighting of the TSDF to the uncertainty of surface measurement. In our implementation, we set $w_i(\mathrm{p}) = 1$, eventuating in a simple average, and $w_{max} = 128$. The local 3D models are reconstructed for all regions using the refined pose estimations retrieved from optimized camera trajectories while a dominant plane is trackable. All key-frames are added to the reconstructed model and the regional segment is created until a significant translation occurs in the camera motion, then the system starts creating a new local model by tracking the new dominant plane and estimating Manhattan frames using local Manhattan frame growing system.

3.5. Final Model Reconstruction

The core of our pipeline is an efficient local Manhattan frame growing system which operates in unison with a local pose optimization algorithm to create local fused segments. In the last stage of this pipeline, we use a geometric registration to integrate these local segments into a global framework to reconstruct the final consistent dense model. We check two consecutive local segments to find overlapping sections and run an iterative point-to-plane surface registration [3] on each point located on these sections to align the set of points in the first model and corresponding points in the second one. The desired registration should minimize the squared metric distance between each source point and the tangent plane to the surface at its corresponding destination point defined below:

$$T_A = \underset{T_A}{\operatorname{argmin}} \sum_i ||(T_A(p_i') - q_i') \cdot N_i||^2, \tag{8}$$

where $p_i' = (p_{ix}', p_{iy}', p_{iz}', 1)^T$ is a point on the surface of the first local model, $q_i' = (q_{ix}', q_{iy}', q_{iz}', 1)^T$ is the corresponding point on the surface of the second local model, $N_i = (N_{ix}, N_{iy}, N_{iz}, 0)^T$ is the unit normal vector at destination point q_i', and T_A is the transformation matrix to align two point sets. After performing this iterative registration, the neighbor segments will be perfectly aligned and the final 3D dense model will be reconstructed.

4. Experimental Results

We evaluated our proposed approach, Local Manhattan Frame Growing (LMFG), on the augmented ICL-NUIM dataset provided by [4]. The original dataset was released by [13] for benchmarking RGB-D visual odometry evaluation, 3D reconstruction and SLAM systems. This dataset includes the synthetic models of two different indoor scenes, a living room and an office.

Our method is used to generate four dense 3D models from depth streams of the augmented ICL-NUIM dataset and is evaluated against Kintinuous [24], DVO SLAM [17], SUN3D SfM [26], Manhattan frame (MF) reconstruction [28], and offline robust reconstruction [4] methods. We have also implemented a reduced version of our proposed approach by combining the MF reconstruction method [28] with a global pose graph optimization (MFR with PGO) to generate dense models by integrating the Manhattan key-frames obtained from the MF estimation framework into a volumetric 3D model using globally refined pose estimations. To evaluate the surface reconstruction accuracy, we have used the `CloudCompare` [11] tool to compute the mean distance of our generated models to the ground-truth surfaces. To provide evaluation for the living room scene, we have used the ground-truth surface provided by [13], and for the evaluation of the office scene, we have used the dense point-based surface model provided by [4].

4.1. Quantitative Comparison

For the quantitative comparison, we have computed the mean distance of the generated models by LMFG, MFR with PGO, and MF reconstruction methods to the ground-truth and compared the errors with the values released by [4] for Kintinuous, DVO SLAM, SUN3D SfM, and offline robust reconstruction systems, as shown in Table 1. It is evident that, LMFG approach outperforms Kintinuous, DVO SLAM, SUN3D SfM, and MF reconstruction frameworks. This comparison confirms that our approach reduces the average mean distance by factors of 3.2 relative to Kintinuous, 2.4 relative to DVO SLAM, 2 relative to SUN3D SfM, 2.4 relative to MF Reconstruction, and 1.4 relative to MFR with PGO. We have also compared our performance with offline robust reconstruction approach [4]. The dense models created by our framework are comparable with this offline pipeline, both quantitatively and qualitatively.

4.2. Qualitative Comparison

The reconstructed models, shown in Fig. 4 show the robustness of our approach for volumetric reconstruction using depth streams and local Manhattan frame growing system. Our proposed approach relies on the local geometric structure of the scene for surface reconstruction. This powerful and precise characteristic helps to preserve the geometry and reconstruct aligned dominant planes of the scene like walls and floor with a high accuracy and precision and to mend holes, gapes, and discontinuities, which are conspicuous in the generated models by other approaches, as shown in Fig. 3.

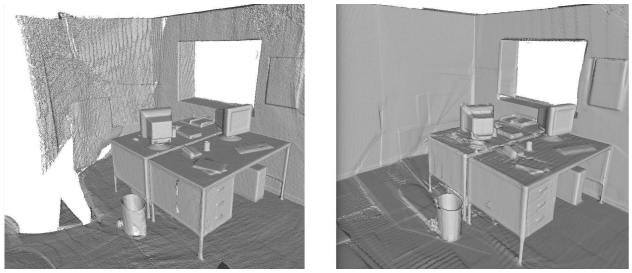

Figure 3. The reconstructions (office 1) generated by offline robust reconstruction (Left) and our approach (Right). Our method preserves the local geometric structure of the planar surfaces.

5. Conclusion

We presented an efficient approach that provides a reliable registration between depth frames to generate robust 3D dense reconstructions of indoor environments. We take advantage of the local Manhattan frame growing system and local pose optimization to reduce the accumulation of the registration error and to increase the fidelity and accuracy of the final reconstructions. The Manhattan frames are extended along the main dominant axis in the scene according to the first identified MF until a significant translation happens in the camera trajectory. These key Manhattan frames are used for a planar pre-alignment, which provide a reliable initialization to facilitate the surface registration. At the same time, the estimated poses for a local region are optimized incrementally and used to create a local model of the scene which is restricted to the dominant plane. Finally, the whole-scene reconstructed model is created by integrating all local models into a global framework.

The experimental results on the augmented ICL-NUIM dataset demonstrate the advantage of our proposed approach to reduce the cumulative registration error and overall geometric drift. Compared with Kintinuous, DVO SLAM, SUN3D SfM, and MF reconstruction, our method is more accurate and reliable for dense modeling and it has a remarkable agility in comparison to the offline robust construction method while providing the same level of quality.

Acknowledgment

This work is supported in part by the US National Science Foundation (NSF) Grant IIS-1427345, the US National Institutes of Health (NIH) Grant R15 AG061833 and the Oklahoma Center for the Advancement of Science and Technology (OCAST) Health Research Grant HR18-069.

Table 1. Mean distance of reconstructed models to the ground-truth surface (in meters). All methods are online and the offline Robust Reconstruction results, which has the best performance are shown as a reference.

Dataset	Kintinuous [24]	DVO SLAM [17]	SUN3D SfM [26]	MF Reconstruction [28]	MFR with PGO	Proposed LMFG	Offline Robust Reconstruction [4]
Living Room1	0.22	0.21	0.09	0.11	0.07	**0.06**	**0.04**
Living Room2	0.14	0.06	0.07	0.09	0.07	**0.07**	**0.07**
Office1	0.13	0.11	0.13	0.12	0.04	**0.03**	**0.03**
Office2	0.13	0.10	0.09	0.17	0.06	**0.05**	**0.04**
Average	0.16	0.12	0.10	0.12	0.07	**0.05**	**0.05**

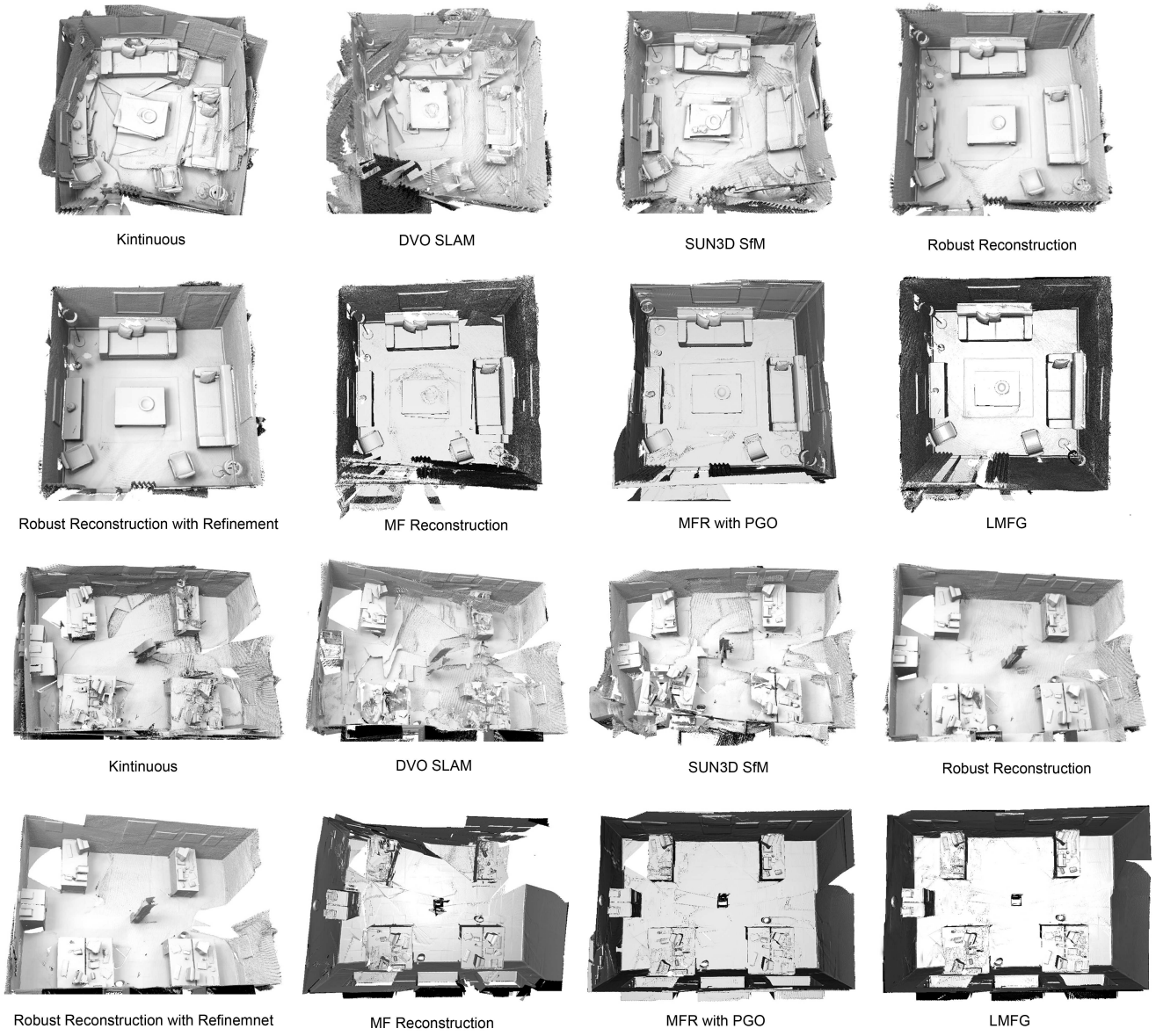

Figure 4. Reconstructed models of Living Room 1 (above) and Office 1 (below), by Kintinuous [24], DVO SLAM [17], SUN3D SfM [26], Offline Robust Reconstruction without and with an optional refinement [4], Manhattan Frame (MF) Reconstruction (without pose optimization) [28], Manhattan Frame Reconstruction (MFR) with global pose optimization, and Local Manhattan Frame Growing (LMFG).

References

[1] Sameer Agarwal, Keir Mierle, and Others. Ceres solver. http://ceres-solver.org.

[2] Jiawen Chen, Dennis Bautembach, and Shahram Izadi. Scalable real-time volumetric surface reconstruction. *ACM TOG*, 32, 2013.

[3] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10:145–155, 1992.

[4] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *Proc. CVPR*, 2015.

[5] Wongun Choi, Yu-Wei Chao, Caroline Pantofaru, and Silvio Savarese. Understanding indoor scenes using 3D geometric phrases. In *Proc. CVPR*, 2013.

[6] James M. Coughlan and A. L. Yuille. Manhattan world: Compass direction from a single image by Bayesian inference. In *Proc. ICCV*, 1999.

[7] Angela Dai, Matthias Nießner, Michael Zollöfer, Shahram Izadi, and Christian Theobalt. BundleFusion: Real-time globally consistent 3D reconstruction using on-the-fly surface re-integration. *ACM TOG*, 36, 2017.

[8] Wei Dong, Qiuyuan Wang, Xin Wang, and Hongbin Zha. PSDF fusion: Probabilistic signed distance function for on-the-fly 3D data fusion and scene reconstruction. In *Proc. ECCV*, 2018.

[9] Yasutaka Furukawa, Brian Curless, Steven M. Seitz, and Richard Szeliski. Manhattan-world stereo. In *Proc. CVPR*, 2009.

[10] Bernard Ghanem, Ali Thabet, Juan Carlos Niebles, and Fabian Caba Heilborn. Robust Manhattan frame estimation from a single RGB-D image. In *Proc. CVPR*, 2015.

[11] Daniel Girardeau-Montaut. CloudCompare: 3D point cloud and mesh processing software open source project. http://cloudcompare.org, 2015.

[12] Giorgio Grisetti, Rainer Kümmerle, Cyrill Stachniss, and Wolfram Burgard. A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine*, 2(4):31–43, 2010.

[13] Ankur Handa, Thomas Whelan, John McDonald, and Andrew J. Davison. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In *Proc. ICRA*, 2014.

[14] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. *Int. J. Robotics Research*, 31:647–663, 2012.

[15] Francisco Heredia and Raphael Favier. Using Kinfu large scale to generate a textured mesh. http://pointclouds.org/documentation/tutorials/using_kinfu_large_scale.php, 2012.

[16] Kyungdon Joo, Tae-Hyun Oh, Junsik Kim, and In So Kweon. Globally optimal Manhattan frame estimation in real-time. In *Proc. CVPR*, 2016.

[17] Christian Kerl, Jürgen Sturm, and Daniel Cremers. Dense visual SLAM for RGB-D cameras. In *Proc. IROS*, 2013.

[18] Rainer Kümmerle, Giorgio Grisetti, Hauke Strasdat, Kurt Konolige nd, and Wolfram Burgard. g^2o: A general framework for graph optimization. In *Proc. ICRA. IEEE*, 2011.

[19] Matthias Nießner and Michael Zollhöfer and Shahram Izadi and Marc Stamminger. Real-time 3D reconstruction at scale using voxel hashing. *ACM TOG*, 32, 2013.

[20] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. *10th IEEE International Symposium on Mixed and Augmented Reality*, pages 127–136, 2011.

[21] Bill Triggs, Philip McLauchlan, Richard Hartley, and Andrew Fitzgibbon. Bundle adjustment—a modern synthesis. *Proc. Vision Algorithms: Theory and Practice*, pages 298–372, 2000.

[22] Hao Wang, Jun Wang, and Liang Wang. Online reconstruction of indoor scenes from RGB-D streams. In *Proc. CVPR*, 2016.

[23] Thomas Whelan, Michael Kaess, Hordur Johannsson, John J. Leonard Maurice Fallon, and John McDonald. Real-time large scale dense RGB-D SLAM with volumetric fusion. *Int. J. Robotics Research*, 34:598–626, 2015.

[24] Thomas Whelan, John McDonald, Michael Kaess, Maurice Fallon, Hordur Johannsson, and John Leonard. Kintinuous: Spatially extended KinectFusion. In *Proc. RSS Workshop on RGB-D*, 2012.

[25] Dominik Wolters. Automatic 3D reconstruction of indoor Manhattan world scenes using Kinect depth data. In *Proc. GCPR*, 2014.

[26] Jianxiong Xiao, Andrew Owens, and Antonio Torralba. SUN3D: A database of big spaces reconstructed using SfM and object labels. In *Proc. ICCV*, 2013.

[27] Hiroaki Yaguchi, Yutaka Takaoka, Takashi Yamamoto, and Masayuki Inaba. A method of 3D model generation of indoor environment with Manhattan world assumption using 3D camera. In *Proc. SII*, 2013.

[28] Mahdi Yazdanpour, Guolinag Fan, and Weihua Sheng. Real-time volumetric reconstruction of Manhattan indoor scenes. In *Proc. VCIP*, 2017.

[29] Qian-Yi Zhou and Vladlen Koltun. Dense scene reconstruction with points of interest. *ACM TOG*, 32, 2013.

[30] Qian-Yi Zhou and Vladlen Koltun. Simultaneous localization and calibration: Self-calibration of consumer depth cameras. In *Proc. CVPR*, 2014.