

# 1 Related Work

Recovering the geometric model of a scene from visual and/or range sensors is a well-known problem that spans across different research domains: in the Computer Vision community it goes under the name of "Structure from Motion" (SfM), while in Computer Graphics it's referred to as "Surface Reconstruction", just to name a few examples. In Robotics, this is a relevant problem in the context of mobile robots that operate in unknown environments, where they have to build a representation of their surroundings to localize themselves in the environment and plan motion to complete their tasks. In this context, the geometric reconstruction is mainly performed by addressing the "Simultaneous Localization and Mapping" (SLAM) problem, since it allows to track the robot trajectory during the exploration and "in parallel" to build a metric representation of the environment for the above-mentioned goals. For 3D geometry, different types of representation have been proposed and, according to [2], can be categorized by the way in which they describe the objects surface.

In the majority of SLAM methods, the scene is represented by a set of sparse 3D landmarks that correspond to relevant features in the environment (e.g. edges, corners) [6, 5]. These methods are attractive because they're based on a lightweight formulation and still they're effective in accurately localizing the robot, but they provide a rather poor representation of the environment since they tend to filter out portions with less features.

A different approach is to describe the 3D geometry by means of an unstructured set of points (i.e., point clouds), which can be obtained by nowadays easily available sensors like RGB-D cameras. Opposed to landmark-based methods, there are also "direct" methods that estimate the trajectory of the robot and a 3D model directly from the intensity and/or depth values of all the image pixels [8, 4]. This type of representation allows to better infer the structure of the scene but it may contain holes in some portions of the environment due to the uneven sensor sampling density.

A reasonable choice to overcome this limitation is to introduce structured representations, such as the ones used in Computer Graphics for Surface Reconstruction [1]. These are called "boundary representations", since they define 3D objects in terms of their surface boundary. In general, b-reps can be classified as: curve-based representations (NURBS or B-Spline), mesh representations (triangulated surfaces) and implicit representations (level-set surfaces). The latter was investigated in the work of Newcombe et al. [7] and received big interest from the SLAM community thanks to its power of representation. The main drawback is the not negligible amount of memory required also in the case of small scale scenes. This limits considerably the scalability of this representation and the development of optimized techniques for large-scale environments is an hot topic in research [3, 10].

Despite all the representations mentioned so far can be considered as a good approximation to the scene geometry, they all suffer from a structural problem which is common to most robotic applications, i.e., occlusions. The natural way to overcome this problem, would be to frame the object to be reconstructed

from as many points of view as possible. Of course, for a robot, this is not always feasible and leads to incomplete reconstruction of objects surfaces. The solution proposed in the seminal work by Moreno et al. [9] is to investigate higher level object-based representations, where the map is made of objects and solid shapes as it is the case in Computer Aided Design (CAD). Of course, building such a map is an expensive procedure and requires many simplifying assumptions. Nevertheless, still this type of representation is very attractive because it naturally lends itself to represent also objects semantic and physical properties, which can be very important for the robot to execute its tasks.

## References

- [1] Matthew Berger, Andrea Tagliasacchi, Lee Seversky, Pierre Alliez, Joshua Levine, Andrei Sharf, and Claudio Silva. State of the art in surface reconstruction from point clouds. In *EUROGRAPHICS star reports*, volume 1, pages 161–185, 2014.
- [2] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016.
- [3] Olaf Kähler, Victor Adrian Prisacariu, Carl Yuheng Ren, Xin Sun, Philip Torr, and David Murray. Very high frame rate volumetric integration of depth images on mobile devices. *IEEE transactions on visualization and computer graphics*, 21(11):1241–1250, 2015.
- [4] C. Kerl, J. Sturm, and D. Cremers. Dense visual slam for rgb-d cameras. In *Proc. of the Int. Conf. on Intelligent Robot Systems (IROS)*, 2013.
- [5] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234. IEEE, 2007.
- [6] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orbslam: a versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015.
- [7] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pages 127–136. IEEE, 2011.
- [8] Richard A Newcombe, Steven J Lovegrove, and Andrew J Davison. Dtam: Dense tracking and mapping in real-time. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2320–2327. IEEE, 2011.

- [9] Renato F Salas-Moreno, Richard A Newcombe, Hauke Strasdat, Paul HJ Kelly, and Andrew J Davison. Slam++: Simultaneous localisation and mapping at the level of objects. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1352–1359, 2013.
- [10] Frank Steinbrücker, Jürgen Sturm, and Daniel Cremers. Volumetric 3d mapping in real-time on a cpu. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 2021–2028. IEEE, 2014.