Large Indoor Scene Reconstruction using Graphs of Primitives

Diego Thomas[1]; Viet-Dung Nguyen[2]; Akihiro Sugimoto[1].

National Institute of Informatics, Tokyo, Japan/JFLI, CNRS.[1]
University of Engineering and Technology, Vietnam National University, Hanoi.[2]

Abstract. We propose to develop an accurate and efficient method to reconstruct large indoor scenes from videos of RGB-D images using graphs of primitives. By projecting the scene onto basic shapes (called primitives) we generate a global model that is light in memory without loss of accuracy and most importantly easy to update with incoming frames. The accuracy of the global model allows accurate reconstruction, while its light representation allows for scalability of the method to very large RGB-D database.

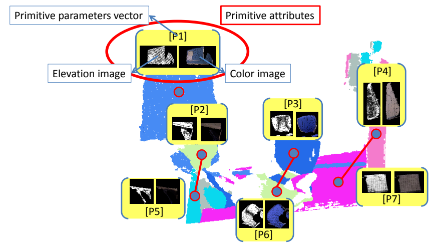
1 Introduction

Access to a large amount of RGB-D images is soon to become a reality thanks to new video depth sensors such as the Microsoft Kinect camera. One interesting application of depth images video stream is the in-hand scanning of very large environments. The user simply walks around inside an environment (a building for example) he wants to scan (this can be the case of real estate companies who want to provide 3D data of buildings) pointing out the camera in all directions and the building is automatically reconstructed and ready to use, without further processing, making the 3D scanning accessible to the non-professional user. While much work has been done on improving the quality as well as computational time of in-hand scanning techniques, only recently a few study has been pursued on how to scale existing techniques to very large RGB-D database. We believe that this is a challenging topic that deserves to be studied deeper in the coming years.

2 Proposed method

We propose to develop a new strategy to automatically build 3D models of large scale indoor scenes from dense sequence of depth images captured with a video-rate depth camera. In general, existing approaches use a frame-to-global-model approach framework to build 3D models from a long sequence of depth images ([1], [2], [3], [4]). In such a framework, single or multiple frames are automatically registered to the global model and then used to update the global model. The global model is either represented as a cloud of points ([1], [2]) or a volumetric signed distance function ([3], [4]).

(a) Input cloud of point.



(b) Graph of primitives.

Fig. 1: Illustration of our scene representation using a graph of primitives.

In spite of the advantages of the current methods, several limitation cases have been reported. In particular, existing methods are limited to rather small scale scenes. By small scale scenes, we mean a scale of the same order of the field of view of the camera. This is because the global model is required to be of at least the same resolution of the input frames to obtain accurate registration. As a consequence, keeping a global model of the scene to which input frames can be registered when the scene is hundreds of times larger than one input frame becomes a challenging and unsolved problem.

We propose to use the recent advances on graph representation of 3D cloud of points to address this issue. A 3D cloud of points can be viewed as multiple simple shapes (called primitives) such as plans, spheres, cylinders and cones, all related together by adjacency relationships such as perpendicular or collinear. Recent research on this field ([5]) has shown impressive results on the accuracy of the reconstructed object while keeping efficient representation of the object (a few nodes with adjacency relationship). We propose to develop an efficient and accurate 3D reconstruction method inspired from the graph representation of 3D scenes. The underlying idea is to update a global graph of primitives representing the scene with incoming input frames. Such a method will allow us to automatically build 3D models of scale hundreds of times larger than the field of view of the camera.

References

1. A.Davison, I.Reid, N.Molton, O.Stasse: Monoslam: Real-time single camera slam. IEEE Trans. on PAMI (2007) 1052–1067
2. P.Henry, M.Krainin, E.Herbst, X.Ren, D.Fox: Rgb-d mapping: Using depth cameras for dense 2d modeling of indoor environments. Proc. of International Symposium on Experimental Robotics (2010)
3. S.Izadi, D.Kim, O.Hilliges, D.Molyneaux, R.Newcombe, P.Kohli, J.Shotton, S.Hodges, D.Freeman, A.Davison, A.Fitzgibbon: Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. Proc. of ACM Symposium on User Interface Software and Technology (2011)
4. H.Roth, M.Vona: Moving volume kinectfusion. Proc. of BMVC (2012)
5. Y.Li, X.Wu, Y.Chrysathou, A.Sharf, D.Cohen-Or, N.J.Mitra: Globfit: Consistently fitting primitives by discovering global relations. Proc. of SIGGRAPH (2011)