

Thesis Proposal

Textured 3D Mesh Reconstruction of Indoor Environments Using RGB-D Camera

Collin Boots

Feb 2, 2014

Background

As robots continue to be incorporated into human environments, the need for intelligent and high-speed reasoning about the objects around them increases dramatically. At the simplest level, mobile robots need to create a map of their environment for navigation. At a higher level, some robots need to recognize distinct objects in their environment, track object movement, and have some intuitive sense of object geometry that is easily stored and processed. Even more important is being able to efficiently generate this environment from sensor data in real time. Like the human brain, the robot should also be able to perform these low level functions with only minimal intervention from higher cognitive functions.

RGB-D cameras provide a great low cost solution for capturing 3D environments. 3D point clouds can very quickly grow into unwieldy dense datasets filled with highly redundant data, especially in structured indoor environments with many planar surfaces. Various methods have previously been proposed to decimate point clouds using triangulation and mesh generation techniques or sparse voxel octrees (SVO). Previous work has demonstrated the diverse capabilities of depth cameras from generating highly accurate 3D surface models [2] to reliable 3D pose estimation [1, 3]. Other approaches have been able to store and merge the surface data more efficiently, but still regard the environment as a unified whole rather than discrete objects.

Proposed Method

A hybrid method is proposed that adapts existing mesh generation and triangulation techniques to online conversion of point clouds into meshes at reasonable frame rates. Each frame from the RGBD camera will be triangulated and converted to sparse meshes or incorporated into previously generated meshes. Only

the decimated mesh and the current frame will need to be stored. Color data will be stored in textures allowing for very convenient processing, storage and rendering of the virtual environment. To achieve the processing speed required by the meshing pipeline, a high percentage of the processing work will be offloaded to the GPU.

By extracting meaningful geometry from the RGB-D in the form of triangle meshes, a large number of advantages can be realized.

1. The storage format of a triangle mesh is very efficient, allowing for large and detailed maps to be stored in memory constrained systems like small mobile robots.
2. Mesh models provide a natural way of isolating distinct objects in the environment in a way that point clouds and surface maps cannot.
3. Meshes are easy to manipulate, modify, and maintain in real time using well established computer graphics technology and open the problem to the wealth of domain knowledge from the gaming graphics industry. By representing the robot's environment in a similar manner to a game world, future robots may be able to more easily leverage the gaming industry's experience in intelligent actor design.
4. By manipulating and moving meshes instead of erasing and redrawing portions of a discrete map or point cloud, map artifacts from more dynamic environments could be handled more gracefully and naturally.
5. Storing objects as distinct, flexible models provides a natural means of handling moving objects in lower level software, freeing higher level processes for reasoning about object movement more abstractly.
6. Meshes provide an efficient and easy to process intuition of geometry to higher cognitive functions that may simplify object recognition and manipulation tasks.
7. Meshes are very flexible in their ability to record complex geometries. By using a single storage format for all object types, many implementation efficiencies may be realized. Also, meshes provide a very straightforward tradeoff between simplicity and accuracy that can be controlled dynamically.

Using an off the shelf RGB-D depth camera like the Kinect to accomplish all of these goals also brings with it all the benefits explored in previous work: hardware simplicity, minimal integration complexity, and low cost. This work will focus primarily on generation of high-accuracy triangle meshes for a variety of scene complexities and difficult edge cases such as touching objects. Time and technology permitting, some of the potential advantages of the approach listed above will be explored and demonstrated.

References

- [1] Felix Endres et al. An Evaluation of the RGB-D SLAM System. *Robotics and Automation (ICRA), 2012 IEEE International Conference on.*, 2012.
- [2] Richard A. Newcombe et al. KinectFusion: Real-time dense surface mapping and tracking. 2011.
- [3] Yuichi Taguchi, Yong-Dian Jian, Srikumar Ramalingam, and Chen Feng. Point-Plane SLAM for Hand-Held 3D Sensors. Technical report.