Map Building via *RTABMap* and *Guidance*

Steve McGuire

University of Colorado at Boulder

October 14th, 2015

Overview

1. Problem Statement
2. Sensor Details
3. Processing Algorithms

Problem Statement

Statement: Given a set of sensors, produce a metrically accurate 3-D map of the world while estimating observation location and motion in that map.

Sensor Suite

- *Inertial* 6DOF inertial data is available with calibration parameters
- *Depth* Dense metric depth imagery is available with a pinhole camera model
- *Visual* Imagery is available with a pinhole camera model

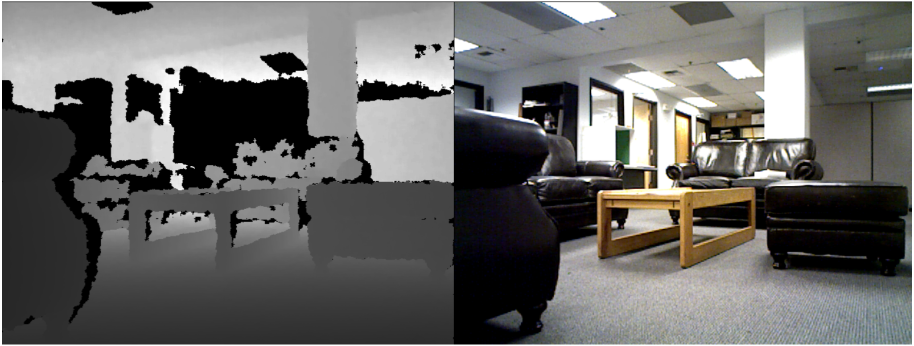

Figure: **Sample RGB-D Image Pair**

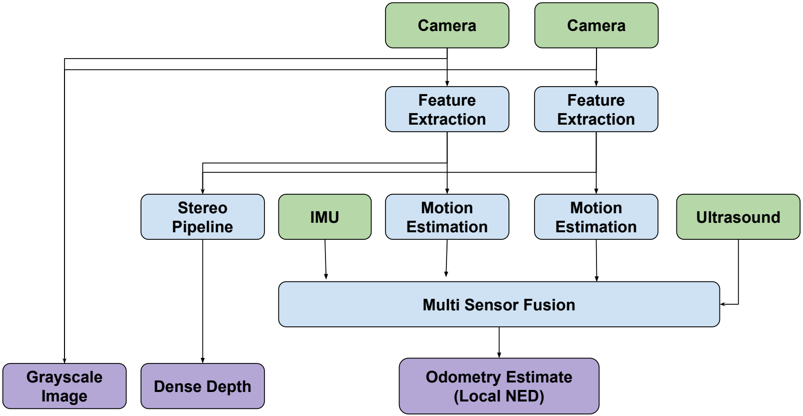

Figure: *Guidance* **System Diagram**

Data Flow Block Diagrams

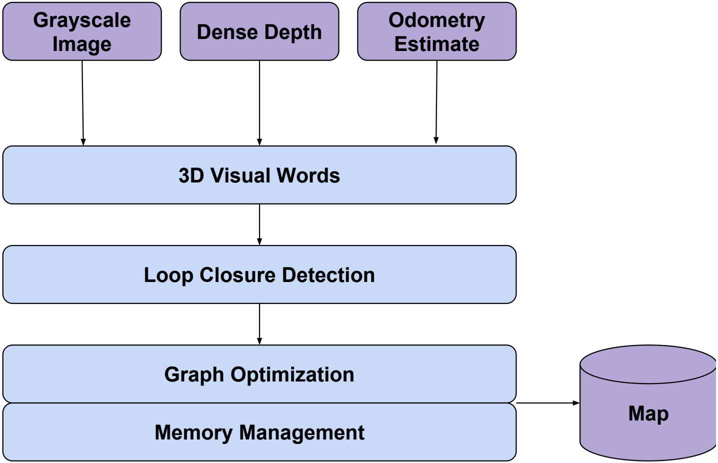

Figure: *RTABMap* **Mapping System Diagram**

Guidance Sensor Detail: Camera

- Five stereo pairs, each rectified to pinhole model with available pinhole parameters (10 total cameras)
- QVGA (320x240) resolution in grayscale
- Global shutter
- Auto exposure
- External 20Hz frame rate, synchronized among all cameras

Guidance Sensor Detail: IMU

- Five Invensense MPU 6050
- Integrated MEMS gyros + accelerometers
- External 20Hz update rate, synchronized to images

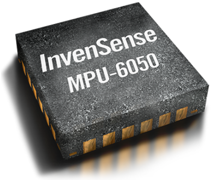

Figure: *Invensense* IMU

Guidance Processing: Feature Extractor / Descriptor

- Independently run on each image (10 total)
- **FAST**: Features from Accelerated Segment Test
- Relies on corner detection to identify feature
- Characterized by a **BRIEF** descriptor: Binary Robust Independent Elementary Features
- Makes a bitmask of pixel intensity comparisons, where pixels are sampled from within a local patch of the feature
- Optimized for memory-constrained applications

Guidance Processing: Motion Estimation

- Uses both 2D→2D and 3D→2D estimation
- Monocular estimation based on least squares fit utilizing epipolar constraints and camera calibration parameters
- Stereo estimation utilizes 3D triangulation from stereo calibration parameters
- Seamless switching between mono/stereo motion sources

Guidance Processing: Stereo pipeline

- Uses a sliding block matching algorithm
- Computes a dense depth maps for two pairs most aligned in the overall direction of travel
- Other three stereo depth maps are unavailable
- External 20Hz frame rate of depth map

Guidance Processing: Sensor Fusion

- Based on metric scale estimator black box of Kneip, Weiss, and Siegwart
- Relies on short-term integration of gyro measurements between successive samples
- Recovers the scale parameter when using mono motion estimates
- Incorporates both stereo and mono motion estimates
- External 10Hz data rate

RTABMap Processing: 3D Visual Words

- **SURF** detector: Speeded Up Robust Features
- Faster than **SIFT**: Scale Invariant Feature Transform
- Not patent encumbered
- Capable of GPU acceleration via OpenCV OpenCL support
- *Interest points*: Cascaded pyramid using Difference of Gaussians (vs Laplacian of Gaussians for SIFT)
- *Local description*: Based on Haar wavelet response (includes an orientation of the feature for rotation invariance)
- Visual Words combine a SURF feature descriptor with a 3D position

RTABMap Processing: Pose Graph Optimization

- SLAM problem represented as a graph: nodes are poses, links are rigid transformations
- Solve by finding minimum residual of estimated poses and estimated transformations that explain measurements
- Optimization only runs using working memory constraints for speed
- Able to run using complete constraints offline for maximum map quality

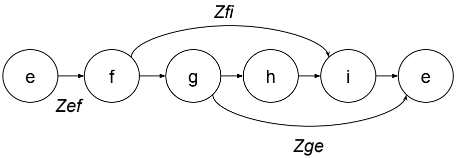

Figure: Pose Graph Representation Showing Rigid Constraints and Loop Closures

- Evaluated by comparing extracted visual words of every N^{th} frame against saved visual word dictionary
- Creates additional pose graph transformation constraints to limit effect of visual drift
- Rigid transformation estimated using RANSAC over matched visual words

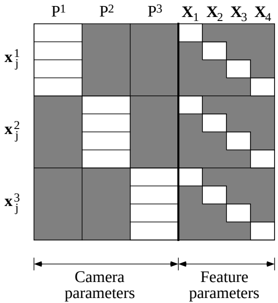

Figure: Pose Graph Sparse Jacobian

RTABMap Processing: Memory Management

- Three levels of storage of visual words: short-term, working, and long-term
- New data percolate towards long-term memory
- Movement of data is determined by similarity ratio; highly similar images are not stored
- Trades off map quality for online execution of graph optimization over working memory

RTABMap Processing: Map Generation

- Map is stored as an OctoMap
- Eight-connected map storing three states: {unknown, empty, known}
- Prepared by projecting depth point clouds from graph optimization into world coordinates

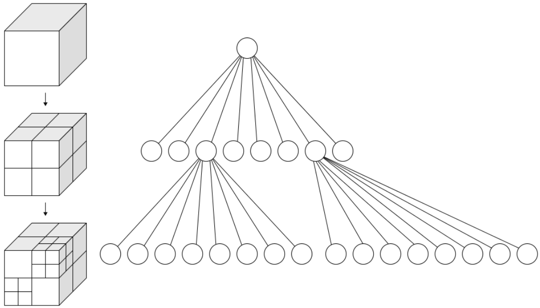

Figure: Octree Occupancy Volume Representation

Comparison to Other SLAM Techniques

- *LSD-SLAM*: Large Scale Dense SLAM: Requires scale parameter estimation external to the system
- *RGBD-SLAM*: Does not provide for loop closure detection (worse map quality)