Wayne Chang
Omer Mano
Eric Wengrowski
Advisor: Dr. Dana

Remotely Processed Visual SLAM for Odometry

Visual Simultaneous Localization and Mapping (SLAM) is a technique used by robots to keep track of their locations while building maps of their environments, all without prior knowledge of their surroundings. It is a collection of algorithms that change depending on the problem being solved.

The technique for determining a robot's location from just motor odometry data is known as dead reckoning. When implemented, this technique is prone to compounding errors due to real world conditions (such as wheel slippage and inclined planes). Visual tracking methods, however, are more accurate in determining robot position after movement *if initial conditions are known*. Visual odometry computes position directly from deltas between frames rather than accumulated historic paths of motion. Additionally, calibrated cameras help ensure the accuracy of the data collected at each frame. These two approaches combined are known as Visual SLAM, and provide an effective solution to the stated problem. Visual SLAM first uses odometry data for a basis estimate of location, then images to accurately refine the position and orientation.

The two major drawbacks of visual SLAM that we have encountered are: (1) circular dependence of initial conditions on a known environment, and (2) low frame rates and noisy images.

To gauge current location and orientation purely from captured images, there must be knowledge of the environment. Unfortunately, knowledge of the environment can only be ascertained given a known location and orientation. This problem is called Structure From Motion, and its circular dependency causes most naive approaches to fail. Current pure-visual approaches that show promise are feedback-based refinement techniques. Two of such techniques are called Bundle Adjustment and Extended Kalman Filters. Bundle Adjustment employs non-linear optimization to reduce reprojection errors between image interest points and 3D coordinates, and Extended Kalman Filters provide a linearization of a non-linear problem (such as estimating future camera pose given acceleration and velocity). The MonoSLAM implementation we use in our project utilizes an Extended Kalman Filter at its core. Both techniques can utilize odometry data to bypass the circular dependency problem by providing basis estimates of location.

Low frame rates may attribute to loss of visual landmarks. If the robot moves too much between frames and loses too many of it's interest points, it will lose visual reference for it's recent motion, and become lost. Imagine spinning too fast, opening your eyes, and not recognising your surroundings. To solve this problem, wide-angle lenses can sometimes be used if greater computing power is unavailable to process a higher frame rate. For instance, the ipod cameras have a very narrow field of view, so external lense attachments may added to widen the viewing angle of each frame. Noisy images are generated when interest points do not match their intended objects either when rectified or between delta frames. This may occur in an environment with many similarly textured features, such as a large brick wall. The frequencies of these errors can often be reduced by removing outlying data points using techniques such as RANSAC filtering.