Large-Scale Robotic 3-D Mapping of Urban Structures

Michael Montemerlo and Sebastian Thrun
Stanford AI Lab
Stanford University
{mmde,thrun}@stanford.edu

Abstract. This article presents results for building accurate 3-D maps of urban environments with a mobile robot based on the Segway scooter. The goal of this project is to use robotic systems to rapidly acquire accurate 3-D maps which seamlessly integrate indoor and outdoor structures. Our approach uses an efficient implementation of the global scan alignment algorithm of Lu and Milios in order to integrate GPS, IMU, and laser data into globally consistent maps. The 3-D models acquired by the robot are analyzed for navigability using a multi-resolution evidence grid approach, and visualized using a meshing algorithm adapted from the computer graphics literature. Results are presented for a number of environments which combine indoor and outdoor terrain.

1 Introduction

In recent years, there has been a number of projects seeking to map physical environments. Classical work includes mapping from the air [12], the ground [7], indoors [6,11], outdoors [24], under water [28] and in the subterranean world [3]. The development of techniques for the acquisition of such maps has been driven by a number of desires. They include photo-realistic rendering [1,2], surveillance [27], measurement [3], and robot guidance [28]. Not surprisingly, the best work in this area has emerged from a number of different scientific fields, such as photogrammetry, computer vision, computer graphics [13,22], and robotics [25].

This paper describes a robotic system designed to acquire large 3-D maps. Urban terrain possesses a number of characteristic features: It combines large open places, such as plazas and roadways, with narrowly confined spaces, such as building interiors. GPS is often inaccurate in outdoor urban areas due to multi-path effects, and unavailable indoors. From a SLAM (simultaneous localization and mapping) perspective, maps of the size targeted by our research involve $10^6$ or more features; gathered over $10^5$ poses. Urban terrain is non-flat, hence the robot has to be localized in 6-D. The overall SLAM problem, thus is orders of magnitude more complex than prior work. While applying equally well in to 6-D data in theory, the vast majority of deployed SLAM algorithms have only been applied in planar environments with 3-D poses, to keep the data sets manageably small. Even those that perform 3-D mapping, often do so via 2-D SLAM [9], with the exception of [21] which offers no provision for closing cycles [8,4]. Further, past work has not provided effective means to incorporate occasional GPS measurements.
2 SLAM in Urban Environments

A key problem in building large-scale urban maps pertains to the ability to integrate information from multiple information sources, specifically GPS (global positioning system), IMU (the inertial measurement unit), odometry, and the LIDAR sensor(a laser-light detection and ranging sensor). This mapping problem is a version of the SLAM problem, short for simultaneous localization and mapping. The SLAM problem is characterized by a necessity to estimate the map of an environment while simultaneously localizing the sensor relative to the map. Outdoors, GPS provides absolute position labels; indoors, it presently does not.

Our approach builds on prior work on SLAM by Lu and Milios, who proposed Kalman filter-based approach that represents SLAM posteriors through collections of local constraints between nearby poses [15] (see also [8]). Also related is recent work in [4,20,26], who propose variants of the information filter for solving the SLAM problem. These algorithms are approximate, and they fail to integrate occasional global position data when available. However, both families of approaches are related in that they represent SLAM posteriors through local constraints—which is in stark contrast to the classical SLAM solution, the EKF [23], which maintains a full covariance between any two features.

Our approach represents the SLAM posterior as an undirected Markov network, where nodes correspond to poses, GPS measurements, and range measurements. The network possesses three types of pairwise node potentials: There are potentials between range measurements and the corresponding pose at which the measurement was required; there are potentials between subsequent poses, governed by the IMU measurements. And finally, there are absolute location potentials for poses at which GPS data was received. All of these potentials are nonlinear-quadratic; they are composed of a deterministic non-linear projective function (e.g., the robot motion model; the measurement model) with a quadratic penalty function that measures deviations from this non-linear projection. This representation generalizes past work on SLAM, most notably [15], in that the resulting sum of potentials can be thought of as a non-normalized log-likelihood function. However, representing them as potentials avoids numerical instabilities of the covariance representation in [15].

The map can be retrieved by finding the most probable set of robot poses given the set of non-linear constraints. This is done by repeatedly linearizing the constraints