High-Resolution Mapping and Modeling of Multi-Floor Architectural Interiors
MIT9904-20

Progress Report: July 1, 2000—December 31, 2000

Seth Teller

Project Overview

In our initial proposal, we proposed to extend and complement our successful automated mapping system for exterior urban spaces, to incorporate mapping capability for interior spaces. We proposed to deploy a rolling sensor equipped with laser range-finder, high-resolution digital camera, and positioning instrumentation (wheel encoders and precise inertial measurement system) to allow acquisition and registration of dense color-range observations of interior environments. We planned to direct the sensor package by remote radio control. We planned to incorporate positioning instrumentation to track both horizontal motion (e.g. along hallways and through rooms), and vertical motion (e.g., up and down wheelchair ramps, and while inside elevators).

We proposed to acquire the following instrumentation: a Nomadics rolling robot; a Canon Optura digital video camera; a K2T scanning laser range-finder; a GEC Marconi inertial measurement unit; and set of Pacific Crest radio modems. We also proposed support for one PhD student, one MEng, and two full-time UROPs. At the end of 1999 we expected to demonstrate the integrated sensor, and data collected from several floors of Technology Square (including LCS and the AI Lab), with geometry and texture data captured to better than 10cm resolution.

Progress Through December 2000

One fundamental technical obstacle has been the development of robust algorithms to determine 6-DOF pose (position and orientation) for the moving sensor. Therefore, we have made extensive efforts over the past year to solve this problem. As of December 2000, we have achieved fully automated exterior calibration (pose recovery) for networks of thousands of images over extended areas spanning hundreds of meters. Our algorithms are accurate to roughly a tenth of a degree of absolute rotation, and five centimeters absolute translation (see http://graphics.lcs.mit.edu/~seth/pubs/pubs.html for relevant publications). However, these algorithms have been demonstrated outdoors, using the Argus sensor. Outdoors, we can reliably detect families of parallel lines from buildings, and exploit GPS for good initial position estimates. Also, the Argus sensor includes a high-resolution camera and mechanical pan-tilt head, enabling the capture of high-resolution spherical mosaics. (A typical Argus mosaic has spatial resolution of about one pixel per milliradian, or 1,000 pixels in a roughly 60-degree field of view.) Our pose estimates are good to within 1-2 pixels rotationally, and 1-5 pixels translationally.

Indoors, the situation is different. GPS is not available, so we have no a priori translation estimates. We can continue to exploit dead reckoning information from odometry, however. Our image sensor, a CycloVision OmniCam, operates at NTSC resolution, about one one-hundredth the resolution of the Argus. (This is about one-tenth the Argus's linear resolution, or about half a degree per pixel.) Whereas outdoors we acquire about one image per minute, with ten-meter baselines between images, indoors we acquire images at 30Hz – a thousand times as fast – and with baselines of only a few centimeters for a slowly moving platform. Finally, whereas outdoors the illumination conditions are uncontrolled and highly variable, indoors we can expect nearly constant lighting. All of these factors amount to a significantly different deployment environment indoors, when compared to the outdoor setting. However, the spherical imagery from the OmniCam is well suited to our algorithms, which exploit vanishing point and expansion center geometry on the sphere. Thus we have begun to map our existing algorithms from Argus onto the new, mobile, platform. We hope to make the registration algorithms run in real-time, rather than offline at the cost of seconds per image.

Automation

This year we achieved complete automation for the sole remaining semi-automatic system module: the camera registration module. This system component takes as input the raw pose-image data from the Argus sensor -- high-resolution images labeled with approximate 6-DOF pose (position and orientation in Earth coordinates), and an absolute time-stamp. The registration module estimates exterior orientation for the images, locking them down into a single coordinate system. Previous automated methods had not been demonstrated for more than a few dozen images, or for camera motions of more than a few meters. Our system achieves rotational registration good to about one tenth of a degree, and translational registration good to about five centimeters, for thousands of images acquired over an area hundreds of meters on a side. The method uses a few seconds of CPU time per image, so it does not currently run in real time.

Scaling

Our original dataset consisted of a few thousand images of a few buildings, in an area about one hundred meters on a side. We are in the process of scaling our dataset by two orders of magnitude, to the entire MIT campus; this involves the acquisition of hundreds of thousands of images of about two hundred buildings, in an area about one kilometer on a side.

Fidelity

Our early reconstructions were simple block models. We now exploit knowledge of building shapes and exterior detail to produce models with greater fidelity. Specifically, we extract models of window grids, and constrain our surface and texture estimation algorithms to match the apparent window structure. This technique makes more assumptions about the world, but it produces much higher-quality models, which are suitable for simulation applications in which detailed synthetic views are necessary.

Research Plan for the Next Six Months

We are investigating a number of techniques for the acquisition system over the next 6-12 months. First is the use of spherical optical flow (ego-motion estimation from dense texture) and feature tracking (ego-motion estimation from edge and corner features). We are exploring a collaboration with Prof. Michael Black (now at Brown University). We expect that this method will complement the feature-based registration methods we have developed for the Argus, allowing the Rover to maintain 6-DOF lock even when near portions of the environment that do not have easily discernible linear or point features, but do have texture.

Our second area of emphasis is to continue to develop a robust sensor fusion architecture which will combine navigation instrumentation (such as odometry and inertial sensors) with the image-based navigation information discussed above. For robustness, the Rover will combine both forms of information, and rely more heavily on each when the complementary form is compromised. For example, when the Rover moves into featureless, textureless regions, the navigation instrumentation can help maintain pose. When the Rover moves very rapidly, or over uneven terrain (so that the inertial sensors and/or odometry produce corrupted readings), the vision-derived information can be used to maintain pose lock.

Third, we plan to continue to augment the Rover with additional sensors, for example continuously scanning laser range finders. Such sensors will give us a rich source of absolute depth information, and should disambiguate many of the geometric features produced by computer vision alone.

Finally, we continue the development of robust, scaleable spatial data structures and algorithms for handling complexity, in the form of very large numbers of observations and output features. In particular, we use spatial data structures that support inverse range queries, so that (for example) given a region of interest, we can rapidly identify the data elements – image and navigation data – which might be relevant to 3D reconstruction and model acquisition in that region.