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

Progress Report: July 1, 2002—December 31, 2002

Seth Teller

Project Overview

Our research has three long-term goals. First, to develop rapid *capture methods* for geometric environments, using autonomous (robotic) sensors. Second, to develop a pervasive *location capability* for indoor environments (without GPS), so that a hand-held mobile computing device can reliably determine its position and orientation. Third, to develop several fundamentally new devices and applications (the *software compass, software marker, software flashlight*) that combine the captured model, and the hand-held device's positioning ability, to enable the user to *interact directly with the environment away from the desktop*.

Our NTT-sponsored efforts focus on three aspects of this long-term goal. First, we are developing *computer-vision capture methods*: algorithms to localize and fuse a low-resolution omni-directional video stream gathered by a rolling or hand-held camera. (Joint work with Prof. John Leonard in the Department of Ocean Engineering.) Second, we are developing *procedural capture methods*: a "compiler" that takes legacy 2D CAD information as input, and produces well-formed 3D architectural models as output. Third, in collaboration with Prof. Hari Balakrishnan, we are *generalizing the Cricket location infrastructure* to support orientation (as well as position) determination, and building early prototypes of the devices and applications mentioned above.

These research goals overlap with the interests of several NTT laboratories (and projects): the Cyber-Space Laboratories (3D Information Processing Systems); the Communications Science Laboratories (GeoLink); the Telecommunication Energy Laboratories (Low-Power Radio-Frequency Devices); and the Network Innovation Laboratories (Software Radios).

Progress Through December 2002 (Updated to February 2003)

Computer-Vision Model Capture. We have achieved the largest-scale autonomous capture tasks with loop closing reported to date in the autonomous robotics literature. These include rolling robot deployment for sessions ranging from tens of minutes to nearly two hours in duration, in various MIT corridor networks with complex topologies (see Figures in accompanying slides). Along with Ph.D. student Michael Bosse, we have formulated the "Atlas Framework" for extended mapping, in which the environment is represented and recovered as a collection of locally rigid sub-maps, each connected to its neighbors by a set of uncertain geometric transformations. This is motivated by our intuition about how humans navigate in complex spaces: we are aware of the local metric structure of our environment, but do not usually have a globally consistent mental model of the entire space.

We have demonstrated the following results. First, we can use the image information to improve the raw navigation data available from the sensor (typically as odometry or inertial integration). This allows us to use relatively less accurate (and therefore cheaper) navigation sensors on boards. Second, we can stabilize the image sequences to persistent structure in the scene (vanishing points), removing the camera rotation introduced

by the operator. This presents a smoother viewing experience to the user, allowing him/her to better maintain a sense of orientation within the simulated environment. Also, stabilized video compresses significantly more efficiently than unstabilized video, due to increased inter-frame coherence.

Third, we have addressed the scaling problem with an approach called "Atlas generation." Rather than attempt to recover a single, globally consistent scene model, we instead produce a set of local maps connected only by their (overlapping) boundaries. This allows us to push uncertainty out of the maps themselves and into the connecting transformations, much as a human exits one room, through a short passageway, and enters another room.   For most applications, this notion of "local" orientation is sufficient. Fourth, we have demonstrated the Atlas framework over environments spanning hundreds of meters, with repeated visits to certain areas of the environment ("Loop Closing").

Procedural Model Generation. We have made significant progress in our effort to extract detailed three-dimensional geometric models from legacy two-dimensional CAD files. MIT's Department of Facilities maintains an extensive corpus of more than 800 floorplans, each including vector (line segment) representations of exterior walls, interior walls, load-bearing columns, doorways and windows. MIT also maintains a "base map" situating each building (ground floor) on campus, and delineating roads, sidewalks, walking paths, grassy areas, and parking areas. Finally, various topographic representations of campus elevations with respect to local sea level are available.

We have combined all of these elements using a series of parsing and interpretation scripts to be run daily, in "batch mode" in the early morning hours. The end-to-end script retrieves the base map from MIT's web site, then fetches all floorplans for each building found on the basemap. The floorplans are then segmented into layers, and each layer is separately extruded into three-dimensional form. The result is exterior geometry with exterior doors and windows, and interior geometry with interior walls, doors, stairwells, elevator shafts etc. All geometry is generated at three "levels of detail" for efficient rendering: the "low-detail" model is a simple vertical prism with no doors or windows; the "medium-detail" model has doors and windows; and the "high-detail" model has full exterior and interior geometry. During interactive viewing, the renderer selects the appropriate level of detail using the viewer's distance from the building. Recent progress on this front includes integration of these floorplans with an existing spatial database for large-scale walkthroughs, and the development (in progress) of new visibility algorithms for indoor-outdoor scenes.

Generalizing the Cricket Position Infrastructure. Our third major area of effort is in generalizing the Cricket location-determination infrastructure to support orientation computations, that is, to determine both position and attitude (bearing and elevation angle) of a hand-held device. We have designed a prototype device, the Cricket "software compass," that uses multiple ultrasonic receivers to infer orientation from phase differentials at the receiver. This device is not yet in fabrication, however. In the interim we have developed an equivalent capability through the use of two ordinary Cricket (position) listeners, attached to either end of a board about 75cm long. From the difference in reported listener positions, we compute the position and attitude of the board. We have also integrated a laser range-finder and VGA projector to produce a prototype "software flashlight." The range-finder yields depth to a modeled projection surface in the environment. The VGA projector allows projection of known model geometry (for example, hidden wires or pipes) onto the projection surface. Together, these components make a fundamentally new device possible, one that allows a kind of "X-ray" vision through the ordinarily opaque surfaces of the environment. At present the accuracy of the software flashlight is rather poor, but we are continuously improving its components.

Recent progress includes the extension of this device to have "software marker" capability, in which the user points at a location in the world and illuminates it using the laser range finder. The software compass integrates position, attitude, and range information to infer the XYZ coordinates of the world point indicated by the user. The user can then attach metadata to this world point to form a kind of "virtual tag" indexed by location.

Research Plans for the Next Six Months

Over the next six months, we plan to continue with each of the research efforts above.

First, we will continue to develop scalable computer-vision methods for model capture. Our next goal is to support generation of Atlases for even larger environments, and environments that intermix indoor and outdoor

elements. We will continue to develop new data structures and rendering algorithms for viewing dense, registered imagery and extracted three-dimensional geometry (typically point and edge features and piecewise-planar models).

Second, we will continue to develop scalable procedural CAD-based methods for model generation. We are working with MIT Department of Facilities to achieve more comprehensive parsing and interpretation of their posted CAD data. (At present we correctly process only a fraction of the available floorplans.) We are also integrating existing procedural algorithms for populating furniture based on space type. We are developing procedural population algorithms to fill in elevator shafts and cab assemblies, stairwells, wheelchair ramps, and pedestrian bridges from icons on the source CAD plans. Finally, we will continue to integrate our network API to serve location-specific data (model geometry, space name and type information, adjacency information) to mobile hand-held computers. This API will support location-aware applications such as route-finding, software marking (including entry of as-built CAD information), resource discovery, and the software flashlight for information overlay.

Finally, we will continue to develop the Cricket and Software Compass architecture. We are actively engineering the first- and second-generation Cricket Beacon and Listener hardware to improve its accuracy, precision, channel efficiency and power usage. For example, the current algorithm to detect the start of the ultrasound pulse is naïve, and could be greatly improved with the addition of a simple transmission pattern and match filter. The current uncertainty in detection produces a rather large spatial uncertainty in the recovered position of the listener, which we hope to reduce significantly. Also, under some circumstances the beacon circuitry can drain a significant amount of power to ground, wasting battery life. We are actively examining these issues.

We have recruited a new UROP student to tackle the problem of developing a reliable software compass using a single Cricket listener with multiple ultrasound transducers. Our current prototype is accurate to about five degrees of bearing. We hope to improve our design in two ways. First, to increase the bearing accuracy to a fraction of a degree. Second, to recover orientation information along multiple axes by using a 3D arrangement of the ultrasound transducers. For example, placing four transducers at four vertices of a cube would enable the listener to recover two, rather than one, orientation degrees of freedom with respect to each sensed beacon.

The software flashlight prototype serves as a proof of concept, but its accuracy is limited at present due to the uncertainty in the underlying Cricket listener position algorithms. In parallel with the above efforts, we will continue to develop calibration algorithms to recover the optical parameters of the VGA projector, for example by aligning projected geometry to equivalent fiducial geometry in the scene. In the next six months we hope to show the software flashlight deployed throughout a large room (tens of meters on a side), and able to faithfully project structural geometry: walls, edges, corners, support beams, doors, and window frames. We also hope to show a proof of concept of an "assisted deployment" method for beacons, in which a few beacons are initially deployed and programmed by hand, then additional beacons are deployed and semi-automatically discover their position with the help of a human operator and software compass.