

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

Progress Report: July 1, 2001—December 31, 2001

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. 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 2001

**Computer-Vision Model Capture.** One fundamental technical obstacle to achieving automated computer vision model capture methods has been the development of robust algorithms to determine 6-DOF pose (position and orientation) for the moving sensor. Most existing “egomotion” algorithms make severe assumptions: that the number of images is small; that the lighting is simple and unvarying; that the excursion or camera motion is

limited. To solve the problem in real-world environments, we must abandon all of these assumptions. Therefore our recent work has focused on gathering challenging input video sequences, and extracting egomotion and scene structure from these sequences.

**Procedural Model Generation.** An alternative model-generation method is to exploit existing information about the structure of the built environment. Many organizations, including MIT and presumably NTT, maintain detailed 2D (floorplan) CAD files describing the coarse room, corridor, and vertical interconnection (stairwells and elevator shafts) that enclose the organization's human-inhabitable space. These CAD files contain valuable information, but are missing critical information: for example, per-floor height and building placement information, detailed geometry, and appearance information. Moreover, these "as-planned" CAD documents are often very different from the "as-built" form of the physical buildings they describe, due to deviations that happen during construction but which are not reflected in the documents. Finally, these CAD documents often contain errors: for example, interpenetrating line segments, or holes, that erroneously describe solid walls (or the absence of walls) in the physical building. Therefore, our recent work has focused on building an effective "compiler" that is capable of parsing the legacy 2D CAD files, adding critical information about building height and placement, and generating "well-formed" 3D CAD geometry through the application of simple rules. For examples, wall segments can be extruded to 3D polygons, and where window segments are present in the input, a window hole can be "cut" in the 3D wall polygon. To date we have collected and parsed more than one hundred buildings, and more than nine hundred distinct floorplans, from the MIT Department of Facilities website.

**Generalizing the Cricket Position Infrastructure.** We are working with Prof. Hari Balakrishnan to develop and exploit sensor-based position and orientation determination capabilities using the Cricket architecture. In analogy to GPS, we instrument the indoor environment with a set of fixed beacons, each of which broadcasts its known position and a unique identifier. A mobile receiver can then combine several received signals to infer its own position. We are extending this capability in two ways. First, by combining multiple ultrasonic receivers, and analyzing the phase variations in pulse arrivals, the hand-held device can infer its orientation as well as position. (This work has been published in the Proceedings of MobiCom (Mobile Computing) 2001.)

**New Devices and Applications.** Fine-grained location and orientation capability, along with an accurate functional model of the environment, will give rise to a number of fundamentally new devices and applications. This capability gives rise to a number of fundamental new devices and applications.

The *software compass* enables a user to navigate indoors, by displaying a functional map of the user's surroundings, a "you are here" dot and "you are facing this way" arrow for the user, and navigation instructions for the user to find what s/he is looking for: a person, place, or thing (resource).

The *software marker* enables the user to mark, or annotate, objects. Rather than writing on the physical object, however, the software marker annotates the model's representation of the object. This makes the annotations persistent over time (rather than ephemeral), and allows annotations to be retrieved from the database using a position-based key (rather than a key based on the object's name or attributes, both of which might be difficult to discover).

Finally, the *software flashlight* enables the user to project geometrically registered (or geo-referenced) information from the environment model into, or onto, the environment itself. So, for example the user could overlay infrastructural information about hidden wiring, plumbing, etc. onto the walls of the environment, to see

where the wall would have to be opened for service. Or the user could overlay instructions (from, for example, a maintenance supervisor) onto the environment to help define or complete a requested maintenance task.

## **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 multi-floor acquisition, in which the video camera is carried not only through planar areas, but also through stairwells connecting multiple floors. This will produce a challenging mix of horizontal and vertical motion components, and an input video stream which exceeds the egomotion recovery capabilities of any current algorithm.

Second, we will continue to develop scalable procedural CAD-based methods for model generation. Specifically, we are integrating height information (from a database of per-building topographic elevations, and per-floor heights) into our CAD-model parser and compiler, so that multiple well-formed 3D floorplates can be combined into full-fledged building models. Moreover, we are integrating processing of exterior building outlines (which, when extruded vertically, give building exteriors) with interior building floorplans (which have icons for things like doors and windows) to produce building exteriors with well-defined 3D door and window geometry.

Third, we will continue to develop the Cricket and Software Compass architecture. We are actively designing the second version of the Cricket Beacon and Receiver hardware, paying special attention to operating considerations such as effective range and power consumption. We have defined an application scenario, which we call "Warmer/Colder" after the child's game. (In this game, the "caller" knows the location of a hidden object. The "seeker" must find the object, using only a series of "clues" in which the caller says "Warmer" when the seeker approaches the object, and "Colder" when the seeker moves away from the object. In our application, all "valuable" objects are tagged with Cricket listeners, and report their location to a network position infrastructure and database. A hand-held device then leads the seeker to any tracked object. We will demonstrate a Warmer/Colder prototype application in 6-12 months.