TravelMATE: A demonstration of SRI's Multimodal Augmented Tutoring Environment

Luc JULIA and Jehan BING
Computer Human Interaction Center (CHIC!)
SRI International
333 Ravenswood avenue
Menlo Park, CA 94025 – USA
{luc.julia,jehan.bing}@sri.com

To facilitate exploration of the augmented reality paradigm, we have been constructing an augmented reality (AR) application framework called the Multimodal Augmented Tutoring Environment (MATE). In this framework, multiple processes for providing sensor readings, modality recognition, fusion strategies and viewer displays, and information sources can be quickly integrated into a single flexible application.

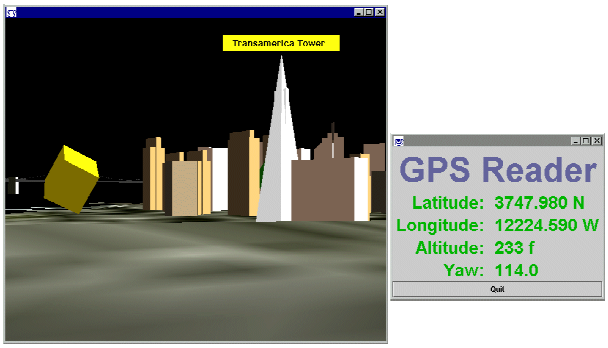

Figure 1: TravelMATE and its GPS reader

Our first AR prototype, "TravelMATE" [Figure 1], makes use of many of the technologies developed in previous 2D and 3D tourist, desktop applications[1]. By adding GPS and a compass sensor, we solved the problem we usually get by trying to navigate a 3D world with a 2D paradigm. Now, as the user walks or drives around San Francisco, her display is automatically updated based on her position and orientation.

The display currently consists of a 3D, virtual, model of San Francisco. When the user is "close enough" to a landmark in this city, a label describing this landmark appears in the virtual world. The "close enough" notion is calculated according to the modeled size of the object in the world. The bigger the object is, the better is the chance to see it from afar. While testing the system in the streets of San Francisco, we were amazed by the efficiency of this simple rule.

The nice thing about a virtual world is that it reflects most of the information contained in the real world and even information that is not accessible in the real world: If the user is in front of the Transamerica tower, by just looking at it, she will not be able to know its height and when it was built. But certainly, somewhere on the Web is a site about this tower, which contains that information. With the appropriate set of agents (e.g., Web search, information extraction), this site is virtually part of the MATE application, augmenting the virtual world and, subsequently, the real world. Most of TravelMATE databases are actually dynamically generated from the Web.

The user can access this handfull of information through a multimodal interface that allows her to ask for details when the system proactively displays labels on the screen. She can talk and gesture naturally to indicate a specific label or object, and the system will answer in a multimedia fashion.

Ultimately, the display in the car will consist only of labels projected on the windshield. There is no need to project the virtual world since the real and the virtual worlds are perfectly aligned, thanks to the sensors. We will have to work out a solution for detecting the gestures (e.g., 3D camera, touch windshield). With the more portable version (for the user who is walking around), the display of the virtual world is still necessary, allowing the user to map the two worlds by herself, until we find reasonable heads up see-through glasses that will work in any lighting condition. Here again, we will still have a problem with the pointing of the objects (eye tracking?).

[1] http://www.chic.sri.com/randd.html