Mirror Worlds for Indoor Navigation and Awareness

Don Kimber, David Lee, Jim Vaughan,
Jacob Biehl, Mathew Cooper
FX Palo Alto Laboratory
3174 Porter Drive
Palo Alto, CA, USA
{kimber,dclee,vaughan,biehl,cooper}@fxpal.com

Jun Shingu
Fuji Xerox
6-1 Minato Mirai, Nishi-ku
Yokohama, Japan
Email: jun.shingu@fujixerox.co.jp

***Abstract*—We describe Explorer, a system utilizing *mirror worlds* - dynamic 3D virtual models of physical spaces that reflect the structure and activities of those spaces to help support navigation, context awareness and tasks such as planning and recollection of events. A rich sensor network dynamically updates the models, determining the position of people, status of rooms, or updating textures to reflect displays or bulletin boards. Through views on web pages, portable devices, or on 'magic window' displays located in the physical space, remote people may 'look in' to the space, while people within the space are provided with augmented views showing information not physically apparent. For example, by looking at a mirror display, people can learn how long others have been present, or where they have been. People in one part of a building can get a sense of activities in the rest of the building, know who is present in their office, and look in to presentations in other rooms. A spatial graph is derived from the 3D models which is used both to navigational paths and for fusion of acoustic, WiFi, motion and image sensors used for positioning. We describe usage scenarios for the system as deployed in two research labs, and a conference venue.**

***Index Terms*—Location Sensing, Localization, Mirror Worlds, Ambient Computing**

I. INTRODUCTION

The applications which became prominent for outdoor navigation and exploration are now moving indoors. These not only include maps, positioning and route planning, but also 3D models, updated images, video feeds, and layers showing important dynamic information such as traffic, path closures, etc. We have developed a system for supporting these kinds of applications for indoor spaces, consisting of dynamic models of the spaces, viewers for web browsers, mobile devices, and large displays such as kiosks, and a back- end which performs sensor fusion over many inputs to dynamically update the model and maintain its history. The applications provide not only navigational assistance, but location based awareness (e.g. is a given conference room in use? are people in the lobby? What did this whiteboard look like yesterday?) and social awareness (e.g. is Joe in his office?)

We have deployed the system at two research lab locations, a 15 story office building, and are in the process of developing deployments for a conference center and exhibition center showroom space. A 3D model for each space is used for viewer applications, and also as the basis for sensor fusion and integration. A spatial graph derived from the model is used to generate paths for navigation assistance, and to provide the structure of a hidden Markov model used to fuse localization sensor information. The Explorer back-end is configurable to accommodate a variety of infrastructure components such as surveillance cameras and trackers, display capture, sound systems, and a variety of sensors such as Cisco WifI-based location sensing, which vary greatly across facilities.

Fig. 1. Mirror world view of our Lab, shown in Google Earth. The color over a person's office indicate whether that person is in the office (green), has a visitor (purple) or is in the building but out of office.

An android App shows a map or model of the space with updated dynamic layers, such as positions of people or occupancy of rooms, and also collects audio samples, WiFi scans, motion data, and images, which are uploaded to a server to determine position of the device and to update the spatial model. Image SIFT features are used to determine device position, and also to allow model textures, such as bulletin boards, to be updated. Audio samples are compared with acoustic background models, or known controlled sounds, to estimate position and to update the background models. The locations estimates are combined using a spatial hidden Markov model and a modified Viterbi algorithm, in which sparse observations are provided by both the device being tracked and environmentally placed sensors.

A video showing our Google Earth viewer is available at https://dl.dropbox.com/u/3031469/IPIN2012.mp4.

II. RELATED WORK

Mirror worlds were described by Gelernter in 1991 [1], and have also been referred to as dual reality or cross reality as

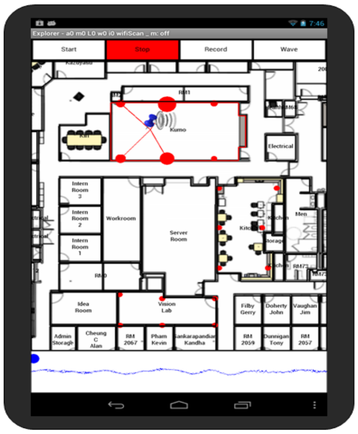

Fig. 2. A mobile app which collects audio, images, WiFi scans and motion data for upload to servers which update the model and estimate the position of the device. Position is shown at room level, or more accurately in zones supporting acoustic localization. The blue pin and sound icon indicate it has a precise ($\approx< 0.5m$) position estimate.

realized in the mirror world built of the MIT Media Lab [2]. Our sensor system includes acoustic background modeling and a Markov framework similar to Tarzia, [4], but with additional sensor types, and with an acoustic TOF component providing 0.5m localization in enhanced zones.

III. MIRROR WORLD VIEWERS AND APPLICATIONS

Applications require a viewer, a dynamic model of the space, and sensors to update the model and determine viewer positions. The viewers run on web pages, on portable devices, or displays placed in the environment, and connect to servers which maintain the model and perform sensor fusion. The sensors may be on board the portable viewers running on phones and tablets, or embedded in the environment.

The viewers show maps or models depicting the space, augmented with layers such as labels, paths, explanatory images, etc. We have built 3D viewers that run in web pages using the Google Earth API, and that run as Windows applications or mobile apps using the cross platform Unity Game engine. We also have an Android app which shows a 2D map, and uploads sensor data for positioning of the device, and for updating the model. (Fig 2.) The app can be used to take pictures which are used to refresh the textures of bulletin boards.

Useful application modes and features include:

Building Overview, Status & Navigation: Fig. 1 shows an overview with the building layout and colors above rooms indicating room status (e.g. occupied, visitor, scheduled meeting). Users may pan and zoom to views details such as contents of meeting room displays. A popup list with names of people and rooms can be used to find paths to desired locations.

Site Tours: Virtual walk-throughs can give a sense of what it is like to be at the location and illustrate navigation paths by showing a 'first person view' of motion along the path.

Virtual Posters: In the research lab setting, these can be drawn over offices to show the faces and recent talks of office occupants. In the conference venue setting these can show session titles or activities associated with rooms.

Timeshifting: Explorer maintains a history of the mirror model, including device tracks, textures associated with displays, etc. A viewer time-line control allows navigation through time. Users may see the display of slides from previous talks, or virtual posters showing the session and talk schedules for each room the following day.

IV. BACK END SENSOR NETWORK AND FUSION

Explorer integrates sensors used for device localization, person tracking and for updating the model (e.g. display contents, room occupancy, etc.) at a variety of scales. Some areas of a building support more detailed tracking, so that the mirror world can indicate the positions of people in a room, or even the body pose of a presenter. For some other areas, the system may indicate which rooms are occupied, but with no detailed tracking. Environmental sensors include Kinect cameras used to track people in some public areas, a Cisco RLTS WiFi System, radio beacons used to determine if people carrying active tags are close to beacons. Portable sensor types include audio, image, WiFi scan, and motion sensor.

The Android app collects WiFi scans which are used to provide room level position estimates with a nearest neighbor classifier. Users may assert their room location by a 'long touch' gesture on the device or their position within a room by double tapping. This is used for to provide training samples, and to determine ground truth for experiments. Users may also request more a more precise localization which causes the App to collect images and audio samples. The audio samples are used for acoustic background based classification similar to [4], and also by a TOF acoustic localization system in some areas. That system, described in [3], plays low-energy pseudo-random sequences (sounding like soft white noise) from speakers at known locations. When correlations are found between recorded audio and any of the sequences, it is known that the device is in the proximity of that speaker. When correlation peaks are found with signals from 3 or more speakers, position is determined by trilateration. In evaluations the system achieves an accuracy of better than a meter over 90 percent of the time in a 20'x17'x8.5' meeting room.

Images are checked for QR codes posted in known locations, and for SIFT feature matches with other images at known location. When sufficiently many feature matches are found in an image the device location can be determined. When sufficient matches are found to determine a homography, the images can be used to update textures in the model.

REFERENCES

[1] D. Gelertner. *Mirror worlds*. Oxford Univ. Press, 1991.
[2] J. Lifton and J. A. Paradiso. Dual reality: Merging the real and virtual. In *Proc. of the First Int.l ICST Conf. on Facets of Virtual Environments*, July 2009.
[3] I. Rishabh, D. Kimber, and J. Adcock. Indoor localization using controlled ambient sounds. *Submitted to IPIN 2012*.
[4] S. Tarzia. *Acoustic Sensing of Location and User Presence on Mobile Computers*. PhD thesis, Department of Electrical Engineering and Computer Science, Northwestern University, 2011.