# Automatic Acquisition of 4D Urban Models and Proactive Auditory Service for Enhanced User Experience

Sang Min Oh, Grant Schindler, Frank Dellaert, Bruce Walker, Jeffrey Lindsay
Georgia Institute of Technology
{sangmin,phlosoft,dellaert}@cc.gatech.edu, bruce.walker@psych.gatech.edu

## Abstract

We introduce 4D-City and System for Wearable Audio Navigation (SWAN) where computer vision and wearable computing technologies are introduced to augment urban user experience. The goal of our work is to provide users with (1) an interactive 4D model of a city which is automatically acquired from a large set of images, and (2) proactive auditory guidance system on a wearable computing platform which localizes robustly even under the noisy sensor signals within the urban environments.

The preliminary results on combination of both works demonstrate the potential impact where the users can interact intimately with one's own urban environments while revealing unknown history of a city with additional enhanced auditory support.

## 1 Introduction

We introduce 4D-City and System for Wearable Audio Navigation (SWAN) where our goal is to provide users with the ability to intimately interact with the surrounding urban environments via the introduction of computer vision and wearable computing technologies.

Our first work, 4D-City, addresses the problem of constructing a spatio-temporal model of a city through which a user can interact in three dimensions and *time*, a 4D interactive immersion in a collection imagery [4, 5]. The proposed system can offer a radically different and much more powerful user experience than the traditional search/browse paradigm.

Secondly, SWAN is a highly functional auditory guidance wearable system which robustly localizes in the urban environments which include GPS, inertial sensors, pedometer, RFID tags, RF sensors and computer vision. The robust localization capability and developed auditory interface let us to address the problem domains where the system can benefit the user experience. For example, the system can aid people to navigate in the unknown environments, let users know interesting or secret stories about historic sites or proactively provide useful information about current location.

Figure 1: A sample of 3 historical photographs from the Atlanta History Center's collection. All images contain the historic Candler building.

The two systems above can potentially be combined and autonomously improve their functionalities by updating the 4D models of cities based on the incoming streams of images and other sensory signals collected from SWAN, which will, in turn, aid the users with endlessly updating 4D models of the environment.

## 2 4D City : Interactive Archive

The 4D-City work focuses on constructing a 4D spatio-temporal model of urban environments. To highlight the educational, cultural and historical value of 4D-City work, consider the three images in Figure 1, which show the sample images of the historic Candler building in Atlanta, the first headquarters of the Coca-Cola Company.

From the sample images, a 3D building structure can be obtained via multi-view computer vision techniques [2], some of which are illustrated in Figure 2.

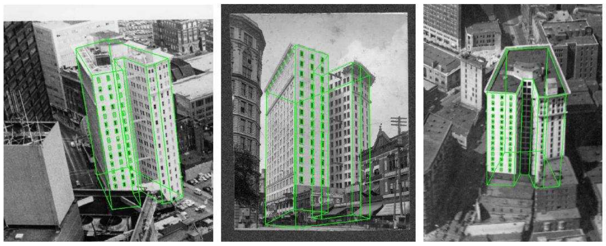

Figure 2: Three images of the Candler building taken from three different view points where the 3D models are overlayed.

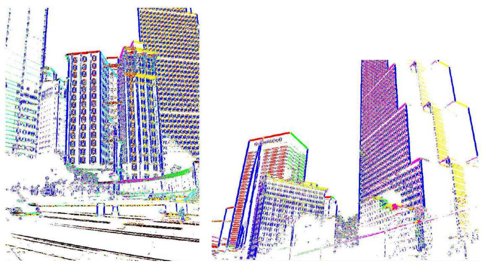

Figure 3: Estimated vanishing directions for two complex urban scenes containing multiple pairs of horizontal orthogonal vanishing directions

In addition, Figure 3 shows that the vanishing directions of complex scenes can be automatically estimated using probabilistic inference method [4]. Iteratively, a rich set of computer vision techniques are applied to correctly estimate the time-varying 3D model of the city from a large set of images available.

The preliminary results of the 4D estimation work can be seen in Figure 4 where it shows a screen shot of the developed prototype kiosk application. The system will be able to allow users to navigate through the model interactively, both temporally and spatially with the advent of powerful graphics processors (GPUs) on every desktop. The potential impact of 4D City work would be even more substantial when we imagine the time point when the built 4D model is open to the public via Internet services, and allow the users to actively contribute new information into the model. Such a model would create a new augmented reality space which would be maintained by the residents of the city, which would truly result in a 4D space with collective intelligence.

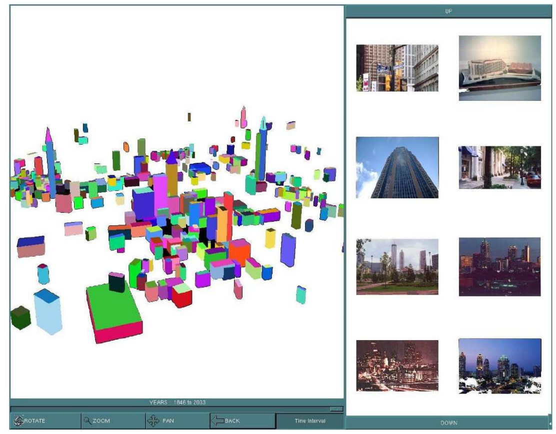

Figure 4: Screen-shot of a prototype kiosk application developed at Georgia Tech, which shows a sample image of the preliminary results on 3D model acquisition

Figure 5: Prototype SWAN system. The helmet is equipped with multiple camera rig and the system aids users through auditory interface.

# 3 System for Wearable Audio Navigation (SWAN)

SWAN is a wearable computing platform where the system is targeted to aid users with location specific information through auditory interface. The system is equipped with multiple types of sensors which include GPS, inertial sensors, multiple camera on the helmet, to name a few. The the prototype system is shown in Figure 5.

The overall functionality of SWAN system critically depends on two main factors which are localization ability and effective auditory interface.

It is well known that the robust localization in urban areas relying solely on GPS signals is very challenging, due to signal occlusion and reflections on the building surfaces. To resolve localization ambiguities, we developed a probabilistic localization algorithm which incorporates the available geometric map information and reliably localizes the system [3]. Although the geometric map information is being obtained through rather manual process within the boundary of Georgia Tech for now, the 3D models acquired as the result of 4D-City work will be considered as plausible inputs once 4D City work achieves results with sufficient accuracy.

Once the user's location and heading direction is determined, SWAN uses an audio-only interface, a series of non-speech sounds called "beacons", to guide the listener along a path, while at the same time indicating the location of other important features in the environments.

To develop the auditory interface, we have also built a virtual reality test environment in which we can try out new audio features, study usability, learnability, and develop training and practice regimes [6]. For example, SWAN includes sounds for the following purposes (for the full listing, please see [6]) :

- Navigation Beacon sounds guide the listener along a predetermined path, from a start point, through several waypoints, and arriving at the destination.

- Object Sounds indicate the location and type of objects around the listener, e.g., furniture, fountains, doorways and etc.

- Surface Transition sounds signify a change in the walking surface, such as sidewalk to grass, carpet to tile, level corridor to descending stairway, and etc.

- Locations, such as offices, classrooms, shops, bus stops, are also indicated with sounds.

In addition to the functionalities mentioned above, we can easily imagine that the robustly localized wearable system can help people to navigate in the unknown environments, tell users interesting/secret stories about historic sites or proactively provide useful information on current location.

Moreover, SWAN might be used to record the user's experience in the form of an automatic diary, as partly studied in [1], which will time-stamp the user's daily activities and potentially generate image diaries of user experience using the cameras equipped on the system.

# 4 Conclusion & Future Work

We introduced our collective work which targets to enhance user's interaction with surrounding environments virtually via 4D-City, and physically via SWAN. The preliminary results of 4D-City system demonstrates that the proposed way of constructing a structured knowledge archive from the exponentially increasing, however densely connected, images can benefit our daily lives by providing us with an alternative way of searching and browsing the information associated with environments.

In addition, the studies on SWAN suggests that the wearable computer users can be greatly benefited through the pro-active auditory interfaces which provides a large range of services based on the location of the users.

There are some future work worth mentioning. First, the two systems above can be combined to produce evolving knowledge archives which can collect new images and additional commentary information from SWAN end-users to update the 4D-City models. On the other hand, SWAN systems will be able to provide enhanced and up-to-date knowledge base by retrieving the necessary information from the newly evolved 4D-City models. The combined model would potentially generate a new ways of information sharing and collective intelligence which are not bound within desktop environments anymore.

# References


[1] D. Ashbrook and T. Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal and ubiquitous Computing*, 7(1), 2003.

[2] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[3] S. M. Oh, S. Tariq, B. Walker, and F. Dellaert. Map-based priors for localization. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2004.

[4] G. Schindler and F. Dellaert. Atlanta World: An expectation-maximization framework for simultaneous low-level edge grouping and camera calibration in complex man-made environments. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[5] G. Schindler and F. Dellaert. Line-based structure from motion for urban environments. In *3D Data Processing Visualization and Transmission*, 2006.

[6] B. N. Walker and J. Lindsay. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors*, 48(2):265– 278, 2006.