2.7 Urban Sensing (URB)

Urban sensing brings innovative measurement systems into everyday life. Sustainable design, healthy living, and effective stewardship of the world’s limited resources require a deep understanding of how countless individual actions generate global effects and how individuals relate to their local environments—natural, built and cultural. Until now, scientists, NGOs, policy-makers, and the public have had to choose between examining the broad characteristics of large populations and looking at small groups in detail. Urban sensing targets technologies and applications that transform our capacity to help individuals, families, and communities to monitor and improve their health behaviors, adopt sustainable practices in resource consumption, and participate in civic processes. Its vision is of distributed data collection and analysis at the personal, urban, and global scale, often using “everyday” technologies like mobile handsets, in which participants make key decisions about what, where, and when to sense.

The area is entering its second generation of technology research and public pilots. We have learned from the initial deployments where our platforms need to be more robust and capable, and how to strengthen their relevance. New mobile handset, server and mote-class platforms have been incorporated into and we have expanded our user interface development efforts to provide more usable tools for individuals and groups. With the encouragement of the CENS External Advisory Board, we continue to deepen our approach to participatory privacy regulation of sensing that involves people, and are building the legal and technical foundations to support ethical design and deployment. Based on last year’s experience in supporting medical and environmental science studies, CENS has developed the expertise to include IRB-authorized human subjects research in three of its projects this year, with another application in progress.

Health, sustainability, and civic engagement have solidified as key themes for our applications research. The technologies we’re working with are well-suited to longitudinal and media-rich data collection needs in these areas. Faculty collaborations include new work in personal health and wellness with the Semel Center for Children and Families, as well as continued research on time-location travel patterns with the Dept. of Environmental Science and on cultural and civic applications with UCLA REMAP. The field continues to expand, as exemplified through a recent workshops and conferences (UrbanSense, MODUS, MobiUS) with submissions from CENS, Dartmouth, Carnegie Mellon, and many others. CENS research remains distinguished by the combination of (1) a participatory focus, which emphasizes the agency of individuals in deciding what to sense, when, and for whom, (2) the use of model-assisted sensing, in which widely-available modalities, such as location, are run through models to generate higher-level inferences, and (3) pilots with domain experts, as described below.

Pilot deployments

CENS has several ongoing pilot deployments of urban sensing systems, including mote-class hardware in the Spotlight project and mobile handset-based sensing in several projects. Over 150 Symbian, Windows Mobile, and Android mobile are currently distributed, with 66 active data plans supplied by CENS, through help from T-Mobile USA. This infrastructure supports real-world testing of a variety of applications. For example, the Personal Environmental Impact Report (PEIR) has been in continuous operation since the Spring of 2008, with a pool of about thirty test users, approximately five of which are active uploaders of location data at any given time. PEIR is an example of a “mobile plus web” application; it performs activity classification and model-based analysis of time-location traces gathered from mobile handsets to generate a personal report on environmental exposure and
impact. In February, 2009, we launched a public pilot of the project with high school students in the Bay Area in collaboration the GoGreen Foundation, Nokia Research, and AT&T Mobility. The project plans to scale to about a hundred users within six months. PEIR was also invited for public demonstration at Wired NextFest last year, through which we learned of the deep interest of the public in such systems.

Experience from these and other deployments has guided our systems and testing approach for a second generation of systems included in this report. Specifically, it has (1) encouraged us to improve our representation of time-location traces from the basic “bag of points” approach (2) required near real-time model calculations in several cases, so that feedback can be provided to users soon after a sensed activity, (3) demonstrated the need for a robust framework for management, debugging, provisioning, and performance analysis of mobile sensing systems to enable more effective and scalable deployment, (4) shown that power consumption for continuous data collection remains a limiting factor for certain types of mobile sensing that will require improved sampling and communication techniques to mitigate, and (5) emphasized how important user interface design is to participant engagement and the ability to understand a system’s operation and data output.

**Technology**

Through support from NSF NeTS-FIND, Google, Nokia, Samsung, and Sun, we continue to develop technology platforms to support three types of urban sensing: (1) **Top-down** data gathering, in which domain experts design experiments or investigations and are the primary recipients of the resulting information. (2) **Bottom-up** data gathering, in which members of a community initiate, manage, and use coordinated data collection (3) **Self-reflective** data gathering, in which an individual gathers and consumes information from the system in a process of self-discovery. As our applications mature, it is clear that many systems are actually combinations of these three approaches.

Reusable platforms developed by CENS and used in many urban sensing projects include: (1) **Campaignr**, an XML-configurable mobile handset based data collection application, whose performance and stability on Symbian S60 was greatly improved this year and was ported from scratch to Windows Mobile; (2) **SensorBase**, described in the Data section of this report, whose query, programmatic control, speed, and storage capacity have been enhanced by its development team in response to urban sensing and other application area needs. Additionally, we are collaborating with Nokia Research Palo Alto on the use and power consumption analysis of **Nokoscope**, a modular background data collection platform for Nokia handsets, and USC has developed the **Urban Tomography** platform for robust mobile video collection. Through these and other application-specific platforms, we incorporated sensing on mobile handsets running Symbian, Windows Mobile, and Android, as well as SMS-based interaction from any phone, GPX data loggers, mote-class hardware, with a bridge to personal vehicle computers (and their sensors) available shortly. All of these data sources and platforms are used for insight into longitudinal assessment relevant to our key application areas described above.

Data collected and stored on these platforms is analyzed using a combination of activity classification, participation reputation and mobility profiling, and other approaches developed at CENS, as well as accepted scientific models (such as the California Air Resources Board’s Emissions Factors (EMFAC) model in PEIR) that have been incorporated into CENS systems. **Activity classification** algorithms are central to several projects. Similar to context awareness research, but intended to generate an annotated data stream for use in model-based processing and/or available to the participant, these apply statistical, signal processing, and geospatial analysis techniques to go from GPS time-location traces to classifications such as “driving”, “walking”, or “bicycling” or bluetooth proximity and audio level data to “at the dinner table” or “out for a walk.” To visualize and interact with raw and derived data we have focused on developing web standards based interfaces (e.g., the use of AJAX and CSS-based web interfaces for CycleSense and Flash for Remapping LA, and both in PEIR). Through the **GeoSIM** project, we have begun to explore next-generation 3D visualization and data fusion for community-based geospatial mapping and modeling.

**Ethics, Law and Policy**

This year, the area received funding from the NSF Ethics Education in Science and Engineering program for continued work to develop a formal ethics framework for participatory sensing, as well as study practices at CENS
and create related educational materials and courses. We have now applied for and received IRB approval for human subjects research in our own projects, including survey of our own participants and designers, and implemented an internal process at CENS to evaluate new pilot and study ideas for human subjects research issues. Our exploratory work on participatory privacy regulation and selective sharing mentioned in the last report has resulted in two new concrete efforts: first, to develop and implement an API for a “personal data vault” for time-location data and second, in collaboration with Professor Jerry Kang of the School of Law, to develop and propose legal principles and technical infrastructure that could help to protect not just data in the vault, but data shared and derived from it.

**Planned Work**

Based on this experience, we will delve deeper into the systems architecture and legal framework needed to support the “mobile personal sensing” that has become central to both participatory, community-oriented data collection and structured data collection in institutionally-organized settings. The data vault and privacy efforts are key to that, as well as work in parsimonious activity classification described in this report, which attempts to derive useful activity information from less intrusive (and less power-hungry) components than GPS. Geospatial data analysis and display continue to be important components for many of these systems, and we are exploring new research collaborations in these areas. Finally, the important opportunity to fuse data collection from multiple sensor classes, including application-specific wireless sensors, mobile handsets, and more complex instrumentation will be explored as we seek multiscale and multimodality sensing to best support our applications.
**URB 01 Campaignr**

**URB 01.1 People**
- Principal Investigators: Jeff Burke, Deborah Estrin
- Faculty: Deborah Estrin, Computer Science, UCLA; Jeff Burke, Theater, Film and Television, UCLA
- Graduate Students: Jason Ryder, Computer Science, UCLA; Nicolai Munk Petersen, Computer Science, UCLA
- Staff: Taimur Hassan

**URB 01.2 Overview**
Campaignr is a software platform originally developed for the Symbian S60 3rd Ed. Operating System and extended to Windows Mobile this year. It enables mobile handsets to act as general purpose sensing devices. The platform is currently being used at CENS in the context of Urban and Participatory Sensing research. It is among the primary platforms with which data collection is performed. Due to stability issues in regards to long running data collections, Campaignr had been replaced by other custom single-purpose programs for use in many data collections. However, over this past year, the stability of Campaignr has been drastically improved.

**URB 01.3 Approach**
Campaignr was designed to be general-purpose so that it would fulfill the sensor needs of a broad range of urban and participatory data collection purposes and be easily configurable by a Campaign designer. It accomplishes this by allowing for a user-configurable campaign xml file to define which sensors, what sampling rate, and what upload destination should be used during data collection.

In order to achieve a high level of customization the design of Campaignr uses many abstract interfaces internally. This, and the complexity of developing native Symbian applications, had led to many unforeseen issues with the software platform. The work focus on Campaignr during this past year has been to make its multi-purpose functionalities very robust (i.e able to run for many days or weeks without requiring a hardware or software reset).

**URB 01.4 System(s) Description and/or Experiments**
Campaignr plays a key role in the systems developed by the Urban Sensing group. It facilitates the collection and transmission of sensor data to a back-end system for post-processing. Data transmission is supported over the cellular network as well as WiFi, where available. Campaignr for S60 has been refactored over the past year to provide robust fault tolerance. Campaignr is capable of recovering from device failures (e.g. Bluetooth, GPRS radio), as well as data corruption. This allows Campaignr to be run for long periods of time without the need for system reset and has been verified through round-the-clock experimentation with the latest version(s) of Campaignr.

The transmission channels for data now support parallel upload to improve performance. Upload is supported over unsecured and SSL/TLS encrypted channels to send data securely. Secure data transport is especially important for the potentially sensitive nature of the data collected.

Another significant contribution to the Campaignr project is the addition of a Windows Mobile port of the original Campaignr’s feature set. The Windows Mobile port of Campaignr will be referred to as CampaignrWM below.

The Windows Mobile port seeks to provide the same features as the S60 version. In addition the port also provides many new experimental modules. Key features include dynamic loading of new sensors, pre-processing on device, processing of incoming SMS and e-mails, and decoupling of upload managers. Hence, it is now possible to quickly write new upload modules for CampaignrWM. As proof of concept of the modular architecture, CampaignrWM has been extended to support both Flickr and Twitter data upload, in addition to the Sensorbase and PEIR formats.

The core components of CampaignrWM was built for Microsoft Compact Framework. The robustness and safety of the framework reduces the complexity of development and makes CampaignrWM a viable alternative for research on the mobile platform. The Compact Framework is a managed environment and has as such certain inherent
resource constraints. So, when necessary CampaignrWM supports the development extensions in native code (C++) - e.g. when one needs to interact with the native API or decrease processing overhead. CampaignrWM supports this through COM and an example of this inter-operation can be readily found in the provided camera sensor.

CampaignrWM was developed using the Footstep project as a driver application, which measures walking activity by leveraging cell phone, GIS, and sensor technologies. We built the system to provide sedentary adults insight into their daily physical walking patterns. The Footstep project was built in parallel to the on-going development of CampaignrWM and was also used to test the end-to-end functionality and reliability of the latter. The Footstep application makes extensive use of pedometer data. It was therefore necessary to integrate an external pedometer with CampaignrWM. A tailored and self-contained driver library was created for CampaignrWM to support a Nike+ iPod wireless piezoelectric sensor. Due to hardware limitations it was only possible to read low-resolution data from the device. Despite this limitation, the modular sensor integrated and loaded into CampaignrWM with minimal effort. We discovered that external wireless sensors are prone to interference and may therefore become disconnected at any time. To solve this problem CampaignrWM incorporates a sensor watchdog, which will attempt to reconnect if a external device does not respond in a timely manner.

Next, Campaignr for S60 will be the primary collection platform for the new Aging in Place project which is a system being designed for elder care through sensor networks and pattern recognition. It also continues as the primary data collection platform for the PEIR (Personal Environmental Impact Report), including it’s public trial with the GoGreen Foundation.

URB 01.5 Accomplishments

During the past year, Campaignr has become a highly reliable platform for data collection. Previously, there existed issues with non-uniform sampling periods due to blocking operations occurring (primarily, cellular network connection and extensive logging to the phone’s external memory). These issues are no longer present in the current version. All network connection is done in a non-blocking manner, so as not to interfere with data collection. Also, logging is now done on the phone’s local memory (which provides a significant performance increase) and moved over to the external memory card during free CPU cycles (this reduces its interference with data collection and upload to almost nothing).

Also, CampaignrWM now offers an additional exciting opportunity for rapid prototyping of mobile and pervasive sensings platform. Future work includes adaptive upload component, device and user authentication, and real-time feedback components.

Public Available Source Code and Binaries. Campaignr also found a new home this past year at campaignr.com. From here anyone can access the free and open-source code base which includes documentation, source, and pre-compiled binaries for both Symbian S60 and Windows Mobile 6.0. CampaignrWM is available as a signed Microsoft Mobile2Market application and can therefore be used on most compatible devices. Campaignr for S60 is available signed for a select number of devices and unsigned for anyone who which to sign and deploy the application. After the launch campaignr.com has had visitors from around the world.

URB 01.6 Future Directions

It is likely that Campaignr development tasks will be largely focus on maintenance for the foreseeable future. The application is at a point where improvements that are needed are sparse and can be accomplished on an as-needed basis. There will be some development with the Aging in Place project in mind. This will consist mainly of better error handling for uploads and multiple-record-uploads to improve performance.

URB 01.7 External Research Partnerships

Nokia
Samsung
URB 02 Personal Environmental Impact Report (PEIR)

URB 02.1 People

- Principal Investigator: Jeff Burke, Deborah Estrin, Mark Hansen, Ruth West
- Faculty: Jeff Burke, REMAP, UCLA; Deborah Estrin, Computer Science, UCLA; Mark Hansen, Statistics, UCLA; Ruth West, CENS, UCLA
- Researchers/Staff: Nicolai Petersen, CENS Researcher; Vinayak Naik, CENS Researcher; Mohammad Rahimi, CENS Researcher; Alexis Steiner, CENS Researcher
- Graduate Students: Frank Capodieci, Computer Science Department, UCLA; Gong Chen, Statistics, UCLA; Andrew Mondschein, Urban Planning Department, UCLA; Min Mun, Computer Science, UCLA; Jason Ryder, Computer Science, UCLA; Katie Shilton, Information Studies, UCLA; Calvin Wang, Statistics; Nathan Yau, Statistics
- Undergraduates: Frank Chen, Computer Science Department; Doris Lam, Computer Science; Stephen Oakley, Computer Science Department; Peng Wang, Design | Media Arts

URB 02.2 Overview

The Personal Environmental Impact Report is a participatory sensing [1] application that uses location data sampled from mobile phones to calculate personalized estimates of environmental impact and exposure. This project focuses on software that runs on the phones to collect the data, server-side components to manage the data, and a user interface to display the data both online via a web dashboard and map interface, and in a social network setting like Facebook.

PEIR’s emphasis on a reciprocal relationship between environmental impact and exposure is operationalized in its indexing of user location-trace data into models and external datasets. We envision future extensions of PEIR through additional modules for different types of environmental impact and exposure models, or through application to multiple geographic regions. This extensibility is made possible by ongoing work in complementary research areas at CENS towards increasing the sensitivity and accuracy of our activity classifiers which comprise the inferencing framework for PEIR, as well as research in software platforms for data collection, adaptive sampling schemas to reduce mobile device power consumption, and designing systems that are privacy-preserving while enabling collection and sharing of user location-trace data.

URB 02.3 Approach

This project proposes and demonstrates how location time series data acquired from GPS-enabled mobile phones can be used as an index into geospatial models to infer personal environmental impact and exposure. Our current implementation of PEIR presents four travel-related metrics (models) to the user: carbon impact, exposure to airborne PM 2.5 particulate matter (smog), emissions of PM 2.5 particulates near sensitive sites such as schools and hospitals, and their exposure to fast food eateries. (See flow diagram below)

These four models were derived from preexisting external data sets such as traffic conditions models and EMFAC, and then transposed into our processing framework. This framework not only serves as a database and store, but also performs the processes of the four models. Additionally, map-matching techniques were used to create activity classifiers that could infer mode of transportation based on location and speed. Then, easily gathered location time series data was used as an index into these geospatial models to infer personal environmental impact and exposure.

URB 02.4 System(s) Description and/or Experiments

A GPS-enabled mobile device (e.g. Nokia N80 with external GPS, N95 or E71) running Campaignr or Nokoscope [2] is used to collect and upload data for PEIR. Experiments were done to test battery life (using CENS engineered extended battery packs), bluetooth connectivity/ stabilization to GPS device, and upload of data over-the-air using
both SIM card and WiFi access points. Campaignr enables automated time location series data collection as well as data upload over-the-air, upgrading/tasking, and automated or manual annotation. GPS sampling occurs at 30-second intervals. We experimentally selected this lowest possible sampling rate that resulted in good activity classification of travel by car while conserving power and battery. We also collect cell tower ID, radio signal strength, several timestamps, phone ID, battery information and manual annotation of activity for debugging and testing of future features. Additionally, we are testing the use of Nokoscope, a software platform with our collaborators at Nokia Research Center to enable collection of data from mobile devices that may provide power consumption improvements over Campaignr. PEIR currently supports Symbian s60 3rd edition phone clients, and Windows Mobile. We are developing location-time series clients for the Android and iPhone platforms. Bulk time-location data can be uploaded in most tracklog formats, such as GPX by integration with GPSBabel software. Location traces are received by a location data multiplexing service (LocMux) that forwards it to both testing and production servers and send a copy to an archival store.

PEIR data flow diagram.

PEIR system architecture
Data Annotation and Analysis are accomplished by web and server-side applications that analyze individual spatio-temporal patterns (location traces), calculate corresponding impact and exposure metrics, and display impact and exposure metrics to users. Location trace processing occurs in four stages as shown in the figure above.

Activity classification enables the exposure and impact models to be applied automatically to user location-traces. The initial implementation of PEIR uses a simple algorithm to classify points as walking, driving or staying in place using GPS-derived speed and map matching. Our classifier users freeway annotation information in addition to speed values as feature inputs to the classifier. Our experiments show that use of this freeway information and map matching technique as part of our classifier increases accuracy from 40% to 82%, based on test data collected while a user was driving in heavy traffic on the highway.

A web-based user interface displays impact and exposure in terms and graphics understandable by a layman. The interface uses a dashboard of activity and associated map-based visualization of user trip data color-coded by level of impact or exposure type. Users explore their data by selecting trips on the map or within a trip log detail view.

Detailed location and impact/exposure metrics are available to users in a private account. Aggregate data as combined impact and exposure scores are made available via a PEIR Facebook widget or PEIR users can also compare their overall impact and exposure scores as well as trends over time with other PEIR users in their online network accessible from their dashboard.

**URB 02.5 Accomplishments**

PEIR exists as an end-to-end functional prototype system with two sets of active users. A private beta-testing user group based in southern California, and a public trial group in northern California. The public trial is being conducted in collaboration with GoGreen Foundation [3], Nokia Research Palo Alto and AT&T. The trial will involve students from four high schools. They and their families are engaging in a sustainability challenge to raise awareness about environmental impact over a six month period starting February 2009. For this system we continue to refine the software platforms for data capture and upload, having created a version for Windows.
Mobile in addition to the existing Symbian s60 version of Campaignr, and are working on a platform for Android. In addition we created a bridge to Nokia's Nokoscope data gathering platform. Within this year we completed the first version of web-based tools for the user interface implementing the PEIR dashboard, map-based visualization interface, trip annotation and calculation detail view. This user-facing system pulls data from a PostGIS data store, and is implemented in PHP and Flash, served by an Apache Httdp webserver, using the Wordpress blog engine. (See figures above for dashboard, network and detail trip log views)

The near real-time modeling and calculation of exposure and impact metrics relies upon a set of modeling assets and distributed data resources. These include: roads and locations, weather, traffic conditions, and vehicle emissions models. From last year to this year, we overcame several technical challenges including context lookup for weather, and performance enhancement to the vehicle emissions model to enable near real-time calculation in addition to the development of a map-based visualization to enable presentation of user data via a web interface as described above.

Maps of local roadways (exposed through PostGIS) are critical for PEIR. These are essential for both map matching in activity classification and in computing exposure to PM 2.5 particulates. We obtain these as well as locations of sensitive sites such as schools and hospitals from Street Pro (ESRI 2006 Maps and Data Collection). For fast food exposure calculations, types of eateries are obtained from the Los Angeles Department of Public Health data. Weather data is obtained through hourly collection of real-time weather from NOAA databases, and served to PEIR's processing modules using a spatial mapping organized by zip codes. To enable geographic scalability of PEIR to regions without real-time traffic measurements, we developed this implementation of PEIR around a traffic flow model from the Southern California Association of Governments (SCAG) that reports bidirectional traffic flow as number of vehicles per hour, types and speeds over six daily time frames. Within our system, this data resides in PostGIS [4].

The vehicle emissions models for PEIR come from the California Air Resources Board (CARB). Their Emissions Factors Model (EMFAC) [5] is a FORTRAN program that computes emissions based on current weather and the speed and type of vehicle. PEIR makes use of EMFAC's PM 2.5 and carbon dioxide emissions estimates. During this year we completed an initial implementation of EMFAC and determined that its performance constituted a

Trip journal with description of calculations for aggregate exposure and impact scores.
bottleneck in our processing pipeline. We overcame this limitation in speed and performance by developing an approximation to EMFAC, computed via a functional ANOVA model. Through phases of repeated fitting and testing, we created a tensor-product spline model that both allowed for fast computation as well as a view in to the dependence of emissions outputs as functions of weather and vehicle calculations. This opened the model for end users and enabled us to display alternative views within the trip log enabling users to see details of the calculations underlying their impact and exposure scores (Figure above). These intermediary calculations are made possible by the above enhancements to model calculation and performance.

Within this year, our work also resulted in improved map matching using intersection-based map matching to give improvements over naïve map matching. We created a GPS/GSM-enhanced activity classifier to enhance accuracy of classification in situations where the user is indoors or moving at slow speeds on a freeway or road. Our experiments show that the GSM-enhanced classifier achieves an overall accuracy of 91% and works well when users are on surface streets and highways, as well as indoors. It will require further analysis to assess its implications for PEIR system performance.

**URB 02.6 Future Directions**

The PEIR project has a great deal of potential for continued development. One key focus of PEIR is the continued development of impact and exposure models as well as further development of a web-based user interface to inform and advise user, provide reports, real-time feedback, visualizations and exploratory data analysis tools for non-professional users. For security and confidentiality of data we developed a secure database that incorporates data encryption through the exchange of keys and also restricting access to the data by allowing only the user to access the data. Additionally we plan to enhance the system by allowing users to delete data remotely through the web-based user interface.

In other areas of the system we plan to continue our efforts for improving power consumption management and will explore adaptive sampling and techniques that determine if users are indoors versus outdoors and adjust sampling accordingly. With the concept of a “trip” functioning as the atomic unit of data for end users, it will be of interest to explore moving from current point-based representation of trips, to non-point based representations and the potential for system optimization and performance improvement as the number of users increases. Given the nature of the data, an additional future goal for PEIR is the explicit inclusion of uncertainty of the captured and computed data along with the primary outputs of the models.

One highly requested feature for PEIR is on-phone status display and feedback. The current implementation is designed to run in the background unobtrusively on the mobile devices. Future enhancements to the user experience will explore the role of on-phone display/feedback components.

While our activity classifiers have demonstrated improvement over this year, an important goal for PEIR is continued improvement in activity classification with more specific modes and higher accuracy. An eventual goal is to enable the system to discern between motorized transportation modes such as bus versus car, and other modes of public transportation. For this we will explore techniques to leverage additional on-board sensor types (e.g. accelerometers and WiFi) and additional GIS and map data such as bus routes.


**URB 02.7 External Research Partnerships**

Nokia Research Center, Palo Alto (current)

Péter Boda, Principal Scientist, Nokia Research Center

Henry Tirri, Research Fellow and Head of Systems Research, Nokia Systems Research Center

Sun/Jim Waldo (Planned)

Topic: Parallelization schema for scaling PEIR processing pipeline.

Centre for Sustainable Communications (planned)

Stockholm Royal Institute of Technology

Topic: Localization of PEIR to Stockholm and user trial for research in persuasive technologies in encouraging sustainability behaviors.
URB 03 Inferring Everyday Mobility States using GSM & Wi-Fi Traces from Mobile Phones

URB 03.1 People

- Principal Investigator: Deborah Estrin, Mark Hansen, Jeff Burke
- Faculty: Deborah Estrin, CS, UCLA; Mark Hansen, STAT, UCLA
- Staff: Jeff Burke, REMAP, UCLA
- Graduate Students: Min Y. Mun, CS, UCLA

URB 03.2 Overview

Inferring mobility states such as being stationary, walking, or driving is critical for many applications in transportation studies, urban planning, health monitoring and epidemiology.

Our focus is on building a pervasive mobility classification system using mobile phones with the goal of large deployments. More specifically, the classification models should provide the following:

- **Low Processing Complexity:** Although mobile phones have advanced drastically in functionality and capability in the past few years, they still have limited resources in terms of CPU and memory (RAM). The classification models for mobile-based applications should be lighter weight and supported alongside everyday applications in a robust and resilient manner.

- **High Energy Efficiency:** People carry mobile phones throughout the day. Running power-expensive applications will significantly reduce the lifetime of mobile phones and burden users; users need to recharge their battery in the middle of the day; otherwise, they would not be able to use their phones for communication usages as well as application services.

- **High User-time Coverage:** To recognize users activities in everyday situations, systems must have ubiquitous user-time coverage. GPS positioning is available as little as 5% of a typical person’s day, while GSM and Wi-Fi coverage is available throughout the day [1].

URB 03.3 Approach

GPS positioning [2-6], external geoindexes such as map information, commercial and custom devices such as pedometers [7,8] and body sensors [9-12] have been widely used to determine people’s fine-grained activities. Unfortunately, it is not always practical to deploy mobility-based sensing applications [13] using such activity classifiers on a large scale due to their cost, complexity, obtrusiveness, coverage and energy profile. In particular, GPS-based mobility characterization raises many issues such as spotty coverage and battery drainage that makes it inadequate to meet all application goals.

Mobile phones are widely deployed and people carry them everyday and all the time, which indicates that the mobile phone is one of the most suitable devices to use in recognizing people’s everyday physical activities.

We propose a new mobility classification method using GSM and Wi-Fi traces that are available on many commercial mobile phones. We adopted C.4.5 Decision Tree as our inference model. It is light and simple, which makes our model suitable to be used in mobile phones for its low processing complexity. Sampling GSM and Wi-Fi data uses relatively low energies, supporting high-energy efficiency. In addition, GSM and Wi-Fi coverage is nearly ubiquitous, which provides a solution to high user-time coverage. Exploring how mobility classification can be performed using GSM and Wi-Fi observations contributes to profiling unconstrained mobility states throughout a typical day on a large scale for the participatory sensing applications including our ongoing application. Our Personal Environmental Impact Report (PEIR) project uses automated mobility-annotated location traces as an input to models of hazard exposure and environmental impacts as described previously. The work described here
could enable PEIR to offer its participants personalized information on environmental interaction throughout the day in an unobtrusive way by sharing only coarse-grained location.

**URB 03.4 System(s) Description and/or Experiments**

**GSM and Wi-Fi**

GSM is the most popular standard for mobile phones in the world. Information from the cell ID provides a rough indication of a person’s position. Features derived from this information, such as the number of changes in the associated cell IDs for certain duration, can substitute for the speed values from GPS data that has been widely used to identify mobility states or activities. Unfortunately, given the nature of large GSM cell sizes, a person’s locations and mobility states within a single cell cannot be identified well from GSM data alone. Therefore we explore the use of data from networks with smaller cell sizes, such as Wi-Fi. We found that the accuracy of our model increases by 10% when we employed both GSM and Wi-Fi information compared to using GSM alone.

In addition, GSM and Wi-Fi coverage is nearly ubiquitous, providing high user-time coverage, while GPS positioning do not work in many situations. And sampling GSM and Wi-Fi uses relatively low energy and provides high-energy efficiency as shown in Table 1.

### Feature Selection

A large portion of mobiles including our experimental device, Nokia N 95, does not have access to cell tower information for cells other than the one it is associated with. In our studies, we use single connected cell ID information to extract features discriminating the mobility states. The intuition is that users see more cells for certain duration, as they move faster. The following features are extracted from GSM traces, where w is the sliding window size:

- **Number of Unique Cell IDs (C unique,w )** is the number of unique cell IDs to which the phone was connected for w second segment.

- **Residence Time in a Cell Footprint (C residence )** is the duration a user spent in the cell that the phone was associated with. Note that C residence is different from the other two features because we divide data into segments by cell IDs. Thus w value can vary.

- Although Wi-Fi data has many features satisfying our design requirements, it should be carefully used due to it’s limitations. In the same area, users may see various Wi-Fi APs at different times. However, a Wi-Fi AP must be visible for the greatest amount of time among other APs in view while being stationary at one place and we call this AP ‘dominant Wi-Fi AP.’ ”Switch dominance” occurs when users move from one place to another. As users move faster, this would happen more frequently and the duration of the most dominant Wi-Fi AP would get lower. The followings are features extracted from Wi-Fi traces.

- **Duration of Dominant Wi-Fi Access point in View (WF dominant,w )** is the amount of time that the most dominant Wi-Fi AP in each segment is seen.

- **Proportion of Duration of Dominant Wi-Fi Access point in View (WF dominant_propor:on,w)** is equal to \((WF\ dominant,w /V) \times 100\) where V is the number of valid points during w second segment.
Building and Evaluating a Model
We computed the four selected features for each data point. We adopted the C.4.5 Decision Tree as our inference model for its simplicity, which enables our model to provide low processing complexity. It also outperforms other inference models such as Bayesian Network, Support Vector Machine (SVM) and Conditional Random Field (CRF) when features are computed over segments. Ten-fold cross validation method is used for evaluation.

Due to different GSM and Wi-Fi densities in various areas, it is important to study whether we need an environment-specific model, and whether our classifier can be effectively applied to new users without additional training procedures, in addition to further evaluating the general performance of our model. We have two data collection scenarios. First, the author collected data in five differently characterized neighborhoods, Wilshire in Westwood (Wilshire), Palms, UCLA, Marina Del Rey (MDR) and East Culver City (E.Culver), chosen based on the Southern California Association of Governments (SCAG) data; for each area, the data collector conducted the three mobility states for twenty minutes each. Second, sixteen individuals, eight males and eight females between the ages of 20-45, gathered one hour data, twenty minutes for each of the three mobility states.

URB 03.4 Accomplishments

General Results
Table 2 summarizes our results. We use the commonly used matrices to evaluate our model: precision (true positive/(true positive + false positive)) and recall (true positive/(true positive + false negative)) confusion matrices. Precision is the percentage of correct predictions and recall is the percentage of correctly identified ground truth cases. Accuracy (true positive/total number of data points) is the percentage of correctly classified data points. To better understand the performance results of our model, we also built mobility classification models using GSM, Wi-Fi and GPS alone.

First, our model successfully identifies the coarse-grained mobility states with overall accuracy of 80% (precision: 80%, recall: 80%). This accuracy is lower than the one of the GPS-based model, 92% (recall: 92%, precision: 91%). But, as mentioned earlier, sampling GPS is accompanied by many drawbacks.

Second, Wi-Fi beacons with smaller cell sizes and GSM data with larger cell sizes are complementary. The classifiers using GSM and Wi-Fi data alone identify mobility states 70% and 68% correctly. The accuracy increases by 10-12% when the combination is used.

![Table 2. Precision and Recall Confusion Matrices of The Classifiers using Various Sensors](attachment:image)

Impact of Environment
Our model can provide better performance by employing the model using only GSM data when Wi-Fi beacons have no coverage or are too sparse. Generally, adding Wi-Fi features to the model improves the accuracy as explained in the previous section. For example, as shown in Figure 1, the LOAO (Leave One Area Out) based model using only GSM data works poorly in characterizing stationary states in UCLA area. Our model built by using the LOAO method increases the values to nearly 80% by taking advantage of Wi-Fi. But, if the technique is used when Wi-Fi densities are too low, computed feature values can be numerically distant from the rest of the data and it could rather degrade the model. As seen in Figure 2, the performance of our model improves and achieves 82% overall accuracy by adopting the GSM-based method when necessary.
User Variation
To evaluate the scalability of our model, we built the model with the data set gathered from one user and tested it with each of sixteen new users whose data is not in the training data set. We achieve a trimmed average accuracy of 78% and the results are promising. Certain might be unique and a training set that has a broad range of activity styles and environmental settings are necessary for a generalized classifier to work properly.

URB 03.5 Future Directions
The current work explores only the first step required to build a pervasive mobility classification system using mobile phones for the large deployment. There are numerous potential methods left unexplored that may be better suited for our system. One example is using accelerometer data. Accelerometer data has been used to recognize activities having similar speed and acceleration features. In addition, sampling accelerometer data is energy-inexpensive. Thus, accelerometer data has much potential to improve the performance of our model by disambiguating being stationary and walking states, where most of the errors occurred from our model because the two states often have similar GSM and Wi-Fi feature values, while satisfying our design requirements. Our future work involves exploring how our mobility classification system would improve with using accelerometer data.

Figure 1. Leave-One-Area-Out Testing

Figure 2. Comparison between Our Model with and without Adopting the GSM-based Model When Wi-Fi AP Densities are Too Sparse, and the GPS-based Model
URB 03.6 External Research Partnerships
Nokia Research Group(current)

URB 03.7 References


URB 04 Ethics in Personal Mobile & Participatory Sensing

URB 04.1 People
- Principal Investigators: Deborah Estrin, Jeff Burke, Mark Hansen, Katie Shilton
- Faculty: Deborah Estrin, Computer Science, UCLA; Mark Hansen, Statistics, UCLA; Jeff Burke, Film, Theater & Television, UCLA; Jerry Kang, UCLA School of Law; Ramesh Govindan, Computer Science, USC.
- Researchers: Jim Waldo, Sun Microsystems
- Staff: Beqa Dawson
- Graduate Students: Katie Shilton, Information Studies, UCLA

URB 04.2 Overview
The mobile phone network will likely become the largest distributed sensing system on the planet. Mobile phone users, however, are generally unaware of the dual use opportunities, in which their communication devices are also information gathering devices. What are the ethics of coordinating this alternative usage mode for research purposes? Can researchers achieve meaningful consent and active participation of mobile phone users? The three-year Ethics in Personal Mobile & Participatory Sensing research and education project will:

- Research, design and assess a participatory approach to managing privacy in personal mobile sensing applications;
- Create both an immersion curriculum and a seminar curriculum to teach participatory ethics for urban sensing to diverse STEM undergraduate and graduate students;
- Evaluate the curricula and disseminate best practices for education in participatory urban sensing ethics to urban sensing, ubiquitous computing, and broader technology education communities.

URB 04.3 Approach
In personal mobile and participatory sensing, everyday mobile devices become a platform for coordinated investigation of the environment and human activity. But transforming phones into data collection instruments raises both technical and ethical challenges. We believe researchers should utilize this network of sensors with the consent and active participation of users. Facilitating responsible, socially trusted, and participatory ethics for data collection and analysis with urban sensing systems remains an open problem, and is the challenge undertaken in this research and education project. We focus on graduate and undergraduate students who are designing systems not just “for the future” but for ongoing pilot projects that have public participation.

During the research component of this project, we are formalizing, designing and assessing a privacy framework we call participatory privacy regulation. The education component of the project will teach participatory ethics such as participatory privacy regulation through development of two curricula for STEM students: a hands-on laboratory approach to designing ethical urban sensing technologies; and an interdisciplinary seminar-style course. The final phase of the project will evaluate and synthesize classroom findings into best practices for participatory urban sensing ethics education. We will disseminate the practices to educational, computing, and technology communities.

URB 04.4 Accomplishments
During the last year, we have honed our core
approaches to privacy in personal mobile sensing systems: participant primacy, participant autonomy, participatory design, and working with minimal and auditable information. These principles have contributed to the design of the PEIR system, which will include data deletion and sharing mechanisms to engage users in privacy decision-making throughout their PEIR experience. The principles have influenced our research into inferring mobility states using parsimonious activity classification. Principles were also incorporated into the design of parsimonious approaches to establishing participant coverage and reputation for projects such as GarbageWatch. Finally, participatory privacy regulation principles are at the center of an ongoing project with Jerry Kang and Ramesh Govindan to design a Personal Data Stream architecture. The architecture is still under development, but will include a personal data vault, metadata, and filters to allow for user engagement in privacy decision-making.

**URB 04.5 Future Directions**

Creating and piloting the Personal Data Stream will provide a unique architecture for protecting and encouraging individual decision-making about sharing and disclosure. When the PDS architecture is compatible with CENS applications like PEIR and AndWellness, we will be able to assess how participants user and respond to the PDS architecture’s features. As CENS builds a core of users, we will engage in interviews to answer the following:

- How deeply, and under what conditions, do participants engage with participatory sensing systems?
- How do participants in urban sensing negotiate decisions to capture, share, and retain their data?
- How well does participatory privacy regulation support privacy and sharing decision-making in participatory urban sensing systems?
- What are other key ethical questions in participatory urban sensing?

Qualitative data documenting interactions between participants and urban sensing systems can suggest answers to contextual questions about when and why participants make decisions to share or withhold data. Interviews with participants can elicit how participants feel while interacting with the systems and how much participants trust the systems. We will use explicit participant critique of our design methods, software, and conclusions to answer our second research question and assess the adequacy of participatory privacy regulation as an ethical framework. Do participants feel comfortable and secure using participatory urban sensing systems? What changes would they recommend? Are any ethical concerns unaddressed? Answering these questions through interviews and focus groups will help us examine participatory privacy regulation from the ethical perspective of those who matter most: the individuals and communities using urban sensing systems.

We will also draw upon previous interdisciplinary approaches to participatory ethics as well as our investigation of participatory privacy regulation to develop two curricula for STEM undergraduate and graduate students: a hands-on pilot project approach to building participatory urban sensing technologies and an interdisciplinary seminar-style course extending and debating participatory ethics in urban sensing and ubiquitous computing. The pilot project curriculum will enhance existing CENS campaigns to draw students into design projects and discussions focused on ethical system development. The seminar course will engage students from CENS as well additional students from fields such as computer science, electrical engineering, statistics, information studies, environmental studies, media studies, geography, law, political science, sociology, and philosophy. The nature of course participation will highlight design as well as other disciplinary approaches to participatory urban sensing ethics.

**URB 04.6 External Research Partnerships**

Collaboration with Dr. Jim Waldo of Sun Microsystems (Current)

Collaboration with Dr. Jerry Kang of UCLA School of Law (Current)

Collaboration with Dr. Daniel Weitzner and his research group in MIT’s Computer Science and Artificial Intelligence Laboratory (planned).
Discovering semantically meaningful places

URB 05.1 People

- Principal Investigator: Deborah Estrin, Donnie Kim
- Faculty: Deborah Estrin, UCLA CSD Professor
- Graduate Students: Donnie Kim, UCLA CSD Ph.D. student.

URB 05.2 Overview

Increasing numbers of mobile devices are now capable of locating themselves based on a multitude of different technologies including satellite, mobile telephony, and 802.11 (WiFi). Chipsets are continuously decreasing in cost and size, making it feasible to integrate them into more mobile devices. Different technologies offer different opportunities and limitations. The Global Positioning System (GPS) provides worldwide coverage except in buildings and underground. Technologies based on WiFi and cellular signals, on the other hand, can potentially provide relatively coarse location estimates anywhere wireless internet and voice services are available.

Several commercialized products have shown that a mixture of GPS and RF-beacon-based location can allow a device to compute its position with high availability throughout a carrier’s day. Raw coordinates provided by these location systems are an excellent resource for current location-aware applications such as navigation and emergency response that require absolute locations for only a short period of time.

Many emerging applications, ultimately will refer to location in terms of colloquial places or collected representations of locations such as “Home”, “My Office”, or “Joe’s plumbing store” instead of a series of raw coordinates. Mobile devices are evolving from telecommunication and personal management tools to smart gadgets that capture and share a user’s context. A stream of location, image, or text data captured by these devices can be used to perpetually understand and record people’s activity and mobility patterns, monitor and report their environment (e.g. traffic, pollution levels), and exchange whereabouts between friends and family. Continuous data collection on mobile devices, however, presents a new challenge: efficiently collecting data on resource-constrained mobile devices and effectively aggregating the large amount of streaming data. Places discovered from underlying location techniques can directly support a variety of location-aware applications, ranging from reminders based on location, to triggering data collection, to effectively aggregate the important information.

URB 05.3 Approach

Place learning algorithms attempt to find a locale that is important to an individual user and carries a semantic meaning. An important locale can be defined as a place where the user spends a substantial amount of time and/or visits frequently. A number of interesting place learning algorithms have been proposed both based on coordinates provided by location system (GPS or Place Lab) or on raw RF-beacon (WiFi Access Point or cell tower) fingerprints [1-4, 7]. In this research, we have illustrated that coordinate based place learning algorithms, which require an intermediate step of acquiring geographic coordinates, may often be inefficient as well as insufficient for discovering places by introducing another layer of error and computation. Moreover, we have demonstrated that existing place learning algorithms based on RF-beacon fingerprints suffer in finding places where beacons are inconsistent or coarse [5] [6][9]. We have focused on providing an accurate place discovery mechanism. This work focused on providing mechanisms for accurately discovering entrance and departure times of physical destinations in someone’s life.

Early works on place learning with GPS used loss of signal to infer the location of important indoor places. Marmasse et al. [1] identify a place as a region, bounded by a certain fixed radius around a point, within which GPS disappears and then reappears as in when a user enters and leaves a building. This approach is sufficient to identify

---

1 e.g., Navizon: Peer-to-peer wireless positioning: http://www.navizon.com; Skyhook wireless: http://www.skyhookwireless.com
indoor places that are smaller than a certain size (e.g. a home), but does not account for larger indoor places (an office complex, multi-floor building, or convention center), and is prone to generating false positives caused by the many possible outdoor GPS shadows. A similar but improved approach was proposed by Ashbrook et al.[2]. Sets of important coordinates are identified as those at which the GPS signal reappears after an absence of 10 minutes or longer. These sets are then clustered into “significant locations” using a variant of the k-means clustering algorithm. Toyama et al. [7] presented a variation of this work that employs multiple radius parameters to detect meaningful locations at different granularities. These allowed overcoming the place-size limitations and most of the false positives that handicap Marmasse's [1] approach, but the use of GPS signal loss to infer place still leaves us unable to infer important outdoor places and multiple places within a single building.

Kang et al. [3] proposed an approach based on distance and time-related heuristics similar to the idea proposed by Hariharan et al. [4] that does not depend on GPS signal losses. Their approach allows using continuous RF-emission based coordinate systems as location sources. Both find a new place when the distance of the new coordinates from the previous place is beyond a distance threshold, and when the new locations span a significant time threshold. However, unlike Hariharan et al., which computes the distance between all pairs of coordinates after every new location measurement, Kang et al. incrementally compares the distance between the mean of the current cluster and the new measurement against the distance threshold. Unlike other clustering algorithms that require offline clustering of complete location traces, their time-based clustering algorithm incrementally extracts stays without expensive computation. However, this approach still does not resolve the inherent problems of GPS or RF-emission based coordinate systems including power issues and discovering places closer than the localization error of the systems.

Krumm et al. [5] measured the variance of the signal strength of the strongest WiFi access point as an input to a simple two-state hidden Markov model (HMM) for smoothing transitions between the inferred states of "still" and "moving". However, while the state of the mobile devices can provide useful hints for learning places, we do not assume that users are immobile. Our early experiments also suggest that even in a single room when the device is steady, signal strength variance of the strongest WiFi beacon can be large due to interference. Hightower et al. [6] defined a time window to find scans stable for at least a predefined time interval and, thus, indicates a significant place. The key is distinguishing beacons seen infrequently while a device is in the same place from new beacons seen as the device is physically departing. To avoid low response-rate beacons erroneously dividing places, they defined a certainty parameter to tolerate infrequent beacons and determines departure when new beacons are continuously found. However, it fails to find places where users are constantly mobile (e.g. markets) or places where severely inconsistent beacons are found due to interference. Ahmad et al. proposed a fair election algorithm that finds the best representative beacons for various length of stays that a recognition method can use [9]. However, they do not address the problem of discovering the place itself.

For better efficiency and accuracy, we have designed PlaceSense algorithm, which incrementally improves the existing place learning techniques. PlaceSense collects WiFi and GSM radio response-rate fingerprints, selects representative RF-beacons, and uses them to discover places more accurately than previous approaches. By concentrating on representative beacons for discovering visits to places, it is more reliable in finding short visits, places where people are mobile, and where inconsistent beacons are prevalent due to interference.

The PlaceSense algorithm is designed to learn places by continuously monitoring the radio beacons surrounding a mobile device. It uses commodity radio beacons such as WiFi access points (APs) or cell towers as its signals, which are pervasive and can be detected by most mobile devices. WiFi access points and cell towers broadcast unique identifiers for both communication set-up and hand-off. For example, WiFi access points transmit periodic frames called Beacon Frames containing the AP’s unique MAC address among other synchronization information. Likewise, cell towers are identified by a unique Cell-ID. We refer to these fixed radio sources as beacons. Mobile devices can periodically scan for these nearby beacons without connecting to or communicating with them. The IDs of these beacons are visible even if the network is encrypted. A timestamped log of these beacon scans is the input to PlaceSense's discovery phase.
The basic approach of PlaceSense is: Define a scan window and the number of windows to wait before deciding that the radio environment is stable. (A radio environment is stable as long as at least one representative beacon is found continuously. A scan window is stable if it contains no new beacons not seen since the examination began.) 

*Entrance to a place* is then indicated by stable scan windows seen continuously for the defined window count. The fingerprint, a list of beacons found during a stay and their response rate, is gathered to define representative beacons of the place. *Departing from a place* is indicated by the loss of all representative beacons.


**URB 05.4 System(s) Description and/or Experiments**

We have demonstrated PlaceSense's effectiveness with a thorough comparative evaluation to two published place algorithms [3, 6] each based on coordinates or RF-beacon fingerprints. To evaluate our algorithm, we gathered radio traces (GPS, WiFi, and GSM cellular) from three volunteers following scripted (for accurate ground-truth) visits to multiple places on campus, and as they went about their normal routines for a week. Each volunteer collected radio traces and kept a written diary of places they visited. Using these two sets of data, we evaluate how accurate the places and their entrance and departure times found by our algorithm are compared to others, and show that it outperforms the other methods in real-life applications.

Following our initial evaluation based on a scripted data set, we further validate our algorithm by running it on multi-day traces. Three data collectors collected these traces for seven days each, following their normal lives. Traces contained various routines from ordinary work and home routines to a multi-day trip to another city. The results of PlaceSense on these real-life traces generated with a representative threshold 0.9 and tolerance depth 3 (optimal configuration obtained from our initial evaluation) is shown in Fig. 1, and compared against BeaconPrint [6] with confidence depth 3 and window size 120 seconds (as suggested). However, we do not evaluate the time boundary accuracy as the error range of time records provided by the data collectors were often more than five minutes.
**URB 05.5 Accomplishments**

By focusing on representative beacons, PlaceSense reduces the number of missed places while also increasing the number of interesting and false places (Fig. 1). We further investigate the distribution of discovered places by their duration length in Table 1. PlaceSense illustrates strength in discovering briefly visited places as well as other long-term places where the radio environment is unstable with many infrequent beacons. Places that BeaconPrint missed, but PlaceSense discovered, includes short visits to a convenience store, gas station, restroom, grocery store, and shops as well as a long stays in meeting rooms and convention centers. False places found by both algorithms include a wait in front of an occupied classroom and some unrecognizable short stays. PlaceSense additionally found a slow walk through the hallway as a place when a strong beacon was found continuously during the walk. Interesting places mostly include bus stops and unrecorded visits to restrooms.

(a) BeaconPrint

<table>
<thead>
<tr>
<th>Duration (minutes)</th>
<th>Missed</th>
<th>False</th>
<th>Interesting</th>
<th>Merged</th>
<th>Divided</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-10</td>
<td>3</td>
<td>7</td>
<td>12</td>
<td>0</td>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>10-30</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>30-60</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td>60-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>71</td>
</tr>
</tbody>
</table>

(b) PlaceSense

<table>
<thead>
<tr>
<th>Duration (minutes)</th>
<th>Missed</th>
<th>False</th>
<th>Interesting</th>
<th>Merged</th>
<th>Divided</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-10</td>
<td>25</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>10-30</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>30-60</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>28</td>
</tr>
<tr>
<td>60-</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>65</td>
</tr>
</tbody>
</table>

Table 1. The distribution of discovered places by their duration

**URB 05.6 Future Directions**

Encouraged with the performance of PlaceSense, we plan to conduct additional data collections and user studies. Data collected from a wider population for an extended duration will allow us to learn more about the systems performance as a function of the number, type, and visit frequency of places people visit in more depth. Furthermore, we plan to explore a number of existing place recognition approaches and benchmark their performance in recognizing the places when they are re-visited.
UBR 06 Smart Power Management on Cellphones

UBR 06.1 People
- Principal Investigator: Deborah Estrin, Ramesh Govindan
- Faculty: Prof. Deborah Estrin (UCLA), Prof. Ramesh Govindan (USC)
- Graduate Students: Hossein Falaki (UCLA)

UBR 06.2 Overview
Modern mobile phones are changing from single-purpose devices to multi-functional programmable computers. As a result, a plethora of mobile applications are emerging, many of which may run in the background and collect context information to process locally or upload to a server. Campaignr and AndWellness are two examples of such applications developed at CENS. With the presence of such applications mobile phones cannot rely on long low-power idle states to conserve energy. Therefore, these background applications negatively affect phone battery life and thus users' satisfaction of their phones. Our experiences with two such applications, Campaignr and Nokoscope, indicate that they reduce the phone battery life to less than 12 hours, and therefore, many users stop running them. This is a major obstacle on wide deployment of Urban Sensing mobile applications. To facilitate wide adoption and deployment of other Urban Sensing applications that run on mobile phones, we started the Smart Power Management project to solve the battery problem in a systematic way.

Several efforts have focused on reducing power consumption of individual phone components. However, there is evidence that significant power saving potential lies at the application layer. This potential can be realized by exploiting the intrinsic trade-off between applications' quality of output and their energy consumption [Flinn99, Zeng02]. In general, decreasing fidelity leads to lower energy consumption. For example, increasing the GPS sampling interval from 15 seconds to 30 seconds reduces the GPS energy consumption by 50%. The effect on quality of service, however, is application specific.

Previous researches in this area are mostly concerned with laptops. The applications on a mobile phone have an intrinsic priority attached to them that the power management system to be aware of. For example, being able to make calls is always more important that listening to music or taking pictures. Existing systems also do not consider legacy applications that have been designed/implemented prior to the introduction of the new power management system.

We argue that urban sensing applications on smartphones should be able to adapt their operation rate or fidelity based on users' battery life expectations. Users cannot be expected to manage how applications run, or to keep track of the battery [Ravi08]. Therefore, the power management system should monitor the system and plan ahead to meet the user's battery life expectations. Such system capabilities would need to rely on models of: battery life, user's charging behavior, energy consumed by "legacy applications" (e.g., calls), and the energy-performance trade-off of each adaptive application.

The initial goal of the project is to construct these models with acceptable prediction accuracy. With these models, we are implementing a prototype of the power management system and modify Nokoscope, the Nokia Research Labs logging tool, to use the prototype. This prototype will be used for initial evaluations and if successful it will be incorporated into current and future urban sensing mobile applications.

UBR 06.3 Approach
We pursue an experimental approach. We initially started collecting system information from a few volunteer smartphone users. We are using these traces to find appropriate models and also test their performance in trace-based simulations. The first prototype will use Nokoscope, and we will evaluate its performance on Nokia smartphones. Future directions of the project will be determined based on the performance of the first prototype.
URB 06.4 System(s) Description and/or Experiments
In a pilot user study we investigated the feasibility and challenges of constructing the three models for battery life, charging behavior, and legacy application usage. For our pilot study we installed Nokoscope, the Nokia Research Labs logging platform, on six volunteers’ N95 smartphones for approximately 30 days each. The subjects used N95s as their primary phones. Nokoscope logged system information: screen inactivity, battery level, and charging status, every 10 seconds. We used the screen inactivity times to infer the length of the users’ interactions with their phones. A user’s interaction with his/her smartphone is a sign of running an interactive application, which can help us profile legacy application usage.

URB 06.5 Accomplishments

User Interaction Model:
We discovered that our users’ interaction times with their smartphones follow a truncated exponential distribution model, with relatively high variance. Figure 1 is a plot of complementary cumulative distribution function of the length of interaction times for one of the users.

We could also find patterns in the length of total phone usage based on the time of the day, but due to high variance it does not seem to be useful as a predicting model. Figure 2 plots the average length of activity vs. each hour of the day for a user. The data is from 30 days. Notice the high variance lines.

Our efforts to find suitable models for other aspects of user interaction times is still in progress.

Charging Model:
Our study shows that individual user’s charging patterns are fairly consistent from day to day. For example, Figure 3 is the histogram of the number of times a user charged his/her phone during each hour of the day.

Battery Model:
Unfortunately the Symbian OS does not report accurate battery level information to user applications. Figure 4 is a plot of battery level changes for a user. Note that only eight battery levels are reported by Symbian.

We are trying to find alternative method for battery life models. We are considering off-line energy profiling of different applications.
URB 06.6 Future Directions
We will continue our efforts to build the necessary models. We will consider further user experiments and also novel approaches to model the data.

We will build the first prototype using Nokoscope. We will evaluate the performance of the new power management system in terms of battery life on Nokia smartphones. Future directions of the project will be determined based on the performance of the first prototype.

URB 06.7 External Research Partnerships
Cisco Research (planned)

URB 06.8 References


Spotlight is a system that aims to profile real-time information about energy and resource consumption by particular end-use to occupants in residential spaces. We develop technologies for an economical and easy-to-deploy monitoring system that is able to provide personalized, real-time, and activity-annotated energy and resource consumption reports. The reports essentially visualize how much resource each individual consumed, at what time (or during what activity), in what form (electric, water, natural gas, or heating oil), and through what appliance or device. The Spotlight system, we expect, would effectively encourage individuals to modify specific energy-wasting habits, compare usage trends within their peer group and neighbors, support behavior modification through auditory and visual feedback, and track that progress over time.

Providing personalized consumption reports to individuals requires solving several technical challenges: (i) economical measurement of the energy and resource being used by various end-points (appliance, faucet, heating unit, etc.) in a household, (ii) identification of the individual(s) using an end-point at a particular time and the activities they were engaged in, and (iii) associating energy and resource consumption at end-points to specific individuals and their activities.

Monitoring end-point level resource consumption is difficult in current technology without expensive or professionally installed sensors for directly metering energy and resource usage at each energy or resource consuming end-point or circuit. Also most of them are not designed for real-time data collection to a central server, thus are not suitable for this application. To develop an economically feasible monitoring system, we develop an architecture that combines pre-existing resource monitoring systems and noninvasive indirect resource sensing mechanisms in a tiered wireless sensing architecture. The key is to develop smart inference and learning algorithms for end-point level resource monitoring and personalized activity inference.

In this architecture, we want to avoid the traditional approach of using complex and carefully trained and calibrated sensing hardware for monitoring resource consumption and occupant. Instead, simple hardware for opportunistic sensing of information about resource consumption and resident activity is used with smart algorithms and models.
to train and calibrate the hardware and to make inferences with minimum user intervention, which is essential for an economically scalable and easily installable system.

To tackle these challenges, we propose several architectural elements. The first element is an auto-calibrating tiered sensing approach to measuring resource consumption at every end-point in the building. High cost, installation and retrofitting difficulty, and aesthetic or ergonomic intrusiveness are inherent in in-line sensors that directly monitor resource usage at every resource-consuming end-point. Our approach is to use a number of cheap less intrusive sensors that detect signals, such as sound, light and/or vibration, emitted by resource-consuming end-points and related to the resource consumption rate, together with a single direct sensor measuring the total consumption of that resource type for the whole building at the main metering point. The key to making this idea work is an autonomous calibration framework that learns the calibration functions that map the readings from cheap indirect sensors into precise resource consumption estimates.

The two other elements of the architecture are occupant activity recognition and resource accounting. To correctly associate resource consumption with specific occupants and their activities is vital to providing personalized resource consumption information. The challenge is to be able to do so without the occupants to continually wear or carry special sensors.

We explore a distributed wireless sensing subsystem using a activity inference engine together with personalized activity models to opportunistically fuse uncertain cues from simple ambient sensors into robust inferences about identity and activity state of users of resource-consuming end-points. By combining this information together with the end-point resource consumption, the accounting subsystem associates resource consumption data with occupant activity.

By creating technology that generates fine-grained personalized energy and resource accounting to be an integral part of all built environments, we hope to engender significant socioeconomic benefits. For example, encouraging conservation among individuals and families; demand side management and assessment of the efficacy of conservation in buildings; resources and waste inventory reporting in organizations; targeting of incentives by utilities; and, auditing for compliance with regulations by government agencies.

**URB 07.4 System(s) Description and/or Experiments**

System Description: to show its overall concept, we developed two prototype systems: a personal power monitoring system and a pipe-level real-time water monitoring system

*Spotlight Prototype (SPOTLIGHT):*

To show its proof of concept design, we have implemented and deployed a prototypical system in a 2-person apartment. The design of the system is centered around instrumentation of participating appliances. In the prototype, four appliances (a television, a coffee machine, and two standing lights) are powered through commercially-off-the-shelf power meters that directly measure the power consumption of the appliances. This power meter provides a serial interface for external logging. The serial interface connects to a MicaZ 'appliance' mote that forms the sensing and communication hub for the appliance. The sampling and reporting rate of the power meter is set at 1 Hz. On receiving power measurements, the appliance mote forwards these messages to the Spotlight server through its radio.

The MicaZ mote is equipped with a 2.4 GHz Zigbee compliant radio that is used for sensing the proximity of participating users and for communicating with the Spotlight server. Users carry MicaZ 'tag' motes that broadcast periodic beacon messages advertising their presence. The appliance motes listen to the beacons and record the radio signal strength (RSSI) for tag beacons. These RSSI values are also forwarded to the Spotlight server for further data processing.

The Spotlight server is a system that receives beacon strengths and power measurements from the various appliance motes and processes them together to identify user proximity and eventually per user appliance energy consumption. The Spotlight server is connected to the appliance motes through a 'base station' MicaZ mote.
attached to the server’s serial port. The Spotlight server uses the SensorBase to record, maintain and retrieve the measurements from the appliance motes.

**Nonintrusive Autonomous Water Monitoring System (NAWMS):**

In order to test and validate our proposed two-tiered sensing system architecture, we constructed a plumbing testbed with three pipes of different materials and diameters (copper ¾”, PVC ¾”, PVC 1¾”). The main water pipe is equipped with a direct water flow meter that meters the flow rate. We used a MicaZ mote for the interfacing medium that provides the fusion center with real-time flow rate measurements. Each branch pipe of the testbed is equipped with an acceleration sensor. In this setup, we use a MicaZ mote with the MTS310 sensor board. The MTS310 sensor board contains a 2-D Analog Device ADXL202 accelerometer that has a range of +/-2g. The node samples the axis perpendicular to the pipe at 100Hz and calculates the variance of acceleration every 50 samples. This variance is then sent to the fusion center for further processing.

In addition to the sensors, each branch has a ball type valve at the end in order to control the flow rate within each pipe. To get the per-pipe ground truth, we added another flow rate meter to the end of each pipe.

**URB 07.5 Accomplishments**

**SPOTLIGHT**

Our prototypical implementation views appliances providing a service to a user and the energy consumption associated with the appliance as a cost for the service. Each participating appliance is defined a service range within which the user can benefit from the service. By using radio receive strength (RSSI) from user-wearable mote tags an appliance can determine the users in its service range. While analyzing energy consumption pattern of users, we found that the system can profile two interesting aspects of energy profiles: first, total energy consumption, which gives us how much energy is consumed by an individual whether it's useful or not, and second, total wasted energy, which implies the energy wasted by an individual unintentionally. Users may consult with the total amount of energy users consume, when they want to reduce their energy consumption. We assume that if occupants are continuously aware of their consumption and their need for energy, users could optimize theirs. Since the system can also give the occupants the information on wasted energy, they also could reduce their unintentional energy waste.

**NAWMS**

In this work, we tried to tackle the first challenge: designing a resource monitoring subsystem that provides the same resolution at a cost and effort that is reasonable for household use. Exploiting the fact that the vibration measure has monotonic relation with water flow rate in a pipe, NAWMS uses inexpensive vibration sensors attached externally to individual pipes, which reduces both cost and effort of installation. Those sensors and their un-controlled installation, however, require to have sophisticated calibration procedure, which may result in typically laborious sensor calibration steps. By developing an autonomous calibration scheme that continuously recalibrates the sensors, we showed that an economical monitoring with reasonable accuracy is possible without labor-intensive sensor calibration and installation.

Under idealized setting where we neither have external noise nor vibration propagation among pipes, we considered various pipe topologies and formulated a numerical optimization problem that yield the system
calibration coefficients at once. It turns out that the autonomous calibration is happening by solving a linear programming problem. In that, the system extensively uses the two-tired architecture and exploits pre-installed main.

Since branch pipes are physically connected to the others, vibration from various pipes may propagate through their connection. Such phenomena either renders the simple linear programming problem a NP-hard problem, when we explicitly account for this, or increases estimation error, when we do not account for this. To solve this, we have derived two types of optimization problems: Two Phase Linear Programming Model and Mixed Linear Geometric Programming Model. Both of them are known to be solvable in known polynomial time. Although both of them relax the original NP-hard problem, they reasonably improve the estimation accuracy.

To be able to regulate system performance, a sensor system needs to recalibrate sensors upon performance degradation, which may increases maintenance cost. To address this challenge, we introduced a well-defined performance metric that helps evaluate the system performance in real-time. We have shown through a case study that it enables the system to adaptively calibrate the calibration functions.

Exploiting the two-tiered sensing architecture, we were able to develop a nonintrusive autonomous water monitoring system that neither requires expensive sensor calibration nor invasive sensor installation. It is possible to have a less intrusive, auto-calibrated, per pipe water monitoring system that provides pipe-level real-time water flow rate which was previously not possible without extensive installation of inline sensors. In our experimental set-up shown in the figure, we showed that vibration based water flow rate estimation is possible. The reported estimation error is less than 10% on average.

**URB 07.6 Future Directions**

We see two main directions, which we want to pursue: (i) generalizing the autonomous calibration method for finer grained resource monitoring system in residential spaces and (ii) developing activity/identity inference mechanisms for resource accounting.

To make the resource monitoring system economically feasible, an autonomous and noninvasive sensing subsystem is essential. With a modification from the NAWMS work, we envision that our proposed framework for water monitoring could be applied to other resource monitoring systems. For example, the power line in a household has a similar architecture, it has one main meter, and magnetic sensors, which can sense electrical power consumption related signals, need to be calibrated in order to estimate the power consumption of an appliance. We expect that the same is true for natural gas and oil for heating and cooking.

We see this problem is a particular case of sensor calibration problems which has one property, that is, one sensor is calibrated and monitors the sum of the resource consumption whereas the other sensors are monitoring resource consumption related signal. Now the main challenge is to find a unique set of calibration functions given a set of measured data. We anticipate, similar to the NAWMS case, that a class of optimization problems yield a unique solution that describes a monotonic input-to-output relationship thus is the unique calibration function, which primarily suggests us that without any human intervention, a system can obtain a complete set of calibration functions by minimizing the class of optimization problems.

In addition, we plan to develop an activity and identity inference technique based on the finer grained resource monitoring subsystem. Correlating resource consumption per occupant is a very essential step for solving the problems associated with detecting and identifying people and their activities. In doing this, we will revisit the problem of human detection and identification in the context of energy and resource consumption by exploiting existing sensing. We seek to find a higher-level algorithm that will decide how to associate end-point resource usage with the occupants and their activities so that they can draw concrete conclusions on their consumption.
URB 08 GeoSIM: An Urban Sensing System for Social Image Mapping of Urban Geolocations

URB 08.1 People

- **Principal Investigator:** Cyrus Shahabi Associate Professor at USC Computer Science Department
- **Faculty:** Cyrus Shahabi (USC; Computer Science Associate Professor; PI), Jeff Burke (UCLA; REMAP Executive Director and CENS Urban Sensing Area Lead; co-PI), Deborah Estrin (UCLA; Computer Science Professor and CENS Director; co-PI)
- **Researchers:** Farnoush Banaei-Kashani (USC; Computer Science Postdoctoral Research Associate), Luciano Nocera (USC; Computer Science Postdoctoral Research Associate)
- **Graduate Students:** Houtan Shirani (USC; Computer Science PhD Student), Leila Kazemi (USC; Computer Science PhD Student), Sanjay Srinivasan (USC; Computer Science MS Student), Penny Bei Pan (USC; Computer Science PhD Student)
- **Undergraduates:** Nicholas Bopp (USC; Computer Science Undergraduate Student)

URB 08.2 Overview

We envision **GeoSIM (Geo Social Image Mapping)**\(^2\) as an urban sensing system with which a group of individuals with camera-equipped mobile phones participate in collaborative/social mapping of the urban image (i.e., the texture of the urban environment) at some target geolocation (see Figure 1). The participating group, which may either consist of dedicated individuals or the general public, are directed to capture geotagged images of the urban environment. The collected images are progressively used for documentation of the dynamic urban scene in multiple spatial resolutions and at different times.

![Figure 1. GeoSIM Vision](http://infolab.usc.edu/projects/GeoSIM/)

We introduced GeoSIM as a motivating application (and system) to pursue the following research objectives:

- Efficient solutions for participatory acquisition of urban visual data; and
- Effective solutions for texture mapping once the visual data is collected.

\(^2\) [http://infolab.usc.edu/projects/GeoSIM/]
URB 08.3 Approach

Below, we summarize our research approach to address each of the two objectives mentioned above:

• For efficient participatory acquisition of the urban visual data, GeoSIM must plan the participation of the GeoSIM users such that the required images are taken efficiently, fast and comprehensively, while ensuring that the constraints and preferences of the participants are satisfied (to provide the social incentive for participation). Deriving the participation plan for each user involves identifying the navigation path(s) that satisfy the user constraints while maximizing the exposure of the user to uncovered sights (for fast and comprehensive texture mapping). Our approach to achieve this objective is to reduce the participatory acquisition problem to a set of novel spatiotemporal queries for multi-destination path planning with constraints (i.e., variants of TSP), as well as complex visibility/line-of-sight queries for exposure analysis along each path.

• It is important to note that path planning and visibility analysis are interdependent problems, i.e., to identify the optimal path first one needs to perform visibility study to find the best destinations from the sight-exposure perspective, while to avoid the complex (and probably infeasible) visibility study of the entire urban area from every point of view, one first needs to find the possible paths and limit the visibility study to the points of view along those paths. Instead of developing a holistic solution for these problems, we start by developing a sub-optimal solution with which we first identify the possible paths only according to the user constraints (disregarding the sight-exposure quality of the path), and thereafter, analyze each possible path for visibility and choose the path with maximum exposure.

For effective texture mapping, GeoSIM must enable merging the collected images in real time, in order to generate the evolving texture of the urban environment as new images arrive. Reconstructing generic scenes using uncalibrated images is a known difficult computer vision problem. To achieve this objective, we take a pragmatic approach by developing a stack of pipelines to be used depending on the available preexisting data. If a low resolution texture is available together with an underlying geometric model, we attempt to merge the incoming images by matching image features and incorporate the existing geometric information. In the simplest case, supposing that locally the surface of an object is planar, a robust estimation of the underlying image homography should suffice in most cases to register an incoming image with and existing texture. We integrate the geotag of the images and use the line-of-sight query to locate the faces of the objects that are visible. This computationally inexpensive approach allows for quick registration of the incoming images for already mapped zones, and also supports merging new images when these overlap with sufficient features. On the other hand, if no texture is available for a given object, but a geometric model is present, we attempt to carry out a Euclidian reconstruction using all the available images on the area and incorporate the preexisting geometrical information. The system handles requirements in terms of how many images are needed to validate a given object texture. Finally, if no information is available we attempt to reconstruct a set of plane surfaces solely based on the image information. This is the most difficult and generic case. In all cases, we use other available data sources in conjunction with the object recognition algorithms on the images to detect and segment objects such as trees that may cause undesirable results in the reconstructions and textures.

GeoSIM Architecture

GeoSIM is being built on top of GeoDec\(^3\), a scalable geospatial data management engine that we have developed at USC. Figure 2 depicts the architecture of GeoSIM. To develop GeoSIM, as the first step we are extending GeoDec to GeoDec\(^+\) to support the requirements of GeoSIM. In the second (and last) step, we will develop GeoSIM as an application on top of GeoDec\(^+\) by introducing two new modules corresponding to the two research objectives mentioned above: 1) a coordination engine that plans participation of the GeoSIM users for efficient, fast and comprehensive participatory texture mapping, and 2) a texture generation engine that depending on the available preexisting data (either low resolution texture together with the underlying geometric model, only the geometric.

\(^3\)**http://infolab.usc.edu/projects/geodec/**
model, or no preexisting data at all), develops the urban image.

Currently, we are at the first step of the development process while planning to develop a prototype for the coordination and texture generation modules based on our research accomplishments (discussed below) to move on to the second step.

**URB 08.4 Accomplishments**

Below, we summarize our research accomplishments toward the two objectives mentioned above:

- **Toward efficient participatory acquisition**, we have investigated and developed solutions for both path planning and visibility analysis. With path planning, we focus on “on-the-fly” users participation planning in a participatory texture documentation (PTD) framework. We term this problem the UPPTD (Users Participation Planning for Texture Documentation) problem. With UPPTD, we investigate solutions to generate participation plans for the users such that all users constraints (e.g., limited participation time) are satisfied while at the same time the participation plans are optimized across users for fast urban texture documentation. We proved that UPPTD is an NP-hard problem by reducing it to the well-known orienteering problem, and accordingly, proposed an individual-based scalable (and efficient) heuristic solution for UPPTD that trades optimality of the plans to achieve on-the-fly users participation planning. With this solution, we strike a balance between optimality of the generated participation plans and the time required to generate the plans. We have studied, profiled and verified our proposed solution by both rigorous analysis and extensive experiments.

  On the other hand, for visibility analysis we formulated the problem of Approximate Visibility Query (AVQ) and developed an efficient solution for answering AVQs in dynamic environments, such as the urban environments we intend to document using GeoSIM. Visibility query is the basis of many analysis and decision-making operations in large environments. Visibility computation is time complex and the complexity escalates in dynamic environments, where the visibility-set of any viewpoint is probe to change at any time. However, exact visibility query is rarely necessary. Moreover, it is inefficient, if not infeasible, to obtain the exact result in a dynamic environment. We formally defined AVQ as follows: given a viewpoint \( v \), a distance \( \varepsilon \) and a confidence \( p \), the answer to an AVQ for the viewpoint \( v \) is an approximate visibility-set such that its “distance” from the exact visibility-set is guaranteed to be less than \( \varepsilon \) with confidence \( p \). Moreover, we proposed an approach to correctly and efficiently answer AVQ in dynamic environments. Our approach is based on the intuitive observation that there exists a significant spatial auto-correlation among visibility-sets of the viewpoints, i.e., the similarity/intersection between the visibility-sets of two points is positively correlated with the distance of the points. Therefore, one can approximate the visibility-set of a point based on those of its proximate neighbors. Accordingly, we devised a novel index structure with which we partition the space into disjoint cells and select a representative viewpoint for each cell such that the distance between the visibility-set of any given viewpoint in the cell and that of the representative viewpoint would not be more than the distance \( \varepsilon \) with confidence \( p \). Thus, the answer to the AVQ for point \( v \) can be approximated with the visibility-set of the representative point of the cell to which it belongs. Our extensive experiments showed validity of our initial observation about spatial auto-correlation of the visibility-sets as well as efficacy of our solution.

  - **Because of the complexity of the problem**, we are concentrating on developing a texturing solution for the case where an initial geometry is known (for the case where an initial either poorly textured or un-textured
model is available from Lidar or aerial imagery data). Once this pipeline is operational we plan to extend it as described above. We have partly developed a mobile solution where a template image is sent to users in the form of an overlay to guide the acquisition. The template is rendered from a virtual camera emulating the users’ mobile phone camera at a virtual location from within GeoDec (our GIS query and visualization software). After the image has been taken by the user, it is transformed/merged in a texture by registering the new picture with previously taken overlapping ones (provided that the lighting conditions were compatible); the registration uses the software we have integrated that computes an optimal projective transformation between planes from features extract by a scale invariant feature detector (SIFT). In parallel we are developing machine learning algorithms capable of providing a measure of quality for the images taken.

**URB 08.5 Future Directions**

Below, we enlist our future directions:

- Proceeding with step two of the GeoSIM system development as described above
- Developing other solutions for path planning and performing comparative study among existing solutions and our solutions
- Integrating our path planning solutions with visibility analysis solutions to develop a consistent solution for efficient acquisition of the visual data
- Extend visibility calculations and planning to take into account resolution, image quality, and occlusions (as detected in user generated images)
- Expand texturing to cases where no prior geometry is available and render the resulting geometry and textures in ways that shows the uncertainty in the data

**URB 08.6 External Research Partnerships**

- **Current partnership:** GeoSIM is a collaborative project between USC and UCLA.
- **Planned partnership:** As part of a large proposal to the NSF Expeditions program, we formed a large alliance including 5 schools at USC, UCLA-CENS, UCLA Joint Institute for Regional Earth System Science and Engineering (JIFRESSE), Cal-State Long Beach, Cal-State LA and Howard University. The topic of the proposal is on developing a new computing paradigm, called *GeoRealism*, and complement and expand the efforts discussed in this report.
URB 09 AndWellness: Improving Wellness with Mobile Personal Sensing

URB 09.1 People

- Principal Investigator: Deborah Estrin
- Faculty: Deborah Estrin, Mary Jane Rotheram
- Researchers: Ruth West
- Staff: Nithya Ramanathan, CJ Cenizal
- Graduate Students: Vids Samanta

URB 09.2 Overview

AndWellness, a new project this year, is a mobile personal sensing application for the Android platform that includes a suite of mobile services, and server-side software to improve personal health and wellness. As a collaboration between the Center for Embedded Networked Sensing and the Global Center for Children and Families, AndWellness will transform mobile devices into tools that uncover a user's behaviors at the heart of personal wellness without violating their privacy, and help users design customized interventions to improve their health. While customizable spatial-, social-, temporal-, and mobility-triggered reminders, assessments, and interventions are relevant to a wide array of behavior change objectives, AndWellness will initially focus on obesity prevention and weight management. (See Figure 1.)

URB 09.3 Approach

As users go about their normal routine, the AndWellness mobile application will profile the user’s behavioral patterns using continuous/periodic sampling of available on-board sensors--radios (e.g. WiFi, GSM, Bluetooth) GPS, and accelerometer. The AndWellness server analyzes this information to identify triggers. Alerts can be based on a user's location, or a more complex context -- authored by the user -- that includes current activity (e.g. whether they are walking, running, or have been sitting for too long), semantically relevant place (e.g. a restaurant, work, or home), and time. When the AndWellness server detects a significant trigger, it will prompt the user to record audio, video, or images for future reflection or it will launch an intervention. For example, when the user is likely to be eating, the application could prompt the user to take a picture of their meal. The picture will automatically be geocoded and time-stamped and uploaded with the user’s contextual descriptors/tags to their private account on the AndWellness server. Traditionally, behavior monitoring relies on retrospective self-reports. Through rigorous side-by-side pilot studies comparing AndWellness with traditional techniques, we plan to evaluate the ability of mobile phones to bring resolution, scale, scope, precision, speed, and diversity of feedback to improving personal health.

Two important characteristics of AndWellness are the need for personalization and privacy. Data collection, journaling, and automated context recognition algorithms must be able to adapt to a user’s environment without undue intrusion. We will explore two services. First, a privacy preserving data collection service will collect data from sensors on-board the mobile platform and give the user the option to modify, obscure, hide, or delete sensitive data before it is uploaded. Second, an adaptive event-detection service will process the sensor data and classify events, such as the user-specified trigger. The event-detection service will incorporate feedback from the user to adapt the algorithm to individuals and their environments.
We will explore two key research questions in implementing these services. First, we will explore how to incorporate feedback from the user into run-time algorithms to improve the accuracy of our event classifier. Second, we will evaluate which functionalities should and can be automated in order to limit the burden on the user. We will leverage our experience in designing partially automated fault detection and diagnosis systems that incorporate feedback from the user [Ramanathan08]. We have found few existing systems that explore this dual approach of incorporating user feedback while automating key functionality; it is an exciting new direction for designing light-weight mobile personal sensing systems that gracefully adapt to a user's context.

![Figure 1 AndWellness will use social, place, time, and activity triggers to prompt the user to collect data and launch interventions. The user will correct the trigger detection, control their information flow, and tailor interventions on the phone.](image1)

![Figure 2 Screenshot of our AndWellness prototype which runs on the HTC G-phone.](image2)

**URB 09.4 System Description**

AndWellness will use the **event-detection service** to classify user-specified triggers and a user's activity. The characteristics of these events vary greatly across individuals. Therefore, the event-detection service uses an automated model that will be adapted to each individual by incorporating their feedback at run-time. In this context an **event** is defined as an occurrence that can be uniquely characterized using a small set of features. Features can take on numerical or string values. For example, the features needed to describe a trigger are place, time, number of co-located people, and activity. Place will be obtained using a combination of GPS and map-matching where possible. Activity will be obtained using the event-detection service to analyze a combination of GPS and accelerometer data as available. When available, the number of co-located friends or family member will be obtained using Bluetooth stumbling to identify people by their Bluetooth devices. These techniques are discussed elsewhere in the report. Users will be prompted to label people or places that are encountered repeatedly. The GPS, Bluetooth, and accelerometer devices will be regularly sampled and uploaded to the server.

The event-detection service will use a simple unsupervised clustering algorithm to group similar feature vectors in a space. Each cluster corresponds to an event. We chose an algorithm that is simple to modify in real-time, making it
easy to incorporate a user’s feedback into the classifier. The clustering algorithm will operate in a space that is defined by the features. In the example described above, triggers would be classified in a 4-dimensional space. The dimensions of the space would be: place, time, number of co-located family, and activity. The features are specified at initiation, and will be selected such that similar events will have similar feature values. Because similar events group together in the space, they can be classified using the clustering algorithm.

The privacy-preserving data collection service will allow users to keep certain events or their constituent data private. Users can either restrict sharing of that data, or take other protection measures (add noise, substitute equivalent data). Services, such as FireEagle [Fire08], provide similar support. Such simple controls are not always sufficient to protect a user’s privacy. Therefore AndWellness will additionally include support for users to modify their data. Local filtering and processing of data can obfuscate sensitive location traces before they are uploaded, further enhancing a user’s control over their own privacy. Data modification, which is not commonly available, is perhaps the most interesting and useful operation provided by our service. We discuss the data modification algorithms in the context of location data because it is arguably one of the most important and difficult modalities to keep private. The service will support additional data types as well. AndWellness will support two methods to modify data. The first method is to add noise to the data [Agrawal00]. Noise addition cannot be applied in all campaigns [Krumm07]. The second method is to replace sensitive original routes or locations with realistic synthesized routes and locations. These methods are discussed elsewhere in the report, and are therefore simply referenced in this section.

**URB 09.5 Accomplishments**

We have an initial implementation of AndWellness running on an Android Phone. A screen shot of the application is shown above in Figure 2. Additionally we have submitted proposals to Google and the NIH to obtain funding for the project.

**URB 09.6 Future Directions**

We will design and implement an AndWellness client on the Android platform, which includes the privacy-preserving data collection service and the responsive event-detection service, and the infrastructure required to locally store and upload data to the server.

We will design and implement the AndWellness server infrastructure to store, process, and visualize data as needed to highlight patterns and design customized interventions.

We will design and implement a scientifically rigorous pilot study of AndWellness, and use the feedback to inform the next version of AndWellness.

Each task will be designed and implemented in tight collaboration with the target population. Therefore, each component will be iteratively designed, implemented, and evaluated. This iterative process is critical to the success of our application, but will extend the overall design cycle.

**URB 09.7 External Research Partnerships**

Global Center for Children and Families (Current)

We plan to reach out to other community groups for the pilot studies.
URB 09.8 References


[Li] Livestrong: Health, Fitness, Lifestyle <http://www.livestrong.com>


URB 10 Remapping-LA Cultural Civic Computing in Los Angeles

URB 10.1 People

- Principal Investigator: Fabian Wagmister, Associate Professor, Department of Film, Television & Digital Media, Director, Center for Research in Engineering, Media & Performance; Jeff Burke, Executive Director, Center for Research in Engineering, Media & Performance
- Faculty: Fabian Wagmister (UCLA Dept. of Film and Television), Jeff Burke (UCLA REMAP and CENS), REMAP Research Staff: Ryan Dorn, Diego Robles, Alessandro Marianantoni, Alejandro Wagmister
- Graduate Students: Francesco Capodieci (CSD), Vidyut Samanta (CSD)
- Undergraduates: Taylor Fitz-Gibbon (TFT), James Dellemonico (TFT), Dustin O’Hara (DMA), Ryan Glennan (DMA)

URB 10.2 Overview

Remapping-LA seeks to engage and work with Los Angeles communities to explore, develop and implement technological tools and mechanisms that: (1) Enable a multiplicity of citizens’ driven investigations of the city’s civic history, structural conditions, urban systems, and developmental process. (2) Promote and facilitate collective forms of expression reflecting in form and content the cultural and social specificity of the producing communities. (3) Instigate multicultural interaction, intersection, and connectivity between the investigations and expressions of the diverse communities of the city.

REMAP has collaborated with CENS to develop sensing and data management technologies for Remapping-LA. The project’s decentralized, participatory approach to technological development aims to follow these principles: (1) Participatory from the ground up. Community groups are involved in every aspect of the design and development of the technological systems and production procedures. In other words the community groups drive the creation and usage of the tools. (2) Awareness of the context in which the systems are being designed. (3) Encouragement of tangible and meaningful interfaces between project members and media.

The core component of the project focuses on a partnership with the California Department of Parks and Recreation to utilize the emerging Los Angeles State Historic Park as a living laboratory. In collaboration with the adjacent Chiparaki Cultural Civic Computing Center, Remapping-LA engages community groups and civic stakeholders in the development of participative interpretive technologies for this important park dedicated to the past, present and future of Los Angeles.

URB 10.3 Approach

Remapping-LA follows a multi-pronged approach: (1) Through the production of a series of interconnected and evolving demonstration prototypes the project explores and demonstrates the possibilities of a multiplicity of current and exploratory technologies and participative technological approaches. These prototypes engage community groups at increasingly deeper levels of research, design, and creation, and have utilized the platforms and technologies (e.g., Campaignr, Sensorbase, and the SMS gateway) created by the urban and participatory sensing group at CENS. (2) The project engages the community groups through a series of dialogues, workshops, training sessions and design discussions centered on their civic specificities, needs, and cultural identities. These
process results in the formulation of the community groups’ own research and expressive interests, and the development and configuration of corresponding tool sets and usage patterns. It also is used to define and adjust technology design direction. (3) The methodology of the research consists of repeating cycles of three steps: first, exploring the city; second, designing and developing new technological systems that express what is discovered; and third, inviting the rest of the city to experience and comment on what is created. (4) An ongoing, evolving interpretive database and data visualization toolset is being developed that is at once expressive of the multiple community driven research/expressive processes and a rich resource for additional ones. Among the contents of this interpretive database are citizens’ generated photography, videos, recordings, maps, keywords, annotations, observations, data gatherings, etc.

**URB 10.4 System(s) Description and/or Experiments**

Remapping-LA investigates and develops urban computing systems for collective applications in collaboration with CENS. We believe Cultural Civic Computing systems emerge by pooling interpretive databases, imaging tools, sensing instruments, and wireless mobile devices. These networks are demonstrating important practical developments, including comprehensive understandings and representations of the cityscape and enhanced interpretive tools for civic development. In 2008, several concepts and systems developed for the public space interactive media piece *Junction/Juncture* were extended further and deployed for the Remapping LA *Hollywood* project.

**3D navigable environment (Hollywood)**

We continue to research the combination of dynamically constructed and curated experience through the use of “interpretive databases” for mobile experiences and media installations, including *Junction/Juncture* (2007), *Imageability* (2006) at the Los Angeles State Historic Park. This approach typically associates with each piece of media a location tag, thematic keywords with relevance rankings, and overall ranking/weighting for themes and media. Most recently, for Remapping LA Hollywood, we built an interactive visualization tool for this type of media database using the Ogre3D open source graphics engine and Python bindings. The system automatically extracted media from the web-based gallery of media, laid it out in geographic space, imported custom maps, and rendered this in the game environment.

**Mobile Tablet Interface (Hollywood)**

We have built a Flash-based mobile application (supported by the maemo ports of sqlite, i, and other tools) that presents location-based “itineraries” of media. The platform will synchronize with the open-source Gallery2 media database used to hold the images, audio, and video gathered by content creators, and enable group members to author rich experiences using a combination of media concept- and geo-tagging that is interpreted by the application. Ongoing research considers the (1) the language and interface used to author what content appears at what time and location, given past viewing and location tracklog history, and (2) a user interface enabling both active navigation and more passive experiences.
The prototype platform will be further developed through continued work on Remapping LA Hollywood, and two other unique opportunities: In January-June, 2009, we will collaborate with the Los Angeles consultancy Public Matters and the Southern California Filipino Workers Center to support their creation of location-based experiences of Filipino culture in Los Angeles, extending PM’s work with the youth of Filipinotown. Finally, in 2009, we expect that this platform will be brought back into our work with California State Parks and Disney Imagineering R&D, and community groups of the Los Angeles State Historic Park, which researches and experiments with new public space experiences of media for this 32-acre park under development for the heart of downtown Los Angeles.

Sensor based contextual aesthetics (Juncture/Junction)
For the interactive public space work Juncture/Junction, we created a computer vision system that could sense the movement of the Metro Gold Line through the park with high spatial precision and low latency. This system then triggered synchronized changes in outdoor lighting and the data visualization. The same system was used to observe the flow of cars on Broadway St. above the location of the installation and manipulate the visualization. The underlying vision system is being developed further for the NSF project Semiotic Pivot Activity Spaces for Elementary Science (SPASES), with PI GSEIS Prof. Noel Enyedy. Finally, we created a web query interface to retrieve the current level of the Sepulveda Dam from a federal online website, and plan to use this data in a future version of the installation.

SMS interactivity interface (Juncture/Junction)
For Juncture/Junction, we developed an SMS-based interface that enabled people in the Los Angeles State Historic Park to dynamically affect the media recalled for the digital mural by texting keywords to the system. These keywords were displayed in the 3D graphics engine, and then tuned the search parameters used to select media.

UCLA / Cisco Metropolitan WiFi Network (MetWi)
In 2007, MetWi was created by UCLA REMAP and CENS with support from Cisco, to provide a metropolitan ‘WiFi’ network testbed at the downtown Los Angeles State Historic Park as well as the UCLA campus. This deployment will complete in 2009, and cover two courtyards on the UCLA campus and the 32-acre Los Angeles State Historic Park site with WiFi access, served by a dedicated, high-speed fiber optic network on campus and extended to the Park via a 45-megabit DS3 connection. The network will enable us field testing of prototype experience.

URB 10.5 Accomplishments
• Remapping-LA Hollywood – in collaboration with Freewaves
  • Produced by Fabian Wagmister, Anne Bray & Jeff Burke
  • Directed by Fabian Wagmister and Jeff Burke
  • Six community groups engage in research about their neighborhood, historical, social, cultural patterns that intersect and influence theirs lives in it. The resulting collective database is a deep and expansive view into Hollywood and its multicultural community. The participating individuals and groups collaborated in the ultimate creation of a navigable interactive 3D environment and a set of mobile tours for the Freewaves Festival of Media Arts.
• UCLA / Nokia Brainstorming Session on the Future of Mobile Entertainment. Organized this full-day session with 35 professionals and academics to develop themes and connections for the new Nokia R&D Laboratory in Hollywood. May 15, 2008

• UCLA / Nokia Summit on Sustainability and Mobile Technology. Co-organized this summit in Nauvo, Finland.

• Two Nokia Research Hollywood seed grants:
  • Integrating Augmented Reality with Location-Based Experience on the Maemo Platform
    • Seeding an open-source ecosystem for location-based mobile experience

![Interactive construction of historical timelines.](img)

**URB 10.6 Future Directions**

• Finalization of a draft Interagency Agreement with California State Parks for a funded collaboration in which REMAP and State Parks will collaborate and explore opportunities for how technology can play a role in the interpretive planning for Los Angeles State Historic Park. When completed the park will be an important public space that will provide unprecedented opportunities in Los Angeles for residents and the public to connect with cultural, recreational, and natural open space along the Los Angeles River and also provide a direct park linkage from downtown Los Angeles to the river as had been proposed over 65 years ago in the unrealized Olmstead/Bartholomew Parkway Plan. This relationship will explore how Los Angeles communities can co-create technological systems for public space with which to express and explore their own, and their community’s cultural heritage and identities.

• Develop analytic and interpretive tool for citizens’ data mining of urban official public data.

• The exploration and integration of different perspectives on community data gathering, cultural experience, and civic engagement. The Remapping-LA project will continue to explore how geographic, social, cognitive, and historical/interpretive models for research and expression can inform (and be embodied within) the tools and processes that are developed.

**URB 10.7 External Research Partnerships**

**UCLA Partners**

School of Theater Film and Television

• Center for Community Partnerships
• HyperCities

Community Partners
• William C. Velasquez Institute
• Anahuak Soccer Federation
• Gay and Lesbian Elder Housing
• LeConte Middle School / LACER STARS
• The Oasis of Hollywood
• Business Improvement District (BID) / Clean Streets
• William Mead Housing Project / Youth Club
• Filipino Workers Center

Other Partners
• California Department of State Parks, Ruth Coleman, Director (current)
• Freewaves – Anne Bray, Director (current)

Industry Support
Cisco Systems
• Nokia Research Hollywood
• Walt Disney Imagineering Research and Development, Inc.
URB 11 Participatory Campaigns for Sustainability (GarbageWatch), Citizen Science (Networked Naturalist), and Active Living (CycleSense)

URB 11.1 People

- Principal Investigator: Deborah Estrin, Jeff Burke
- Faculty: Deborah Estrin, Computer Science, UCLA; Mark Hansen, Statistics, UCLA; Mani Srivastava, Electrical Engineering, UCLA; Jeff Burke, Film, Theater & Television, UCLA
- Researchers: Eric Graham, Eric Yuen, Nithya Ramanathan
- Staff: Beqa Dawson
- Graduate Students: Sasank Reddy, Electrical Engineering, UCLA; Olmo Maldonado, Electrical Engineering, UCLA; Vidhyt Samanta, Computer Science, UCLA; Min Mun, Computer Science, UCLA; Katie Shilton, Information Studies, UCLA; Peter Capone-Newton, Public Health, UCLA
- Undergraduates: Edi Rocha Guerrero, Adam Brenner, Saro Meguerdichian, Guadalupe Hernandez

URB 11.2 Overview

Participatory sensing is distributed, real-time data collected by a community of users, aggregated across space, time, and diverse datasets, and visualized in engaging and accessible ways to uncover and highlight the patterns that shape individual and community behaviors. Participatory sensing introduces familiar technologies, such as mobile phones, in innovative ways to aid people in observing and discovering unexpected patterns through real-time, prompted or automatic data collection.

Three current CENS participatory sensing campaigns include GarbageWatch, Networked Naturalist, and CycleSense.

GarbageWatch is a project that asks members of the UCLA community to perform a coordinated waste audit using their mobile phones. Individuals use their phones to collect and upload images of the contents of garbage bins to help UCLA Facilities determine where new recycle bins should be placed. Students from the Education for Sustainable Living (ESLP) Program at UCLA along with interested students involved with CENS were recruited to collect data. The use of mobile phones made it easy for many concerned students to contribute data to this campus-wide effort.

Networked Naturalist is a data collection system to enhance citizen science learning for users of Project BudBurst [Budburst-URL, Henderson09], a national citizen science field campaign aimed at collecting phenology data across the U.S. Participants are volunteers who record the events of plant life history stages, including first leaf, first flower, etc. We are working to ensure data quality and improve data submission techniques, thereby transforming Project BudBurst as well as general Citizen Science campaigns for the digital era. Participants can submit text and images by web, email, or SMS, thereby enhancing participation in Project Budburst. Additionally, access via
the Networked Naturalist website to expert and peer-based discussions will engage participants, add to their knowledge, and reinforce the community of participants.

*CycleSense* is a system to help bikers plan safe routes and collect data to improve those routes. *CycleSense* differs from standard mapping applications by harnessing the wide array of technologies available on mobile phones, such as a GPS, accelerometer, microphone, and imager, for real-time distributed data collection. A mobile *CycleSense* application running on the phone automatically gathers and shares data with the Cyclesense server such as location; starts and stops; roughness of road; and noise level of route.

**URB 11.3 Approach**

Our three campaigns tackle three different technical and usability issues that aim to increase participation in campaigns and are fundamental to designing successful participatory sensing campaigns. *GarbageWatch* will provide a system to aid in recruiting participants by matching an individual’s past reliability and spatial mobility with the needs of the current campaign. *NetworkedNaturalist* aims to improve the quality of data collected by citizen scientists by providing almost real-time feedback to participants using automated data quality analysis and immediate data visualization. *CycleSense* aims to engage target communities in a participatory design process and automate data capture as much as possible to reduce the burden on the user. We elaborate on each approach below.

*GarbageWatch*: In traditional sensor systems, one of the fundamental problems concerns the placement of sensors. The analogous problem in participatory sensing is choosing users to perform a particular data collection task. Ideally, volunteers cover areas and times of interest (e.g., daytime hours and major campus walking paths) and have a high participation rate. The system considers the availability of the user to participate in terms of spatial and temporal contexts, the reputation of the user as a data collector, and the cost associated with the user as elements involved in the process of choosing data collectors.

This recruitment system can be used by a campaign designer or sponsor, such as the Facilities Department at UCLA or a group concerned with sustainable practices (ESLP), to quickly help identify individuals that would best provide the necessary coverage and participation rate needed to assess waste usage and help in providing feedback to the volunteers to obtain higher utility information.

*Networked Naturalist*: We are creating a data collection system to enhance citizen science participation by improving the quality, feedback, and submission of data. We wish to increase participation and retention in Citizen Scientist campaigns through three avenues afforded by the use of new technologies:

- Opening the methods for successful data collection, as through the web, email, mobile phones, and automatic data uploads from user-owned hardware.
- Providing immediate feedback on data through automated analysis and real-time display of graphical and map-based representations of participant’s data.
- Leveraging the tremendous participation in social networking sites to further enhance the user experience, as well as expand and retain participation.

*CycleSense*: *CycleSense* is being developed with the advice and participation of the city’s biking advocacy groups to design a system that will be useful for bikers. Pilot testers from the Los Angeles County Bicycle Coalition and BikeSage will have input into what data to capture, how to capture it safely, and how they would like data analyzed and shared.
CycleSense also aspires to capture as much data as possible without user intervention to reduce the burden on the user. While photographs can help document a route, bikers may not wish to stop to pull out their phones. Instead, phones stowed in a backpack or mounted on the bike can automatically detect location, as well as features such as roughness (using the accelerometer). This data can then be matched against existing sources of data, such as elevation maps, traffic maps, or accident maps. By combining existing information about conditions in Los Angeles with biker contributed data (such as pavement roughness estimations), CycleSense will enable area bikers to plan routes with the least probability of traffic accidents, best air quality, or according to personal preferences, such as best road surface quality or connections with public transportation. CycleSense will also encourage bikers to contribute information to improve the safety and well-being of the Los Angeles bike community.

**URB 11.4 Systems Description**

Each of our three systems uses different mechanisms to increase participation. GarbageWatch aims to model participants past participation as accurately as possible using statistical analysis techniques. NetworkedNaturalist uses mechanisms that are simple and accessible in order to be as user-friendly as possible. CycleSense, unlike the other two campaigns, provides some software to run on the mobile phone that will automatically collect data, even if the rider never clicks a button.

**GarbageWatch:** To model the availability of participants, we explore obtaining different resolutions of “mobility” (location and activity traces) from individuals. This mobility information can come in the form of a detailed GPS trace annotated with activity, a description of building level occupancy sent via text message, or a submitted time-activity form. From the mobility information, which is an “association matrix” that captures the amount of time spent in a particular context and determines coverage, the system creates a profile. Solving a cost constrained set-cover problem, the system finds the ideal set of individuals to perform the campaign. Since availability can change from the generated profile, system components compare the similarity of profiles for different periods. By using Singular Value Decomposition (SVD) to obtain eigen-behaviors for the association matrix, we are able to get a sense of whether individuals are following their profile or behaving differently.

To calculate reputation, participants perform calibration tasks where ground truth information is known. The reputation of individuals as data collectors is calculated by comparing their individual contributions to what could have obtained in an ideal case. This reputation score, which is modeled using the Beta distribution due to the ability to get a sense of epistemic and stochastic uncertainty, is used as an additional factor in the process of choosing data collectors for the campaign.

To support GarbageWatch and explore some of the research challenges outlined above, a system was developed that involved three basic components. First, a data collection client was created for Nokia N95 phones to collect mobility traces and submit images. Second, an image-processing module was created that handles obtaining images from the mobile phones (via email) and sending it to Flickr for storage and future retrieval. Third, a set of modules were written that analyzed where data is contributed, who contributed data, and under what context to validate both availability and participation models during a campaign. The system was used by the ESLP students along with members of CENS for a period of 1 month to perform data collection.

**Networked Naturalist:** Networked Naturalist is an open-source platform to help manage the Project Budburst Citizen Science data collection campaign. Our primary goal is to design Networked Naturalist to be easy for anyone to use, and easy for developers to extend. Networked Naturalist supports campaign creation and recruitment, user-friendly interfaces to submit data, and data processing, aggregation, and visualization to make data accessible and legible to all users.

Networked Naturalist will need to manage terabytes of data, hundreds-to-thousands of participants, occasional software revisions, and changing data analysis algorithms. These features combined will support many simultaneous campaigns. Currently, Networked Naturalist users can:

Use the website to set up plants that they would like to observe as part of Project Budburst.

- Submit SMS, MMS, and email messages with observations, images, and personal notes on their plants.
- Receive confirmation messages, error messages, or query responses over SMS and email.
- Access, sort, filter, and visualize individual data or everyone's data over a web interface.
- Modify or delete old observations.

**CycleSense:** CycleSense bikers carry a GPS-enabled mobile phone during their commute. The phone automatically uploads bikers’ routes to a secure and private website. The phone also uses its accelerometer to register bumps in the route, and takes sound samples to calculate a relative comparison of noisiness along the route. Participants will be able to log in to see their route combined with existing data, including elevation change along the route, air quality, time-sensitive traffic conditions, and traffic accidents. Participants can also use the system to share information about their routes with other riders. Bikers can document impediments by taking photos with the mobile phone or sending a text message to CycleSense.

**URB 11.5 Accomplishments**
All three of our systems have been used in pilot studies.

**GarbageWatch:** We have created a system that enabled 30 users to contribute over 2000 images. The existing runtime system continually operates during a campaign and provides feedback on needed coverage and also individual participant statistics. Furthermore, we have tested our initial algorithms in regards to recruitment both on the fronts of mobility models for coverage and the reputation framework for participation.

**Networked Naturalist:** We have established a working relationship with Sandra Henderson at UCAR, the leader of Project BudBurst. Through this contact, Project BudBurst has expanded on some of the concepts outlined in Networked Naturalist for their 2009 campaign. Indeed, we have jointly applied for funding through several agencies for further collaboration. We have also arranged meetings with the Santa Monica branch of the National Park Service, in order to expand on Networked Naturalist to incorporate local participation in invasive species data collection. Our general pilot campaigns have indicated that campaign organizers and system designers are able to easily and systematically manage all of the entities, system components, and information flow involved in supporting a campaign from start to finish. Networked Naturalist is extensible, built on top of Code Igniter (http://codeigniter.com/), an open source PHP Framework for web applications. Code Igniter uses the Model-View-Controller approach, which allows separation between logic and presentation. By separating data management from the web pages, Networked Naturalist is easier to manage and modify.

**CycleSense:** We have developed software that collects and uploads GPS data, accelerometer readings, and sound samples from Nokia N95 mobile phones. A rudimentary user interface currently displays this data, and a more complex user interface is currently under construction. Using this software and interface, we have pilot tested collecting location, sound, and accelerometer data in anticipation of beta testing the CycleSense system as a whole. In addition, we have established partnerships with leaders in the Los Angeles bicycle community, including the Los Angeles County Bicycle Coalition and BikeSage, an emerging cyclist mentoring group. We have also discussed a potential partnership with MapMyRide.com to pilot test CycleSense with their 30,000 Los Angeles-area riders.

**URB 11.6 Future Directions**
All three systems aim to do more pilot studies in the next year. In order to get there, GarbageWatch will incorporate different forms and reliability levels of data; NetworkedNaturalist will incorporate data from different sources, improve participant privacy, and include social forums; CycleSense will complete the user interface.

**GarbageWatch:** Future directions for the project include exploring methods to deal with different resolutions and quality of mobility information given for availability profile creation. Project plans also include considering rating availability information based on how it was collected and submitted and also considering completeness of the information (e.g. does it represent a whole week or just certain parts of week). Calibration tasks seem to be a good method to estimate reputation for individuals before a campaign occurs, but it is unclear how we can make these tasks convenient for users. We are in the process of creating algorithms that indicate areas of maximum mutual coverage for this purpose. Availability of information at larger scales (weeks or months) can be reliably compared
using the SVD mechanism mentioned above. But further exploration is needed to see if these algorithms would be appropriate in shorter time-scales (days as opposed to weeks). Finally, planning data collection tasks for users based on queries for possible participation areas when the campaign is running.  

**Networked Naturalist:** We will soon adopt a system for incorporating existing web-based services and systems into our data flow. Near future capabilities include:

- Methods to protect participant privacy. Sensitive information submitted to the project includes location coordinates of personal plants; personal login and contact information.
- Social forums on the website, and a better interface to other social networking sites such as MySpace, Facebook, and Friendster.
- Incorporating GIS visualizations that will draw on external data-streams, from freely available meteorological data to user-uploaded spatial data sets.
- Allowing alternative data inputs, such as from personal webcams or from image sharing websites such as Flickr, with some automated analysis of such data.
- Continuing with increasing robustness features to save data during crashes and identify faulty inputs.
- Storing data in the database with an open API so that anybody can access their own data and “anonymized” versions of others’ data using a programmatic interface.
- **CycleSense:** In the next year we will complete the user interface for the CycleSense system and begin beta testing the system with experienced Los Angeles cyclists.

**URB 11.7 External Research Partnerships**

- **GarbageWatch:** UCLA Education for Sustainable Living Program (current)
- **Networked Naturalist:** Sandra Henderson at UCAR, lead on Project BudBurst (pending grant proposal, future collaboration is also planned)
- **CycleSense:** Los Angeles County Bicycle Coalition (current), BikeSage (current), MapMyRide.com (planned)

**URB 11.8 Citations**


URB 12 Urban Tomography

URB 12.1 People

- Principal Investigator: Ramesh Govindan, Martin Krieger
- Faculty: Ramesh Govindan (USC), Martin H. Krieger (USC)
- Graduate Students: Moo-Ryong Ra (USC)

URB 12.2 Overview

This project is developing an “Urban Tomography” system for capturing geo-tagged videos on video-capable cellphones and automatically sending them to a back-end server infrastructure using wireless networking technologies such as EDGE/GPRS or 802.11. Our system is designed to enable pervasive dense audiovisual documentation of city life. As our neighborhoods become increasingly diverse and complex, such documentation can enable a better understanding of social interactions and the use of urban spaces. It can also help urban planners re-structure existing cities in order to improve their quality of life.

URB 12.3 Approach

We call our system Urban Tomography, a technological framework that enables the collection and creative navigation of a large corpus of audiovisual urban documentation. Underlying this framework is the observation that relatively cheap, ubiquitous, internet-connected, mobile personal devices such as cellphones and PDAs are increasingly equipped with good video capture technologies and GPS. These technological advances promise pervasive sensing of urban phenomena: we envision swarms of trained volunteers spreading out across a cityscape and regularly recording urban processes for extended periods of time. Casual users can also contribute to the corpus. These audiovisual records can be transmitted over the network for near instantaneous analysis by anyone with access to the global Internet.

To enable Urban Tomography, two advances are necessary. At the “back-end”, the acquisition, transmission, storage, and indexing of audiovisual records must be made as automatic and transparent to the user as possible. A simple back-end will require less training and thereby enable more pervasive sensing by more people. However, this is technologically challenging because of the scale of the problem: a potentially large number of video clips introduces technical challenges in robustly distributing the corpus to ensure high availability, and organizing information to enable intuitive navigation and fast searching of the corpus. The second set of advances requires audio-visual display technologies that simultaneously present inter-related urban processes with the aim of challenging the viewer to find previously unsuspected relationships between the perspectives. This is analogous to tomography, which attempts to simultaneously explore multiple perspectives (“slices” or “cuts”) of an object (an...
organ, a geophysical feature) in order to reconstruct the object in its entirety. In our case, we are exploring techniques such as split-screen and multi-channel sound to present these multiple views in a creativity-enhancing way.

**URB 12.4 System Description**

Our prototype system consists of two major components: the capture and transmission software on the cellphone, which captures video files and automatically (without user intervention) transfers them to a server. The server contains a database that stores all video files and its metadata, as well as a simple web viewer which displays a group of videos as “small multiples” using tiles.

Our capture subsystem is designed for the Nokia N95 phone. This phone has a 1GB micro-SD card, supports 640x480 high resolution video and has “Assisted” GPS functionality. Our videos are stored on a commodity server machines running the Apache web server and a mysql relational database backend. In turn, the capture subsystem includes a user interface process which allows the user to access the camera, configure the application, or examine the application log. The main user interface process on the capture subsystem also associates the current video recording with metadata. This metadata is stored as a separate file, and contains timestamp and GPS location associated with the captured video clip (more precisely, the GPS location of the phone at the end of the clip). A background process is responsible for transferring video files to the server. It periodically scans a designated location on the local file system, and if it discovers any video files, attempts to transfer the video to the server system. This happens automatically and without user intervention. Since wireless connectivity can vary dynamically, and large file uploads are more likely to be interrupted by intermittent wireless connectivity, the background process breaks large video files into chunks to ensure forward progress for file upload. In addition, the background process monitors available forms of network connectivity (whether the GPRS network is available, or which of many hotspots are available). It prefers to use a WiFi hotspot if one is available, but otherwise attempts to use the GPRS network.

**URB 12.5 Accomplishments**

Our system has been operational for several months. In this time, we have completed several field experiments of the system. Here is a partial list of such field experiments.

- May 08: Swarming on the UCLA Campus, CS219 by Prof. Estrin’s students.
- Oct. 08: SPPD Parents Weekend, USC by Prof. Krieger.
- Oct. 08: METRANS 10th Anniversary.
- Dec. 08: OSU employment in Katrina sites.
- Dec. 08 ~ Jan. 09: Tokyo.
- Jan. 09: Beijing by Prof. Govindan.
• Jan. 09: SPPD 80th Anniversary, USC by Prof. Krieger.
• Mar. 09: Initiation of security application by a transportation site in Los Angeles.

The project webpage http://tomography.usc.edu also includes several other field experiments.

Related works
QIK (http://www.qik.com/)
QIK supports streaming service from the phone. Users can create their own content and stream it directly to the server. Other users can see them in real time with small delay. But they support lower quality than we do and do not support to show multiple videos at one place.

The Visual History Archive (VHA) of the USC Shoah Foundation Institute. (http://college.usc.edu/vhi/)
The Visual History Archive (VHA) of the USC Shoah Foundation Institute for Visual History and Education is a video archive of testimonials from Holocaust survivors and witnesses offered through the USC Shoah Foundation Institute. The archive includes nearly 52,000 video testimonies of Holocaust survivors, witnesses and liberators collected in 32 languages and from 56 countries by the USC Shoah Foundation Institute. The archival purpose is suggestive for our project, although our use of sensors is distinctive. They are not using cellphone as a collecting device, and also don’t support multiple display.

URB 12.6 Future Directions
There are several research challenges in this project.

The first is that the system does not use the best performing network available. It sometimes uses an unreliable wireless access point even if there are better ones available. Large file uploads are adversely affected by this. A simple solution is to use some form of “link estimation” to continuously ensure association with high quality access points. One way to do this would be to obtain access point signal strength (this is currently difficult to do on the Nokia cell phones because, even though the OS tracks access point strength, third party applications do not have enough privilege to access this). Failing this, we could empirically determine link quality by determining the rate of progress on a connection, and decide to switch if this rate of progress is deemed inadequate.

The N95 has a very short battery life. Under continuous video file upload, it loses its complete charge in about two hours. While battery life will improve with future generations of phones, there is scope for energy-saving techniques: putting the phone into energy-saving modes when a network is unavailable, choosing good quality network connections to avoid energy wasted in lost-packet retransmissions, and so forth.

A well-known problem with GPS, of course, is that it does not work inside buildings and other obstructed structures. Our software currently tags captured videos with the last obtained GPS reading, which is currently acceptable for our purposes.

There are interesting challenges in being able to synchronize tiled 6 to 9 videos — OS overhead adds noticeable delays between tiles. These shortcomings motivate several interesting future directions of research, some of which we intend to pursue in the coming year.

URB 12.7 External Research Partnerships
Prof. Jennifer Cowley, Ohio State, documentation of post-Katrina Mississippi. (current)

Supported by METRANS grant, Project No. 09-05, “Thousand Eyes on California’s Streets, Roads, and Infrastructures”. (current)