Learning Significant User Locations with GPS and GSM

by

Xiao Yu

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2006

Author ....
Department of Electrical Engineering and Computer Science
August 22, 2006

Certified by.............................
Larry Rudolph
Principal Research Scientist
Thesis Supervisor

Accepted by.............................
Arthur C. Smith
Chairman, Department Committee on Graduate Theses

Learning Significant User Locations with GPS and GSM

by

Xiao Yu

Submitted to the Department of Electrical Engineering and Computer Science
on August 22, 2006, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

This thesis addresses the tasks of *place discovery* and *place recognition* — learning and recognizing places significant to a user — by analyzing GPS location and GSM cell tower data collected from the user's mobile phone. Location provides valuable context into the user's environment, and place-discovery and recognition algorithms enable human-centric systems to communicate with the user in human terms. In this thesis, we introduce a novel two-phased approach to place-discovery and recognition that combines the advantages of GPS and GSM cell data. We design and implement a system that produces a compact travel summary from the user's daily GPS logs. We then use computational geometry to investigate the aspect ratios of GSM cell coverage polygons as an optimization to place recognition. Finally, we conclude by presenting a one-month empirical study to demonstrate the effectiveness of our two-phased approach, and identify a set of anomalies in our experiment that can direct further development of place-discovery systems.

Thesis Supervisor: Larry Rudolph
Title: Principal Research Scientist

Acknowledgments

I would like to thank Larry Rudolph for inspiring me to find my passion for research and for allowing me the freedom to define and shape this thesis. He has provided me with invaluable guidance and encouragement, and his enthusiasm as the subject of the case study presented in this thesis was highly infectious.

Thanks to Albert Huang for his boundless patience in teaching me vast amounts of technical knowledge and for always being there when I needed help.

Thanks to Greg Little for his willingness to discuss and explore technical ideas at odd hours of the evening.

Finally, I would like to thank my father, Hao Yu, and my mother, Xiaodan Zhao, to whom I owe everything for the myriad of opportunity they've blessed me with through their personal sacrifices.

Contents

List of Figures

List of Algorithms

List of Tables

Chapter 1

Introduction

1.1 Motivation

We design, implement, and evaluate a system that learns places significant to a user by analyzing a log of the user's GPS locations and cell tower observations. Location can provide valuable contextual clues about a user's environment, and research in spatial cognition shows that people naturally structure their experiences in terms of socially and personally meaningful places[9]. In order to communicate with the user on more human terms, human-centric systems stand to benefit from *place discovery* — the ability to analyze a user's everyday routines and extract the abstract notion of a significant place, and *place recognition* — the ability to recognize when the user has returned to a significant place.

Place discovery and recognition systems provide valuable functionality with which to manipulate the user's past, present, and future. Place discovery in the past provides a complete, accurate record of places where the user has spent time; querying the past can provide context to deferred events that the user may not remember in detail (e.g. such as what the user was doing during a missed call). Place recognition and discovery in the present localizes the user and enables a device to make intelligent decisions based on the user's current context (e.g. a mobile phone silences itself when it detects that the user has entered a conference room/auditorium; displaying a to-do list at the shopping center[15]). The past and present can then give context about the user's future, enabling an entire class of "just-in-time" services such as traffic and schedule updates in navigation systems[14], location-based reminders, and schemes for reducing communication cost[6].

Table 1.1 summarizes place discovery in the past, present, and future. While prior work has provided compelling proof-of-concept implementations for place discovery in the present and future, this thesis focuses on place discovery in the past using mobile phones as the user's representative agent.

	Search	Localization / Context	Navigation, Prediction, Interruption management	Example Systems
Past	✓			Reality Mining[8]
Present		✓		Jimminy[18], PlaceLab[1], BeaconPrint[11]
Future			✓	Social Serendipity[7], LeZi-Update[6]

Table 1.1: Representative place-discovery systems for the user's past, present, and future.

Mobile phones are a natural platform for tracking user location for the purposes of place discovery and recognition. Though conceptually simple, it is challenging in practice to rely on any single location-sensing technology for accurate place discovery due to a number of trade-offs inherent in the location data (e.g. loss of GPS signal, location ambiguity due to large cell tower coverage, lack of wireless infrastructure). And a useful place-discovery system must be effective subject to the constraints on the mobile phone's hardware resources (limited battery and storage).

This thesis describes a novel approach to place-discovery that meets the following requirements. Our approach uses GPS to provide a first estimate of the state space for a user's significant places and then combines this estimation with GSM (Global System for Mobile Communication) cell information for more resource-efficient place recognition.

- **Automatic collection of location data**

The collection of location data must be passive and must not hinder the user's everyday activities with the mobile device.

- **Compact data storage**

  Location data can be huge if stored naively, and so we must store a summary of location data as well as statistics of the user's significant places in compact ways.

- **Minimize the usage of additional service infrastructure**

  Battery life is a major resource constraint on mobile devices. Factors such as continuous computation and broadcasting on radio frequency mediums affect battery drain rate.

- **Robustness to noisy data**

  Certain kinds of data (e.g. GPS) are subject to signal loss and lack of precision (cellular coverage). Our place-discovery algorithm must minimize the effect of noisy data in producing false-positive points of interest.

1.2 Types of location data

The type of location data directly influences how place-discovery works within the system, and systems typically express location in one of two ways: by coordinate or by known beacons; each approach has its own benefits and drawbacks.

Coordinate-based systems specify absolute location by a `(latitude, longitude)` coordinate pair, while beacon-based systems express location by a user's relative proximity to fixed beacons such as cell towers, wireless access points, or fixed Bluetooth beacons. There are trade-offs between coordinate and beacon location data: the globally-available infrastructure of GPS makes coordinate-based systems easier to implement, while shorter-ranged beacons provide higher resolution and location-awareness in places where GPS signal is often not visible.

The hybrid approach developed for this thesis combines the best aspects of GPS coordinate-based place-discovery with beacon-based place-recognition using natural beacons (e.g. GSM cell towers) for mobile phones. Before introducing this approach, we examine the pros and cons of the four most common types of location data available to mobile phones subject to the requirements in Section 1.1.

1.2.1 Global Positioning System (GPS)

GPS is the de facto standard for location data. It is a satellite navigation system consisting of 27 low-orbit satellites with the guarantee that 4 are visible to receivers anywhere on earth at any given time. The satellites broadcast precise timing signals by radio to GPS receivers, which then use triangulation to determine the receiver's location.

The main advantages of GPS is that its infrastructure is already globally available, and it provides accurate absolute locations (within 2 meters) and rich metadata with its location traces (e.g. latitude/longitude, altitude, speed).

Unfortunately, GPS coverage has limitations in practice, and its main disadvantage is its inability to function indoors and in "urban canyons" that occlude satellite signals. In urban environments, it can take up to 15 minutes to acquire a GPS signal, during which the receiver may miss important location data.

One final disadvantage of GPS is its effect on mobile phone battery life. Continuous GPS readings is infeasible, as it will reduce a mobile phone's battery life to only two to three hours.

1.2.2 Global System for Mobile Communication (GSM)

GSM is the most widespread cellular communication standard in the world. A GSM mobile phone communicates with a *base station* equipped with directional antennae that define sectors of coverage called *cells.* Each GSM cell broadcasts its own unique id, and a mobile phone may communicate with a number of different cells due to overlapping base station coverage.

Cellular networks are present almost everywhere, including indoors where GPS signal is not visible; its coverage far outreaches those of other radio frequency technologies (e.g. 802.11). The biggest advantage to GSM cell data is that it requires no external service or infrastructure. Every mobile phone already maintains a list of cells from which it currently receives signal, thus place-discovery based on GSM data leverages the phone's existing hardware and removes the need for additional radio interfaces.

On the other hand, recognizing locations from GSM cell data is challenging for the following reasons:

- The actual location of the GSM base station may be unknown, so GSM cell data do not provide precise location information.
- Cells can span kilometers in diameter, thus being in a particular cell may not provide enough resolution in terms of location.
- Base station coverage overlaps, so several cells whose base locations are significantly far apart may be observed from one location.
- Cell signal strength is unreliable — reflections, people, and other obstacles may attenuate cell signal in non-linear ways.

1.2.3 802.11 Wireless & Bluetooth

Location systems such as PlaceLab[1] and RADAR[4] that use ambient 802.11 signals from nearby wireless access points and base stations can provide location estimates within 3 meters in both indoor and outdoor environments. Wireless access points broadcast their unique MAC addresses periodically, and WiFi devices may scan for these unique IDs without connecting to the actual network.

Bluetooth is a low-power, low-range radio frequency technology that supports decentralized data transmission and detection between devices in close proximity to one another. Each device has a unique Bluetooth hardware ID, and static devices may act as beacons for proximity-based location detection on an even smaller scale.

The main disadvantages of location systems based on short- or medium-ranged RF technology such as 802.11 and/or Bluetooth are that they require extensive calibration (density and spacing of access points affect location accuracy) and infrastructure (known fixed locations of the access points). As a result these systems do not scale well over large areas without stringent infrastructure demands.

1.3 Combining GSM/GPS: A Hybrid Approach to Place Discovery and Recognition

Technology	Precise Location	Low Infrastructure Demands	Low Power Usage	Scalable Coverage	Works Indoors
GPS	✓	✓		✓	
GSM		✓	✓	✓	✓
802.11 WiFi					✓
Bluetooth			✓		✓

Table 1.2: Comparison of existing location technologies

Table 1.2 summarizes the performance of various location technologies against the requirements in Section 1.1 for a place-discovery system. Thus, to meet the criteria raised in Section 1.1, we propose a novel two-phased approach that performs place discovery from GPS data traces, and then maps this information to GSM cells so that place recognition can occur in a resource-efficient manner. This approach has the following benefits of GPS-based analysis: full knowledge of user location and rich metadata associated with each location. It has the following benefits of GSM: instantaneous access to location data and the ability to recognize a learned place without external infrastructure.

The complexity of place discovery in the absence of absolute location coordinates is very high. The two-phase approach has the advantage of using precise location knowledge only when necessary to build the initial state space of a user's significant places. We show that obtaining this estimate of the state space allows battery-efficient GSM-based place recognition to take over.

1.4 Thesis Outline

The rest of this thesis is structured as follows:

Chapter 2 describes prior research in location-aware systems with emphasis on clustering algorithms for place-discovery and prediction. We evaluate existing approaches to place-discovery using the criteria of robustness of place discovery, infrastructure demands, and power usage.

Chapter 3 describes our novel two-phased approach that combines GPS and GSM technologies for place-discovery and recognition. Our approach does not rely on knowing the exact network topology of GSM base stations, and maps GSM data to discovered places using computational geometry.

Chapter 4 provides the results of an empirical investigation of our two-phased approach to place-discovery. In this chapter, we present a case study in which a user collects GSM and GPS traces over a one month period in urban Cambridge. Despite noisy data traces, we show that our approach is able to successfully identify the user's significant places.

Chapter 5 lists possibilities for future work and summarizes the contributions of this thesis.

The key contributions of this thesis are:

- A survey and critique of clustering algorithms used in place-discovery systems
- The development of a novel place-discovery scheme combining the best aspects of GPS and GSM location detection
- A prototype system to automatically collect, analyze, and build travel summaries of GPS and GSM location data.

Chapter 2

Related Work

This thesis examines the effectiveness of combining GPS and GSM location data to learn locations that are significant to the user. While this combination is novel, place-discovery and localization have been explored before. In the literature, previous researchers have developed two general classes of place-discovery algorithms: geometric algorithms and fingerprinting algorithms. Geometric algorithms are most commonly used in coordinate-based systems and identify significant places in terms of circles and polygons in reference to absolute location coordinates. Fingerprinting algorithms are typically used in landmark/beacon-based systems (in which no explicit geographic topology is known) and identify significant locations in terms of unique "signature" sequences of metrics such as cell tower id or signal strength patterns. In this chapter, we survey the literature on clustering strategies for place-discovery.

2.1 Geometric Place Discovery

Perhaps the pioneering work on place-extraction is Marmasse and Schmandt's work on the *comMotion*[15] system. Their key insight is that the loss of GPS signal signifies an important place because it indicates that the user has arrived inside of a building. Unfortunately, this approach does not identify significant places in outdoor settings (e.g. parks, camping trails), and is prone to produce false-positives due to "urban canyon" effects occluding the GPS signal.

Ashbrook and Starner[3] improves place-extraction by pre-processing GPS data to filter for only when user movement is detected and identifies points as significant if the user's dwell-time is greater than 10 minutes. In their study, a user's significant places are identified using a variant of the well known k-means clustering algorithm, and these clusters are subsequently examined in temporal sequence to build a Markov model to predict future transitions between locations. However, k-means clustering requires the number of clusters to be known before clustering, favors symmetrically-shaped clusters (e.g. circles, spheres) and is sensitive to noise in the data. We follow a similar approach to place-discovery in this thesis, but incorporate the predictive model to further distinguish between clusters that are significant locations and "clusters" where the user is actually in transit between two locations.

Density-based clustering approaches such as Density-Join clustering (DJ-Cluster) [20] use the density of local neighborhoods of points for place discovery, and is an improvement over k-means clustering due to its robustness to noise and anomalous points. In contrast to k-means clustering, density-based clustering can identify clusters of arbitrary shape. Here, density is defined by two parameters — the radius of a circle and the minimum number of points within the circle. Density-based clusters have a notion of "reachability" that allows a chain of connected, merged circular neighborhoods to form a cluster of arbitrary shape, and clusters identified by DJ-Cluster have considerably higher density than points outside of the cluster.

Zhou et al. later devised the Time Density Join (TDJ) and Relaxed Time Density Join algorithms as improvements to DJ-Cluster by adding additional parameters such as relaxed time constraints on the formation of a neighborhood/cluster to capture often-visited locations with small dwell-time.

Kang et al. [12] simplifies place extraction by contributing an incremental clustering algorithm that identifies places with arbitrary shape from traces of Wi-Fi location data observed from PlaceLab[1]. The drawback of this algorithm is that its temporal clustering requires frequent data sampling of one GPS reading per second, which can result in significant battery life depletion for a mobile device.

2.2 Fingerprint Place Discovery & Localization

Laasonen et al.[13] uses the transition between GSM cell towers to build a graph representation of the user's travels. In contrast to the previous geometric coordinate-based algorithms, Laasonen's approach does not attempt to estimate absolute location, and instead uses groups of cell tower id transitions to approximate a significant location.

Otsason et al.[16] built the first accurate indoor GSM localization system and demonstrated that GSM towers and signal strength can be used to accurately locate a user. Their system uses signal-strength fingerprints of up to 29 GSM channels in addition to the 6-strongest GSM cells detected by a mobile phone to locate the user. Although their primary goal is not GSM-based place-discovery, their system shows that wide-area GSM fingerprints can achieve median accuracies of 2.5 to 5 meters indoors, and that metropolitan environments exhibit significant signal diversity to enable high localization accuracy.

BeaconPrint[11] continually logs Wi-Fi Access ID's and GSM tower ID's from its environment to discover significant places. The system defines a significant place by recognizing "stable scans" — in which no unexpected unique ID's appear within a small time window — and associates a histogram of GSM and Wi-Fi Access Point ID's with each significant place. During place recognition, BeaconPrint compares a live fingerprint with the place histograms and presents an ordered list of the device's most likely locations.

Chapter 3

Landmark/Coordinate combined clustering

In this chapter, we describe an approach to find locations that the user considers significant and to maintain succinct records of the user's travels. We combine GPS and GSM cell data in a two-phased approach as follows: (1) partition the GPS data into cluster regions using a variant of the k-means clustering algorithm and build a Markov model and travel summary of the data to identify significant locations, (2) analyze the geometric properties of the GSM cells' coverage areas to define clusters in terms of GSM cell data as a potential optimization to on-device place recognition.

3.1 GPS Place Discovery & Prediction

3.1.1 GPS Place Discovery

We are interested in learning locations with high *dwell-time* — e.g. where the user spends the most time. We use a k-means clustering approach similar to that of Ashbrook and Starner[3] in order to learn important locations from a user's GPS log traces (see Algorithm 3.1.1).

A *cluster* is a circular region defined by a center and a radius. The k-means clustering algorithm takes as input a radius r and the set of points to be partitioned into clusters of the input radius. The algorithm iteratively forms new clusters and assigns points to each cluster until all points are covered by a cluster.

Cluster formation starts by choosing a data point p at random and defining a cluster with center at p. The algorithm marks all points within r distance of p and calculates the average center of all of the points in this cluster. If the average center changes, the cluster has moved from its initial position, and the algorithm iteratively scans the dataset to mark new points within r of this new center, re-caculating the average center at the end of each scan. When the average center is stable, the formation of the current cluster is complete, and there are no unmarked points in the dataset that also belong to the current cluster. The algorithm then removes members of the completed cluster from consideration, and repeatedly forms new clusters until there are no points to consider.

Algorithm 1
Variant of the k-means clustering algorithm. Iteratively partitions the input data into clusters of radius *radius*.

```
compute_clusters(gps_data, radius)
1: Filter gps_data for only points when user speed > 1 mile/hour
2: places $\Leftarrow$ Filter gps_data for points with dwell-time > 10 minutes
3: clusters $\Leftarrow$ []
4: while places not empty do
5:   center $\Leftarrow$ choose a random point $p$ from places
6:   neighbors $\Leftarrow$ find all points within radius of center
7:   avg_center $\Leftarrow$ mean center point of neighbors and center
8:   if avg_center == center then
9:     add neighbors to clusters
10:    remove all members of neighbors from places
11:  else
12:    center $\Leftarrow$ avg_center
13:    Goto line 6
14:  end if
15: end while
16: return clusters
```

Similar to Ashbrook and Starner, our clustering approach preprocesses the GPS log data according to the following assumptions in order to reduce the number of irrelevant points in our partitioning:

1. The user does not move at high speeds when at a high dwell-time location. Thus we need not consider points that are far below the average human's walking speed (e.g. less than 3 miles per hour)

2. Filtering by walking speed may create gaps in the dataset where the user stays at a significant location. Our preprocessing step identifies these gaps as points of interest by filtering for points with dwell-times longer than 10 minutes.

One drawback of this algorithm is that it must iterate until all points in the dataset are assigned to a cluster. As a result, the algorithm produces frequently used transition paths between two significant places as clusters. Figure 3-1 shows several such transition clusters. In order to determine which clusters are meaningful vs. transitional, we employ metrics such as probability of visits and total/average dwell-times in our prediction modeling (Section 3.1.2) to reduce the number of erroneous clusters identified as significant locations.

3.1.2 User Movement Prediction

The clustering algorithm gives a list of clusters that account for all GPS points in our data log. Many of these clusters account for transitions between important locations, and we take further steps to identify locations where the user spends the most time by sequencing the clusters into a *travel sequence* and building a Markov model for the user's movement.

To obtain the travel sequence, assign each cluster a unique ID, and substitute each point in the original chronologically-ordered GPS trace with the ID of the enclosing cluster. This gives a chronological sequence of visited clusters during the user's travels.

During construction, the following statistics are maintained for each cluster:

- `total_dwell_time` - total amount of time spent in the cluster

- `average_dwell_time` - average dwell time (helps to identify transitional clusters)

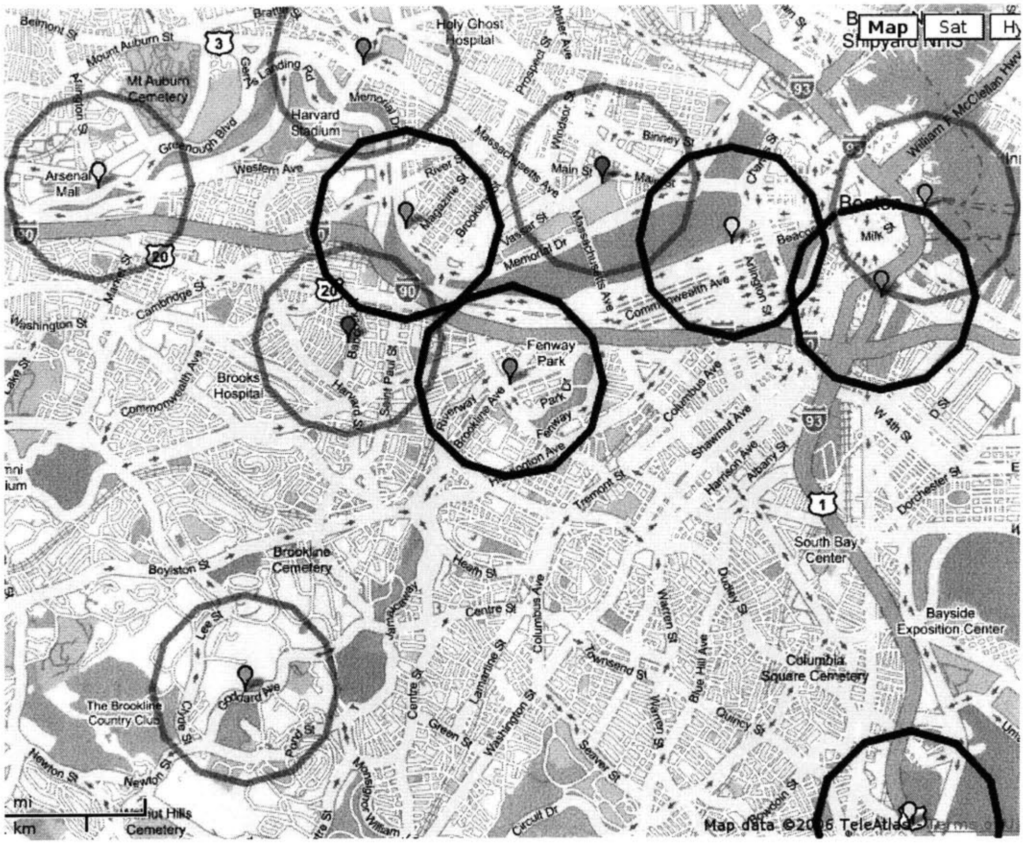

Figure 3-1: Transitional (black) clusters along the highway identified by the clustering algorithm. Movement prediction and dwell-time ranking in Section 3.1.2 help to distinguish between transition clusters and clusters that represent significant locations.

- `nvisits` - number of visits to this cluster (transitions between the same cluster do not increment `nvisits`

- `intervals` - the list of `(arrival_time, exit_time)` pairs denoting periods when the user was in the cluster

Next, a Markov model is built from the chronological travel sequence, in which clusters are states, and transition probabilities between clusters are calculated via frequency counting (see Table 3.1). The Markov model assists in defining clusters that represent significant places — transitions to and from significant places occur with greater than random probability among all observed transitions in the travel sequence, and clusters that represent significant places have much higher total and average dwell-times.

Transition	Relative Frequency	Probability
$Home \rightarrow Home$	16/44	0.3636
$Home \rightarrow c1$	1/44	0.0227
$Home \rightarrow Work$	12/44	0.3636
$Home \rightarrow c5$	1/44	0.0227
$Home \rightarrow c9$ (false-positive)	2/44	0.0455
$Home \rightarrow c10$	1/44	0.0227
$Home \rightarrow c11$	1/44	0.0227
$Home \rightarrow Pool$	3/44	0.0682
$Home \rightarrow c15$	1/44	0.0227
$Home \rightarrow c16$	1/44	0.0227
$Home \rightarrow Store$	2/44	0.0455
$Home \rightarrow Airport$	2/44	0.0455
$Home \rightarrow c27$	1/44	0.0227
$Pool \rightarrow Home$	3/5	0.6000
$Pool \rightarrow Pool$	1/5	0.2000
$Pool \rightarrow Work$	1/5	0.2000
$Work \rightarrow Home$	12/27	0.4444
$Work \rightarrow Work$	10/27	0.3704
$Work \rightarrow c9$	1/27	0.0370
$Work \rightarrow Pool$	1/27	0.0370
$Work \rightarrow c13$	1/27	0.0370
$Work \rightarrow Theatre$	1/27	0.0370
$Work \rightarrow Store$	1/27	0.0370
$Store \rightarrow Home$	4/4	1.0000

Table 3.1: Probabilities for transitions in a first-order Markov model for Cambridge-area clusters. Transitions to "transitional" clusters (denoted by cx) occur much less frequently than transitions to significant places (denoted by a human-friendly name).

Finally, the `intervals` data for each cluster summarizes when the user has spent time in each cluster. This information yields a *travel summary* that represents the user's travels much more succinctly than the logged GPS coordinates. The travel summary may then be kept on the mobile device and updated as the user goes about normal traveling routines. In the next section, we discuss optimizations to query the travel summary using GSM cell tower data.

3.2 GSM Cell Tower Coverage

Once the clustering algorithm in Section 3.1.1 partitions the GPS data into clusters, efficiently recognizing when a user enters a cluster during active travel allows the travel summary to be updated without constraining the resources of the mobile device. Mobile phones constantly scan for GSM cell tower presence, making GSM cells are an attractive and low-cost option for place-recognition.

Prior approaches to GSM place-recognition and localization rely on two general approaches: fingerprinting algorithms that define significant places in terms of unique cell id sequences (independent of actual geographic information); and signal strength variation (when the location of the cell base station is known explicitly). In contrast to these prior approaches, we combine GSM cell data with GPS data to uniquely identify clusters using geometric properties of the GSM cell observations.

The main geometric property of interest is a GSM cell's *coverage polygon*. Each timestamped GPS data point is accompanied by the unique ID of the GSM cell tower observed at the time of reading. We define a cell's coverage polygon to be the convex hull of GPS points from which we observe its signal. We calculate the convex hull for each GSM cell using the well-known Graham Scan algorithm[10], and investigate the following hypothesis in Chapter 4:

- A high dwell-time cluster exhibits stable GSM cell coverage. The user moves at relatively slower speeds when inside of a building or significant location, resulting in observed coverage polygons to be fairly compact.

- A high dwell-time cluster contains GSM cells with large % area overlap with the area covered by the cluster; these cells can uniquely identify the cluster.

- Routes along which the user travels between high dwell-time clusters locations exhibit much narrower GSM cell coverage polygons, since the user is traveling in a relatively fast, directed manner.

We define compactness of a cell coverage polygon in terms of its *aspect ratio* and provide an empirical case study and analysis of our approach in Chapter 4.

Chapter 4

Case Study Results

To evaluate our two-staged approach, we construct a system to record and model an individual's travels over a one-month period. In its current form, the system estimates the initial state space for significant places using an approach similar to the k-means clustering of Ashbrook and Starner's system[3]. Our system then uses computational geometry to define the clusters in terms of GSM cell signatures. Our results show that the aspect ratios of GSM cell coverage polygons while the user is in transit are significantly lower than those when the user is at a high-dwell-time location.

4.1 Data Collection

We collected continuous GPS and GSM trace logs over a one-month period as a research group member went about his normal life routines. The case study subject strove to carry a Nokia 6680 mobile phone and a Bluetooth GPS receiver at all times. The mobile phone ran a simple data collection client that collected timestamped location data from the GPS receiver and GSM cell id's from the phone's built-in cell signal scans.

The data collection client scanned for GPS and GSM cell readings at a rate of 1 reading per 2 minutes. The client uploaded the timestamped log entries to a remote server where the entries were amalgamated for post-processing using our two-staged approach.

In total, we collected approximately 2,000 data points over 840 hours of data logging, and encountered 282 unique GSM cells. Our data is quite noisy, however, since the data collection client suffered occasional crashes, GPS data fixes were not always available due to urban canyons in the Cambridge area, and our subject occasionally chose not to log his locations due to privacy.

4.2 Data Collector Demographic

The subject of our case study typically drives or bikes to and from work every day, but frequently takes walks around his residential area to shopping centers, the library, and nearby parks. He is also an avid traveler. During our logging period, he made a 1,000-mile trip to Urbana, Illinois for a conference as well as a number of recreational trips (e.g. rafting and whale-watching) far from the routine Cambridge, MA location.

4.3 Experimental Results

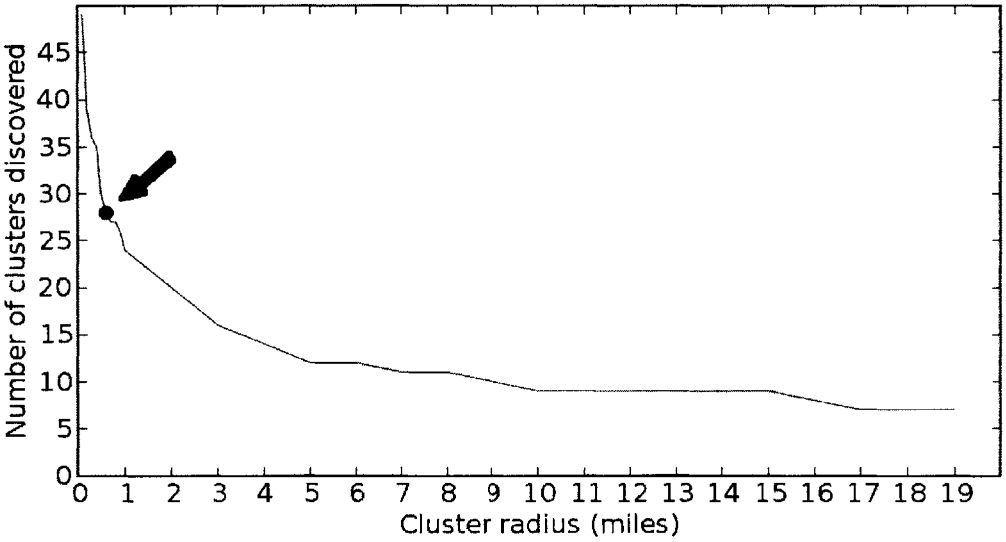

Figure 4-1: Number of "significant places" found as the cluster radius parameter changes. The arrow denotes a "knee" at radius=0.6 miles — the radius before the number of clusters converges to the number location points.

4.3.1 GPS Clustering

To obtain the optimal cluster radius, we calculate the number of clusters found for a range of input radii and look for the radius just before the number of clusters converges to the number of points in our dataset. Figure 4-1 shows the number of clusters found against the clustering algorithm's input radius parameter. We observe that 0.6 miles is the optimal clustering radius for our subject's travel patterns. Naturally, larger cluster radii may result in place discovery that is too coarse-grained, and small radii may produce too many individual clusters (see Section 4.4.1 for an analysis of our clustering algorithm's performance).

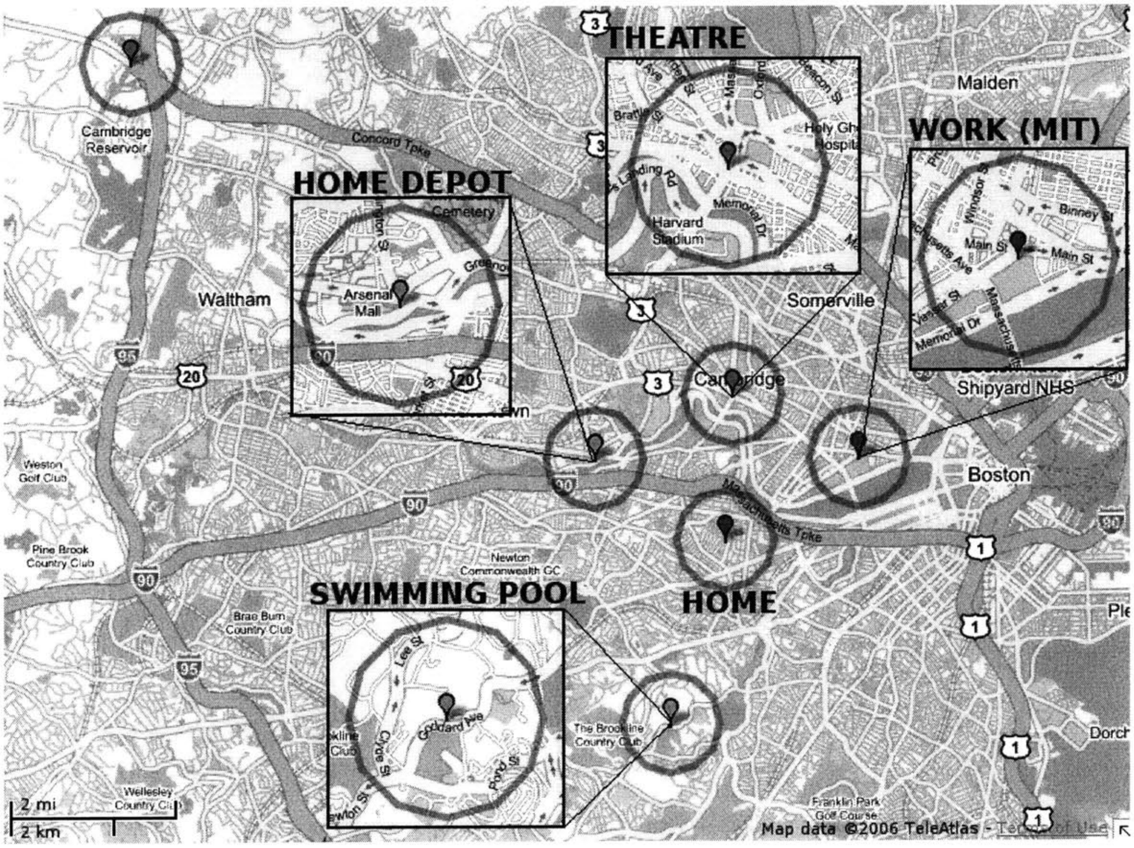

Figure 4-2: Partial map view of the highest dwell-time clusters in Cambridge, MA found by our variant of the k-means clustering algorithm. Ranking the clusters by total dwell-time successfully eliminated noisy transition clusters from these results.

Our algorithm with cluster radius of 0.6 miles defines 27 clusters of interest. However, the k-means clustering algorithm was quite susceptible to noise and identified several areas where the user frequently traveled in transit between significant places as significant. Our system successfully used total dwell-time as a metric to eliminate the majority of noisy clusters from the final results (partially shown in Figure 4-2).

4.3.2 Geometry of GSM Cell Coverage

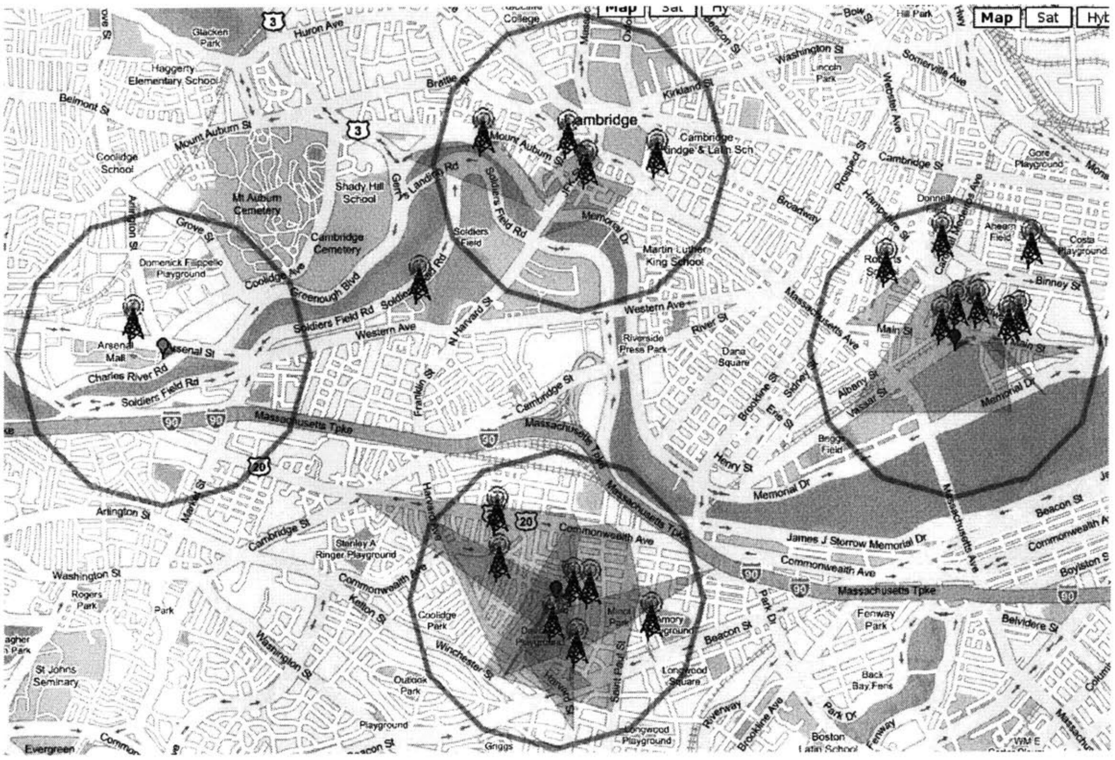

Figure 4-3: Cell coverage polygons for cells that uniquely identify significant places have relatively large aspect ratios. The coverage polygons that uniquely define each cluster have mean aspect ratio=0.49883. The clusters for place (mean aspect ratio) in this figure are as follows: *Home* (0.4398), *Work/MIT* (0.4402), *Home Depot* (0.68741), and *Harvard Theater* (0.4279).

Our algorithm calculates *cell coverage polygons* — e.g. the convex hull of GPS coordinates from which the GSM cell was observed — for each of 282 unique GSM cells. Our results show that the observed cell coverage polygons closely mimic the user's state of motion. In contrast to cell coverage polygons in high dwell-time clusters, cell coverage polygons in clusters where the user is moving from one place to another tend to be much narrower. Our results suggest that cell coverage polygon aspect ratios may augment conventional fingerprinting approaches in distinguishing between a significant cluster and a transitional cluster.

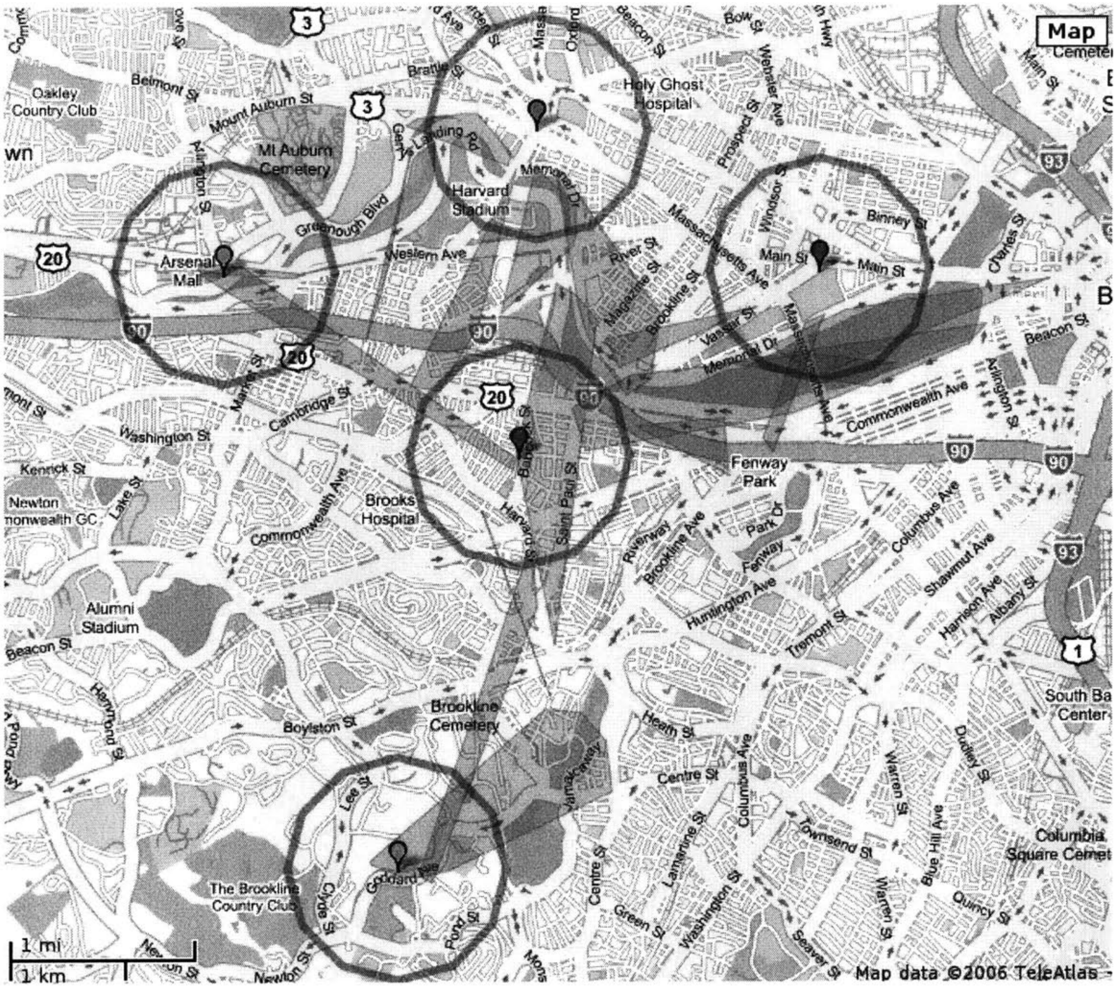

Figure 4-4: Cell coverage polygons for cells observed during transit between significant places have significantly narrower aspect ratios (Mean=0.1836). The coverage polygons closely map the user's path, and are characterized by long, narrow strips.

High-dwell time places result in several GSM cell coverage polygons contained entirely within the cluster. Thus, observations of these *cluster-identifying cells* are sufficient to define a cluster. Our data reveals a large difference in shape between clusters that represent significant places and the transitional areas connecting them. Coverage polygons for cluster-identifying cells and transition cells of key locations in Cambridge are shown in Figures 4-3 and 4-4, respectively. Due to continued movement and higher speeds during transitions, transition cell coverage polygons (mean aspect

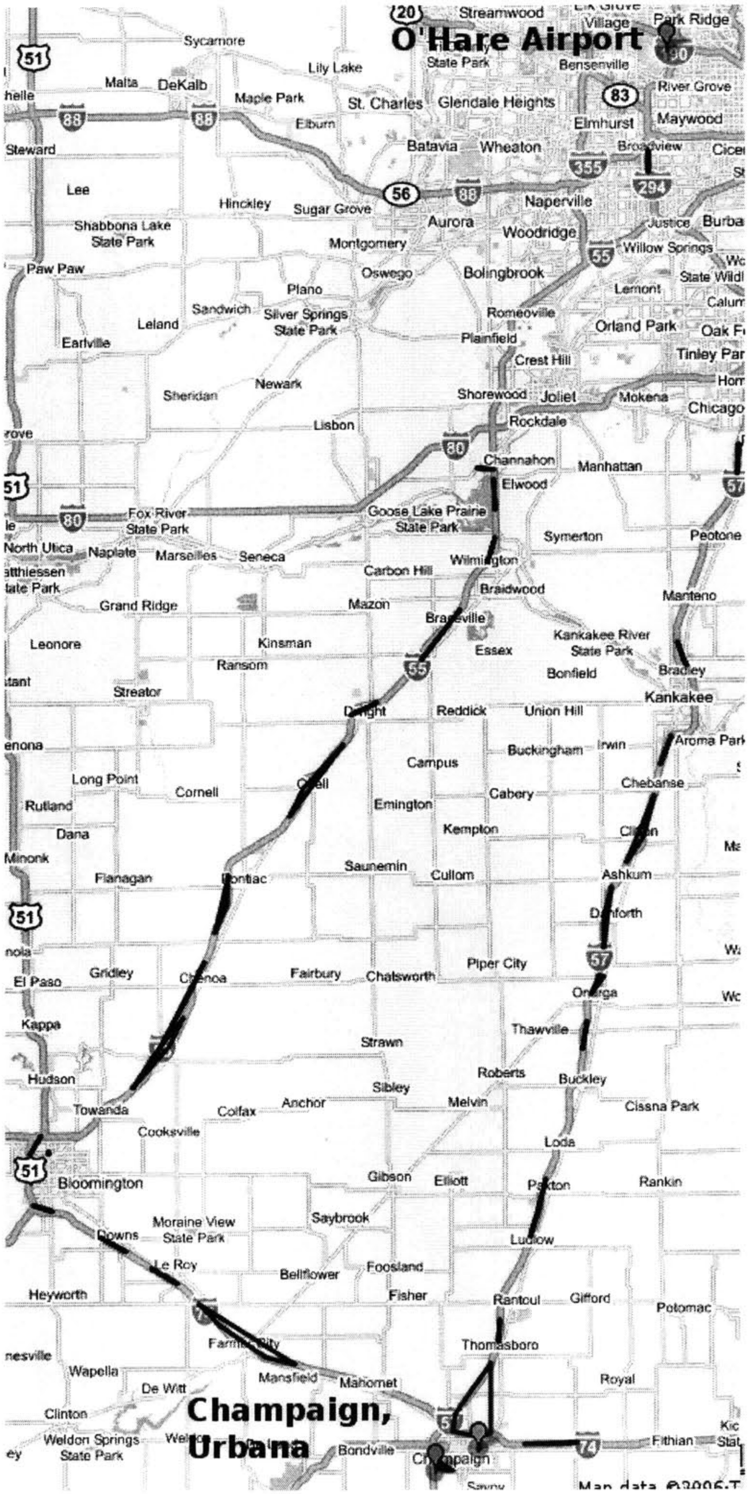

Figure 4-5: Cell coverage polygons for cells observed during the user's conference in Urbana, IL in which the user spent the majority of his transit time on the highway.

ratio=0.1836) have much smaller aspect ratios than those of cluster-identifying cells (mean aspect ratio=0.49883).

The effect of travel speed on coverage polygon aspect ratio is most apparent in data collected during the user's trip to Illinois (Figure 4-5). Here, the narrowest strips of GSM cell coverage exactly align with the path along the interstate highway.

Aspect ratio is not applicable to all of our GSM cell data — namely, several GSM cells' coverage polygons consist of exactly one or two points. The sparseness of these GSM cell observations also suggest that the user was most likely traveling. The rest of our dataset verifies this hypothesis, since the majority of single-point and two-point convex hulls for GSM cell coverage are on paths along a major express way. Table 4.1 summarizes thc characteristics of GSM cell coverage polygons with respect to their convex hull size among the 282 unique GSM cell.

Convex Hull	# Transient	Avg. %Area Overlap	Avg. Aspect Ratio	
Size	Cells	(non-transient)	transient	non-transient
1	121/149 (81.20%)	100%	N/A	N/A
2	30/37 (81.08%)	99.78%	0.0003 (for all 2-pt. hulls	
3+	32/96 (33.33%)	56.08%	0.15684	0.30703

Table 4.1: Sparseness of GSM cell observations indicates that travel speed is relatively high; thus, GSM cells with coverage polygons consisting of one or two points occur outside of high dwell-time clusters with high probability.

Our investigation provides some preliminary foundations for using cell coverage polygon shape to differentiate between transition and cluster-identifying cells. Figure 4-6 plots the coverage polygon aspect ratios of cluster-identifying cells against those of transition cells with coverage polygons that overlap a significant cluster. A GSM cell is considered cluster-identifying if at least 85% of its coverage area overlaps with a cluster region. Currently, we only consider GSM cells with coverage polygons of three or more vertices. Due to the sparseness of our data, we are able to show comparative data only for the most heavily-visited locations.

It may seem peculiar that the cluster-identifying cells for left-most three transitional clusters have narrower aspect ratios than cells that only partially overlap with these regions. This is due to the GSM coverage polygons produced by transporta-

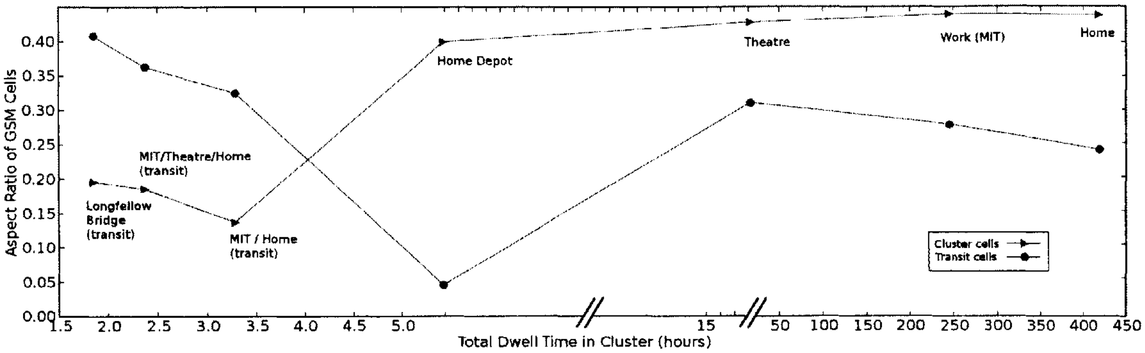

Figure 4-6: A comparative plot of aspect ratios for cluster-identifying and transition cells (with coverage polygons of at least 3 vertices) for high dwell-time clusters and frequently-travelled transition areas.

tion paths along the Charles River. The user spends a great deal of time traveling around (but never entering) the river, thus the cell coverage polygon is wider than the actual paths the user uses within its coverage area. On the other hand, GSM cells whose coverage polygons fall within the transitional clusters are direct land routes, and therefore have small aspect ratios similar to the coverage polygons produced by direct highway routes during the user's trip to Illinois. Thus, these results are consistent with our hypothesis that comparing the aspect ratios of the identifying and non-identifying cell coverage polygons can differentiate between significant places and transitions between them.

4.4 Evaluation & Analysis

4.4.1 Clustering Performance

Our clustering algorithm is a simple k-means clustering variant that automatically derives k, the number of clusters to account for the dataset. For some input configurations, the worst-case running time for the k-means clustering algorithm is superpolynomial (a lower bound of $2^{\Omega(\sqrt{n})}$ as shown by Arthur and Vassilvitskii[2]). In practice, this algorithm converges very quickly — e.g. the number of iterations is much less than the size of the GPS dataset to be clustered (see Figure 4-7).

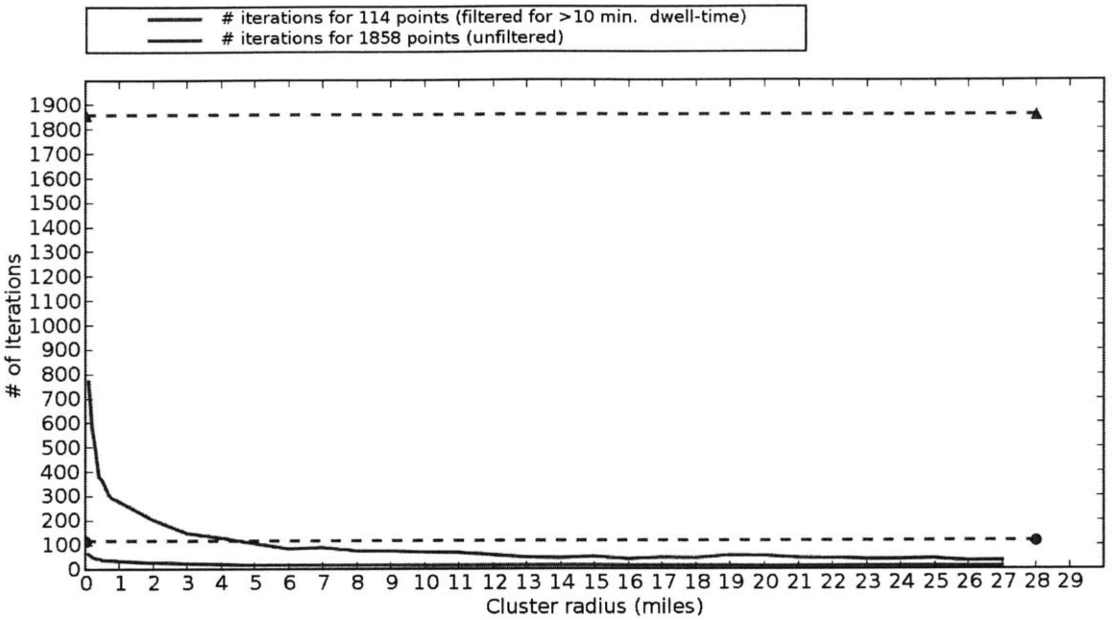

Figure 4-7: Performance (in terms of # of clustering iterations) for the raw unfiltered GPS data (1858 total points), and for pre-processed GPS data (114 total points). In practice, the number of clustering iterations is much less than the total number of points to be processed.

While the clustering algorithm works especially well when spherical clusters are naturally available in the data, the amount of overlapping clusters discovered by our clustering algorithm suggests that location data may not conform to spherical clusters. Overlapping clusters introduce ambiguity into the chronological sequence with which the clusters are visited (e.g. how to classify a new location coordinate that is within range of two or more cluster centers). Ideally, the clustering algorithm discovers clusters that are not only significant in terms of the user's dwell-time, but are also disjoint. Figure 4-8 shows that the percentage of overlapping clusters is sensitive to the input radius; even our optimal cluster radius of 0.6 miles still results in 25% of the clusters overlapping. These cluster overlaps suggest that the clusters may have the wrong shape or radius; this suggests an investigation to determine whether running our clustering algorithm recursively on areas of heavy cluster overlap with smaller input radii improves on this problem.

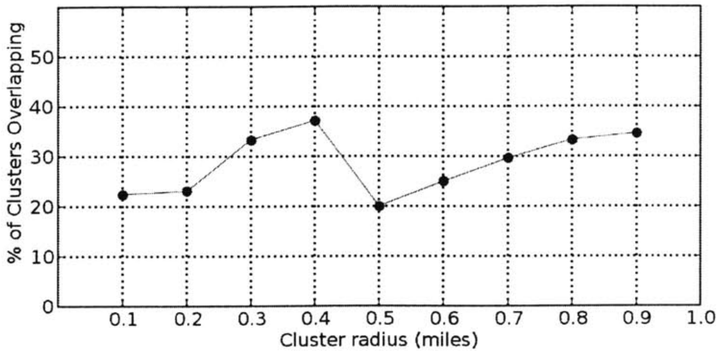

Figure 4-8: Percentage of clusters that overlap as cluster radius changes.

4.4.2 Travel Summary Data

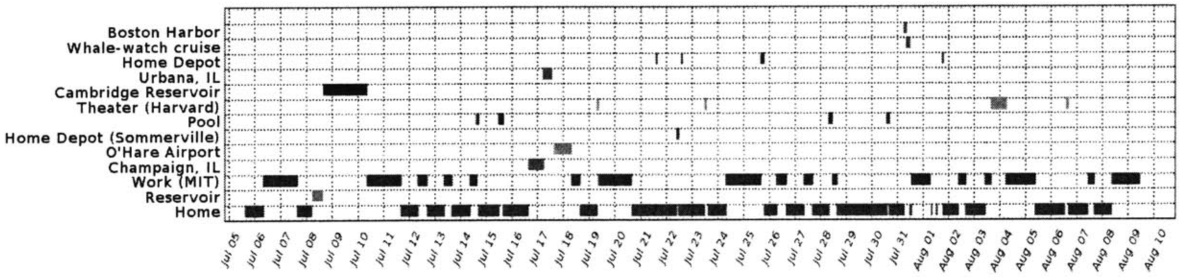

Figure 4-9: Travel summary indexed by location (transitions omitted). The temporal sequence of visited clusters yields a mapping from each cluster location to the set of (t_{arrive}, t_{exit}) time intervals when the cluster was visited.

When combined with the original GPS dataset, the clusters extracted by the algorithm yield the temporal sequence of cluster visits. This sequence can be used to map each cluster to a list of *(arrival, exit)* timestamp intervals. These mappings form a *travel summary* that is much more compact—approximately one order of magnitude less storage for our month's data—than keeping the entire GPS dataset on record.

The travel summary for our case study shows that the subject's daily routines revolve around Home and Work with preferred hours of travel activity between 9AM and 8PM (visualized in Figures 4-9 and 4-10; detailed dwell-time intervals for the Cambridge locations can be found in Appendix A). Our GPS data traces are extremely noisy — the dotted lines in Figure 4-10 indicate time intervals when the user enters a cluster without exiting it in the same day. In practice, the user did not

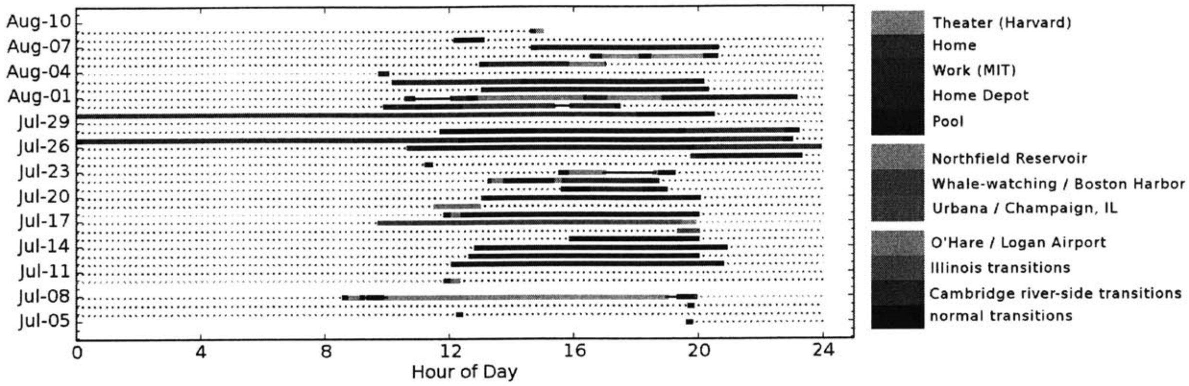

Figure 4-10: Travel summary indexed by date shows a per-day breakdown of the user's travels. The dotted lines denote time intervals in which the user enters and exits a cluster on *different* days — an indication that the user lost GPS signal, turned off the GPS device, or experienced a crash.

always remember to collect location data, and daily GPS traces often ended prematurely (For example, on 7/10/06, the GPS traces show the subject moving from Home to Work with the last recorded location at Work. The next day, the first recorded GPS location is at Home, suggesting that the user forgot to turn on the GPS device until the next morning). Discontinuity in data collection tends to exaggerate the dwell-times of clusters visited across the discontinuity. While this does not affect the dwell-time ranking on the anchors of the user's routine (e.g. Home and Work), it can produce misleading results for less commonly-visited places. In future research, we are interested in exploring methods to identify discontinuities in data collection and use them to quantify a confidence metric for our results.

The travel summary gives the foundation for answering queries such as "Who called me on the way from Home to Work?" As the user continues to log location data, the travel summary provides a compact way to maintain and update *(arrival, exit)* intervals for each cluster. Furthermore, our analysis of GSM cell coverage polygons suggest that the mobile device may recognize important locations with its built-in GSM scanning capabilities. Once clusters are extracted from the GPS data, maintaining their dwell-time and *(arrival, exit)* interval statistics using GSM cells instead of GPS can serve as a significant optimization to resource usage on the mobile device.

Chapter 5

Future Work & Conclusion

5.1 Future Work

5.1.1 Hierarchical Place-Discovery

One shortcoming of our approach is that the k-means clustering algorithm relies on a fixed radius. As can be seen from the mapped clusters and GSM cell coverage polygons in Chapter 4, the user actually covers a very small amount of area within some clusters (e.g. *HomeDepot*'s cell coverage is incredibly small compared to the overall area of its enclosing cluster). One possibility for future work would be to use the cluster-identifying cell coverage polygons to adaptively learn the "right" radius for each cluster. We would like to investigate the effect of clustering algorithms that identify clusters with arbitrary shape on cluster-identification with GSM cells.

Our data over the one-month collection period describes a user that travels on multiple scales (foot, bike, car, plane, cruise ship), but our clustering algorithm does not have an automatic way to detect significant location on different scales. For example, the user walks around his neighborhood to the library, but our clustering algorithm was not able to pinpoint the library as a specific waypoint. Our current methods identify clusters with high self-loop transition probability, and in the future, we may use high self-loop transition probabilities in the Markov model as a heuristic to indicate whether finer-grained clustering is required. We would then recursively run the clustering algorithm with smaller radii on clusters with high self-loop transitions.

5.1.2 Signal Strength Variability

We currently use cell coverage polygon aspect ratio to determine whether the user is transitioning between significant places. Prior work due to Sohn et al. [19] has indicated that GSM cell signal strength varies considerably when the user is in motion. We believe that signal strength variance can augment the ability to identify transition clusters.

5.1.3 Road Network Information

The user's traveling around a natural obstacle such as a river or lake can produce cell coverage polygons with fairly high aspect ratio for transition GSM cells. Combining coverage polygon information with road and terrain information would improve the certainty with which we identify transition GSM cells.

5.1.4 Semantic Labels for Significant Locations

An important area of future research is to devise a scheme to automatically name clusters where the user spends significant amounts of time. One approach to doing this may be to associate entries in the user's address book or calendar with latitude/longitude coordinates using web-based translation services. This approach has potential because our two-phased approach provides location information along with clustering-identifying GSM cells, so a mobile phone can obtain an approximation of its latitutde/longitude position when it detects several identifying GSM cells by calculating the intersection of their coverage polygons.

5.1.5 Personalized Navigation Routes

Patel et al.[17] have began preliminary work on navigation systems that produce user-specific driving routes. Their approach uses step trees and a user profile representing the user's *a priori* knowledge to produce routes with reduced route complexity by compressing well-known steps along familiar routes into a single contextualized step. Their system requires users to manually input landmarks and specify how the landmarks are connected to each other, and it also uses a boolean notion of familiarity for known landmarks. The place discovery and prediction modeling described in this thesis complements the goal of personalized navigation routes by automatically extracting significant places and providing a spectrum of familiarity based on metrics such as dwell-time and transition probability.

5.1.6 Time-based Decay of Familiarity

In our current system, the number of significant locations for a user increases monotonically over time. However, this may not truly reflect reality, as a user's travel patterns may change significantly over time (e.g. changes in schedule results in the user's life being structured around a different set of locations). We would like to investigate the notion of time-based decay so that older high dwell-time clusters that have not been visited for a long time do not dominate more recently-discovered clusters. This may provide a useful foundation for examining how a user's clusters change over time.

5.2 Conclusion

In this thesis, we have made the following contributions:

- Two-phased place discovery and recognition using computational geometry on GPS and GSM data

  In Chapter 3, we describe a novel two-phase approach to construct and maintain a user's travel summary using timestamped GPS location and GSM cell data. We depart from past approaches by analyzing the geometric properties of GSM cell coverage polygons to distinguish between significant places and areas of transition.

- Case study and analysis of GPS and GSM place-discovery in practice

  In Chapter 4, we investigate the effectiveness of our two-phased approach in a one-month user case study. Our results are encouraging and show that travel summary and GSM cell coverage polygon aspect ratios are compact representations for place discovery and recognition. Our analysis identifies some anomalous characteristics of data collection on mobile devices in practice. We believe further exploration of these results will lead to significant improvements in the accuracy and reliability of place-discovery systems.

Appendix A

Tables

LOCATION	ARRIVE	LEAVE
Pool	2006-07-14 19:12:53	2006-07-14 20:49:23
	2006-07-15 15:53:07	2006-07-15 19:54:26
	2006-07-25 23:03:25	2006-07-25 23:15:15
	2006-07-27 22:45:50	2006-07-27 22:57:42
	2006-07-28 11:43:24	2006-07-28 14:51:17
	2006-07-28 22:48:15	2006-07-28 23:10:24
	2006-07-29 23:50:37	2006-07-30 00:03:34
	2006-07-30 18:05:19	2006-07-30 20:25:14
	2006-08-05 13:01:23	2006-08-05 13:09:05
Home Depot	2006-07-21 18:13:27	2006-07-21 18:53:43
	2006-07-22 18:12:00	2006-07-22 18:37:50
	2006-07-25 19:56:51	2006-07-25 22:54:22
	2006-08-01 22:03:03	2006-08-01 23:05:23
Home	2006-07-05 19:44:49	2006-07-06 12:14:25
	2006-07-06 12:15:26	2006-07-06 12:16:28
	2006-07-07 19:46:17	2006-07-08 08:36:12

LOCATION	ARRIVE	LEAVE
Home	2006-07-10 11:52:13	2006-07-10 11:54:20
(continued)	2006-07-10 11:55:22	2006-07-10 11:56:24
	2006-07-11 20:15:53	2006-07-12 12:05:04
	2006-07-12 20:41:43	2006-07-13 12:30:15
	2006-07-13 19:54:37	2006-07-14 12:49:40
	2006-07-14 20:49:23	2006-07-15 15:53:07
	2006-07-15 19:54:26	2006-07-16 19:21:55
	2006-07-18 11:50:09	2006-07-18 11:52:24
	2006-07-18 19:55:38	2006-07-19 11:30:27
	2006-07-20 13:04:15	2006-07-20 13:08:05
	2006-07-20 19:58:13	2006-07-21 15:30:42
	2006-07-21 16:24:40	2006-07-21 18:13:27
	2006-07-21 18:53:43	2006-07-22 13:15:56
	2006-07-22 15:37:12	2006-07-22 18:12:00
	2006-07-22 18:37:50	2006-07-23 15:32:25
	2006-07-23 15:34:40	2006-07-23 15:36:55
	2006-07-23 19:09:30	2006-07-24 11:14:18
	2006-07-25 19:48:15	2006-07-25 19:56:51
	2006-07-25 22:54:22	2006-07-25 23:03:25
	2006-07-25 23:15:15	2006-07-26 10:38:52
	2006-07-26 10:41:07	2006-07-26 10:43:46
	2006-07-26 19:44:59	2006-07-26 23:51:26
	2006-07-27 00:06:11	2006-07-27 00:06:57
	2006-07-27 00:07:43	2006-07-27 00:08:29
	2006-07-27 00:09:14	2006-07-27 12:19:39
	2006-07-27 20:16:43	2006-07-27 20:17:29
	2006-07-27 20:18:02	2006-07-27 20:18:53
	2006-07-27 20:19:38	2006-07-27 22:45:50
	2006-07-27 22:57:42	2006-07-28 11:43:24

LOCATION	ARRIVE	LEAVE
Home	2006-07-28 19:30:48	2006-07-28 19:31:34
(continued)	2006-07-28 19:32:07	2006-07-28 19:32:52
	2006-07-28 19:33:50	2006-07-28 22:48:15
	2006-07-28 23:10:24	2006-07-29 23:50:37
	2006-07-30 00:03:34	2006-07-30 16:43:11
	2006-07-30 17:11:12	2006-07-30 18:05:19
	2006-07-30 20:25:14	2006-07-31 09:53:39
	2006-07-31 15:46:46	2006-07-31 17:13:23
	2006-07-31 17:13:55	2006-07-31 17:14:24
	2006-07-31 17:15:03	2006-07-31 17:15:34
	2006-07-31 17:16:04	2006-07-31 17:16:35
	2006-07-31 17:17:06	2006-07-31 17:17:36
	2006-07-31 17:18:15	2006-07-31 17:18:46
	2006-07-31 17:19:16	2006-07-31 17:19:47
	2006-07-31 17:20:18	2006-07-31 17:20:47
	2006-07-31 17:21:26	2006-07-31 17:22:27
	2006-08-01 10:37:53	2006-08-01 10:41:05
	2006-08-01 12:05:36	2006-08-01 12:35:41
	2006-08-01 12:36:12	2006-08-01 12:36:43
	2006-08-01 12:37:25	2006-08-01 12:37:56
	2006-08-01 16:18:55	2006-08-01 16:57:02
	2006-08-01 23:05:23	2006-08-02 13:03:07
	2006-08-02 13:03:38	2006-08-02 13:04:09
	2006-08-02 13:04:50	2006-08-02 13:05:20
	2006-08-02 20:14:15	2006-08-02 20:14:54
	2006-08-02 20:15:24	2006-08-02 20:15:55
	2006-08-02 20:16:25	2006-08-02 20:16:55
	2006-08-02 20:17:25	2006-08-03 10:11:13
	2006-08-03 10:28:16	2006-08-03 10:29:31

LOCATION	ARRIVE	LEAVE
Home	2006-08-03 10:30:48	2006-08-03 14:17:36
(continued)	2006-08-05 13:09:05	2006-08-05 15:51:58
	2006-08-05 16:56:21	2006-08-05 16:57:36
	2006-08-05 16:58:52	2006-08-06 16:33:14
	2006-08-06 16:34:30	2006-08-06 16:35:45
	2006-08-06 16:37:08	2006-08-06 16:38:24
	2006-08-06 20:15:57	2006-08-07 14:38:06
	2006-08-07 14:39:21	2006-08-07 14:40:36
	2006-08-07 20:34:33	2006-08-07 20:35:07
	2006-08-07 20:35:51	2006-08-07 20:36:24
	2006-08-07 20:36:58	2006-08-08 12:10:43
	2006-08-08 12:11:32	2006-08-08 12:12:20
Theater	2006-07-08 08:38:16	2006-07-08 08:50:55
	2006-07-19 11:30:27	2006-07-19 12:56:22
	2006-07-23 15:43:47	2006-07-23 16:56:01
	2006-07-23 18:35:01	2006-07-23 18:44:17
	2006-08-03 20:06:05	2006-08-04 09:46:05
	2006-08-06 18:23:41	2006-08-06 20:12:11
Work	2006-07-06 12:20:35	2006-07-07 19:42:10
	2006-07-10 12:15:25	2006-07-11 20:15:53
	2006-07-12 12:11:48	2006-07-12 20:36:47
	2006-07-13 12:30:15	2006-07-13 19:50:08
	2006-07-14 12:56:25	2006-07-14 18:57:00
	2006-07-18 11:54:39	2006-07-18 12:06:10
	2006-07-18 12:13:49	2006-07-18 19:51:09
	2006-07-19 12:56:22	2006-07-20 13:04:15
	2006-07-20 13:14:50	2006-07-20 19:32:37

LOCATION	ARRIVE	LEAVE
Work	2006-07-22 13:20:25	2006-07-22 13:22:40
(continued)	2006-07-24 11:21:01	2006-07-25 19:48:15
	2006-07-26 10:45:02	2006-07-26 19:40:32
	2006-07-26 23:58:01	2006-07-27 00:02:32
	2006-07-27 12:25:29	2006-07-27 20:12:45
	2006-07-28 15:00:36	2006-07-28 19:26:51
	2006-07-31 17:24:39	2006-08-01 10:37:53
	2006-08-01 12:42:58	2006-08-01 12:50:05
	2006-08-01 16:02:43	2006-08-01 16:14:03
	2006-08-02 13:14:17	2006-08-02 20:12:40
	2006-08-03 14:27:27	2006-08-03 19:54:49
	2006-08-04 09:57:35	2006-08-05 13:01:23
	2006-08-06 17:58:57	2006-08-06 18:09:12
	2006-08-07 14:45:47	2006-08-07 20:23:33
	2006-08-08 13:01:40	2006-08-09 14:36:05

Appendix B

Figures

Bibliography

[1] Place lab.

[2] David Arthur and Sergei Vassilvitskii. How slow is the k-means method? In *SCG '06: Proceedings of the twenty-second annual symposium on Computational geometry*, pages 144–153, New York, NY, USA, 2006. ACM Press.

[3] Daniel Ashbrook and Thad Starner. Learning significant locations and predicting user movement with gps. In *ISWC*, pages 101–108. IEEE Computer Society, 2002.

[4] Paramvir Bahl and Venkata N. Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *INFOCOM*, pages 775–784, 2000.

[5] Michael Beigl, Stephen S. Intille, Jun Rekimoto, and Hideyuki Tokuda, editors. *UbiComp 2005: Ubiquitous Computing, 7th International Conference, UbiComp 2005, Tokyo, Japan, September 11-14, 2005, Proceedings*, volume 3660 of *Lecture Notes in Computer Science*. Springer, 2005.

[6] Amiya Bhattacharya and Sajal K. Das. Lezi-update: An information-theoretic approach to track mobile users in pcs networks. In *MOBICOM*, pages 1–12, 1999.

[7] Nathan Eagle and Alex Pentland. Social serendipity: Proximity sensing and cueing.

[8] Nathan Eagle and Alex Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2006.

[9] K. A. Franck and L. H. Schneekloth, editors. Van Nostrand Reinhold, 1994.

[10] Ronald L. Graham. An efficient algorithm for determining the convex hull of a finite planar set. *Inf. Process. Lett.*, 1(4):132–133, 1972.

[11] Jeffrey Hightower, Sunny Consolvo, Anthony LaMarca, Ian E. Smith, and Jeff Hughes. Learning and recognizing the places we go. In Beigl et al. [5], pages 159–176.

[12] Jong Hee Kang, William Welbourne, Benjamin Stewart, and Gaetano Borriello. Extracting places from traces of locations. In *WMASH '04: Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, pages 110–118, New York, NY, USA, 2004. ACM Press.

[13] Kari Laasonen, Mika Raento, and Hannu Toivonen. Adaptive on-device location recognition. In Alois Ferscha and Friedemann Mattern, editors, *Pervasive*, volume 3001 of *Lecture Notes in Computer Science*, pages 287–304. Springer, 2004.

[14] Lin Liao, Dieter Fox, and Henry A. Kautz. Learning and inferring transportation routines. In Deborah L. McGuinness and George Ferguson, editors, *AAAI*, pages 348–353. AAAI Press / The MIT Press, 2004.

[15] Natalia Marmasse and Chris Schmandt. Location-aware information delivery with *ommotion*. In Peter J. Thomas and Hans-Werner Gellersen, editors, *HUC*, volume 1927 of *Lecture Notes in Computer Science*, pages 157–171. Springer, 2000.

[16] Veljo Otsason, Alex Varshavsky, Anthony LaMarca, and Eyal de Lara. Accurate gsm indoor localization. In Beigl et al. [5], pages 141–158.

[17] Kayur Patel, Mike Chen, Ian Smith, and James Landay. Personalizing routes, 2006.

[18] Bradley Rhodes. Using physical context for just-in-time information retrieval. *IEEE Trans. Comput.*, 52(8):1011–1014, 2003.

[19] Timothy Sohn, Alex Varshavsky, Anthony LaMarca, Mike Y. Chen, Tanzeem Choudhury, Ian Smith, Sunny Consolvo, and William Griswold. Mobility detection using everyday gsm traces. In *Ubicomp*, 2006.

[20] Changqing Zhou, Dan Frankowski, Pamela Ludford, Shashi Shekhar, and Loren Terveen. Discovering personal gazetteers: an interactive clustering approach. In *GIS '04: Proceedings of the 12th annual ACM international workshop on Geographic information systems*, pages 266–273, New York, NY, USA, 2004. ACM Press.