

Using GSM CellID Positioning for Place Discovering

Filipe Meneses, Adriano Moreira
Dept. of Information Systems
University of Minho
Guimarães, Portugal
{meneses, adriano}@dsi.uminho.pt

Abstract— Mobile phones can be used not only for voice and data communications but also as a computing device running context-aware applications. In this paper we present a model that, based on GSM cell identification, identifies places visited by a user and provides a user familiarity level for each of these places. This information can be used by context-aware applications to adapt their behaviour accordingly to the knowledge its user has about the current location. The achieved results are assessed by overlapping the discovered places with manual collected data, showing that GSM cellID positioning data can be used to identify places that are closer to each other than the average cell radius.

Context; Location; Positioning; Context improvement; Place discovery; Familiarity; GSM cellID

I. INTRODUCTION

The spread of location-based services and context-aware applications is dependent on the existence of free and pervasive positioning technologies to acquire the user or device location. GPS is one of the most popular positioning technologies because it provides world wide coverage, high accuracy, is free and a wide number of non-expensive GPS receivers are available. However, the main flaw of GPS technology is that it only works in open spaces where the satellites signal can be received, failing in narrow streets or inside buildings.

A number of technologies were developed to allow indoor location, mostly based on short-range signals like Bluetooth, WiFi, infrared or RFID. Active Badge [1] (based on infrared sensors) and Active Bat [2] are two examples of those technologies.

WiFi networks [3], can also be used to retrieve the user position. Rough positioning information on WiFi networks can be acquired just by identifying the access points used by the mobile device while more acute positioning information can be provided through the analysis of the radio signal level received from the different access points that cover a geographic area.

GSM (Global System for Mobile Communications) is the standard for the second generation of mobile phones cellular networks, adopted by all the European countries. Currently, these networks cover almost all the territory of many countries and are used by almost 100% of the population of many countries. Assisted-GPS (A-GPS), Enhanced Observed Time Difference (E-OTD) and Time of Arrival (TOA) are three technologies developed to acquire the position of the handsets on the GSM networks [4]. However, the deployment of these

technologies demands the installation of new network equipment in each network cell and, in some cases, to upgrade the mobile devices. Moreover, most of the operators do not make the location information available to third part applications or it is only made available to applications that run on their backbones.

Besides those three technologies, there is a very basic form of positioning in GSM networks known as cellID positioning, in which the handset is located by identifying the network cell to which it is associated. CellID positioning is simple, can be done on the network or on the user device and does not require any upgrade to the handsets or network equipment. However, cellID accuracy depends on the cell size.

Development of universal location-based services and context-aware applications is dependent on the existence of a truly ubiquitous positioning system which works everywhere, without demanding the use of specific equipment. The wide coverage of the GSM cellular networks in the European countries and the very large number of its users makes these networks very appealing to support a universal positioning service. Moreover, mobile phones are small and people take them everywhere, being common the presence of a mobile phone on every people's pockets.

In this paper we present a model that, using the cellID information (freely available on the handsets), and without any prior knowledge of the network topology, and without requiring the intervention of the users, is capable of identify places in a personal symbolic referential. With the proposed model, contextual information about individual users can be inferred from the cellID positioning data, by identifying and characterizing places corresponding to areas most frequently visited by them. Moreover, a familiarity index can be estimated for each identified place which can be used by context-aware applications to dynamically adapt their behaviour based on the knowledge a user has about a place.

In section II we present the mathematical model, which includes the detection of the user movement inside the GSM network, the identification of the visited places and the estimation of the familiarity index. Section III describes the achieved results based on data collected by two independent users. Section IV presents the related work. The last section presents the conclusions and directions for future work.

II. DISCOVERING FAMILIAR PLACES

Although a cellular network is often represented as a set of joined hexagonal cells, which side-by-side covers one area, it is in fact made of cells with different and irregular shapes, with variable dimension, and that overlap partially.

Each cell supports a limited number of simultaneous communications. In urban areas, with higher density of subscribers, the networks are made of a larger number of smaller cells in order to support all the network users and to provide a strong network signal inside the buildings. In some cases, two or more cells are overlaid to support all users of a densely populated geographic area. In contrast, in rural areas where the number of users is smaller, the area covered by each cell is bigger and the number of cells per unit area is smaller.

Identifying the cell that is being used by a mobile terminal is the most basic positioning mechanism for a GSM network. It is also the easiest to implement because this basic information is needed by the network core to keep track of the mobile phones and to be able to forward phone calls or deliver data. ITU assigns to each country a Mobile Country Code (MCC) and within each country a Mobile Network Code (MNC) is assigned to each cellular network operator [5]. Each operator is responsible for creating the Location Area Codes (LAC) for his network and to assign a numeric identification to each cell (cellID). Whenever a mobile terminal is connected to the network, it is associated to one of these cells. Therefore, the absolute location of a terminal can then be expressed by the set {cellID, LAC, MNC, MCC}.

This symbolic positioning information based on the cell identification can be transformed into geographic information by considering the coordinates of the cell centre and its radius. However, a map with the geographic location of the cells and the corresponding identification is usually not made public available by the operators due to privacy reasons. Moreover, a mobile terminal can use networks other than its home network while in roaming. On the other hand, the cellID, LAC, MNC and MCC information of the serving cell is available at the mobile terminal, independently of the network or service provider.

In this paper we present a model that supports the continuous inference of a list of familiar places in a personal symbolic referential for each user. Here, a personal symbolic referential is defined as a list of places that have already been visited by the user, and that can be used to characterize his/her context. This referential is symbolic because each place is described by a symbolic identifier and not by a pair of coordinates or geometric region. The proposed approach is based on detecting time periods in which the user is not moving or is moving slowly within a restricted area, and on identifying and clustering fingerprints corresponding to the GSM cells serving the corresponding area.

A. Movement Tracking

While a user is moving within the coverage area of a cellular network, his terminal is constantly “jumping” from one cell to another. This is called the handover process, and the change from the current cell to a new cell usually occurs when

the signal level of the current serving cell drops below a pre-defined threshold. Consequently, when a user is moving, one expects to observe changes into the cellID based position information. However, since a particular location is often served by multiple cells, handover occurs even when the terminal is stationary. This means that the movement of a terminal can not be assumed just from a change in the cellID.

In a populated area, with several GSM cells, the mobile terminal can change from one cell to another after a few seconds or only after several minutes. Thus, if within a certain period of time the terminal stays in the same cell or changes between the same set of cells, it can be assumed that the user is immobile or moving slowly. This assumption is supported by experimental data that confirms that the correlation between the user movements and the changes in the CellID is, on average, higher the 0.9 (see section III-A).

Given a set of consecutive records (cellID, timestamp), the current level of mobility of a user can be estimated by calculating the Mobility Index over a pre-defined time period (TimeMin), from the current time instant back to TimeMin seconds ago (sliding window). The Mobility Index is the sum of the distance between each record and the previous ones, where the distance is the inverse of the time spent on each cell.

Figure 1 shows the Mobility Index calculated for two different sizes of the sliding window over the same set of records collected during three consecutive days.

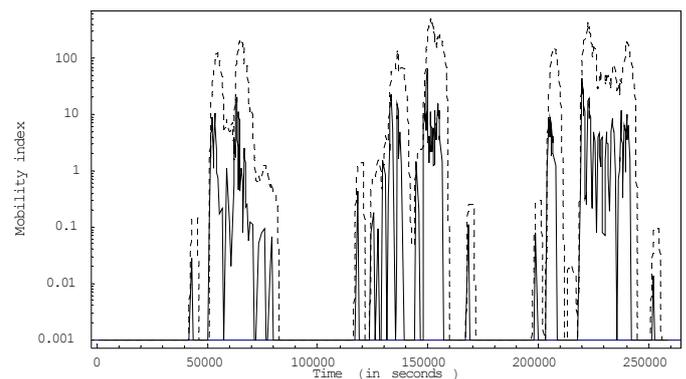


Figure 1. Mobility index.

Solid and dashed line in figure 1 represents the Mobility Index for a sliding window of 10 and 30 minutes respectively. Smaller values of the sliding window produce lower Mobility Index values (solid line in Figure 1) that react faster to the start/stop pattern of the mobile user (the Mobility Index decreases faster). For the purpose of detecting places, it is considered that a user is not moving when the Mobility Index is smaller than a pre-defined threshold defined according to the sliding window size. In Figure 1, the three time periods during the night are clearly visible as the periods where the value of the Mobility Index is very low.

B. Cell Fingerprint

While the Mobility Index is lower than a pre-defined threshold, we can assume that the user is not moving or is moving within a restricted area. When the user starts to move significantly, the Mobility Index rises above that threshold,

meaning that he is leaving the place where stayed for some time. That place can be characterized by a cell fingerprint defined as the list of cells used by the mobile terminal while in that place, and the corresponding percentage of time spend in each one of these cells. A fingerprint is then represented as follows:

```
{ { {cellID1, timePercentage1},
  {cellID2, timePercentage2}, ...,
  {cellIDk, timePercentagek}, totalTime,
  timeStamp }
```

A fingerprint is composed of three elements: a) the cellIDs and the percentage of time spent in each cell; b) the total time spent in this group of cells, and; c) the data and time of the last record used to create the fingerprint.

C. Finding Places

Every time the user visits a place, a new fingerprint representing that place is created based on the data collected during the time spent on that place. In order to create a personal map with the places known by the user, it is necessary to identify which of the created fingerprints represent the same place.

Mobile terminals change between the available cells in an unpredictable way which results, for different visits to the same place, into different time percentages spent on each available cell. Thus, for the same place, different fingerprints are created in each visit.

Although fingerprints for a same place may be different, they have some similarities, being composed of similar lists of cells. Therefore, by clustering the similar fingerprints it is possible to create a set of clusters with each one representing a place known by the user.

Since places (clusters of fingerprints) are to be discovered in real-time (while data is being collected in the mobile terminal), and because the data to be clustered is symbolic (fingerprints), a new clustering algorithm had to be developed. Clustering algorithms can be divided into partitioning, hierarchical or density-based [6]. A wide number of different clustering algorithms have been reported in the literature, but most of them require that all records to be clustered are available in memory, and that the total number of clusters to be created be known in advance. However, cells can be collected indefinitely, creating an endless number of records, and the total number of clusters to be created cannot be pre-defined because as the time goes by the number of places known by a user grows and can be very different from one person to another. To deal with an unlimited number of fingerprints (resulting from an endless number of records) and to deal with an unpredictable number of clusters, a new clustering algorithm was developed. It takes as input the fingerprints as they are created and returns a list of places.

For the clustering we have adopted an approach based on the distance between fingerprints. In order to measure the similarity between two fingerprints, we define a distance function (HDistance) based on an adaptation of the Hamming distance. For a cellID present in both fingerprints we calculate

the absolute difference between the percentages of time spent in that cellID in each fingerprint. If the user spent the same percentage of time in both fingerprints then there is no difference and the calculated value is zero. For two completely different fingerprints the distance is one while for two perfectly equal fingerprints it is zero.

Joining two similar fingerprints creates a new fingerprint in which the total time spent on the new fingerprint is the sum of the total time spent on both fingerprints. The percentage of time spent in each cell is calculated proportionally between the time spent in each fingerprint and the total time spent on both fingerprints. Therefore, a cluster has the same data structure as a fingerprint but results from the union of similar fingerprints instead of being created directly from the collected data.

The clustering algorithm creates the first place (cluster) from the first detected fingerprint. After that, for each new fingerprint, the similarity between that new fingerprint and every existent cluster is calculated. Because a cluster has the same structure as a fingerprint, the similarity is calculated based on the adapted Hamming distance. If a fingerprint is similar enough to any of the existing clusters, it is joined to the closest one. Otherwise, a new cluster is created. A parameter defines the minimum similarity level that must be achieved to join the new fingerprint to an existing cluster.

Another parameter of the algorithm defines the maximum number of clusters that can be created, thus imposing a limit on the number of discovered places. This parameter is useful to limit the amount of memory used by the algorithm on a mobile device. After the maximum number of clusters has been reached, a fingerprint can replace an old cluster or be discarded. The decision whether an existing cluster is replaced or not by a fingerprint is based on a Cluster Importance Ratio (CIR) that measures the relative importance of the fingerprint in relation to a cluster (equation 1). If this importance ratio is higher than one for more than one of the existing clusters, the cluster that produced the highest ratio is replaced by the new fingerprint. The CIR is defined as:

$$CIR(cl, fp) = \frac{KldgIdx(tt(fp))}{FgIdx(Age(cl, fp)) \times KldgIdx(tt(cl))} \quad (1)$$

where $Age(cl, fp)$ is the time difference between the fingerprint timestamp and the cluster timestamp (relative age), $tt(cl)$ is the total time spent in the cluster, and $FgIdx(x)$ and $KldgIdx(x)$ are a Forget Index and a Knowledge Index as defined next.

The importance ratio (equation 1) is calculated using the total time spent in the cluster and the time elapsed since the last fingerprint was added to the cluster (cluster age). Clusters recently updated or corresponding to places where the user spent more time are more valuable, while clusters that are unaltered for a long time represent places that are no longer visited by the user and have a more strong potential to be replaced by new ones as time goes by.

The Forget Index (FgIdx), represented by equation 2, models the percentage of forgetfulness as time goes by, and is defined as follows:

$$FgIdx(x) = \begin{cases} 0 & \text{if } (x < a) \\ \frac{1}{2} \times \left(\sin\left(\frac{\pi \times (x-a)}{b-a} - \frac{1}{2}\pi\right) + 1 \right) & \text{if } (a \leq x < b) \\ 1 & \text{if } (x \geq b) \end{cases} \quad (2)$$

The Forget Index was modelled with parameters a and b of 1 and 60 months respectively.

D. Familiarity Level

Some places are more familiar to a user than others. “Home” and “working place” are two places where people spend more time and thus the familiarity level with such places is enormous.

The familiarity level with a place varies accordingly to the total amount of time spent on that place and with the amount of time elapsed since the last visit to that place. Besides identifying places known by the user (represented by the clusters) we need to assess the user familiarity level to each of the discovered places. Based on the total amount of time spent in a place and on the time elapsed since the user visited that place for the last time, we define a Familiarity Index (FmIdx) function (equation 3) that measures the user familiarity level for a cluster.

$$FmIdx(cl) = KldIdx(cl) \times FgIdx(cl) \quad (3)$$

A cluster Familiarity Index models the process of learning about a place as more time is spent in it, and also the process of forgetting about a place as the time goes by. The learning process is described by a Knowledge Index (KldIdx) function that is computed based on the total amount of time spent on that place.

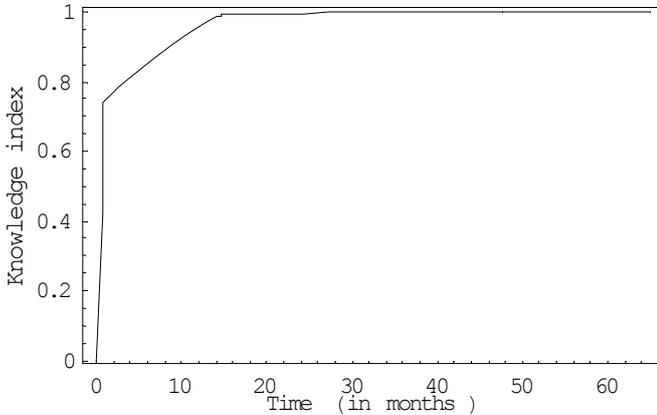


Figure 2. Knowledge index.

Some persons tend to know a place better and faster than others. Similarly, some people forget things faster than others. Thus, the creation of a universal mathematic model to express the knowledge of a place based on the time spent on that place or to express how good a persons’ memory about a known place is, is not an easy task. Moreover, the knowledge a person have about a place can be influenced by a wide number of factors like the mean of transportation used, the visit purpose (work or holidays), etc. Therefore, we modelled the knowledge

of a place as depicted in figure 2 and the forget index as defined in equation 2.

Although those indexes were created based on our perception of the reality, without any scientific support, they allow us to show that is possible to compute a familiarity index based on the information available for each cluster.

E. Model Usage

The proposed model provides a way to estimate the familiarity of a user about his current place using just the information provided by the mobile phone. At any moment, given the current cellID record, it is possible to identify if the user is in a known place and how familiar he is with the surrounding area.

The last record processed by the algorithm represents the last known position of the user. By searching the clusters which contain this last cellID positioning data, it is possible to identify in which cluster (place) the user is at. Because the same cellID can be present in more than one cluster, then it may identify more than one cluster as a possible place where the user is at. To solve this ambiguity, we associate a probability of being the correct place to each cluster, based on the percentage of time spent on that specific cell on each cluster. The cluster with the highest probability is assumed to represent the place where the user is at.

III. RESULTS

A Symbian [7] application was developed and was run on two users’ mobile phones for several months. It creates a log file with the user movements inside the GSM cellular networks, creating records with timestamp and {cellID, LAC, MNC, MCC} information.

CellID location is checked every eight seconds creating a record every time it changes or after every fifteen minutes if no change was detected since the last recorded position. The creation of a record at least every fifteen minutes permits to distinguish between when the user mobile phone was associated to the same cell for a long period from when the application was turned off.

Although data was collected to a log file and later processed according to the process described in the previous section we must emphasize that records were processed by the order they were created. This sequential processing of data shows that the developed algorithms work conveniently when applied to data collected in real time.

A. Mobility Index

The Mobility Index is a measure of how much a person is moving and, in this work, is used to detect the time periods while the mobile user is stopped or moving slowly within a restricted area. As illustrated in Figure 1, the value of the Mobility Index depends on the value of the window size (timeMin) - lower values of the window size lead to lower values of the Mobility Index and also make it more sensible to the user’s movements (reacting faster to the begin and end of the user’s movements).

In order to assess how well the Mobility Index models the moving pattern of a mobile user, we compared the information extracted from the Mobility Index (the detected time periods where the user is assumed to be stopped or moving within a restricted area) with similar information obtained from the user trajectory measured using a GPS receiver. For this, we simultaneously collected GSM cell data and GPS positions during a period of about a week. During this period, the user moved in both urban and rural areas, and visited (stayed during more than 10 minutes) 17 different places.

From the GPS data, the velocity of the user was calculated (using a moving average over a 5 minutes period to remove spurious movements and minimize GPS errors) and the moving periods were calculated as the time periods where the speed was higher than 3 km/h. The moving periods obtained from the Mobility Index were also calculated as the time periods where the Mobility Index values were higher than a specified threshold. We then calculated the correlation between the two moving periods as the percentage of time they were time coincident. This correlation value was then used to estimate the optimum values for the window size and the threshold referred in section II-B.

We used a sliding window of 10 minutes and a mobility index threshold value of 6. The obtained results show a good correlation between the moving periods calculated from the Mobility Index and from the GPS data. They are in agreement during more than 90% of the time.

B. Clustering Parameters

Our clustering algorithm is dependent on two basic parameters: the similarity threshold of the fingerprints and the maximum number of clusters.

If only the very similar fingerprints are joined then the system creates different clusters for the same place. Otherwise, joining very different fingerprints into the same clusters leads to clusters representing more than one place. Good results were achieved joining fingerprints that have up to 75% of similarity (0% means equal fingerprint).

During assessing of the proposed algorithm, one of the trial users spent most of the days at the University campus, which is located in a city with thousands of students. To support the communications for a big number of students and population around the campus, the mobile network operators have installed a considerable number of cells. Thus, the set of cells available on the different areas of the campus are not the same, but not different enough to allow the model to distinguish if the user is at laboratory, having lunch at canteen or at the bank (all inside the campus). However, for the same user in a rural area, it was possible to identify two different places distanced 300 meters.

We limited the number of clusters to 50 but this limit was not reached by any of the trial users. It is necessary to run the trial for an even longer period in order to reach this number of clusters and have some with a CIR small enough to be replaced by the newer ones. Limit the number of clusters to a very small number is not necessary because the algorithm is efficient and a cluster can be stored in a few bytes.

C. Places Results

In order to access the validity of the proposed model we compared the results of the clustering process with manually collected data. While collecting positioning data with the mobile phone we also have manually identified the visited places.

During a two months period one of the users made 194 visits across 39 different places. It includes places very often visited like his home (63 times) or the university where he studies (34 times) and, simultaneously, places visited few times like friend's houses, stores or a pharmacy. It includes yet a one week trip to a different country to participate on a conference.

Table 1 shows the results obtained after processing the collected data with the algorithms presented on the previous sections.

The results show that most of the time the application performed well. In 147 occasions (75.8% of 194 visits done to different places) it was able to correctly identify that the user was visiting a place and assigned the created fingerprints to a cluster that represents that place.

TABLE I. RESULTS ACHIEVED AFTER PROCESSING RECORDS COLLECTED DURING TWO MONTHS

Visits to places	194	
Correctly detected	147	(75,8%)
Clustering errors	5	(2,6%)
Partially correctly detected	6	(3,1%)
Misses	36	(18,5%)
False positives	4	

Clustering errors occur when the user was conveniently detected as visiting a place but the fingerprint(s) created during that time were incorrectly clustered and associated to a cluster that does not represent the place in fact visited by the user. In two months data, this error was detected 5 times and corresponds to two different places. For three occasions the user was classified as being at the University when he was visiting his friend Ana who lives in the city residential area, 500 meters away from the University. In the two other occasions the user was classified as being at home when he was in a place named "Fonte longa" which is located 300 meters away from his house. Nevertheless, in other thirteen occasions the user went to "Fonte longa" and four times to Ana' house and in all those occasion the system detected the user movements conveniently. Thus, the clustering errors happened only in a small percentage of visits made to places geographically very near.

A visit to a place can be detected by more than one fingerprint. If at least one of those fingerprints is incorrectly clustered then we classify that visit to a place as being only partially correctly detected. One of the six of these errors happened when the user went to Ana' house. Other error corresponds to the final minutes of a trip made to a place inside a natural park, where few GSM cells are present because it is an uninhabited area and the narrow and twisted roads lead to slow speed driving. In this case the system classified the user as being in a place when in fact he only arrived there some

minutes later. The last four errors correspond to partially misclassification of fingerprints created during the week in which the user participated on a conference. The conference took place in a hotel located five minutes walking distance from the hotel where the user was staying and in four occasions the user presence in one hotel was classified as being in the other.

Not detecting the user visited a place ends to be the most common error. However, 18 of the 36 of those errors correspond to places where the users stayed fifteen or less minutes. Short periods of time in a place may not be sufficient to detect the visit because it is not sufficient to decrease the mobility index below the defined threshold. Decreasing the threshold would probably allow to detect those visits but it would increase the number of situations in which the user is moving and is classified as being stopped in a place (false positives). In our experiment we had 4 false positive (four fingerprints where created during trips made by the user). However, one of the false positive corresponds to a traffic jam on a road where the user took 25 minutes to drive one kilometre.

Figure 3 presents the Familiarity Index measured for a set of clusters (A to D) during a two months time period. As expected, the Familiarity Index grows as the user spends more time in those places.

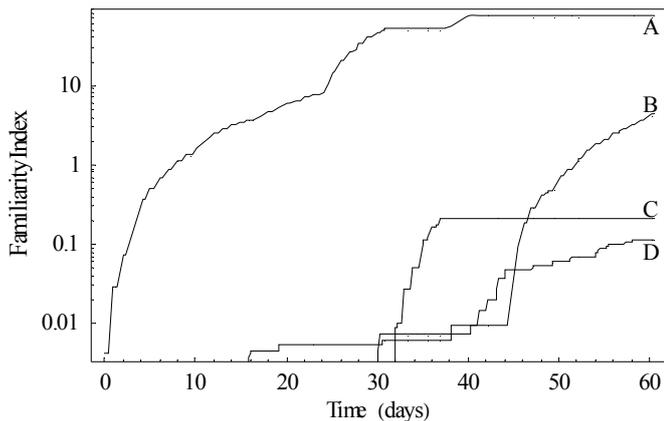


Figure 3. Familiarity index for different clusters.

Between day 32 and day 39, the user made a trip to a different country, and this episode is visible in Figure 3. During this time period, curves A, B and D, corresponding to the places "Home", "University1" and "University2", remained constant as these places were not visited. In contrast, curve C, corresponding to the place abroad, increased during this time period: as more days were spent in that place, the familiarity of the user with the surrounding area increased and the model was able to reflect this knowledge.

Figure 4 presents the familiarity of the user with his current location, computed as the position data was being collected and processed. We can observe the generic trend for the familiarity level to increase as the time goes by, except for the week spent by the user in a place where he has never been before (abroad). Figure 4 represents the data typically available for the

context-aware applications. In this example, a generic guide application, running on the mobile terminal, would be able to explore this data to be aware that, during a few days, the user was in an unfamiliar place and, therefore, needing assistance.

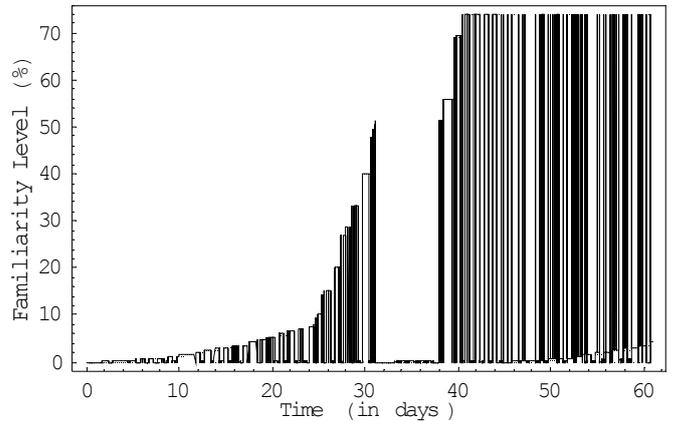


Figure 4. Familiarity index of the current place.

For the second user, the same process was executed and similar results were achieved.

D. Application Scenario

To have the location or position of a user is fundamental to create location-based services and applications. Computing the familiarity index is also useful to some applications that can adapt their behaviour accordingly to the user familiarity with the surrounding environment. For example, a guidance application can provide useful information when the user is travelling, providing information to guide the user to his destination without getting lost. However, if the user is moving in an area that is familiar to him, he does not need any assistance. In a familiar place a guidance application should not provide any help, specially if the application does not use any service to be updated with the recent changes in the traffic or to be updates with the traffic jams at that time of the day.

IV. RELATED WORK

There are a wide number of technologies that can be used to acquire the user location. However, in many cases it is important to recognize the place where the user is at ("home", "office", etc.) instead of his position (a set of coordinates in a geographic referential). A considerable effort is being done to automatically recognize the places visited by the users, which shows the interest of this kind of information for context-awareness.

BeaconPrint [8] is a place learning and recognition algorithm that uses IEEE 802.11 and GSM radio fingerprints. It continuously gathers statistics about the radio environment around the user mobile device and uses the collected data to learn, merge and recognize waypoints. Unlike BeaconPrint we do not gather statistics for WiFi and GSM radio. A mobile phone is linked to a cell but also has a list of up to 6 other cells existent on that area and the use of these six other cell available could be used to improve the accuracy of the achieved results.

However, just using the current cellID we were able to correctly identify 76% of the visited places, which is good when compared with the results of BeaconPrint that range from 96% to 56% computing a more complete data set composed by the WiFi cells and all the GSM cells detected by the mobile phone.

Kang [9] developed a clustering algorithm to extract significant places from a trace of coordinates acquired using the Placelab platform [10], considering a significant place an area where the user spends a considerable amount of time. A new algorithm was developed to cluster the locations along the time axis and extract clusters without any prior knowledge of the number of clusters. Placelab uses the WiFi network interfaces to estimate the user position using the radio beacons transmitted by the surrounding WiFi access points, and by transforming these beacons into geographic coordinates by querying a database which contains the positions of the WiFi access points.

Place Lab was created to bootstrap the dissemination of location-based services and applications. It can be run on different platforms and is able to acquire the user position through several mechanisms. Place Lab can use a GPS receiver (acquiring a pair of geographic coordinates); can acquire the BSSID of the neighbourhood WiFi access points; retrieving the cellID of the current GSM network cell and; scanning for Bluetooth devices in the neighbourhood. Although it uses several positioning technologies it provides only the user current location without any other contextual data. Adding our system as a new layer, on top of the existent Place Lab architecture, it is possible to provide a richer user context, adding a familiarity level for the current location of the user. Yet, the addition of this new layer does not compromise the user privacy because the computation is done locally, without the use of any network service.

Ashbrook [11] also developed a system to automatically cluster GPS data acquired over a period of time and obtain a set of meaningful locations. It defines a "place" as the locations where the user spends more than a certain amount of time.

BeaconPrint uses 802.11 and GSM radio response-rate fingerprints to learn and recognize places. However, the algorithms described in [9] and [11] can only be applied to GPS points which express the user position in an absolute referential.

In [12], Laasonen presents an adaptive framework for recognizing personally important locations in cellular networks, using cell-based location data. Although capable of identify all the visited places it is not a pervasive system (demanding the user to give a name to each place) neither was able to distinguish between the more important places and those the user visited occasionally.

Trevisani experimental study [13] argues that cellID location technique is not appropriate to deploy very simple location-based services because it does not provide acute positioning data. Nevertheless it shows that a Voice Location-based service can take advantage from the knowledge of the approximate user position.

In this paper we show that cellID of GSM networks can in fact be enough to fulfil the needs of many context-aware applications, providing a basic level of knowledge about the user environment, computing a index that expresses the familiarity of the user with the visited place.

V. CONCLUSIONS

In many areas, many applications can benefit from the availability of the user' context to find and use location-based services. GSM cellular networks cover a significant part of many countries but unfortunately the user' location is not made freely available by the operators. Only CellID can be acquired on the user devices, without querying any network operator service, but it lacks the precision necessary by most of the services and applications.

In this paper we have proposed a model to infer a personal map of known symbolic places, based on cellID position. It is a model that can run on platforms with limited resources, like a Symbian phone, without compromise the user privacy. The obtained results show that places known by the user can be identified, with a precision superior to the basic cellID positioning, without any prior knowledge of the network topology and without any user intervention. Moreover, a Familiarity Index associated with each discovered place was estimated, allowing context-aware applications to query the system in order to identify, in real time, where the user is and how familiar he is with the surrounding area.

REFERENCES

- [1] A. Harter, and A. Hopper, "A distributed location system for the active office", IEEE Network, vol. 8, pp. 62-70, 1994.
- [2] A. Harter, A. Hopper, P. Steggle, A. Ward, and P. Webster, "The Anatomy of a Context-Aware Application", Wireless Networks, Vol. 8, February 2002, pp. 187-197.
- [3] IEEE computer Society LAN MAN Standards Committee, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications", IEEE Std 802.11.1 1999. The Institute of Electrical and Electronics Engineers, 1999.
- [4] "Location Technologies for GSM, GPRS and WCDMA Networks", White Paper, SnapTrack, A QUALCOMM Company, November 4, 2001.
- [5] ITU - International Telecommunication Union, "Recommendation E.212 (05/04) - The international identification plan for mobile terminals and mobile users", <http://www.itu.int/>, 2005
- [6] J. Han, and M. Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, 2001.
- [7] Symbian OS Web site, <http://www.symbian.com>. 2005.
- [8] J. Hightower, S. Consolvo, A. LaMarca, I. Smith, and J. Hughes, "Learning and Recognizing the Places We Go", Proceedings of 7th International Conference Ubiquitous Computing (UbiComp 2005), M. Beigl, S. Intille, J. Rekimoto, and H. Tokuda (Eds.). Lecture Notes in Computer Science, Vol 3660. Springer (2005), pp. 159-176.
- [9] J. Kang, W. Welbourne, B. Stewart, and G. Borriello, "Extracting places from traces of locations". Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots (wmas2004). ACM Press, USA (2004), pp. 110-118.
- [10] A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, and et. al, "Place Lab: Device Positioning Using Radio Beacons in the Wild", Proceedings of the 3rd International Conference on Pervasive Computing (Pervasive 2005), Munich, Germany, May 2005. Lecture Notes in Computer Science 3468 Springer, pp 116-133.

- [11] D. Ashbrook, and T. Starner, "Using GPS to learn significant locations and predict movement across multiple users". *Personal and Ubiquitous Computing*, Vol. 7, Issue 5 (2003), pp. 275-286.
- [12] K. Laasonen, M. Raento, and H. Toivonen, "Adaptive On-Device Location Recognition", *Proceedings of the Second International Conference on Pervasive Computing (Pervasive 2004)*, Vienna, Austria, April 21-23, 2004. *Lecture Notes in Computer Science 3001*, Springer-Verlag, pp 287-304.
- [13] E. Trevisani, and A. Vitaletti., "Cell-ID location technique, limits and benefits: an experimental study", *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications, 2004 (WMCSA 2004)*, Dec. 2004, pp 51- 60.