Discovering Personal Places from Location Traces

Muhammad Umair*, Wan Seok Kim**, Byoung Chul Choi**, Sung Young Jung**

*Dept. of Engineering, University of Science and Technology (ETRI Campus), Daejeon, South Korea

**Electronics and Telecommunication Research Institute (ETRI), Economic Strategy Research Dept., Daejeon, South Korea

umair@etri.re.kr, wskim@etri.re.kr, cbc@etri.re.kr, jsykt@etri.re.kr

Abstract— Sensing and computational capabilities in smartphones are enabling attractive new applications in the area of location aware systems. Location aware devices can accurately compute their physical location in the form of latitude and longitude. However places contain much more valuable information to user rather than coordinates. A place is important to a user's personal daily life and carries socially important meanings to a user such as a place where one studies, works, eats, lives etc. GPS enabled devices like smartphones and tablets use location as a context in applications to share their current location. A GPS log of a moving object contains time stamped latitude and longitude information. The discovery of a user's places is a key challenge and involves mapping of GPS data log to meaningful personally important places. In this paper, we propose an algorithm that extracts a user's personally important places from location traces, and evaluate the algorithm with real user's data.

Keywords── Place discovery, Place recognition, Location aware computing, Significant places, Personal places

I. INTRODUCTION

Ubiquitous computing was first articulated by Mark Weiser [1], forecasting an era where computing would be available everywhere embedded in light switches, cars walls, chairs, clothing etc. Weiser anticipated that ubiquitous computers will be available in different sizes, each suited to a particular task that would take place at all scales, while making them effectively invisible to the user.

With the growing number of connected devices, context has gained significant importance in the field of ubiquitous computing and different frameworks for development and execution of context aware applications for ubiquitous environments have been developed [20]. Ubiquitous computing devices are now embedded with many sensors that are capable of detecting variety of contextual information. This imposes challenges on applications and services to be aware and adapt to context such as location, time, activity, neighboring devices and the state of physical environment.

Smartphones and tablets are ubiquitous mobile devices. They are central computing and communication devices and integral part of people's lives nowadays. In addition to Smartphones as communication device, Sensors in smartphones have become much common in the recent years making smartphones a sensing device. Market research has shown a significant growth of sensors in the mobile market in the past years and foresees the sensor demand growing within the next decade. Sensor enabled smartphones are contributing towards an emerging area of interest for researchers called mobile phone sensing. Incorporating sensors into mobile phones empowers them to sense, process and distribute data through various communication technologies [2], [3].

Today's Smartphones are equipped with variety of embedded sensors, such as accelerometer, GPS, proximity sensor, gyroscope, digital compass, camera, microphone, gesture , ambient light sensor, barometer, temperature sensor, humidity sensor etc. Sensing and computational capabilities in cell phones are enabling attractive new applications across a wide variety of domains such as healthcare [11], social networking [12], environment monitoring [13], traffic monitoring [14], [15], commerce [17] and human behavior monitoring [18] etc. Changing location in mobile phones allows sensors to generate enormous amount of data, hence these devices opens up exciting new areas for data mining research and its applications.

In the past few decades there has been a massive emergence of location aware systems. Nowadays such systems are embedded in Laptops, smartphones, tablets to wearable gadgets. Location aware devices can accurately compute their physical location in variety of forms and provide user a rich set of location aware applications. GPS enabled devices like smartphones and tablets use location as a context in applications for position tracking services [19] and recommendation systems [21]. Many people have started recording their physical movements with GPS enabled devices for life experience sharing [4], [16], sports activity analysis [5] and geo-tagging multimedia content [6].

Location in such applications is expressed in the form of coordinates (latitude and longitude). However places are more valuable to a user rather than coordinates. A place carries socially important meanings to an individual such as "home", "office", "school", "lab", "restaurant" etc. The key challenge faced by these applications is to figure out what constitute a user's personally important place and how to map GPS data to a personal place.

In this paper we propose an algorithm for automatically discovering a user's personally important places from his/her GPS logs. Our devices can make more intelligent and efficient decisions when they have access to high level information. In future as part of our main project, we intend to implement a recommendation system using such high level information. The implementation of the algorithm proposed in this paper is done by developing an application in android platform.
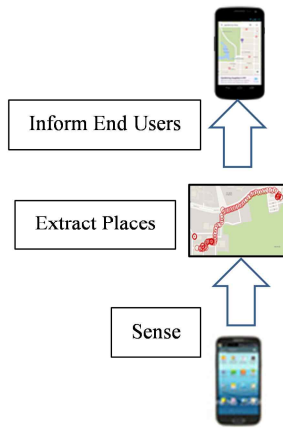
Figure 1. Architecture of our application

The remainder of the paper is organized as follows. Section II discusses the application design architecture. We present our place discovery approach in Section III. Experiment carried out is described in Section IV. Section V presents the related work and finally we conclude and present future work in Section VI.

II. APPLICATION DESIGN ARCHITECTURE

In this section we describe the design characteristics of our application. Figure 1 shows the architecture of our place discovery application. The current implementation of the application extracts user's personally important places in three steps. Initially, it collects a user's location data using GPS sensor available in smartphone. Next, it uses our proposed place discovery algorithm to extract his personally important places. Finally, it informs user about his places by showing them on map. The operational phases of the application are described below.

A. Sense

The application automatically infers personal locations using GPS sensor embedded in smartphone. We collects raw GPS log data is at given intervals depending upon user's activity. Our application not only gathers location coordinates of outdoor locations but also collects indoor locations. In collecting GPS data we must obtain sufficient GPS points to represent common user's places.

B. Extract Places

Simply collecting GPS logs is insufficient. Raw GPS data points are not important to a user, so our aim is to obtain physical locations for a person's places that are important to his daily life routine. The input to our place discovery algorithm is the raw GPS data points and output is location geometrics of personally important places.

C. Inform End Users

Equipped with the ability to extract important personal places in people's life, our application informs end users by showing them their personally important places on maps.

III. OUR APPROACH

Figure 2 shows the flow of our system, which is comprised of the following steps; getting GPS data, neighbourhood establishment, DLN formation, stable neighbourhood selection, significant place discovery and informs end users.
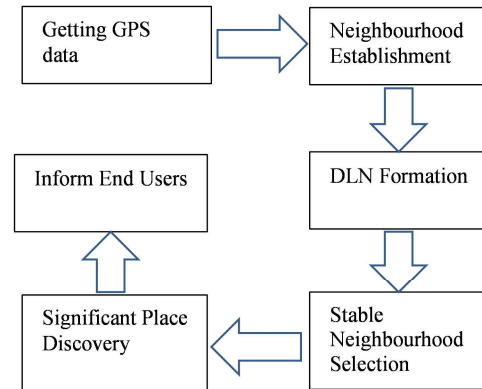

Figure 2. Flow of place discovery system

We present an algorithm for extracting significant places from trace of coordinates. First, we will clarify some terms; including GPS Point (pa), Distance threshold (d_{thresh}) and Minimum time threshold (t_{min}).

Definition 1. GPS Point: A GPS Point (pa) represents a user's location at any time and it contains latitude (Lat), longitude (Long) and timestamp (T) information.

Definition 2. Distance Threshold (d_{thresh}): It is the minimum threshold distance in which we find the neighbours of a given point pa.

Definition 3. Minimum Time Threshold (t_{min}): t_{min} is the minimum time required for a user to stay in a particular place.

We start to define a GPS point pa's logical neighbourhood: the set of logical neighbours Na is the set of points from which the point pa is at distance of less than or equal to d_{thresh}. Each point maintains its own table of logical neighbours. Each line of the table represents one logical neighbour and contains information about pa neighbour's (py ∈ Na), their latitude (py,lat), their longitude (py,long) and the time instant when the point was located (t_i(py)). The number of the logical neighbours can be quantified by its density value: the density ρ(Na) measures the number of the logical neighbours of point pa. The neighbourhood of a point is identified through the observation of its density. A logical neighbourhood is said to be Dense Logical Neighbourhood (DLN) if its ρ(Na) is longer than a given k_{min} value:

$$\rho(Na) \geq k_{min}$$

The neighbours with whom a point maintains a dense neighbourhood are more suitable to be a geographic area. A combination of neighbours of pa, (py ∈ Na) either forms a newly created dense neighbourhood or if any point within neighbourhood of pa contains a point which is also a part of any other neighbourhood then the current neighbourhood is merged with existing neighbourhood.

710

A. Stable neighbourhood selection

The more amount of time a person spends at a place, the more stable that place would be and more important that place would likely to be for the person. Given the time stamped GPS data points we can easily determine the time spent by a user at a particular place. At time instant $t_i(Na)$, when a user firstly enters a neighbourhood the duration of the visit can be specified by its stability value: the stability $\eta(Na)$ measures the duration of time a user spends in a neighbourhood. $\eta(Na)$ is computed by the point pa at instant t by applying the following expression:

$$\eta(Na) = (t - ti(Na))$$

The overall neighbourhood duration time for m visits is sum of all individual visit times

$$\eta(Na) = \eta 1(Na) + \eta 2(Na) \dots \eta m(Na)$$

A logical neighbourhood is said to be stable if it lasts longer than a given t_{min} value:

$$\eta(Na) \geq tm\dot{n}.$$

We present our place discovery below:

Algorithm:

    Input: pa, Na, ρ(py) (∀py ∈ Na), lat(py) (∀py ∈ Na)
    (∀py ∈ Na), ti(py) (∀py ∈ Na)
    Output: Important places list

1.   for a given point pa
2.   **if** (ρ(Na) >= k$_{min}$) **then**
3.   insert(Na,list_DLN)
4.   **for** each neighbour py ∈ Na **do**
5.   insert neighbour_list(lat(py), long(py), t$_i$(py))
6.   **if** (py ∈ existing list_DLN) **then**
7.   remove(Na, list_DLN)
8.   merge both DLNs
9.   insert(merged DLN,list_DLN)
10.   **end**
11.   **end**
12.   **end**
13.   **if (**η(Na) ≥ tmin) **then**
14.   insert(Na,stable_neighbourhood)
15.   **end**
16.   **if** (Na ∈ list_DLN) && (Na ∈ stable_neighbourhood)) **then**
17.   insert(Na,important_places_list)
18.   **end**
19.   prioritize(user_stay_time,important places list)
20.   **return** important_places_list

B. User's Place Determination

A user's place is a geographical region which is physically important to a user. It can be office, home, restaurant, shopping mall etc. We use the DLN in important place discovery algorithm and propose an algorithm to compute the personally important places. The main idea behind this algorithm is to select as personally important place the DLN, which is also stable indicating that it has more user stay time. The purpose of this rationale is to select those DLNs with the highest probability of being an important place. After the discovery of important places, we prioritize a user's important places based on user's stay time at a place. From a user's list of important places, the place which has higher user stay time is ranked at the top of important places list.

IV. EXPERIMENT

In this section we describe the experiment performed in detail. At first we discuss the data collected during the experiment, experiment procedure and finally we present results produced.

A. Trace collection

In order to collect location traces, we used our smartphone application running on android platform. Traces were collected indoors and outdoors with smartphone running our application during the daily activities of the first author as shown in figure 3. The raw data points shown in figure 3 represents the author's location points scattered along routes and walkways between places. During trace collection smartphone was placed in author's front pocket. We set the application to take a GPS reading every 10 seconds when the subject is walking and every 30 seconds when the subject is stationary. The location traces were collected within ETRI (Electronics and Telecommunication Research Institute) campus and outside campus with total duration of about 3 hours. We collected data at three basic locations labelled in figure 3. The log trace was started when the author was in his office as shown in figure. In the office he stayed at his desk and also attended a meeting in conference room. After about 40 minutes in his office, at dinner time he headed to a restaurant few blocks away within campus. After having dinner for about 30 minutes, he then left for his dormitory. On his way off campus he walked for about 10 minutes. After reaching his dormitory, he cooked food in kitchen and also visited a friend's room. Later on he stayed in his room for the rest of night. Table 1 describes the description of three basic locations along with the time during of author stay.

Figure 3. Location traces collected

TABLE 1. DESCRIPTION OF PLACES VISITED

Place	Duration	Description
Office	40 min	Indoor: 2nd floor in a building.
Restaurant	30 min	Outdoor: 1st floor in open air restaurant.
Dormitory Room	100 min	Indoor: 3rd floor of a building

We evaluate our algorithm using author's trace data and present results below.

B. Experimental results

In this section, we report the results generated by our algorithm. We evaluate our algorithm on raw location data shown in figure 3. Figure 4 and 5 shows a closer view of collected GPS points in office, restaurant and dormitory respectively. Finally figure 6 shows significant places extracted by our algorithm. Algorithm prioritizes discovered locations and labels them as place 1, place 2 and place 3 depending upon the amount of a user spends at a particular location. Three significant places are discovered by the algorithm. Scattered points between the places have been eliminated from the final result and only those GPS points are shown which marks a significant place. The results shows that places extracted by our algorithm matches the three basic locations of author as described in Table 1. During the computation of logical neighbourhood, we set d_{thresh} to be 20 meters approximately equal to the accuracy of smartphone GPS.

Figure 4. GPS points collected in office and restaurant
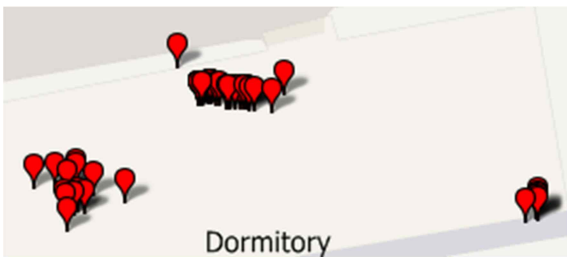

Figure 5. GPS points collected in dormitory

Figure 6. Important places discovered by algorithm when $k_{min} = 15$ and $t_{min} = 30 \, min$

V. RELATED WORK

Marmasse and Schmandt [10] used loss of GPS signal to infer important indoor locations. This approach identifies a place within a certain radius where GPS signal disappears and then appears. This kind of approach can identify small indoor places but it is not suitable to identify large places and suffers from false positives.

Authors in [7] presented a time based clustering algorithm to extract places using access point's MAC address of a WIFI network to capture location data. The algorithm clusters the locations along time axis. In the algorithm as a new place is discovered then the distance between new and old locations is measured. If the new location is moving away from the old location and it spans over a time threshold then it is considered as a different cluster. However the algorithm does not differentiate between same places visited several times. Each time the algorithm discovers a place it is considered to be a different place. This approach makes it hard to discover places that are visited frequently but for short time. Finally this approach also requires continuous data sensing with fine intervals, which requires large data storage.

K-means is a well-known efficient clustering approach used by authors in [9] to extract important places from location history data. However k-means need to know exact number of clusters before the beginning of clustering. This kind of approach could not be suitable for users since they do not know their visiting places in advance.

DJ-Cluster, a density based clustering approach has been presented in [8]. This approach makes a cluster if there are at least MinPts points with a distance of Eps. These clusters are then critical in important place discovery procedure. However this approach does not considers the amount of time a user spends at a particular location, which makes it difficult to extract locations where a user spends more amount of time.

Our approach takes into account the existing clustering techniques and presents an approach which not only takes into account density of GPS points but also time spent by a user at a particular place. Our application collects GPS traces indoors and outdoors intelligently depending upon the user's activity and presents a novel approach to extract a user's personally important places based on real user's GPS tracks.

VI. CONCLUSION AND FUTURE WORK

In the paper we presented a significant place discovery algorithm from GPS traces. A place where user collected enough GPS readings and spends considerable amount of time is considered to be a significant place. Our algorithm not only considers density of data points but also considers time spent by a user in a particular place without having prior knowledge of number of places. We evaluated our algorithm with real trace data collected by our GPS data collection application and ignored outlier points in significant place discovery process.

In future we intend to evaluate our algorithm with more number of data points scattered across large area and compare our algorithm with existing techniques. We also intend to test our algorithm with varying minimum points (k_{min}), distance threshold (d_{thresh}) and time threshold (t_{min}) parameters and analyse their impact on number of significant places discovered.

REFERENCES

[1] M. Weiser, "The Computer for the 21st Century," *Scientific American*, vol. 265, no. 3, 1991, pp. 66-75.

[2] X. Bao and R. Roy Choudhury, "Movi: mobile phone based video highlights via collaborative sensing," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, ser. MobiSys '10. New York, NY, USA: ACM, 2010, pp. 357–370. [Online]. Available: http://doi.acm.org/10.1145/1814433.1814468

[3] M.Bilal, "An Intelligent Future Mobile Terminal Architecture," in *Proceedings of Information and Emerging Technologies*, Pakistan, 2010.

[4] Y. Zheng, L. Wang, R. Zhang, X. Xie and W. Ma, "GeoLife: Managing and understanding your past life over maps," in *Proceedings of Mobile data Management* ser. MDM'09, Beijing, China, April 2008.

[5] (2002) SPORTSDO. [Online]. Available: http://sportsdo.net/Activity/ActivityBlog.aspx/

[6] Scott Counts, M. Smith, "Where were we: Communities for sharing space-time trails," in *ACM GIS,* 2007.

[7] Jong Hee Kang, William Welbourne, Benjamin Stewart, and Gaetano Borriello, "Extracting places from traces of locations," *in Proc. WMASH*, pages 110–118, 2004.

[8] Changqing Zhou, Dan Frankowski, Pamela Ludford, Shashi Shekhar, and Loren Terveen, "Discovering personal gazetteers: An interactive clustering approach," in *Proc. ACMGIS*, pages 266–273, 2004.

[9] Daniel Ashbrook and Thad Starner, "Using GPS to learn significant locations and predict movement across multiple users," in *Personal Ubiquitous Computing*, 7(5):275–286, 2003.

[10] Matalia Marmasse, Chris Schmandt, "Location-aware information delivery with comMotion," in *Proceedings of the 2nd International Symposium on Handheld and Ubiquitous Computing* (HUC 200), Bristol, UK, September 2000.

[11] S. Consolvo et al., "Activity Sensing in the Wild: A Field Trial of Ubifit Garden," in *Proceedings of 26th Annual ACM SIGCHI Conf. Human Factors Comp. Sys.*, 2008, pp. 1797–1806.

[12] E. Miluzzo et al., "Sensing meets Mobile Social Networks: The Design, Implementation, and Evaluation of the CenceMe Application," in *Proc. 6th ACM SenSys*, 2008, pp. 337–50.

[13] M. Mun et al., "Peir, the Personal Environmental Impact Report, as a Platform for Participatory Sensing Systems Research," in *Proc. 7th ACM MobiSys*, 2009, pp. 55–68.

[14] A. Thiagarajan et al., "VTrack: Accurate, Energy-Aware Traffic Delay Estimation Using Mobile Phones," in *Proc. 7th ACM SenSys*, Berkeley, CA, Nov. 2009.

[15] UC Berkeley/Nokia/NAVTEQ, "Mobile Millennium"; [Online]. Available: http://traffic.berkeley.edu/

[16] Yu Zheng, Yukun Chen, Xing Xie, Wei-Ying Ma. "GeoLife2.0: A Location-Based Social Networking Service," in *Proceedings of international conference on Mobile Data Management*, ser. MDM'09, 2009.

[17] S. Sehgal, S. S. Kanhere, and C. T. Chou, "Mobishop: Using mobile phones for sharing consumer pricing information," in *In Demo Session of the Intl. Conference on Distributed Computing in Sensor Systems*, Santorini, Greece, June 2008.

[18] E. Miluzzo, C. T. Cornelius, A. Ramaswamy, T. Choudhury, Z. Liu, and A. T. Campbell, "Darwin phones: the evolution of sensing and inference on mobile phones," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, ser. MobiSys'10. New York, NY, USA: ACM, 2010, pp. 5–20. [Online]. Available: http://doi.acm.org/10.1145/1814433.1814437.

[19] K. Min, K. Nam, and J. Kim, "Multilevel Location Trigger in Distributed Mobile Environments for Location-Based Services," in *ETRI Journal*, vol.29, no.1, 2007, pp.107-109.

[20] A. Moon, H. Kim, H Kim and S. Lee, "Context-Aware Active Services in Ubiquitous Computing Environments," in *ETRI Journal*, vol.29, no.2, 2007, pp.169-178.

[21] M. Park, J. Hong, S. Cho, "Location-Based Recommendation System Using Bayesian User's Preference Model in Mobile Devices," in *Proceedings of 4th international conference on Ubiquitous Intelligence and Computing*, ser. UIC'07. Hong Kong, China, July 2007.

Muhammad Umair received the BS degree in Information and Communication System Engineering from National University of Sciences and Technology, Pakistan, in 2012. Currently he is a Masters student in University of Science and Technology, South Korea, in Electronics and Telecommunication Research Institute (ETRI) campus. His research interests include context aware computing, artificial intelligence, networking and performance evaluation.
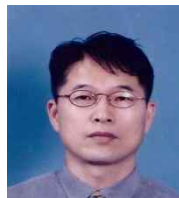
Wan Seok Kim is a principal researcher in the Electronics and Telecommunication Research Institute (ETRI), South Korea, and an associate professor in the University of Science and Technology, South Korea. His current research interests include real time kernel (OS), database management system, technology forecasting in ICT field, and Intuitive service of computing objects.

Byoung Chul Choi received the BS and MA in Economics from University of Seoul and PhD in Economics from Chungnam National University, South Korea, in 1995, 1999 and 2013, respectively. In 1999, he joined Electronics and Telecommunications Research Institute (ETRI). Currently, he is a director in the Technology Policy Research Section, ETRI. His research interests include innovation policy, technology policy and strategy in ICT field.

Sung Young Jung graduated from Chungbuk National University, South Korea in 1984 with a BS in Economics, received Ph.D. in Institutional Economics from Chungbuk National University in February, 1998. He has worked for ETRI for 25 years in the area of IT policy, institutional infrastructure, business strategy of telecommunications companies, and national innovation system.