Autonomous Place Naming System using Opportunistic Crowdsensing and Knowledge from Crowdsourcing

Yohan Chon
Department of Computer Science
Yonsei University
Seoul, Korea
yohan@cs.yonsei.ac.kr

Yunjong Kim
Department of Computer Science
Yonsei University
Seoul, Korea
yjkim@cs.yonsei.ac.kr

Hojung Cha
Department of Computer Science
Yonsei University
Seoul, Korea
hjcha@cs.yonsei.ac.kr

ABSTRACT

A user's location information is commonly used in diverse mobile services, yet providing the actual name or semantic meaning of a place is challenging. Previous works required manual user interventions for place naming, such as searching by additional keywords and/or selecting place in a list. We believe that applying mobile sensing techniques to this problem can greatly reduce user intervention. In this paper, we present an autonomous place naming system using opportunistic crowdsensing and knowledge from crowdsourcing. Our goal is to provide a place name from a person's perspective: that is, *functional name* (e.g., food place, shopping place), *business name* (e.g., Starbucks, Apple Store), or *personal name* (e.g., my home, my workplace). The main idea is to bridge the gap between crowdsensing data from smartphone users and location information in social network services. The proposed system automatically extracts a wide range of semantic features about the places from both crowdsensing data and social networks to model a place name. We then infer the place name by linking the crowdsensing data with knowledge in social networks. Extensive evaluations with real deployments show that the proposed system outperforms the related approaches and greatly reduces user intervention for place naming.

Categories and Subject Descriptors

C.m [**Computer Systems Organization**]: Miscellaneous-*Mobile Sensing Systems*

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Location-based Services, Location Naming, Smartphone Sensing

IPSN'13, April 8–11, 2013, Philadelphia, Pennsylvania, USA.

1. INTRODUCTION

People typically carry mobile phones all the time, and a user's location can practically be tracked in everyday lives. The acquired location information is widely used in emerging mobile applications, such as location-based social networks, place recommendations, and location-based searches. However, most services have a limitation in recognizing the actual name of a place (e.g., Starbucks or Apple store) because of the noises in physical location information. For example, mobile services provide a list of nearby places based on the estimated latitude and longitude, and a user manually decides the present place. In addition, reverse geocoding, which obtains a readable address from a physical location, often generates inaccurate output due to localization error. Meanwhile, people normally refer the places by name (e.g., McDonalds), not address (e.g., Yonsei Street 100). This indicates that a gap exists between people's understanding about the place and the raw coordinates of a location. Therefore, an advanced system is necessary to provide a place name from a user's point of view, not just raw coordinates on a map.

Active studies have recently been conducted to understand and provide place naming from a user's viewpoint. Lin et al. [19] analyzed the preference of place naming: how people to refer places when interacting with others. They defined a taxonomy on place naming, such as *functional name* (e.g., shopping, food place), *business name* (e.g., Walmart, Burger King), or *personal name* (e.g., my home, my workplace). Most works in information retrieval have studied similarity functions to provide a ranked list of places. To estimate the rank, previous works have employed various features, including location [11], check-in histories at social networks [18, 27], web-based popularity [18, 24], and search histories [16]. However, previous studies still required manual user intervention, such as searching a place with additional keywords or explicitly choosing a place in a list. We believe that leveraging mobile sensing systems to solve this problem greatly improves the accuracy of place naming and reduces user intervention.

In this paper, we present an autonomous place naming system to provide a place name according to three different types, as defined in [19]: *functional, business*, and *personal name*. The proposed system uses previously untouched resources that can be collected by individual smartphones, such as user mobility, phone usage, and images captured by users. The main idea is to integrate crowdsensing data obtained from smartphone users with information in social network services (SNS) to provide the place name. Smartphones can opportunistically collect sensing data about the places in users' daily life, but manual labeling of the place name is necessary. Meanwhile, social networks have a broad set of location information manually built by business providers and users. The information typically includes physical location, place name, and place type, but location-based lookup [1] is often limited in recognizing a place, due to incorrect location estimation. Our system mines sensing data to extract semantic features of the places, which include publicness, behavior patterns of visitors, and captured pictures in places. In order to provide a place name, our system then matches extracted features with location information in SNS, which include check-in histories, posted images, and texts.

The main contributions of our paper are as follows:

- We present an autonomous system for place naming that combines opportunistic sensing data from smartphones with location information in social networks. We greatly improve the accuracy of place naming.

- We thoroughly validated the proposed system using large-scale data collected by 70 smartphone users and 31,000 social networks users in Seoul, Korea. The results show that the proposed system outperforms the related approaches and reduces manual user intervention for place naming.

- The proposed system enables us to understand human behavior patterns in city-scale by linking a place to semantic meaning. The results of automatic place naming can lead to advanced understanding of user behavior in places.

2. PRELIMINARY STUDY

We motivate our work by discussing the limitation of location-based lookup and the potential of applying a sensing system to this problem. As a preliminary study, we investigated the coverage and quality of location information in social networks. We explored the following questions:

- How many places does an SNS location information cover in real life?

- How much user intervention does location-based lookup require for place naming?

To answer these questions, we analyzed approximately 3,800 places visited by 70 smartphone users, and 130,000 locations crawled from Foursquare [11]. We deployed mobility monitoring tool [3] from November 2011 to September 2012, to collect real user traces in Seoul, Korea. We also crawled social network data from May 2012 to September 2012. The details regarding data collection are further described in Section 4.1.

2.1 Coverage of Location-based Lookup

We first investigate the coverage of location information in social networks. Intuitively, the location information in SNS would contain a subset of the places in real life, as social media captures a relatively small fraction of our lives by people who are active participants of SNS. To estimate the coverage, we manually matched the places visited by users in daily life to the locations in SNS. We used the labeled names and offline feedback from the participants. Figure 1(a) shows that SNS misses about 22% of the places users visited. The missing places are mostly food or shopping places with low popularity, or recently built places, as shown in Figure 1(b). Considering that our dataset is collected with 70 smartphone users, the portion of missing places would increase with more participants since the manual registration of places in SNS cannot contain all the places in real life.

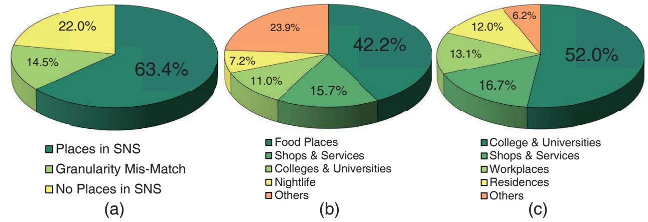

Figure 1: (a) Coverage of location information in SNS; (b) the categories of missing places and (c) granularity mismatch.

Another finding is that mismatch exists in user's understanding of the actual place and the locations provided by SNS. For example, a user stays at the NexOne Inc. office, but SNS provides this place as the name of the building. A user would think that the building name could not express the place she visited, since the building has many offices, stores, and restaurants. Figure 1(a) shows that the mismatch ratio is about 14.5% of the places in our dataset. The mismatched places were mostly located in college and university regions, shopping districts, or office buildings, as shown in Figure 1(c).

In summary, our preliminary analysis indicates that crowdsensing using smartphones has the potential to expand the coverage and improve the granularity of location information that are missed by social media.

2.2 Limitation of Location-based Lookup

We now discuss the limitation of place naming method that uses location-based lookup in SNS. Most services provide a list of nearby places since direct mapping of a user's location to a place name is challenging. We calculated the distance between the estimated location by mobile phones and the location coordinates of places in SNS. Figure 2(a) shows that only 8.6% of places are within 10 meters, and 80.4% of places are within about 100 meters. The list of nearby places should show the places within 180 meters in order to cover the actual place for 90% of the cases.

The questions to be answered, then, are *how many places does the list contain within error distance?* and *what is the rank of the actual place in the list?* If the list contains many places and the rank of the actual place is high, manual selection of a place is burdensome for some users. We statistically analyzed the location information regarding the Seoul region to answer these questions. The results show that the list contained 87.8 places within 100 meters error bound (median was 44, first percentile was 20, and third percentile was 121 places); the rank of 35% of places was higher

[1] Location-based lookup is the process of finding nearby places in a location database based on a estimated location (i.e., latitude and longitude).

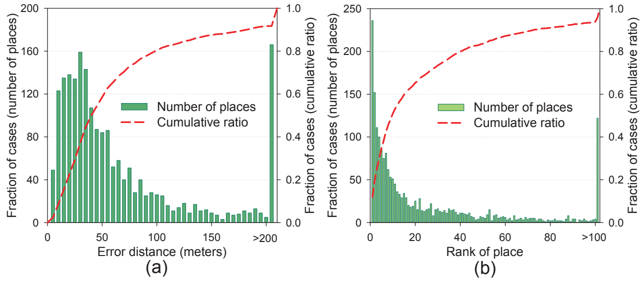

Figure 2: (a) Error distance between estimated locations in mobile phone and places in SNS and (b) the rank of actual place in the list of nearby places.

than 20, as shown in Figure 2(b). This means that manual selection of a place in the list is a serious burden for the users, due to the high density of places. A user had to enter further keywords to find his/her current place by location-based lookup. In addition, 17.9% of places were registered as more than two places, with slightly different names, in the location database. The reason for this discrepancy is that some users failed to find a place in the list and registered a new place instead, even though the place already existed in the database. Our findings indicate that to provide a place name, we need robust and efficient features beyond simple raw coordinates of location.

3. AUTONOMOUS PLACE NAMING

We propose autonomous place naming system that integrates crowdsensing data from mobile phones with location information in social networks. We first describe the usage scenario, and then present the technical details.

The proposed system automatically provides the name of places visited by users in daily life according to three types [19]: functional name (e.g., food place, nightlife, shopping place), business name (e.g., Starbucks, Apple Store), and personal name (e.g., my home, my workplace). We assume that a user does not manually provide information about the place name. When a user stays at a place for a certain period of time, the system opportunistically collects sensing data and phone usage to generate the place characteristics. The system then extracts features from sensing data collected by crowd users. The system finally matches extracted features with location information in SNS to infer the place name.

Figure 3 illustrates the overall process of the proposed system, which consists of three parts: data collector, feature extraction, and name provider. In data collector, we use GPS, WiFi, and cellular sensors to collect mobility data, along with images captured by users. On the server side, the crawler collects location and interaction data from social networks, including places, check-in histories, images, and posts in places. In feature extraction, the system applies a set of classifiers to extract the characteristics about place. We estimate the familiarity of place, residence time, and stay duration. We also analyze images to extract features about environments. Based on the extracted features, the name provider infers the place name by linking crowdsensing data with knowledge in social networks.

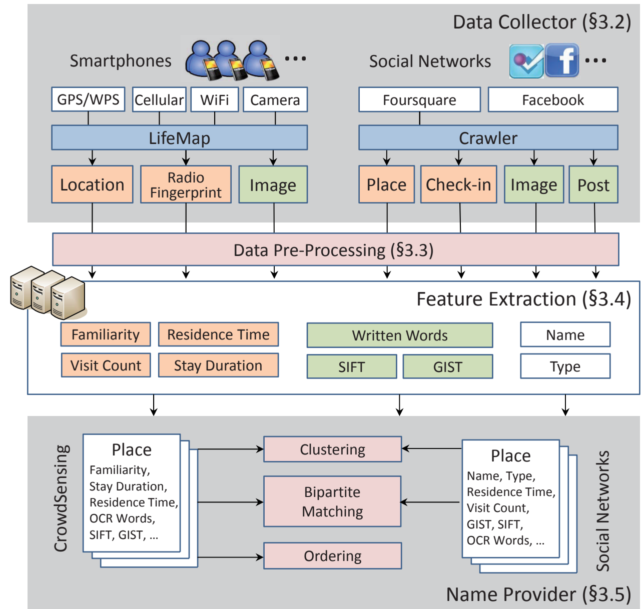

Figure 3: Overall process of the proposed system.

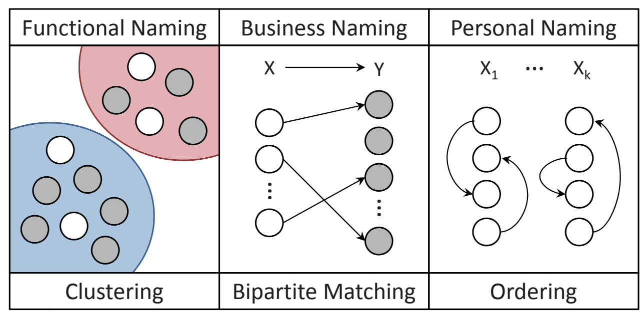

Figure 4: Conceptual description of place naming problem.

3.1 Problem Statement

We first formulate the problem of place naming. Specifically, crowdsensed data X contains a set of n places $X = \{x_1, \cdots, x_n\}$. Each node x includes four elements (L, S, R, I), in which L is a set of estimated locations, S a set of stay behaviors, R a set of radio fingerprints, and I a set of captured images. Meanwhile, location information Y in SNS includes a set of m places $Y = \{y_1, \cdots, y_m\}$. Each node y consists of six elements $(name, type, l, T, P, I)$, in which $name$ is the name of the place, $type$ the type of place, l the location information, T the set of check-in time, P the set of postings, and I the set of uploaded images.

Now, the problem is to infer the functional name, business name, and personal name of node $x \in X$. Figure 4 illustrates the conceptual descriptions of each problem. First, functional naming is defined as the inference of place type. Given X and Y, the system clusters all nodes $x \in X, y \in Y$ into different clusters $C_1, \cdots, C_k$, where k is larger than the desired number of place types. Here, the places in the same cluster are assumed to contain unique characteristics derived from the same place types. Then, the functional name of node x in C_i is a place type derived from nodes $y \in C_i$.

(a) (b) (c)

Figure 5: LifeMap visualizes collected data as (a) map, (b) daily trajectory, and (c) list form. LifeMap is available in the Android market.

Second, business naming is defined as a bipartite matching problem in graph theory. The system generates a direct edge e_{ij} from x_i to y_j. The capacity of each edge is 1, and the cost e^c_{ij} is the similarity between x_i and y_j. The goal is to find a set of edges $E \in X \times Y$ for maximizing the number of flows $|E|$ with a minimum cost $\sum_{e \in E} e^c$. In other words, the system matches nodes in crowdsensing with nodes in social networks to minimize the matching cost globally.

Last, personal naming is accomplished by ordering node $x \in X$ to choose private places. The system splits X into $X'_1, \cdots, X'_k$ depending on individual users. X'_i is a set of visited places by the i-th users. We should order $x \in X'_i$ elements to find the most private place for each user.

3.2 Data Collector

We extended the mobility monitoring tool LifeMap [3, 7], for data collection in smartphones. LifeMap continuously collects a user's mobility every two minutes using WiFi, GPS, and cellular sensors. We added an image-capturing functionality to collect images captured by users. LifeMap visualizes the visited places along with collected data on map and list form, as shown in Figure 5. Users can confirm and delete collected data at each place for privacy issues.

We also developed a crawler to collect location, as well as interaction data from social networks, such as check-in histories, posted tips, pictures, and basic user information. To preserve privacy issues, we only used public data without private authorization.

3.3 Data Pre-Processing

The system first defines *a place* in a user's trajectories. We segment the stream of collected data into places with room-level accuracy and aggregate the data at identical places. Here, we describe the place segmentation technique and the node generation for pre-processing.

Place Segmentation. We used a radio fingerprint-based place learning to segment places with room-level accuracy in a user's trajectories. Let (l_t, r_t) be sensing data from data collector at time t, in which l_t is a location and r_t is a radio fingerprint. The system considers that a user stays at same place if the similarity of the received signal strengths from WiFi APs is larger than a certain threshold. We used the Tanimoto coefficient [12] to estimate the similarity of radio fingerprints, defined as:

Table 1: Summary of used features.

Features	Sources	Terms in nodes	Usage
Residence time	Stay behavior Check-in history	Discretized into 48 bins	Functional name Business name
Stay duration	Stay behavior	Discretized into 9 bins	Functional name Personal name
Words	Image Posting	Raw words	Functional name Business name
GIST	Image	Clustered ID	Functional name Business name
SIFT	Image	Clustered ID	Functional name Business name
Familiarity	Visit frequency	-	Personal name

$$\mathcal{S}(r_a, r_b) = \begin{cases} \text{stationary} & \text{,if} \frac{r_a \cdot r_b}{\|r_a\|^2 + \|r_b\|^2 - r_a \cdot r_b} \geq \varphi \\ \text{move} & \text{,else} \end{cases}$$

where φ is the similarity threshold and the output is a similarity estimated between 0 to 1. When the system continuously detects a stationary state from time t^s to time t^e (i.e., $\min_{t^s \leq t < t^e} \mathcal{S}(r_t, r_{t+1}) \geq \varphi$), we generate a stay behavior (L, R, t^s, t^e), in which L is a set of estimated locations, R a set of radio fingerprints, t^s the start time of stay, and t^e the end time of stay. Each node x then includes a set of stay behaviors. In other words, we group the collected dataset into different places based on radio fingerprints. We emprically set the threshold φ to group the places with room-level accuracy.

For social network data, place segmentation is straightforward since the location database provides the unique ID of each place.

Node Generation. The system constructs places as nodes with a bag-of-words model. One node contains all features extracted from the sensing data in an unordered manner. We denote the term as an attribute of extracted features. We discretize continuous features (e.g., the distribution of residence time or stay duration) into a series of discrete terms by certain intervals. To handle the noisy measurements in the collected dataset, the system modifies the frequency of terms within a node, based on the confidence and uniqueness. We first filter out outputs with low confidence scores, using an empirically determined threshold (see Section 4.4). The system then applies the term frequency-inverse document frequency (tf-idf) [23], which is frequently used in conventional document analysis. This technique emphasizes unique terms and reduces the weight of non-discriminative terms. For example, tf-idf decreases the frequency of common terms across all nodes, as those terms are meaningless for discrimination. The tf-idf of term w at node x in set X is defined as:

$$\text{tf-idf}(w, x, X) = \frac{f(w,x)}{\max\{f(w,x) : w \in x\}} \times \log \frac{|X|}{|\{x \in X : w \in x\}|},$$

where $f(w, x)$ is a frequency of term w in node x.

3.4 Feature Extraction

We describe the features we used to model the place characteristics. Each node contains features about a user's behavior and environments, obtained from both crowdsensing and social network data. Table 1 summaries the extracted features and their usage at the name provider.

Residence Time. Residence time indicates stay behavior of users at a place tied with time-of-day (e.g., 10am or 4pm). The underlying assumption is that both the stay duration and the time-of-day imply meaningful patterns when

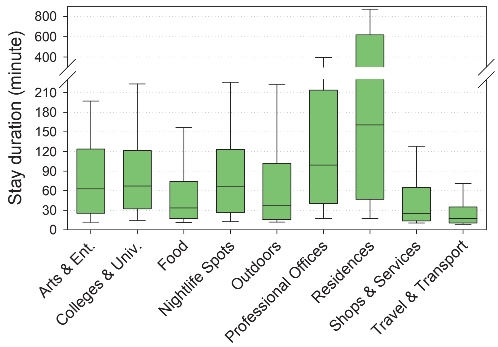

Figure 6: Distribution of stay duration according to the functional name.

people visit a certain place. Intuitive examples include people spending meal times at food places or students staying at schools from 9am to 5pm. From the stay behaviors S^x at node x, the system generates the distribution of residence time at place R^x, based on the discrete histogram of residence time set $(t_1^s, t_1^e), \cdots, (t_k^s, t_k^e)$. The probability of R^x at time t is defined as:

$$R^x(t) = p(t_i^s \leq t \text{ and } t \leq t_i^e | (t_i^a, t_i^e) \in S^x),$$

where t is a specific hour and minute of a day (e.g., 8am or 4pm). We normalize R^x as $\int_t R^x(t)\mathrm{d}t = 1$ and discretize R^x into 48 bins (i.e., 30 minutes for each bin). We use the overall distribution of residence time instead of probability at a specific time.

For social network data, we generate R^x based on check-in history T. Compared to stay behavior sensed by a smartphone, T contains a set of instant check-in time t^c, not the duration. Therefore, $R^x(t)$ is a fraction of observed check-ins at time t to overall check-in histories, defined as $p(t_i^c = t | t_i^c \in T)$.

Stay Duration. The system produces the distribution of stay duration D^x at node x in each place. This feature takes the pattern of stay behavior without time-of-day. For example, many college students visit specific classrooms with uniformly distributed patterns from 9am to 6pm, and their stay duration may show consistent patterns—i.e., the length of a class hour. D^x is a form of discrete histogram based on the set of stay duration at node x, expressed as $D^x = \{s_i = t_i^e - t_i^s | (t_i^s, t_i^e) \in S^x\}$. We discretized the stay duration into 9 bins (i.e., 15, 30, 60, 90, 120, 150, 210, 400, 400+ minutes) based on the distribution of the stay duration according to the functional name, as shown in Figure 6.

Social networks do not provide stay duration at places because the check-in operation indicates only an instant time instead of the duration. Previous work estimated stay duration approximately by using the time difference between consecutive check-ins of each user [8], but such approach requires private information (i.e., user's identification of each check-in) which is not acquired in our system. The stay duration is therefore a missing value in social network data.

Written Words. The written words found in menus, store signage, posters, or postings imply the characteristics about the place. For example, we see the brand name (e.g., Starbucks or Apple) on signs or words related to coffee (e.g., Americano or roast) in a cafe. Similarly, people post texts about places for sharing purposes, such as recommendations of menus at restaurants or reviews of services at stores in location-based social networks.

The system mines words from the captured images in smartphones, and also from the uploaded images and texts in social networks. To mine such words in a set of captured image I, the system incorporates optical character recognition (OCR) technique [10]. From each image, the engine provides recognized words and confidence score in each result. We utilized confidence score to modify the weight of recognized words. We set the high threshold of confidence score to filter out the noisy results.

Similar to written words mined from images, the texts in social network contains meaningful words related to place characteristics. The system selects postings associated with places and parses texts into an unordered set of unigram words. To reduce the noises in extracted words, we used the dictionary to filter out non-grammatical words.

Local Features in Images. The local features describe the unique characteristics of specific objects in images. Practical examples include a brand logo on a wall or a mark on the signage. We incorporate scale-invariant feature transform (SIFT) [20] to mine local features in images. SIFT is commonly used in computer vision for object recognition or image stitching. This technique produces a set of key points in each image that describes position, scale, and orientation of each interesting pattern in images. We cluster the extracted SIFT features from all images in a dataset and generate the distribution of cluster ID at each node.

Scene Features of Images. Scene feature describes the overall scene characteristics of an image, such as major shape or visual patterns. For example, we see strong horizontal lines in supermarket shelves or round shapes of bowls in images capturing dishes in restaurants. To leverage scene features, we chose GIST feature [22], which is widely used in the literature to capture scene characteristics. We produce GIST-based feature vectors for each image. Similar to local features, we cluster GIST vectors to group images into similar sets. The system then takes the frequency of images in each cluster into the features of a place.

Familiarity. Familiarity indicates how frequently a user visits a place. We define the familiarity of node x that contains k stay behaviors, motivated by the entropy in [9], as follows:

$$\text{Familiarity}(x, S) := -\sum_{i=1}^{k-1} p(i; S) \log p(i; S),$$

where $p(i; S)$ is $\frac{t_{i+1}^s - t_i^s}{t_k^s - t_1^s}$. A place will have high familiarity if a user visits it with certain regularity (e.g., every day). Conversely, a place will have low familiarity if the distribution of stay at a place is biased or randomly observed. We use familiarity to determine the privateness of a place for each user. For example, employees at a restaurant exhibit high familiarity, as they regularly come to the restaurant for work. Conversely, customers at a restaurant show relatively lower familiarity, due to the randomness of their stay behavior.

3.5 Name Provider

We now describe the principles we applied to integrate all extracted features for inferring place names according to three types (i.e., functional, business, and personal name).

Functional Name. Functional name indicates the semantic type of place, such as food place, nightlife, or shopping place. We applied a clustering technique to infer the functional names of places. The intuition is that the functional name is a category of multiple places that exhibit the similar characteristics of sensing data. The system clusters all nodes into different groups with consistent patterns of terms. To handle no prior knowledge about the number of groups, we adopted Dirichlet process mixture (DPM) suggested by [2, 25]. DPM describes infinite Dirichlet distribution as a prior of groups, and provides a flexible scheme to estimate the number of clusters and parameters. Our system used residence time, stay duration, words from images or postings, GIST and SIFT features to compute the distance between nodes. Given a node a and b, the distance between a and b is defined as

$$u(a,b) = KL(R^a \cap D^a, R^b \cap D^b) + \alpha \frac{1}{|W^a \cap W^b|},$$

in which $KL(\cdot)$ is the symmetrized Kullbeck-Leibler (KL) measurement [15] of the residence time and the stay duration distributions, α the scaling factor, and $|W^a \cap W^b|$ the overlapped number of mined words, GIST, and SIFT features. The KL divergence measures the expected amount of information required to transform samples from a distribution into another one. The system calculates the KL divergence without stay duration if node a or b does not contain the information. We set the minimum value of $|W^a \cap W^b|$ as 1 if two nodes do not have overlapped terms. The metric indicates that the distance between two places is close if the stay behaviors at two places are similar and two places sufficiently include the same written words or features from images. For example, the residence places would be grouped in the same cluster because of shared terms in nighttime (e.g., 0am to 7am) or long stay durations (e.g., 400+ minutes). Or, the coffee places can be differentiated from the shopping places due to different observations in mined words. Given a cluster C_i, the functional name of node $x \in C_i$ is the type of node $y \in C_i$ with highest probability $\max_{type \in y} \frac{|type|}{|C_i|}$.

Business Name. Business name is the actual name of a place such as Starbucks, Apple Store, or Marche. To infer the business name, the system should generate one-on-one links between a place generated by crowdsensing and a place in SNS. We applied the bipartite matching which finds a flow between two sets with a minimum cost. The problem is to match node $x \in X$ to node $y \in Y$ with a minimum matching cost. The system generates a graph with four layers, U, X, Y, V, as illustrated in Figure 7. The system generates direct edges from U to V: $U \rightarrow X \rightarrow Y \rightarrow V$. Initially, the capacity and the flow of all edges are 1 and 0, respectively. The edges in $U \rightarrow X$ and $Y \rightarrow V$ have no cost, and the cost of edges $X \rightarrow Y$ is the similarity between two nodes, x and y, defined as:

$$\mathcal{H}(x,y) = d(x,y) \times u(x,y),$$

where $u(\cdot)$ is the similarity between two nodes and $d(\cdot)$ is the cdf form of relation between error distance and location mapping we determined empirically (see Section 2.1). In other words, we consider the difference between the distributions of terms at two nodes along with the distance between locations. $\mathcal{H}$ outputs low value if the distance between two nodes is close and the nodes shared many terms. We then employed the minimum-cost maximum-flow algorithm [1] to find an optimal flow F that derives maximum flow from U to V with minimum cost. In F, the business name of node x_i is the name of node y_j if the flow of edge e_{ij} is 1. This approach finds the match between places with a globally-minimum cost. In the list of business name, the name of node y_j is ranked at first, and the rest nodes are ordered by the similarity value.

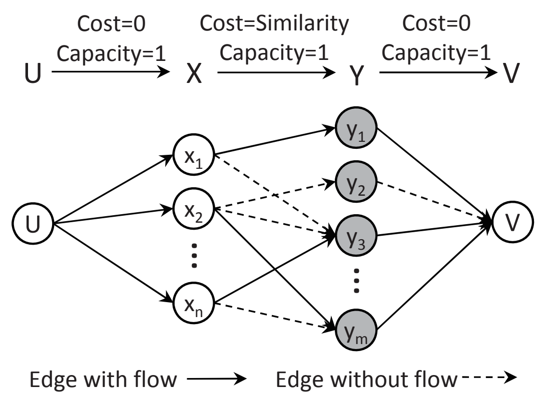

Figure 7: Conceptual description of business naming.

Personal Name. Personal name is the personalized name of a place given by an individual user, such as *my home* or *my workplace*. Given set X' of a particular user, the problem is to order nodes in X' to ultimately find the most private nodes x. The system considers the places generated by mobile phones since the personal name should be provided in a personalized manner. We ordered nodes $x \in X'$ by familiarity and stay duration to infer the personal names of places. Among the places inferred as residence type, the system considers a place with the highest familiarity as *my home*. Then, excluding the residence type places, the system considers the next place with the highest familiarity as *my workplace*. We cannot directly link to *professional offices* to my workplace. For example, Starbucks should be inferred as my workplace for an employee at Starbucks, but the functional name of the place is *food place*, not *professional offices*.

4. EVALUATION

4.1 Data Collection

We deployed the data collector tool from November 2011 to September 2012 to collect data traces from 70 participants on their primary smartphones. The participants labeled the place names and categories using a user interface for ground truth. The participants also provided the place information offline at the last day of data collection. The average collection period was 94 days, and the median was 60 days. Table 2 presents the description of collected dataset. The collected traces included 3,300 images and 3,800 places over a sampling period of about 174,000 hours. Note that we are currently sharing the collected dataset in the CRAWDAD research communities [14].

We performed the crawling in Foursquare [11] for four months (from May 2012 to September 2012) and restricted spatial regions within Seoul, Korea. The dataset contains 9,200 pictures, 101,200 tips (description about places), and 1,078,100 check-ins at 130,000 locations by 31,000 unique

Table 2: Description of collected data.

Place Type	CrowdSensing		Social Networks		
	# of Place	# of Image	# of Place	# of Image	# of Posting
Arts & Entertainment	298	152	6,892	786	9,145
College & Universities	815	1,392	5,296	157	3,870
Food Place	1,337	952	46,449	4,750	58,193
Nightlife	129	84	6,212	421	4,188
Outdoors & Sports	136	86	11,909	911	7,758
Residences	183	209	1,768	38	543
Shops & Services	376	260	18,962	699	11,827
Travel & Transport	283	68	5,983	413	1,254
Professional Offices	208	117	21,117	767	10,629
Others	71	21	5,445	233	1,827

users. We only collected the public data that are accessed without private authorization.

4.2 Implementation

We implemented the data collector client in Android SDK 2.2. The WiFi scanning intervals and window size are 2 minutes and 30 seconds, respectively. The tool monitors captured images as event-driven methods. The crawler and backend server were implemented on a Windows 7 server. The crawler collected location and interaction data using Foursquare APIs [11] every 5 minutes. For feature extraction, we set the similarity threshold of WiFi vector to 0.7 for segmenting places with room-level accuracy, as suggested in [13, 7]. We built two sets of residence-time histograms, one for the weekend and one for weekdays, as suggested in [6]. Each histogram bin represents a 30-minute period during a single day (i.e., 48 bins). The dataset derived approximately 12,000 GIST vectors and 60 million SIFT vectors from images. To discretize the large number of features from images, we applied the fast-approximate k-means clustering, provided by [21]. We followed the definition of the functional name provided by Foursquare [11]. We used 10 types of the functional name as presented in Table 2.

4.3 Macro Benchmarks

We validate the overall performance of the proposed system, compared to the related approaches below. Note that we denote the proposed system as *crowdsensing-aided*.

Location-based Lookup. This method estimates similarity between places based on physical distance. We chose the type of closest places for functional naming, and for business naming, we generated a list ordered by distance.

Popularity-aided. This method uses the quantity of interaction data in social networks [18, 24]. The basic assumption is that a user may frequently visit popular places. We used the number of check-ins and posted tips, as suggested by [18], and further utilized the number of uploaded images to determine popularity. The method chooses the most popular place among the nearby places. We considered places within 180 meters error to cover 90% of cases (see Section 2.2).

We used *precision* and *recall* metrics to evaluate the performance of functional naming. The classified type is considered correct if the ground truth is equal to the type with the highest probability. Figure 8 shows the precision and recall of the proposed systems compared with the related methods. Our system outperforms the related methods by 27% precision and 13% recall. The results indicate that user behavior

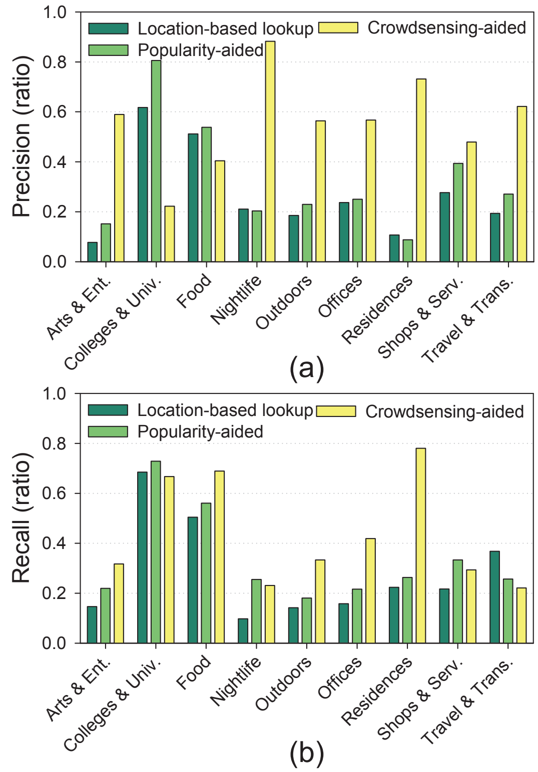

Figure 8: (a) Precision and (b) recall of related systems in functional naming.

and environmental features are more effective than distance or popularity for inferring the functional name. Our system shows relatively worse precision in colleges and food places, as shown in Figure 8(a). The reason is that (1) the outputs of location-based lookup and popularity-aided method tend to be biased as food places; and (2) the distance metric works well for college and university places since those places are located densely in certain regions. This phenomenon is presented as a confusion matrix in Figure 9. Location-based lookup and popularity-aided methods infer about 34% of places as food places, due to the large number of samples of food places in social networks. Compared to the related approaches that showed 33±20% accuracy, the proposed system correctly infers the type of 56±18% places.

Regarding the performance of the proposed system in business naming, we define the accuracy of business naming as the order of places in the generated list. For example, 20% accuracy for top-5 indicates that the order of the actual place is less than five in the list for 20% of cases. The proposed system generates 29±8% higher accuracy than the related methods, as illustrated in Figure 10(a). Our system correctly inferred the business name in 38% of the cases, and 84% of places are within 10 places in the generated list. The result indicates that the semantic features from sensing data enable the precise recognition that is missed by using explicit features, such as distance or popularity. The proposed system provides less than 6.7 ranks for 80% of cases, but location-based lookup and popularity-aided output less than 22.8 and 69 ranks for 70% of cases, respectively, as shown in Figure 10(b). Considering that higher rank means greater burden on the user, the results reveal that our system

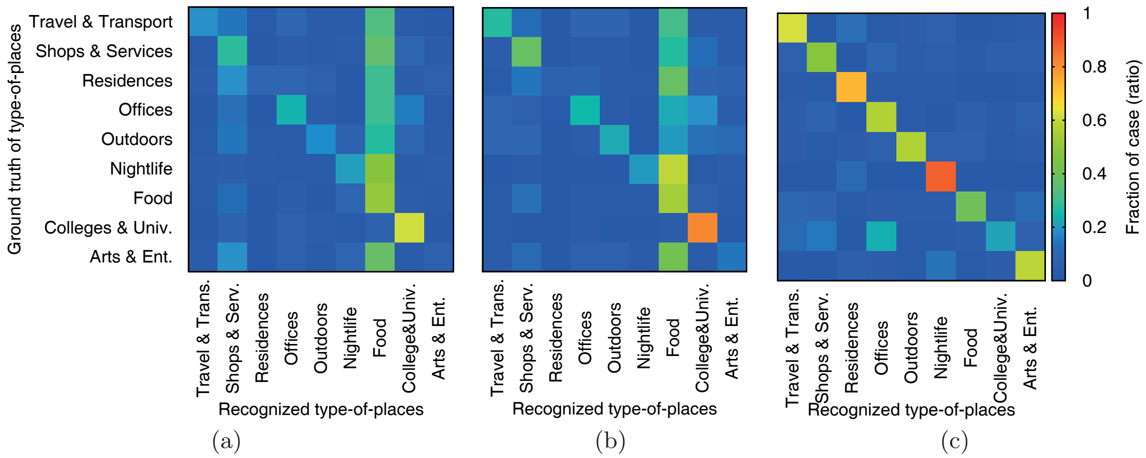

Figure 9: Confusion matrix of functional naming in (a) location-based lookup, (b) popularity-aided, and (c) the proposed system.

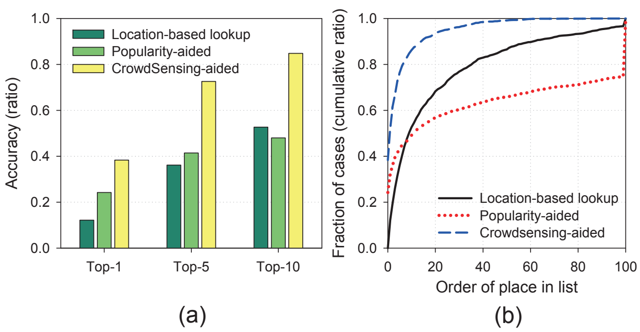

Figure 10: (a) Accuracy of related systems in business naming and (b) the CDF form of order of all places. Left-most curve indicates better results for business naming.

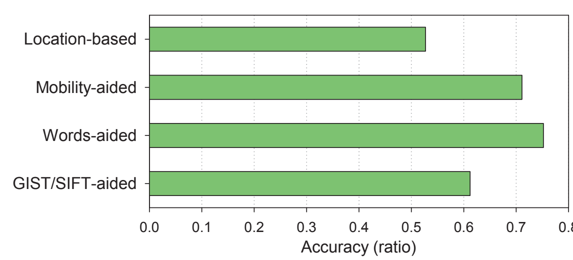

Figure 11: Accuracy of business naming in top-10 list used by different features.

greatly reduces user intervention by integrating the sensing system for place naming.

Figure 11 shows the accuracy of business naming, using different features and sources of data in isolation. Based on the use of physical location, we additionally used the features such as mobility data, words from OCR classifier and posting, and GIST/SIFT features from images. The performance gains obtained with using mobility and words are relatively larger than using GIST/SIFT features. The result indicates that the written words in the place are meaningful to differentiate the places, and people exhibit similar stay behaviors at the same place. The indoor scene classification (i.e., GIST/SIFT features) operates poorly in business naming, although the features are effective to recognize the place types [5].

For personal naming, we ignored the comparison with the related methods, as we needed private authorization to collect personal information from social networks. The proposed system correctly infers 89±8% of personal places, as shown in Figure 12(a). *My home* shows relatively higher precision and recall than *my workplace/school* because of unique stay patterns at home. The errors in recognizing my workplace/school occurred in participants with multiple places considered as workplace or school. Familiarity

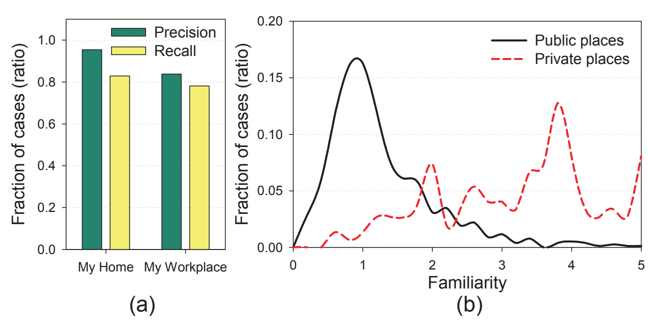

Figure 12: (a) Precision and recall for personal naming and (b) distribution of familiarity in public and private places.

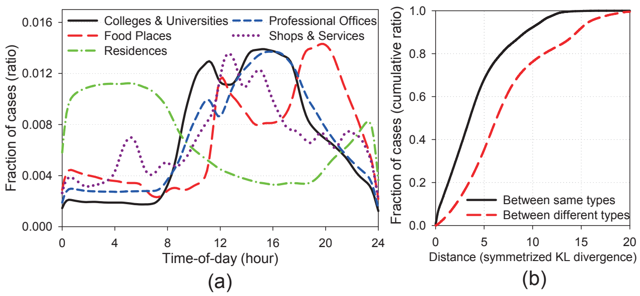

Figure 13: (a) Stay behavior pattern in several types of places and (b) similarity of patterns. The left-most curve indicates higher similarity.

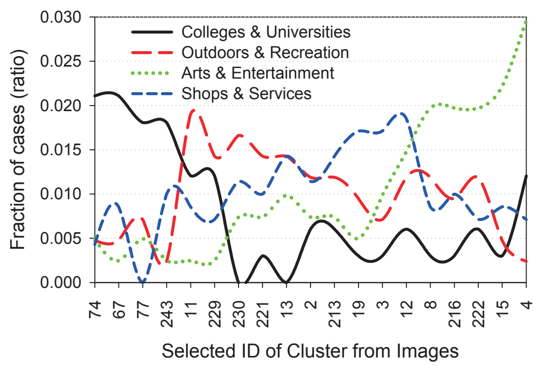

Figure 14: Distribution of terms mined from images.

effectively measured the privateness of places, as presented in Figure 12(b). The private places show relatively higher familiarity than the public places, because of regular visit patterns on private places.

4.4 Micro Benchmarks

In exploring the discriminative power of each feature on the performance of the proposed system, we find that features regarding a user's behavior are robust for functional naming, and the environmental features mined from images or postings are effective for business naming.

Figure 13(a) presents the residence-time distribution of several place types on weekdays. We found that the stay behaviors have strong features that can differentiate some types (e.g., residences and food places), but several types (e.g., colleges and professional offices) are confused, due to similar stay patterns. We further investigated the difference of distributions between places, as shown in Figure 13(b). The figure illustrates the Kullback-Leibler divergence [15] between the residence-time distributions. The results indicate that places of the same type have higher similarities compared to places of different types.

We highlight the characteristics of the environmental features mined from images (i.e., GIST, SIFT, and OCR words). Figure 14 presents the distribution of attributes mined from images. We presented a few selected attributes since the semantic features from images were sparsely distributed over a large number of attributes. The result shows that clustered terms were observed differently on some types of places that can differentiate places. The example images in the same cluster inferred as food place are presented in Figure 15(a).

Figure 15: Example images (a) in the same cluster and (b) at the two places.

The images contains cups or bowls with round shapes that derived similar terms. Similarly, the terms mined from images at the same place show consistent patterns since those images contains similar scene or local objects, as illustrated in Figure 15(b).

In OCR words cases, the system observed a high frequency of OCR words in *food places* and most images captured at *shopping* and *travel and transport* contained written words, as illustrated in Figure 16. The words were mostly mined from images capturing store signs, menus, or a direction boards in a place. To filter out noisy results, we utilized the confidence score provided by the OCR engine. Figure 17(a) shows that false positives decrease as the threshold of confidence score increases. The result shows that even though the classifier produces noisy results, reliable results can be obtained using high threshold of confidence. The side-effect of this approach is a significant drop in true positives, but the system still extracts sufficient amount of results due to the large amounts of data collected by crowdsensing and crowdsourcing. We set the threshold of confidence as 300 to obtain high precision with small loss in recall, as illustrated in Figure 17(b). The low recall indicates that the classifier would require further investigation to obtain more classified features from noisy images. Table 3 indicates that major words on different types of places reflected the characteristics of places.

Finally, we explored the relationship between the number of samples and the performance of the proposed system. Intuitively, as shown in Figure 18, the more data leads a more accurate result. We found that a larger number of visits induced better accuracy in functional naming, and a greater number of images is effective for business naming. Places with more than 100 visits showed 80% of accuracy, while more than 20 images derived 4.3±4.9 ranks in the list. The reason is that the number of matched features of the same places increases with the number of images. The visitations would increase as time goes by in daily life, but the collection of images requires the active user participation. Considering that 9.7% of places in SNS and 34% of places in crowdsens-

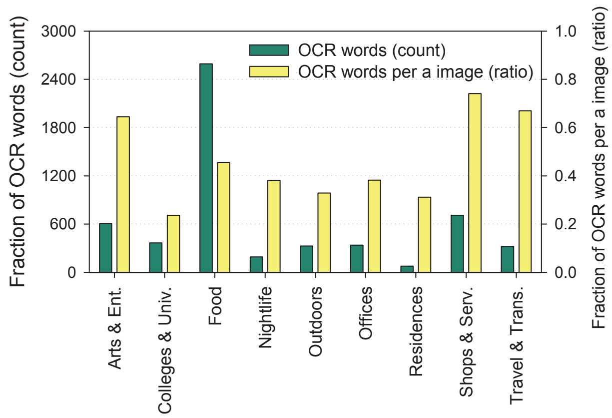

Figure 16: Distribution of recognized OCR words in several types of places.

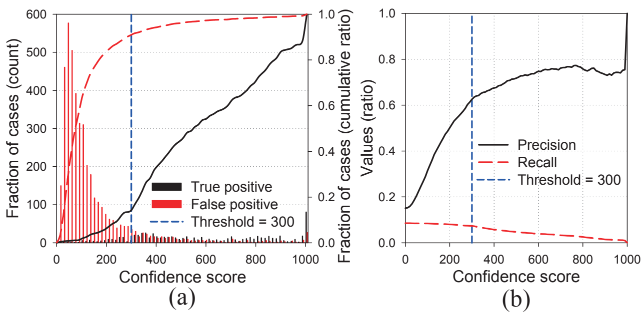

Figure 17: (a) Distribution of confidence score in OCR and (b) precision and recall according to confidence score.

ing contain image data, the reward for data collection is required to induce user participation.

5. RELATED WORK

Extensive studies have been conducted recently to extract advanced information on places, beyond simple raw coordinates. We discuss related work in three different domains: place naming, local search and recommendation, and place segmentation.

Place Naming. Lin et al. [19] analyzed people's preferences in place naming and defined the hierarchical relation of place naming taxonomy. They designed a model to predict the desired types of place names that people would use when naming in a given situation. The system, however, did not solve the problem of actual place naming. We followed their taxonomy and solved the problem of place naming in practice.

The work by Lian et al. [18] is the most relevant to our work. They designed a model to infer the current place in a given GPS point. They used check-in histories at social network websites and the quantity of interaction at location review websites, including popularity, visit frequency, and temporal information of stay. Our work improves the past works in many ways. We exploit more attributes captured by mobile phones that have not been previously covered. Our system also provides place names without user intervention.

Table 3: Top OCR words on several types of places.

Place Type	Top OCR Words
Arts & Entertainment	size, movement, sequence, lotte, mayor, world, march, admission, cctv, cinema, fantasy, films
Food Place	coffee, set, menu, café, open, lunch, free, rice, special, tea, all, close, noodles, take, chicken
Nightlife	wine, band, bar, cinnamon, glass, non, open, sparkling, available, blending, cocktail, dinner
Outdoors & Sports	forest, park, playground, cafeteria, center, children, fee, holidays, information, operation
Shops & Services	keep, new, shop, sale, thinking, fax, free, lucky, open, point, smart, women, accessory, best
Travel & Transport	airport, express, seoul, airways, america, center, departure, Incheon, seat, transfer
Professional Offices	room, public, rest, consulting, information, korea, library, service, smart, team, www

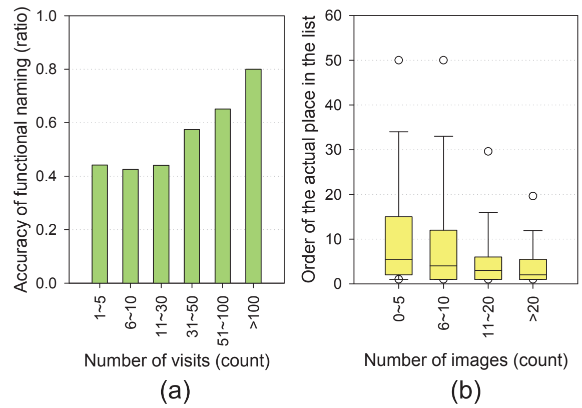

Figure 18: (a) Relation between number of stays and accruacy of functional naming and (b) box plot of relation between number of images and the order. Box indicates lower quartile, median, upper quartile and whisker indicates 10% and 90% of observation.

Recently, several works have focused on the inference of functional name of a place. Works in [4, 28, 5] used various features, such as number of radio beacons [4], phone usage [28], stay duration [4, 5], or captured images [5]. Ye et al. [27] employed check-in histories on social networks. They focused on three types of functional name (nightlife, food place, shopping place). Our previous work [5] used the topic modeling approach to infer the functional name of places. The system used image and audio data captured by mobile phone to characterize places into 7 categories. The system is, however, required to label the place name manually by smartphone users. In this paper, we utilized the feature extraction method in [5] and integrated the knowledge found in SNS to provide autonomous place naming without user intervention. Additionally, we designed the place naming method from a person's perspective (i.e., functional, business, personal name). Our system now provides autonomous naming even for business and personal name of places. This was not covered in [5].

Local Search and Recommendation. Research on local search and place recommendation produced the ranked list of relevant places based on user queries or contexts. Diverse approaches are used to estimate relevance between places. Shankar et al. [24] used public interaction data in social networks, CLR [17] employed similarity between users, and Hapori [16] used time, weather, and activity of users as given contexts. We utilized these works to estimate relevances, and we further provided the actual name of a place automatically, by combining smartphone data and SNS data.

Place Segmentation. Research on place segmentation recognized point-of-interest in user trajectories. Most studies have used the fingerprint method of surrounding radio beacons, such as Bluetooth [26] or WiFi access points [26, 7]. The basic idea is that the signal strengths of radio beacons are similar when a user stays in a certain place. We utilized radio-based place learning to segment places at room-level accuracy and generated stay duration in places.

6. DISCUSSION

We discuss the limitations of the proposed system, along with future research directions. We also highlight a number of areas that require further investigation.

Privacy. Although our data collector enables the deletion of collected data, uploading raw images into the server has privacy concerns in practice. Here, one possible solution is to process the classifiers locally in smartphones. For example, instead of uploading raw data, smartphone extracts OCR words, GIST, or SIFT features and uploads these features to server in an unordered manner. To realize this local processing technique, we need an efficient classifier with low complexity and also a job scheduler that would execute classifying/uploading operations when the device is relatively under usage (e.g., night time with battery charging). Meanwhile, one thing we counter-intuitively found in our study is that, many participants are willing to share pictures publically with others. In fact, social network users are already sharing many images, when interacting with others. Therefore, we believe that local data processing may accelerate the active and voluntary participation of end users in crowdsensing framework.

Incremental Learning. The proposed system required an non-trivial amount time of data processing in our server (several hours of processing in Intel i7 CPU 860 server with 8GB RAM). The major workload in the server is to process the learning components (i.e., feature extraction and clustering). Considering that we, in our work, used data collected in only one city and the amount of data is rapidly increasing with a widespread use of smartphones and SNS, the data processing issue is practically a big challenge in crowdsensing framework. We believe that an incremental learning scheme would be necessary to handle a large volume of data efficiently in real life.

User Participation. The proposed system requires knowledge from crowdsourcing to infer the place name. Our work expands the coverage problem on place naming by applying the SNS knowledge to information collected with the crowdsensing approach. However, the system is not able to learn additional names that are not in SNS. To solve this limitation, the system should somehow support generating new names, or at least provide a method to encourage users to put their knowledge back into SNS. Generating new names is beyond the scope of our research. We, instead, plan to investigate a reward system to induce active user participation in crowdsensing approach. Considering that SNS is missing about 22% of the places visited in daily life (see Section 2.1), an appropriate rewarding mechanism would greatly improve the coverage of place naming.

7. CONCLUSION

In this paper, we proposed an autonomous place naming system for providing place names according to three types. Our system leverages the place characteristics mined from opportunistically crowdsensed data using smartphones and crowdsourced information in social networks. By integrating sensing system and exploiting crowdsourcing to gather large volumes of data, our system is able to provide place name from a person's perspective, beyond raw location coordinates.

Our study presents the possibility of linking sensing systems to social networks, toward the advanced understanding of human life. The proposed system provides rich awareness of the places that have been annotated manually by users in previous works. Such advanced information is a building block for many context-aware applications, such as city-scale activity recognition, mobile advertising, or enhanced place recommendations. Our future work includes study on a crowdsensing framework that will explore effective incentives to encourage active participation of end users.

8. ACKNOWLEDGMENT

We sincerely thank the 70 data donators among LifeMap users in Korea. This work was supported by the National Research Foundation of Korea grant funded by the Korean government, Ministry of Education, Science and Technology (No.2011-0006464, No.2012-0005522).

9. REFERENCES

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Inc., 1993.

[2] C. Antoniak. Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174, 1974.

[3] Y. Chon and H. Cha. Lifemap: A smartphone-based context provider for location-based services. *Pervasive Computing, IEEE*, 10(2):58–67, April-June 2011.

[4] Y. Chon, Y. Kim, H. Shin, and H. Cha. Topic modeling-based semantic annotation of places using personal behavior and environmental features. In *Proceedings of Nokia Mobile Data Challenge Workshop, in Conjunction with Pervasive'12*, 2012.

[5] Y. Chon, N. D. Lane, F. Li, H. Cha, and F. Zhao. Automatically characterizing places with opportunistic crowdsensing using smartphones. In *Proceedings of the 14th International Conference on Ubiquitous Computing*, Ubicomp'12, pages 206–212. ACM, 2012.

[6] Y. Chon, H. Shin, E. Talipov, and H. Cha. Evaluating mobility models for temporal prediction with high-granularity mobility data. In *Proceedings of the 10th International Conference on Pervasive Computing and Communications*, PerCom'12, pages 206–212. IEEE, 2012.

[7] Y. Chon, E. Talipov, H. Shin, and H. Cha. Mobility prediction-based smartphone energy optimization for everyday location monitoring. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, SenSys'11, pages 82–95. ACM, 2011.

[8] G. Colombo, M. Chorley, M. Williams, S. Allen, and R. Whitaker. You are where you eat: Foursquare

checkins as indicators of human mobility and behaviour. In *Proceedings of the 10th International Conference on Pervasive Computing and Communications Workshops*, PerMoby'12, pages 217 –222. IEEE, 2012.

[9] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh. Bridging the gap between physical location and online social networks. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 119–128. ACM, 2010.

[10] J. Du, Q. Huo, L. Sun, and J. Sun. Snap and translate using windows phone. In *2011 International Conference on Document Analysis and Recognition*, ICDAR'11, pages 809 –813. IEEE, 2011.

[11] Foursquare Inc. Foursquare. `http://foursquare.com`.

[12] P. Jaccard. The distribution of the flora in the alpine zone. *New Phytologist*, 11(2):37–50, 1912.

[13] D. H. Kim, Y. Kim, D. Estrin, and M. B. Srivastava. Sensloc: sensing everyday places and paths using less energy. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, SenSys '10, pages 43–56. ACM, 2010.

[14] D. Kotz and T. Henderson. Crawdad: A community resource for archiving wireless data at dartmouth. *IEEE Pervasive Computing*, 4(4):12–14, Oct.-Dec. 2005.

[15] S. Kullback and R. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[16] N. D. Lane, D. Lymberopoulos, F. Zhao, and A. T. Campbell. Hapori: Context-based local search for mobile phones using community behavioral modeling and similarity. In *Proceedings of the 12th International Conference on Ubiquitous Computing*, Ubicomp'10, pages 109–118. ACM, 2010.

[17] K. W.-T. Leung, D. L. Lee, and W.-C. Lee. Clr: a collaborative location recommendation framework based on co-clustering. In *Proceedings of the 34th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '11, pages 305–314. ACM, 2011.

[18] D. Lian and X. Xie. Learning location naming from user check-in histories. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11, pages 112–121. ACM, 2011.

[19] J. Lin, G. Xiang, J. I. Hong, and N. Sadeh. Modeling people's place naming preferences in location sharing. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 75–84. ACM, 2010.

[20] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the 7th IEEE International Conference on Computer Vision*, ICCV'99, pages 1150–1157. IEEE, 1999.

[21] M. Muja and D. G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *International Conference on Computer Vision Theory and Applications*, VISAPP'09, pages 331–340, 2009.

[22] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[23] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5):513–523, 1988.

[24] P. Shankar, Y.-W. Huang, P. Castro, B. Nath, and L. Iftode. Crowds replace experts: Building better location-based services using mobile social network interactions. In *Proceedings of the 10th International Conference on Pervasive Computing and Communications*, PerCom'12, pages 20–29. IEEE, 2012.

[25] S. Tominaga, M. Shimosaka, R. Fukui, and T. Sato. A unified framework for modeling and predicting going-out behavior. In *Proceedings of the 10th International Conference on Pervasive Computing*, Pervasive'12, pages 73–90. Springer, 2012.

[26] L. Vu, Q. Do, and K. Nahrstedt. Jyotish: A novel framework for constructing predictive model of people movement from join wifi/bluetooth trace. In *Proceedings of the 9th International Conference on Pervasive Computing and Communications*, PerCom'11, pages 54–62. IEEE, 2011.

[27] M. Ye, D. Shou, W.-C. Lee, P. Yin, and K. Janowicz. On the semantic annotation of places in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 520–528. ACM, 2011.

[28] Y. Zhu, E. Zhong, Z. Lu, and Q. Yang. Feature engineering for place category classification. In *Proceedings of Nokia Mobile Data Challenge Workshop, in Conjunction with Pervasive'12*, 2012.