# URB 05 Discovering semantically meaningful places

## URB 05.1 People

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

## URB 05.2 Overview

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

Several commercialized products have shown that a mixture of GPS and RF-beacon-based location can allow a device to compute its position with high availability throughout a carrier's day.[1] Raw coordinates provided by these location systems are an excellent resource for current location-aware applications such as navigation and emergency response that require absolute locations for only a short period of time.

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

## URB 05.3 Approach

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

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

---

[1] e.g., Navizon: Peer-to-peer wireless positioning: http://www.navizon.com; Skyhook wireless: http://www.skyhookwireless.com

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

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

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

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

The PlaceSense algorithm is designed to learn places by continuously monitoring the radio beacons surrounding a mobile device. It uses commodity radio beacons such as WiFi access points (APs) or cell towers as its signals, which are pervasive and can be detected by most mobile devices. WiFi access points and cell towers broadcast unique identifiers for both communication set-up and hand-off. For example, WiFi access points transmit periodic frames called Beacon Frames containing the AP's unique MAC address among other synchronization information. Likewise, cell towers are identified by a unique Cell-ID. We refer to these fixed radio sources as beacons. Mobile devices can periodically scan for these nearby beacons without connecting to or communicating with them. The IDs of these beacons are visible even if the network is encrypted. A timestamped log of these beacon scans is the input to PlaceSense's discovery phase.

The basic approach of PlaceSense is: Define a scan window and the number of windows to wait before deciding that the radio environment is stable. (A radio environment is stable as long as at least one representative beacon is found continuously. A scan window is stable if it contains no new beacons not seen since the examination began.) *Entrance to a place* is then indicated by stable scan windows seen continuously for the defined window count. The fingerprint, a list of beacons found during a stay and their response rate, is gathered to define representative beacons of the place. *Departing from a place* is indicated by the loss of all representative beacons.

[1] N. Marmasse and C. Schmandt. Location-aware information delivery with commotion. In *HUC '00: Proceedings of the 2nd international symposium on Handheld and Ubiquitous Computing*, pages 157–171, London, UK, 2000. Springer-Verlag.

[2] D. Ashbrook and T. Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Comput.*, 7(5):275–286, 2003.

[3] J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello. Extracting places from traces of locations. In *WMASH '04: Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, pages 110–118, New York, NY, USA, 2004. ACM.

[4] R. Hariharan and K. Toyama. Project lachesis: Parsing and modeling location histories. In M. J. Egenhofer, C. Freksa, and H. J. Miller, editors, *GIScience*, volume 3234, pages 106–124. Springer, 2004.

[5] J. Krumm and K. Hinckley. The nearme wireless proximity server. In *In Proceedings of Ubicomp: Ubiquitous Computing*, pages 283–300. Springer, 2004.

[6] J. Hightower, S. Consolvo, A. LaMarca, I. E. Smith, and J. Hughes.Learning and recognizing the places we go. In *Ubicomp*, pages 159–176, 2005.

[7] N. Toyama, T. Ota, F. Kato, Y. Toyota, T. Hattori, and T. Hagino. Exploiting multiple radii to learn significant locations. In *LoCA*, pages 157–168, 2005.

[8] N. Song. Discovering user context with mobile devices: location and time. Thesis, Massachusetts Institute of Technology. Dept. of Electrical Engineering and Computer Science., 2006.

[9] U. Ahmad, B. J. d'Auriol, Y.-K. Lee, and S. Lee. The election algorithm for semantically meaningful location-awareness. In *MUM*, pages 55–63, 2007.

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

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

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

### URB 05.5 Accomplishments

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

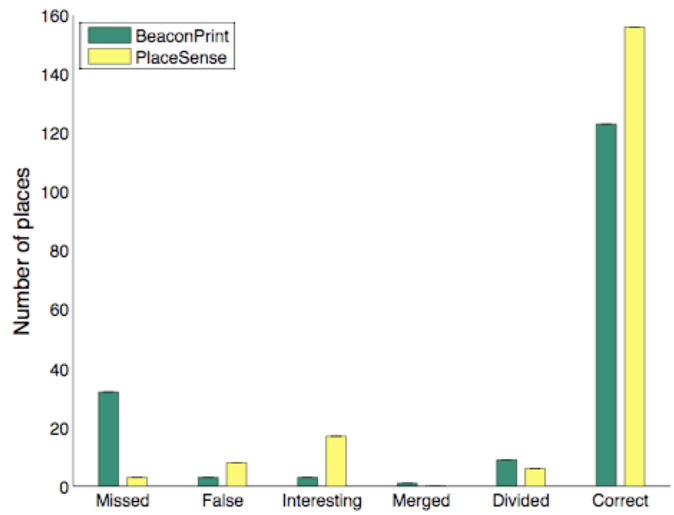


Fig 1. Number of places found from real-life traces by PlaceSense and BeaconPrint. PlaceSense reduces the number of missed places while also increasing the number of interesting and false places.

(a) BeaconPrint

| Duration (minutes) | Missed | False | Interesting | Merged | Divided | Correct |
|---|---|---|---|---|---|---|
| 5-10 | 3 | 7 | 12 | 0 | 2 | 29 |
| 10-30 | 0 | 1 | 5 | 0 | 1 | 25 |
| 30-60 | 0 | 0 | 0 | 0 | 2 | 31 |
| 60- | 0 | 0 | 0 | 0 | 1 | 71 |

(b) PlaceSense

| Duration (minutes) | Missed | False | Interesting | Merged | Divided | Correct |
|---|---|---|---|---|---|---|
| 5-10 | 25 | 3 | 2 | 0 | 0 | 9 |
| 10-30 | 5 | 0 | 1 | 0 | 0 | 21 |
| 30-60 | 1 | 0 | 0 | 1 | 3 | 28 |
| 60- | 1 | 0 | 0 | 0 | 6 | 65 |

Table 1. The distribution of discovered places by their duration

### URB 05.6 Future Directions

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