

Exploring the Use of Ambient WiFi Signals to Find Vacant Houses

Abstract. The population decline is creating the urgent need to address the problems of vacant houses in many local communities. However, it is often costly to keep track of the up-to-date information about the locations and the conditions of vacant houses. In this paper, we propose a technique to infer the locations of occupied houses based on ambient WiFi signals. Our technique collects georeferenced RSSI (Received Signal Strength Indicator) data based on opportunistic smartphone sensing, constructs hybrid networks of WiFi access points, and analyzes their geospatial patterns based on statistical modeling. We show that the technique can successfully infer occupied houses in a suburban residential community, and argue that it can substantially reduce the cost of field surveys to find vacant houses.

Keywords: Ambient WiFi signals, vacant houses, civic computing, localization

1 Introduction

The population decline is creating the urgent need to address the problems of vacant houses in many local communities. For example, the number of vacant houses is increasing rapidly in Japan and more than 30 percent of Japanese houses will likely be vacant in 2033 [1].

Vacant houses can: (1) decrease the quality of landscapes, (2) decrease safety and peace of mind related to crimes and disasters, (3) induce illegal dumping of garbage, (4) increase the risk of fire, (5) produce bad smell, (6) collapse when hit by strong winds, heavy snow, or earthquakes, (7) decrease vitality of community life, and so on. Thus, there exists an urgent need to find and manage vacant houses effectively based on the collaboration among local government, experts, and citizens. However, it is often difficult to keep track of up-to-date information about the locations and the conditions of vacant houses due to the high cost for conducting field surveys as well as the limitations to the current information processes and practices around local communities.

In this paper, we propose a technique to infer the locations of occupied houses based on ambient WiFi signals. Our technique collects georeferenced RSSI (Received Signal Strength Indicator) data based on opportunistic smartphone sensing, constructs hybrid networks of virtual and real WiFi access points, and analyzes their geospatial patterns based on statistical shape modeling. We show that the technique can successfully infer occupied houses in a suburban residential community, and argue that it can substantially reduce the cost of field surveys to find vacant houses.

2 Related Work

Chi, et al. [3] use location records of Baidu users to analyze spatial distribution of vacant housing areas. Their analysis focuses on the issue of “ghost cities” in China and shows that location records of a large number of mobile users can reveal areas in which most houses are vacant. As our focus is wormhole-like sporadic vacant houses rather than entirely vacant city blocks, we must perform a much finer-grained analysis aiming to infer the locations of individual vacant houses. Thus, we make inferences about individual houses based on the assumption that houses that contain active WiFi access points are likely occupied by people. More than half (53.6%) of the households in Japan use WiFi according to the government’s survey in 2014 and

Our approach relies on a technique to determine precise locations of WiFi access points based on ambient WiFi signals. Existing research projects on indoor positioning show that locations of mobile users can be determined by using ambient WiFi signals. For example, Bahl and Padmanabhan [2] proposed a fingerprint-based algorithm that can determine users’ indoor locations. Place Lab [6] uses estimated and know locations of WiFi access points and GSM cell phone towers in Seattle to compute users’ locations at a city scale. Koo and Cha [5] proposed a multidimensional scaling-based approach based on relative distances between pairs of WiFi access points to estimate locations of WiFi access points in indoor spaces. Other researchers proposed to use probabilistic techniques for localizing indoor access points [4] and road-side access points [7]. Unlike existing approaches, we propose to use a hybrid network model along with multidimensional scaling to support mixed uses of various devices with or without the location sensing capability, thereby making it easier to deploy in local communities with sparadically distributed vacant houses.

3 Localizing WiFi Access Points in Local Communities

This section describes a method to infer occupied houses in local communities in Japan based on ambient WiFi signals.

3.1 Data Collection

There are existing databases of ambient WiFi signals, including the ones owned by private organizations and the others collected by the wardriving community (e.g., Wigle.net [8].) Even when local communities cannot easily access existing databases, they can collect data by using off-the-shelf smartphones, tablets and notebook computers relatively easily. Our method requires that WiFi mac addresses and RSSI (Received Signal Strength Indicator) data be recorded along with GPS-based location information if available.

We believe that a small number of local volunteers may carry such devices in their pockets or bags while taking a walk. In this way, they could collect a sufficient amount of data relatively quickly (e.g., a few or several hours in total depending on the area of a local community.)

3.2 Constructing Hybrid Networks

We could infer the locations of WiFi access points (APs) by triangulation using the locations of mobile devices. However, this approach could not determine the locations of APs very accurately when individual access points are measured less than three times, or when mobile devices are not equipped with accurate GNSS receivers. We thus employ a localization method based on multidimensional scaling (MDS). Unlike existing MDS-based localization methods for indoor WiFi APs [5], which model collected measurements as a weighted graph of WiFi APs, we model collected data as an Euclidean distance-weighted hybrid network of real APs (RAPs) and virtual APs (VAPs). A RAP corresponds to an actual WiFi access point while a VAP represents a measurement point by a mobile device. A mobile device creates n VAPs when it measures WiFi signals n times.

To construct a hybrid network with RAPs and VAPs, we first create vertices with all RAPs and VAPs based on mac addresses and timestamps in the collected data. We next instantiate two types of edges, one connecting pairs of georeferenced VAPs (i.e., V^2 edges) and the other connecting pairs of a VAP and a RAP (i.e., VR edges).

For all VAPs with GPS location information, we simply compute their mutual distances on an Earth ellipsoid, and use them as weights for corresponding V^2 edges.

For each WiFi measurement point, we connect the corresponding VAP vertex with the RAP vertices that represent observed WiFi APs, thereby instantiating VR edges. Their weights are determined according to Euclidean distances d , which are estimated based on RSSI measurements $rssi$ using the following model.

$$d(rssi) = \frac{MAX_DIST}{L} * floor(10^{\frac{\log_{10}(L+1)*(MAX_RSSI-rssi)}{(MAX_RSSI-MIN_RSSI)}}) \quad (1)$$

This is based on an oft-adopted model of the relationship between the distance and RSSI, however, with the additional floor function to quantize the distances at L levels so as to reduce the effects of irrelevant small variations of radio signal strengths. MAX_RSSI and MIN_RSSI are the maximum and minimum RSSI values, respectively. MAX_DIST is the maximum distance. It is the distance at which the smallest level of RSSI would be observed. Preceding to the computation of $d(rssi)$, we can eliminate the effects of unreliable weak radio signals by removing RAPs with smaller RSSI than a predefined threshold value.

3.3 Computing Relative and Absolute AP Locations

We first compute shortest-path distances between all vertices in the resulting hybrid network \mathbf{G} to obtain a distance matrix \mathbf{D} . We next apply multidimensional scaling to distance matrix \mathbf{D} to produce relative positions of all vertices including RAPs and VAPs in a two dimensional space. We then scale, rotate and translate the positions of RAPs as follows to obtain their absolute geographical locations.

$$\mathbf{P}_a = s * \mathbf{P}_r \cdot \mathbf{R} + \mathbf{T} \quad (2)$$

\mathbf{P}_a represents absolute geographical locations of RAPs, which are obtained by scaling and rotating the relative positions of RAPs, denoted as \mathbf{P}_r , with the scaling factor s and rotation matrix \mathbf{R} , and translating it by adding \mathbf{T} . We can use procrustes analysis to match the relative positions of VAPs to the longitude and latitude values of the corresponding VAPs, there by deriving s , \mathbf{R} , and \mathbf{T} .

4 A Field Trial

This section presents our field trial in a suburban residential community. We have collected ambient WiFi signals, inferred occupied houses based on the estimated locations of WiFi access points, and compared the results with the ground truth provided by the local community members.

4.1 Data Collection

A researcher collected ambient WiFi signals in a local community by walking along all the streets in the community, which took about an hour. The signals have been recorded by using WiGLE WiFi app [8] on two Android smartphones (Nexus 5 and Xperia Z Ultra) in the researcher’s backpack pockets. Geolocations have been recorded when GPS signals were available. Consequently, we captured data about 962 WiFi access points (RAPs) and 610 virtual access points (VAPs).

4.2 Constructing Hybrid Networks

We have ignored weak WiFi signals with their RSSI values less than -87 dB, and computed the distances with $L = 3$, $MAX_RSSI = -57dB$, $MIN_RSSI = -87dB$, and $MAX_DIST = 27m$ according to equation (1). The resulting distances have been used as the weights of VR edges to construct a hybrid network by using the *igraph* package of *R*.

4.3 Computing AP Locations

We have computed shortest-path distances between all vertices of the hybrid network by using *distances* function of the *igraph* package, and applied the multidimensional scaling technique by using *cmdscale* function of the *stats* package of *R*. We next derive scaling factor s , rotation matrix \mathbf{R} and translation matrix \mathbf{T} by using the *procrustes* function of the *vegan* package of *R*. Not surprisingly, as shown in Figure 1, the matching was not perfect and the differences (i.e., arrows) are somewhat large for VAPs (i.e., circles) near the bottom left. Finally, we use s , \mathbf{R} , and \mathbf{T} to compute the geolocations of the WiFi access points.

4.4 Results

We simply judge that houses containing at least one WiFi access point are occupied. The yellow houses in Figure 2 have been judged as occupied in this manner. Figure 2 also shows actual vacant houses in purple. The white houses were not inferred as occupied but they are actually occupied (i.e., false negatives). It is not surprising that the number of false negatives is not small since nearly half households may not have WiFi access points. However, false negatives are not too problematic for our application scenarios. Without exploiting the proposed method, field workers need to check all the houses. The more yellow houses it finds, the less workload would be imposed on field workers.

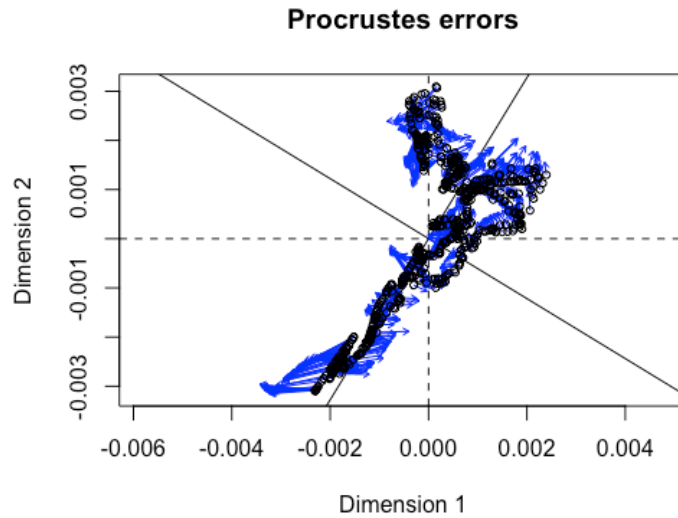


Fig. 1. Procrustes errors of our dataset

5 Conclusion and Future Work

We have proposed a technique to infer the locations of occupied houses based on ambient WiFi signals. Our technique collects georeferenced RSSI (Received Signal Strength Indicator) data based on opportunistic smartphone sensing, constructs hybrid networks of WiFi access points, and analyzes their geospatial patterns based on statistical modeling. We have shown that the technique can successfully infer occupied houses in a suburban residential community.

- argue that it can substantially reduce the cost of field surveys to find vacant houses.



Fig. 2. Comparative geovisualization. Yellow shows the houses inferred as occupied, and the purple shows actual vacant houses.

- Privacy issues
- more field test needed
- how to decrease false negatives

References

1. Nomura Research Institute (2016). News release, June 7, 2016. http://www.nri.com/Home/jp/news/2016/160607_1.aspx
2. Bahl, P., & Padmanabhan, V. N. (2000). RADAR: An In-building RF-based User Location and Tracking System. Proc. IEEE INFOCOM 2000. The 19th Annual Conference on Computer Communications, 2, 775–784. <http://doi.org/10.1109/INFCOM.2000.832252>
3. Chi, G., Liu, Y., & Wu, H. (2014). “Ghost Cities” Analysis Based on Positioning Data in China. ArXiv.
4. Ji, M., Kim, J., Cho, Y., Lee, Y., & Park, S. (2013). A novel Wi-Fi AP localization method using Monte Carlo path-loss model fitting simulation. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC, 3487–3491. 2013.6666752
5. Koo, J., & Cha, H. (2012). Unsupervised locating of WiFi access points using smartphones. IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, 42(6), 1341–1353.
6. LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., ... Schilit, B. (2005). Place Lab: Device Positioning Using Radio Beacons in the Wild. Pervasive Computing, 3468, 116–133.
7. Wu, D., Liu, Q., Zhang, Y., McCann, J., Regan, A., & Venkatasubramanian, N. (2014). CrowdWiFi: Efficient Crowdsensing of Roadside WiFi Networks. Proceedings of the 15th International Middleware Conference on - Middleware '14, 229–240.
8. Wigle.net. (2017) <https://wigle.net/> Accessed: January 3, 2017.