[accepted manuscript before copyediting]
Shin'ichi Konomi, Tomoyo Sasao, Simo Hosio, Kaoru Sezaki (2017). Exploring the Use of Ambient WiFi Signals to Find Vacant Houses. Proceedings of the 2017 European Conference on Ambient Intelligence (AmI 2017), Malga, Spain, April 26-28, 2017. Lecture Notes in Computer Science, Springer, Berlin/Heidelberg, 2017. The final publication is available at link.springer.com.

Exploring the Use of Ambient WiFi Signals to Find Vacant Houses

Shin'ichi Konomi[1], Tomoyo Sasao[1], Simo Hosio[2], and Kaoru Sezaki[1]

[1]Center for Spatial Information Science, The University of Tokyo
[2]Center for Ubiquitous Computing, The University of Oulu

Abstract. In many countries, the population is either declining or rapidly concentrating in big cities, which causes problems in the form of vacant houses in many local communities. It is often challenging to keep track of the locations and the conditions of vacant houses, and for example in Japan, costly manual field studies are employed to map the occupancy situation. In this paper, we propose a technique to infer the locations of occupied houses based on ambient WiFi signals. Our technique collects 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 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 as the number of potential houses to be inspected decreases.

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

1 Introduction

The population decline and movement to big population hubs is creating the urgent need to address the problems of vacant houses. A particularly challenging case can be found in Japan, where the number of vacant houses is increasing rapidly, and more than 30 percent of Japanese houses are projected to be vacant already in 2033 [6].

Vacant houses can be problematic, as they (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) are more prone to damage by strong winds, heavy snow, or earthquakes, if improperly managed, (7) decrease vitality of community life, and so on. In certain places, local officials lack efficient means of keeping track of house vacancy situation, and thus there exists a clear need to find vacant houses before their condition deteriorates and to find new use for them. This should effectively happen via collaboration between local government, urban planners, and citizens. In Japan, typical means to do this is conducting costly field surveys to verify occupancy status.

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 thus reduce the cost of field surveys to map vacant houses.

2 Related Work

Chi, et al. [2] 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 find 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. Currently, more than half (53.6%) of the households in Japan use WiFi according to the government's survey in 2014, and the expanding market of household IoT devices can cause a rapid increase of WiFi and other radio signals in residential communities.

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 [1] proposed a fingerprint-based algorithm that can determine users' indoor locations. Place Lab [5] uses estimated and known locations of WiFi access points and GSM cell phone towers in Seattle to compute users' locations at a city scale. Koo and Cha [4] 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 [3] and road-side access points [8]. 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 sporadically distributed vacant houses.

3 Localizing WiFi Access Points in Local Communities

There are existing databases of WiFi signals, including the ones owned by private organizations and the others collected by the wardriving community (e.g., Wigle.net [7].) WiFi data can be collected easily by using off-the-shelf smartphones, tablets and notebook computers. Our method requires that WiFi mac addresses and RSSI (Received Signal Strength Indicator) data be recorded along with GPS-based location information if available. Volunteers can carry such devices in their pockets or bags while taking a walk, thereby collecting a sufficient amount of data relatively quickly (e.g., a few or several hours in total depending on the area of a local community.)

3.1 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 GPS receivers. We thus employ a localization method based on statistical shape modeling. Unlike existing localization methods for indoor WiFi APs based on multidimensional scaling, which model collected measurements as a graph of WiFi APs [4], we model collected data as a hybrid network of real and virtual APs. A real AP (RAP) corresponds to an actual WiFi access point while a virtual AP (VAP) represents a measurement point by a mobile device. A mobile device creates n VAPs when it measures WiFi signals n times in a local community.

To construct a hybrid network with real and virtual APs, we first create vertices with all real and virtual APs based on mac addresses and timestamps. We next instantiate two types of edges, one connecting pairs of virtual APs, which we call V^2 edges, and the other connecting pairs of virtual APs and real APs, which we call VR edges. For all virtual APs that have location information, we simply compute their mutual distances on an Earth ellipsoid, and use them as weights for corresponding V^2 edges. The weights for VR edges are determined based on the RSSI values of WiFi APs (RAPs) as measured by mobile devices (VAPs). More precisely, we determine weights for VR edges according to Euclidean distance d, which is calculated as follows.

$$d(rssi) = \frac{MAX\_DIST}{L} * floor(10^{\frac{log10(L+1)*(MAX\_RSSI-rssi)}{(MAX\_RSSI-MIN\_RSSI)}}) \tag{1}$$

This is based on an oft-adopted model of the relationship between the distance and RSSI. The floor function quantizes the distances at L levels to mitigate the effect of unstable signals. 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.

3.2 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 in a two dimensional space. We then scale, rotate and translate the positions of real APs as follows to obtain their absolute geographical locations.

$$\mathbf{P_a} = s * \mathbf{P_r} \cdot \mathbf{R} + \mathbf{T} \tag{2}$$

$\mathbf{P_a}$ represents absolute geographical locations of real APs, which are obtained by scaling and rotating the relative positions of real APs, denoted as $\mathbf{P_r}$, with the scaling factor s and rotation matrix $\mathbf{R}$, and translating it by adding $\mathbf{T}$. Procrustes analysis is used to match the relative positions of virtual APs to the longitude and latitude values of the corresponding virtual APs, thereby deriving s, $\mathbf{R}$, and $\mathbf{T}$.

4 A Field Trial

This section presents our field trial in a suburban residential community. We collected ambient WiFi signals, inferred occupied houses based on the estimated locations of WiFi access points using the proposed method, and compared the results with the ground truth provided by the local community members. One researcher collected ambient WiFi signals in the community by walking along all the streets in the neighborhood (see map in Figure 1), which took approximately an hour. The signals were recorded using WiGLE WiFi app [7] on two Android smartphones (Nexus 5 and Xperia Z Ultra) in the researcher's backpack pockets. Location data were captured via GPS, when available. Consequently, we captured data about 962 WiFi APs and 610 virtual APs. This implies that there are quite a few houses having multiple APs as there are only 286 houses in this community.

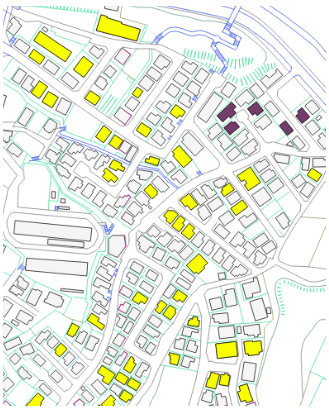

Fig. 1. Comparative geovisualization. Yellow shows the houses inferred as occupied, and the purple shows actual vacant houses.

4.1 Constructing Hybrid Networks

As weak WiFi signals may be caused by multipath and obstructions, we ignored weak WiFi signals with their RSSI values less than -87 dB, which approximately corresponds to the 80% cutoff point with the maximum RSSI of -57 dB and the minimum of -95 dB. We then 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.

4.2 Computing AP Locations

We computed shortest-path distances between all vertices and applied multidimensional scaling using *cmdscale* function of the *stats* package of R. We then derived 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, and as depicted in Figure 2, 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 used s, $\mathbf{R}$, and $\mathbf{T}$ to compute the locations of the WiFi APs.

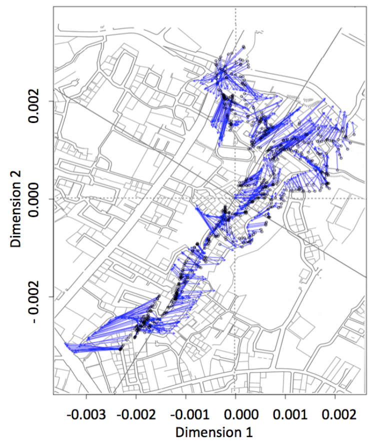

Fig. 2. Procrustes errors of our dataset

4.3 Detecting Occupied Houses

We establish that houses containing at least one WiFi access point are actively occupied. The yellow houses in Figure 1 have been judged as occupied in this manner. Figure 1 also shows actual vacant houses in purple. The white houses were not inferred as occupied but they are actually occupied (i.e., false negatives). Of the 286 houses in the community, 278 are occupied and 8 are vacant. Our method detected 57 (20%) occupied houses without falsely detecting vacant houses as occupied houses (i.e., no false positives), suggesting to reduce the workload of field surveys accordingly. The number of detected houses is fairly lower than the number of detected APs, and improving the accuracy of AP localization could further reduce the workload for field surveys.

5 Conclusion and Future Work

We have proposed, implemented and verified 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.

While the detection accuracy is far from perfect (cf. false negatives in Figure 1), the amount of correctly detected occupied houses is not small and the method did not produce false positives in our field trial. Potential explanations for not detecting a house occupied can be too weak signal strengths or no WiFi access point installed in those buildings (cable/mobile or no connection). Without exploiting the proposed method, field workers need to check all the houses. Using Figure 1 as an example, the more yellow houses the method finds, the less workload is imposed on field workers to manually visit all houses multiple times.

In some cases, individuals may be wary of the SSIDs of their WiFi APs being used for finding their community's vacant houses. Encrypted hash IDs can be used to minimize such concerns, and they can also choose to disable the beaconing feature of their WiFi APs.

Naturally, more tests are warranted to further verify and optimize the method. The next step of our work seeks to find volunteers from local communities who are interested in urban development and public safety in their neighborhood to carry devices provided by us for logging purposes.

References

1. Bahl, P., & Padmanabhan, V. N. (2000). RADAR: An In-building RF-based User Location and Tracking System. Proc. IEEE INFOCOM 2000, 775–784.
2. Chi, G., Liu, Y., & Wu, H. (2014). "Ghost Cities" Analysis Based on Positioning Data in China. arXiv:1510.08505 [cs.SI]
3. 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. Proc. IEEE PIMRC, 3487–3491.
4. Koo, J., & Cha, H. (2012). Unsupervised locating of WiFi access points using smartphones. IEEE Trans. Systems, Man and Cybernetics Part C: Applications and Reviews, 42(6), 1341–1353.
5. LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello, G., & Schilit, B. (2005). Place Lab: Device Positioning Using Radio Beacons in the Wild. Proc. Int. Conf. Pervasive Computing, 116–133.
6. Nomura Research Institute (2016). News release, June 7, 2016. `http://www.nri.com/Home/jp/news/2016/160607_1.aspx` (in Japanese)
7. Wigle.net. (2017) `https://wigle.net/` Accessed: January 3, 2017.
8. Wu, D., Liu, Q., Zhang, Y., McCann, J., Regan, A., & Venkatasubramanian, N. (2014). CrowdWiFi: CrowdWiFi: Efficient Crowdsensing of Roadside WiFi Networks. Proc. Int. Middleware Conf., 229–240.