International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 12 (2015) pp. 31503-31512

http://www.ripublication.com

Intense User Tracking Incorporating Mobility Mapping Technique In Shotgun Clustering For Wireless Signal

Franklin George Jobin[1] and Dr. M. Rajaram[2]

[1]*Assistant Professor, Dept. of Electronics and Communication Engg., Lord Jegannath College of Engg & Tech, Kanyakumai District*
franklingj@gmail.com
[2]*Professor, Dept of Electrical Engineeing, Anna University, Chennai,*

Abstract

Mobility mapping in ubiquitous computing environments plays an important role in identifying human activities. In an indoor environment, the geographical position of mobile device such as smart phones not easy to be estimated without the use of specific infrastructures. This paper is based on a robust adaptive mobility map construction scheme for large scale which does not require any offline fingerprinting efforts. Using a large number of users, the method works by collecting a large data-set of received Wi-Fi signal strength taken during their normal daily activities. The system design is based on the method by which a mobility map is constructed using randomly selected and un-labelled sequences of Wi-Fi received signal strength (RSS).The obtained signal sequences are treated as shotgun reads. The output in a DNA sequencing is a set of linear sequences of genomes, the output in our system is a directed and weighted graph. This paper focuses on the construction of this graph.

Index Terms— Mobility mapping, Wi-Fi, Shotgun reads, Spectral clustering.

I. Introduction

The widespread availability of wireless networks (Wi-Fi) has created an increased interest in harnessing them for other purposes, such as localizing mobile devices. While outdoor positioning has been well received by the public, its indoor counterpart has been mostly limited to private use due to its higher costs and complexity for setting up the proper environment. In this paper, we use local Wi-Fi network to localize a mobile user in an indoor environment. Wi-Fi (or 802.11 networking) works on the basic principle that data packets are sent using radio waves. These radio waves can be received by any compatible receiver placed in a pc, mobile phone, tablet pc or any other circuit. Through Wi-Fi, one may be able to track objects or people in real time, while adapting to changes in both the environment, and the Wi-Fi network, in a reliable manner.The data distribution may vary based on changes in temperature and humidity, as well as the position of moving obstacles, such as people walking throughout the building. This uncertainty makes it difficult to generate accurate estimates of signal strength measurements. The Received Signal Strength (RSS) values measured by most radio transceivers can be used to estimate the distance between nodes and implement range based localization schemes. Received signal strength indicator (RSSI) is a measurement of the power present in a received radio signal. RSSI is an indication of the power level being received by the antenna. Therefore, higher the RSSI number (or less negative in some devices), stronger will be the signal. The Received Signal Strength (RSS) values measured by most radio transceivers can be used to estimate the distance between nodes and implement range based localization schemes. These schemes are popular because no additional hardware is required on the nodes to localize.The transmitter of the signals is known as an Access Point. In computing networking, a wireless access point (AP) is a device that allows wireless devices to connect to a wired network using Wi-Fi, or related standards. The AP usually connects to a router (via a wired network) if it's a standalone device, or is part of a router itself.

In the past decades, the LBS market has grown significantly and is expected to reach a size of $12.7 billion in 2014. The location based services provides the information such as the user's current location, knowledge about where the user is expected to go etc. In an outdoor environment, GPS provide precise locations of mobile devices with world-wide coverage, whereas in indoors, Wi-Fi and cellular signals are used.

One successful approach for indoor user tracking is a Wi-Fi based fingerprint [2], [3]. The technique builds a radio map by measuring the Wi-Fi signals at each reachable calibration point, a priory and tracks a mobile device based on run-time observation of the Wi-Fi signals. But it involves some cumbersome work of fingerprint landmark calibration. Previous works in mobility mapping includes assisted GPS (A-GPS), Cell-ID, Bluetooth and Wi-Fi fingerprinting technologies [4]-[6]. In contrast to the previous tracking systems which are based on GPS measurements and geometric clustering, the adaptive mobility mapping is designed to adaptively construct a mobility map of environment using randomly selected and un-labelled sequences of Wi-Fi received signal strength (RSS).The RSS sample vectors are recorded by many individuals moving around the environment as they conduct their daily activities. Each signal typically covers only a small part of the coverage area and the idea is to piece these sequences together by treating them like shotgun reads. Shotgun read is a term used in DNA sequencing [7].

The output in DNA sequencing is a set of linear sequences of genomes. The output in the present system is a directed and weighted graph. This graph is called the mobility map. It abstracts the environment under coverage into a finite set of un-labelled location point (LP's). After the map is created, labeling of location can be done [8]. Although the user can label the place on the map, simple automatic methods can be used to construct semantically meaningful place names from the observed data.

II. SYSTEM DESCRIPTION

The objective of the work is to collect Wi-Fi RSS sample vectors to create a mobility map. This can be said as

$$G= (V_C, Ev)$$

Where, V_C represents an Location Point.

An LP is formed when there is a large similarity between RSS sample vectors. The weight of each edge $V \in E_v$represents the transition matrix between two neighboring LP's. The location of a user can be determined by fitting the real time RSS trajectory of user into this mobility map during mobility tracking.To carry out hand-held drive-tests with our Android phone an RF signal tracker is used[9]. With this, we can monitor the RF and Wi-Fi signal strength for the device as well as serving cell locations and hotspots, describe a cell site's zone of coverage, identify changes in technology and handover points and save and playback that data. Many of the phone stats in the app can be displayed on the phone already (go to Settings-> About-> Status to see them). The advantage of the app is we can map, record, and analyze and share that data in a meaningful way.

Wi-Fi signal measuring tool is shown below:

Fig 1: Signal Analyzer

The process for mobility map construction consists of 3 steps: Shotgun read Collection, Spectral Clustering on shotgun reads and estimation of first order Markov transition probabilities. The block diagram of the mobility map construction is shown in figure2.

A. Shotgun Read Collection

In this scheme, a subset of users is equipped with an Android phone. It will start recording RSS measurements at a regular time interval of 1 second by default, while the user is moving. The starting and ending of recording is triggered manually. A V1.4 pedometer is equipped with the phone in order to detect the movement of the user. A pedometer is a motion sensor which is an application of the Android phone.

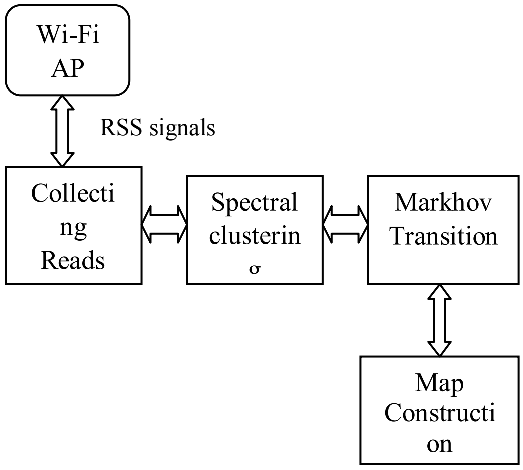

Fig 2: Mobility Map Construction scheme

Each measurement obtained as sample vectors contains the measured RSS's from a set of Wi-Fi AP's. Any sequence i.e.) too long can be fragmented in to shorter sequences of 600 samples to limit the computation time. The clustering and mapping procedures can be done, once a sufficient number of shotgun reads have been collected [11].

Assume N raw shotgun reads have been collected. The l^{th} shotgun read is represented as $R^l = \{\mu_1^{\,l} ... \mu^l_{N^l}\}$ where $\mu_i^{\,l}$ is the collected RSS sample vector and N^l is the number of sample vectors in the read. Also, the total number of Wi-Fi AP's in the coverage is assumed to be M. Each raw RSS sample is considered to be M-dimensional vector. Clustering based on RSS value alone leads to unstable results due to random variations in RSS measurements.

B. Spectral clustering

In recent years, spectral clustering has become one of the most popular modern clustering algorithms. It is simple to implement, can be solved efficiently by standard linear algebra software, and very often outperforms traditional clustering algorithms such as the k-means algorithm. Clustering is one of the most widely used techniques for exploratory data analysis, with applications ranging from statistics, computer science, and biology to social sciences or psychology. In virtually every scientific field dealing with empirical data, people attempted to get a first impression on their data by trying to identify groups of 'similar behavior' in their data.

Given a set of data points X_1, ….$X_n$and some notion of similarity $S_{ij}\geq 0$ between all pairs of data points X_i and X_j, the intuitive goal of clustering is to divide the data points into several groups such that the points in the same group are similar and points in different groups are dissimilar to each other, if we do not have more information than similarities between data points [12]. If we do not have more information than similarities between data points, a nice way of representing the data is in the form of the similarity graph G= (V, E). Each vertex V_i in this graph represents a data point x_i. Two vertices are connected if the similarity S_{ij}between the corresponding data point's X_i and $X_j$is positive or larger than a certain threshold, and the edge is weighted by s_{ij}. The problem of clustering can now be reformulated using the similarity graph: we want to find a partition of the graph such that the edges between different groups have very low weights (which means that points in different clusters are dissimilar from each other) and the edges within a group have high weights (which means that points within the same cluster are similar to each other).

One of the important methods of vector quantization is K-means clustering. It is originally developed from signal processing. The aim of this method is to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. As a result of this, the data space is partitioned into Voronoi cells. It is similar to expectation-maximization algorithm[13]. Both of the algorithms employs iterative- refinement approach to model the data. K-means algorithm also known as Lloyd's algorithm uses an iterative refinement approach. Often it works by assigning objects to the nearest cluster by distance. Due to its slight inaccuracy, it assigns by least sum of squares. The algorithm may stop from converging using a different distance function other than Euclidean distance. It is correct that the smallest Euclidean distance yields the smallest squared Euclidean distance and thus also yields the smallest sum of squares. The modifications of k-means include spherical k-means and k-mediums, which use other distance measures.

K-means algorithm works by a two-phase algorithm which minimizes the sum of point-to-centroid distances, summed over all k-clusters. In the initial phase, it uses batch updates where each iteration corresponds to reassigning points to their nearest cluster centroid and also the recalculation of cluster centroids is done. This is often used in small data sets. The batch phase is fast, but potentially only approximates a solution as a starting point for the second phase [14].

The second phase uses online updates, where points are individually reassigned which leads to reduction of sum of distances, and also cluster centroids are recomputed after each reassignment. Each iteration consists of one pass through all the points. This phase will converge to a local minimum. The global minimum problem can be solved by an exhaustive choice of starting points.In basic mean shift clustering algorithms, the data points are maintained the same size as the input data set. The data points are initially copied from the input set. Then the mean of those data points are calculated and then those points are replaced by the mean within a given distance of that point. This updated set is restricted to k points which are lesser than the number of input data points.

Two steps are involved in spectral clustering-

- Use of low dimensional space to map each RSS sample vectors.
- Use of K-means clustering in low dimensional space to congregate similar vectors into clusters.

Step1: Objective Function minimization

We construct a weighted graph $\psi^l = (V^l, E^l)$ for each shotgun read. The similarity between two nearby vertices is represented as the weight of an edge. To represent the RSS data in a low dimensional space, map the weighted graph ψ^l onto a line ensuring that any vertices which are nearby will have corresponding mapping points.

Within R^l similarity between the i-th and j-th sample vectors can be calculated by

$$s_{ij}^l = e^{-\left[\alpha_R F_R\left(\mu_i^l,\mu_j^l\right)+\ \alpha_T F_T\left(T_i^l,\ T_j^l\right)\right]}$$

Where,

α_R and α_T are tunable weighting coefficients for the RSS and for timestamps.

F_R $(\mu_i^l,\ \mu_j^l)$ and F_T $(T_i^l,\ T_j^l)$ are the features of RSS and of timestamp respectively. They are calculated as:

$$F_R\left(\mu_i^l, \mu_j^l\right) = \frac{\left\|\mu_i^l,\mu_j^l\right\|}{\max_{s,t}\left\|\mu_i^l,\mu_j^l\right\|} \quad (1)$$

$$F_T\left(T_i^l, T_j^l\right) = \frac{\left|T_i^l,T_j^l\right|}{(N^l-1)\delta} \quad (2)$$

The features are symmetric and hence so are the similarities can be said as

$$S_{ij}^l = S_{ij}^l \ where \left(S_{ij}^l \ \geq 0\right)$$

Step 2:

We run the K-means clustering on each read $\acute{R}^l = \{C_1^l \ldots C_l^{\varphi l}\}$,

Where, C_r^l and $\varphi^l$are the r[th] RSS cluster and the number of clusters in R^l. Since the clustering for each read is based on the similarity of both RSS and relative timestamp values, the clustering outcome is more deterministic and stable than using RSS values alone as random variations in RSS measurements are filtered out.

C. First Order Markhov Transition Probabilities

The term "Markov chain" refers to the sequence (or chain) of states in which a process moves through. Conventionally a Markov chain is defined for a discrete set of times (i.e., a discrete-time Markov chain). The state transmutations in the Markhov chain are described as the transitions and the probabilities cognate to the transmutations in the state of the system are referred as transition probabilities. In Markhov process, a transition matrix describes the probabilities of particular transitions and an initial state is characterized across the state space. Here, all the possible states and transitions are included in the definition of the process, and hence there is always a next state. Thus the process does not terminate.

In a discrete-time arbitrary process, system will be in a certain state at each step and the state will be transmuting arbitrarily between steps. The steps are considered as the integers or natural numbers and the mapping of these numbers to states is considered as the desultory process. The conditional probability at next step depends only on the current state and not on the precedent states. It is generally infeasible to soothsay the state of a Markhov chain at a given point in the future since the system changes arbitrarily.

Predicated on mean relative timestamp of the LP's order the LP's composed by reads in chronological order. The transition probabilities among LP's are calculated by the first order Markhov model. The transition probability matrix is engendered utilizing the below formula

$$\varphi_{uv} = \frac{Number\ of\ transitions\ from\ L_u\ to\ L_v}{Number\ of\ transitions\ from\ L_u} \quad (3)$$

Where, φ_{uv} denotes the transition probability from L_u to L_v.

D. Naming Policy

By labeling the location points, the user could identify and recognize the current place. Labeling can be done manually by the user or by automatic extraction of meaningful labels. In this system, although the user can manually assign a name to the place after the map is built, it supports automatic extraction of semantically meaningful name of the place. Initially, we look for the strongest access points from the current location. The access point, which emits a strong radio signal, is simply assumed as being installed in the current location. We can assume unlabelled places with the name of the nearest access point where the access point has strong signal strength over a threshold -56dBm in our experiment.

With this method, we extracted meaningful words from our data set "restaurant", "school gate" and so on. Many access points were operating in the areas, hence sophisticated filters are required to prevent assigning meaningless names.

III. PERFORMANCE ANALISIS

We obtain the results for RSS data on two different paths and it is recorded by the suitable application at a scanning rate of 1 sample/second. The similarity between the shotguns read collection on the LPs is shown by a cluster gram as shown in figure 5 and the corresponding connecting graph of vectors is shown in figure 6.Then the clustering of sample vectors is done using K-means clustering algorithm as shown in figure 7.

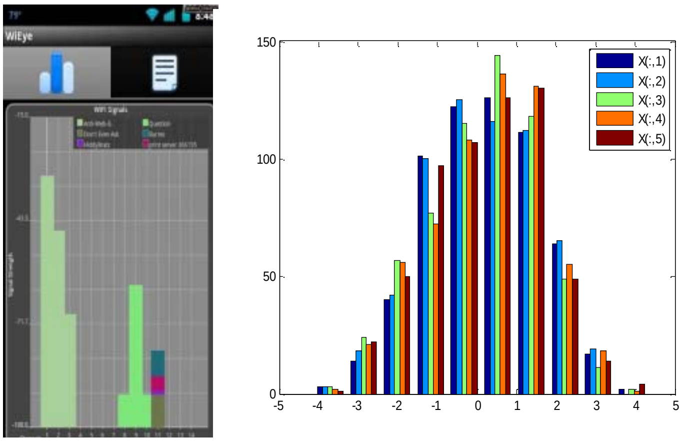

Fig 4: Shotgun read collection

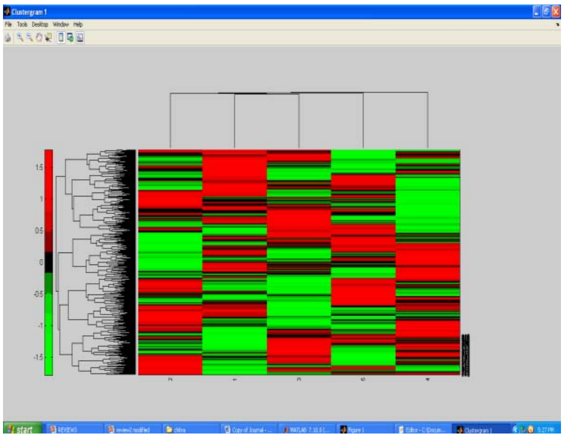

Fig 5: Cluster gram-Similarity between Vectors

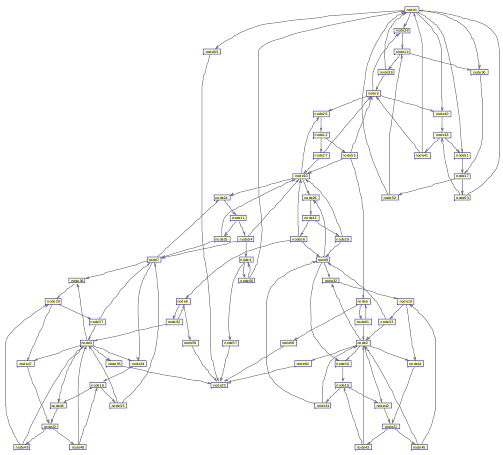

Fig 6: Graph construction

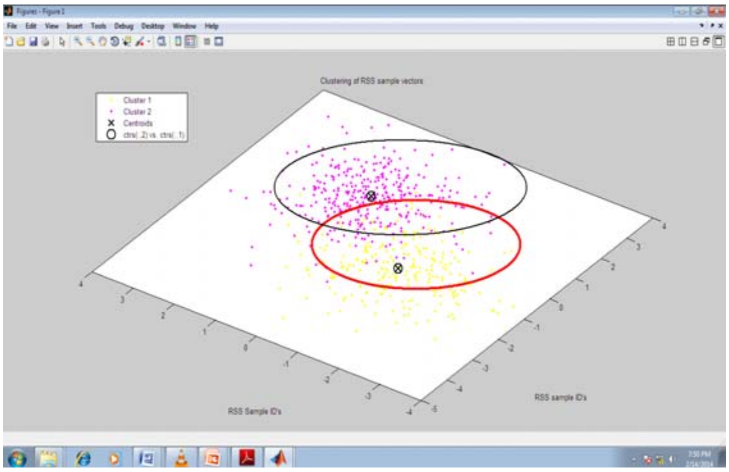

Fig 7: Clustering between sample vectors

IV. CONCLUSION

We described an approach to use ubiquitously and randomly recorded Wi-Fi RSS data to construct an adaptive mobility map for mobility tracking. With the use of unlabeled Wi-Fi shotgun reads, our approach can avoid the cumbersome work of collecting labeled data for fingerprinting. The mobility map created is a graph consisting of LP's, which are formed automatically through the clustering process. The mobility map will provide guidance to people tracking during the online phase.Main obstacle to designing an algorithm for smart phones is the limitations of the phones themselves. One major limitation for phones is battery life. This is in fact the single largest obstacle to designing more complex algorithms, we cannot insist that the user constantly be scanning Wi-Fi networks in order to localize, as that would be an enormous strain on battery consumption. One aspect of phones that we could not account for at all was dealing with the wide variety of phones and corresponding hardware, putting limits on how much we could trust the data given to us. Finally, the computing and memory requirements have to be taken into consideration. While it is true that smart phones are highly capable machines, the users themselves don't want an application that takes gigabytes of data just to improve accuracy in localization. Indoor localization on smart phone is critical to enable novel features for location based applications. However, existing approaches have yet to prove that they can satisfy what is desired in many business scenarios. Due to the prevalence of Wi-Fi infrastructure, it is imperative to study the accuracy that Wi-Fi localization can practically achieve on smart phones.

REFERENCES

1) D.Ashbrook and T.Starner, "Using GPS to learn significant locations and predict movement across multiple users", Personnel and Ubiquitous Computing, vol.7, no.5, pp.275-286, Oct.2003.

2) M.Zhou, P.Krishnamurthy, Y.B.Xu and L.Ma, "Physical distance vs. signal distance: an analysis towards better location fingerprinting, " in Proc.2011 IEEE HPCC, pp.977-982.
3) A.K.S.Wong, K.T.Woo, A.T.L.Lee, X.M.Xiao and V.W.H.Luk."An AGPS-based elderly tracking system", in Proc.2009 IEEE ICuFN, pp.100-105.
4) M.Anisetti, C.A.Ardagna, V.Bellandi, E.Damianiu and S.Reale, "Map based location and tracking in multipath outdoor mobile networks, "IEEE Trans.Wireless Commun., vol.10, no.3, pp.814-824.
5) M.Y.Lee and D.S Han, "Voronnoi tessellation based interpolation method for Wi-Fi radio map construction, " IEEE Commun.Lett., vol.16, no.3, pp.404-407.
6) X.M.Xiao, A.K.S.Wong, K.T.Woo and R.S.Cheng, "Energy efficient elderly tracking algorithm", in Proc.2011 IEEE ICC, pp.1-5.
7) S.Anderson, "Shotgun DNA sequencing using cloned DNase I-generated fragments, " Nucleic acids research, vol.9, no.13, pp.3015-3027, July 1981.
8) Paolo Barsocchi, Francesco Furfari, Paolo Nepa and Francesco Potort, "RSSI localisation with sensors placed on the user", 2010 International Conference on indoor positioning and indoor navigation (IPIN), 15-17 September 2010, Zurich, Switzerland.
9) Eduardo Navarro, Benjamin Peuker, Michael Quan "Wi-Fi Localization Using RSSI Fingerprinting", California Polytechnic State University
10) ZakariaBouguettaya, "Exploration Of Relative Localization Using Wireless Fidelity Technologies", COMP3740: Project work in computing.
11) David A.Case, "Deploying WLAN in a Hospital Setting: Understanding the Issues." Deploying WLAN in a Hospital Setting: Understanding the Issues. CE Magazine. June 2006.
12) Solomon Chan, GunhoSohn, "Indoor Localization Using Wi-Fi Based Fingerprinting and Trilateration Techniques for LBS Applications", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXVIII-4/C26.
13) Gianni Giorgetti, Sandeep K. S. Gupta, Gianfranco Manes, "Localization Using Signal Strength:To Range Or Not To Range?", MELT'08, September 19, 2008, San Francisco, California, USA.
14) Anton Hansson, LinusTufvesson, "Using Sensor Equipped Smartphones to Localize Wi-Fi Access Points", Department of Automatic Control, Lund University, May 2011
15) David Törnqvist, "Estimation and Detection with Applications to Navigation", Linköping studies inscienceand technology dissertationsno.1216, LinköpingUniversity, November2008