Wireless Indoor Location Estimation Based on Neural Network RSS Signature Recognition (LENSR)

3rd IEEE Conference on Industrial Electronics and Applications

Kurt Derr
Milos Manic

June 2008
Wireless Indoor Location Estimation Based on Neural Network RSS Signature Recognition (LENSR)

Kurt Derr  
Idaho National Laboratory  
2525 Freemont Avenue  
Idaho Falls, ID 83415, USA  
derr584@uidaho.edu  
Kurt.Derr@inl.gov

Milos Manic  
Department of Computer Science  
University of Idaho at Idaho Falls  
1776 Science Center Dr., Ste.306  
Idaho Falls, ID 83402, USA  
misko@uidaho.edu

Abstract—Location Based Services (LBS), context aware applications, and people and object tracking depend on the ability to locate mobile devices, also known as localization, in the wireless landscape. Localization enables a diverse set of applications that include, but are not limited to, vehicle guidance in an industrial environment, security monitoring, self-guided tours, personalized communications services, resource tracking, mobile commerce services, guiding emergency workers during fire emergencies, habitat monitoring, environmental surveillance, and receiving alerts.

This paper presents a new neural network approach (LENSR) based on a competitive topological Counter Propagation Network (CPN) with k-nearest neighborhood vector mapping, for indoor location estimation based on received signal strength. The advantage of this approach is both speed and accuracy. The tested accuracy of the algorithm was 90.6% within 1 meter and 96.4% within 1.5 meters. Several approaches for location estimation using WLAN technology were reviewed for comparison of results.

Keywords—RSS, localization, neural network, CPN, k-nearest neighbor, signature recognition, GPS.

I. INTRODUCTION

Both indoor and outdoor location estimation is a significant problem posing serious technical challenges. Location estimation, also known as localization, concerns the positioning of mobile devices in some physical space. Although location estimation represents an active area of research, proposed solutions are generally cost prohibitive, inaccurate, or infeasible due to practical issues.

GPS and wireless technologies are useful for determining location. Although GPS supports LBS, the number of mobile devices using GPS technology available in the marketplace today is limited. This is due to high cost, power requirements, and the inability to work in certain environments, such as indoors, underground, and in city canyons (an area of a city with narrow streets and high buildings) [1, 2].

This paper focuses on indoor location estimation from radio signal strength (RSS) values received by a mobile device with wireless (WiFi) capabilities as the device moves around an area of a building; e.g., a smart phone, personal digital assistant, equipment or packages with WiFi sensors, or a robot with WiFi capabilities. The ubiquity and low cost of 802.11 technology makes localization based on wireless local area network (WLAN) technology a viable alternative to GPS, enhancing the value of the wireless network.

The building has wireless access points (APs) acting as anchors deployed at various locations. First, a radio grid map is obtained offline. This way attenuation and reflection of signals in an urban environment is recorded as it is, resulting in both computational time savings and precision of recorded signals. The mobile device then estimates its position algorithmically using RSS values received from access points and the grid map. This is entirely a client-based system in that the mobile device does not send packets to a server in order to determine the location of the device. WiFi coverage of the complete area of interest is necessary for accurate location estimation.

In this paper we present a new Counter Propagation neural Network (CPN) with k-Nearest Neighbor (k-NN) algorithm for location estimation using WiFi received signal strength. To the best of our knowledge this algorithmic approach has not been previously used for location estimation. The paper is organized as follows. Section 2 discusses related work, section 3 reviews the principles of mobile device location estimation, section 4 describes the proposed CPN with k-NN algorithm, section 5 presents test results computed in a MATLAB environment, and section 6 states our conclusions.

II. RELATED WORK

Wireless localization schemes are generally categorized as deterministic or probabilistic [3]. The deterministic techniques are range or proximity based. The range based approach uses the characteristics of the channel, such as Received Signal Strength (RSS), to find the distance from a mobile device to wireless access points. Alternatively RSS fingerprinting techniques may used to locate a mobile device in a building. Neural networks, specifically a generalized regression neural network, have been used as the pattern matching algorithm in geo-location systems [4].

The probabilistic technique [5,6,7,8,9] constructs a conditional probability distribution over some area of interest to determine the likelihood of a mobile device being at some position at a specific point in time. Probabilistic techniques are computationally more expensive than deterministic techniques but provide a higher degree of accuracy (90% within 2 meters [10]).

Some examples of a wireless localization systems and
techniques follow. These systems and techniques typically have an offline training phase and an online location determination phase [11]. Some location estimation techniques build a radio map in the offline phase that represents the RSS values to each reachable AP from every location in the area of interest. A location estimation algorithm run on a mobile device requires a copy of the radio map. The advantage of running the algorithm on the mobile device is the preservation of user privacy and improved scalability.

RADAR [12, 13] uses RSS measurements obtained from multiple locations to triangulate a user’s position in an area. An experimental radio map is built offline by taking measurements in all possible grid locations of the area of interest. The system performs both location estimation and user tracking to within 2 to 3 meters of the actual location.

The Clustering and Decision Tree-based [14] method (CADET) selects the set \( S \) of wireless access points that give the best performance for each location of a wireless area in the offline phase. The grid space is then partitioned into clusters and a decision tree is built for each cluster. In the online mode the RSS values from selected access points are used to determine which cluster and coarse location is associated with the device. Next, the decision tree for the identified cluster is evaluated resulting in a specific grid location. The best accuracy of CADET is 83.4% within 1.5 meters.

The Joint Clustering (JC) [10] technique uses clustering of location maps and probability distributions. A cluster, which represents a set of locations that shares a common set of access points, is calculated offline, as well as the joint probability distributions of the signal strength of different access points. During the online location estimation phase RSS values are acquired from some set of APs, which are used to determine the cluster to search for the probable location. The radio map and Baye’s theorem are used to determine the most probable user location within a cluster. The accuracy of this system is 90% to within 7 feet.

Uncertainty in RSS signal measurements can be modeled as fuzzy sets. [15] divides the area of interest into zones. A radio map is developed offline and is used to train the fuzzy inference system. There are six fuzzy sets for RSS: Excellent, Very Good, Good, Low, Very Low, and None. The degree of membership of a mobile device to a specific area is used to determine the location estimate, providing an accuracy of near 90%.

A generalized regression neural network (GRNN) is used for a pattern matching algorithm in [4]. A measured RSS value for each AP and the corresponding grid location in the radio map are inputs to the neural network during the training phase. The GRNN has one hidden layer, and an output layer corresponding to two neurons representing the \( x \) and \( y \) locations in the radio map. During the online phase a set of RSS values are input to the GRNN and the output is the estimated user’s location in \( x \) and \( y \) grid map coordinates. The maximum error between estimated and true positions for the test data was 43.2 meters. The location accuracy for the test data is 45% to within 5 meters.

III. MOBILE DEVICE LOCATION ESTIMATION

The approach to mobile device location estimation presented in this paper is based on a comparison of RSS signal vectors recorded by a mobile device and RSS vectors from a radio grid map. The radio grid map can be created offline in two ways. One way is either to have a person manually collect and record RSS signal strength values for each grid location, or to do this task automatically with a robot. With this approach, both precision (actual signal is recorded) and computational time savings are achieved (no analytic determination of attenuation and reflection is needed). Another way is the creation of a theoretical propagation model representing the RSS signal levels that are calculated for every location of the radio map based on propagation equations. For the sake of simplicity, the latter approach is taken in this paper. The effectiveness of the presented algorithm is the same regardless of the way in which the radio grid is created. The advantage of a recorded map is that complex, analytic modeling [16] of signal attenuation and reflection in an indoor environment can be effectively avoided, resulting in more correct, actual radio map. The use of a theoretical model allows the algorithm to be deployed in a new environment without having to physically acquire signal strength readings for building a radio map. Alternatively, the radio map may be constructed by 1) manually recording RSS readings from access points in each designated grid location in the area of interest, or 2) using an autonomous vehicle/robot to collect signal strength information. The CPN with k-NN AP approach to location estimation is valid regardless of how the radio map is built. Locations and WiFi transmitters and receivers power are all that is needed to build the radio map. The APs are expected to be homogeneous; i.e., same transmitter power. Cable and connector losses are omitted to simplify the model. The theoretical model is based on the following equations for received signal strength \( R_s \) and path loss \( L_p \): [17]

\[
R_s = T_x - L_p
\]

where \( R_s \) is received signal strength value, \( T_x \) is AP power in dB, and \( L_p \) is path loss, and

\[
L_p = 33dB + N \cdot \log_{10}(D) + 20 \cdot \log_{10}(f)
\]

where \( f \) is a frequency in gigahertz. \( N \) is a path loss exponent, and \( D \) is a distance in meters.

Based on equations (1) and (2), for factory environment path loss and access point power (\( N = 5.5, T_x = +20dBm \)), and AP frequency \( f = 2.4 \), received signal strength \( R_s \) is:

\[
R_s = 20 - 33 - 5.5 \cdot \log_{10}(D) - 20 \cdot \log_{10}(2.4)
\]

Equation (3) shows how the signal strength decreases exponentially with distance from a WiFi access point regardless of transmission power and antenna gain.

The theoretic model is developed by using equation (3) for calculating the received signal strength for each access point in every location of the radio map. Some test values are calculated by adding noise to the theoretical model. The theoretical map of RSS values for a 20 meter square area \((n \times n\) grid...
cells) with 4 WiFi access points is shown in Figure 1 for each individual access point.

IV. COUNTER PROPAGATION NETWORK (CPN) WITH K-NEAREST NEIGHBOR

A feed-forward CPN network generally consists of two layers. This modular neural network [18, 19] uses other neural networks as building block components and provides capabilities that a single monolithic neural network is not capable of providing [20].

The presented approach (LENSR) is based on a symbiotic algorithm of a CPN and k-nearest neighbor approach with multiple winning neurons in the first layer (see Figure 2). CPN networks were introduced by Hecht-Nielsen in late 1980s as two layer, vector-to-vector hetero-associative memory networks [21, 22, 23]. This network was selected for a number of advantages. CPNs are very fast and easy to use: training is performed by design, hence training is much faster than a typical 2-layer network; there is no feedback or delay during recall mode.

In addition to being simple to use and control, these networks can recognize patterns that have not been seen by the network before. In the case of no perfect matching, the network will find the closest match possible and provide the output accordingly. In the case of multiple closest matches, the network will provide a composite output, proportional to the degree of similarity between unseen and previously stored patterns. Input vector recognition is done simultaneously against all previously learned vector patterns. The advantageous speed of these networks is a consequence of the parallelism, inherent to topologic organized CPNs.

The area of interest is divided into a grid of \( n \times m \) cells as shown in Figure 1. Each access point, plotted as a hump in the diagram, produces an RF signal detectable by other WiFi devices within radio range. Each RSS signature is a \( k \)-dimensional vector with signals measured from each of \( k \) access points. These Signal Space Vectors (SSV) are augmented with the recorded spatial coordinates for each specific cell \( (x, y) \):

\[
SSV_i = (AP_1, AP_2, ..., AP_k, x_i, y_i), \quad i = 1, 2, ..., n \times m
\]
where \( p \) is a predicted location and \( a \) is the actual location.

The LENS algorithm presented in this paper can recognize an unseen RSS signature by matching it against the table of RSS vectors measured previously. If an exact match is found, the corresponding location or \((x, y)\) pair is returned. If the exact match is not found, a number of similar patterns (patterns within certain distance threshold) are returned. The minimum distance threshold is defined as a cell size (the test examples section presents results with various threshold size and various distance metrics). Further, the inverse normalized proportion of distances is used to produce a cumulative location of unseen pattern in following manner:

1. Calculate distance between radio map RSS signal vectors and test vector. This is computing a distance in the signal space to other samples based on a distance metric.

If distance is within a previously defined threshold, record distance and vector coordinates. The result of this comparison against all RSS vectors from a radio map is:

2. Calculate weights, \( w_i \), (proportional normalized distances) of these nearest neighborhood vectors as:

\[
\frac{1}{\sum_{i=1}^{k} d_i} , \quad i=1,1,...,k
\]

\[
\sum_{i=1}^{n} w_i = 1
\]

where \( d_i \) is a distance to \( i^{th} \) neighborhood vector, and \( k \) is the number of nearest neighbors. Distance metrics may be the Euclidean, Manhattan, or other. The Euclidean distance between \((x_1, y_1)\) and \((x_2, y_2)\) is defined as:

\[
d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}
\]

while the Manhattan distance is defined as:

\[
d = |x_1 - x_2| + |y_1 - y_2|.
\]

The sum of all weights, \( w_i \), equals 1. This approach to calculating of weights ensures that the neighbors that are closer in signal space get weighted more heavily than those far away.

3. Calculate test vector location as the output of the CPN, \( O(x, y) \):

\[
O(x, y) = \sum_{i=1}^{n} (w_i * p_i)
\]

where \( p_i = (x_i, y_i) \) is the location of neighborhood vector.

V. TEST RESULTS

Several tests were performed to access the accuracy of the model. Test 1 used a small number of test points so the effect of RF signal attenuation on nearest neighbors could be visualized. Tests 2 and 3 used test points at \( x \) mid-grid locations (1.5, 2.5, ..., 20.5) along each \( y \) point (1 to 20). Test 4 used 8 access points to calculate the theoretic map. Test 5 used the Manhattan distance metric. Tables 1 and 2 summarize the test results.

**Test 1:**

Input: 5 unique test SSVs different from measured SSVs,
Output: 4 calculated positions with their neighborhood plots
Discussion: The calculated neighborhoods and predicted test positions are shown in Figure 3 with access points marked with an X. An estimated location is not available for one of the test vectors because no neighbors were within the threshold. The symmetry of the neighbors around one side of the calculated location is due to the similarity of RSS vectors within the distance threshold. Neighbors do not appear on the other side of the calculated location because the measured RSS values increase as the access point is approached. This effectively increases the distance to the access point, falling outside the threshold boundary and resulting in no neighbors for this area.

**Test 2:**

Input: 400 unique test SSVs different from measured SSVs,
Output: Calculated positions for each test SSV
Discussion: Several thresholds for k-NN were tested with the smallest threshold yielding the best results in terms of cumulative error. Error increases, as expected, as the distance between RSS test and measured SSVs increases.

Locations at the periphery of the grid map have significantly higher error (see Figure 4) due to distance from the closest access point. The tested accuracy of the algorithm was 90.6% within 1 meter and 96.4% within 1.5 meters at a threshold of 6. 

**Figure 3. Neighborhoods and Estimated Positions for Five Test Points**

**Figure 4. Error Landscape of Complete Radio Map with Threshold of 6**

Some spatially close neighbors to the RSS test vectors provide the optimal data for estimating the position of the mobile device. As the distance to the neighbors increases, the error also increases due to the inclusion of points further away, which are inaccurate. Some signal strength components of the RSS vector will increase while others will decrease as a mobile device moves across the landscape. Therefore, even some test points close to the access points, which measured RSS values are based on, will have a higher error rate than others. Figures 5 and 6 plot the predicted test locations where the error was less than or equal to 1, and the thresholds were 6 and 15, respectively. As the threshold is increased, more RSS vectors will fall within the threshold and be used to calculate the predicted location of the mobile device. This effectively increases the error in the calculation since some of those vectors are close in signal space but not in spatial distance.

**Test 3:**

Input: Same test locations as test 2 minus the test points on the periphery of the grid map.

Output: Calculated positions for each test SSV

Discussion: The error landscape using a distance threshold of 6 and 15 are shown in Figure 7 and 8, respectively. Signal strength from an access point decreases exponentially with distance. Therefore, points closer to an access point will provide a more accurate relationship between distance and signal strength. The number of measured RSS vectors used to calculate the location of the mobile device increases with threshold, which increases the error. The error rate between access points (dark blue peaks) in Figure 8 is relatively constant indicating that measured RSS vectors in this region have similar values. The tested accuracy of the algorithm was 94.5% within 1 meter and 100% within 1.5 meters at a threshold of 6.

**Test 4:**

The density of the access points for the radio map was increased from 4 per 20 square meters to 8 and tests 2 and 3 were rerun. The best tested accuracy of the algorithm was 96.9% within 1 meter and 100% within 1.5 meters at a threshold of 6.

**Test 5:** Test 2, 3, and 4 were rerun using the Manhattan distance metric.
increased for the radio map area of coverage. 

Increases, as expected, as the density of access points is clustering approaches reviewed in this paper. Accuracy estimation than other neural network, nearest neighbor, and neighboring locations.

Based pattern matching and linearization of obtained algorithm is increased. The third advantage is the attenuation calculations are avoided and hence the precision of distribution is recorded as it is. Complex signal deflection and CPN architecture. Another advantage is that the actual signal estimation are multifold. The first is the inherent parallelism of based on RSS values. The advantages of the CPN for location estimation are:

T<=6

<table>
<thead>
<tr>
<th># Points</th>
<th># Access</th>
<th>Error &lt;= 0.5 meter</th>
<th>Error &lt;= 1.0 meter</th>
<th>Error &lt;= 1.5 meter</th>
</tr>
</thead>
<tbody>
<tr>
<td>392</td>
<td>4</td>
<td>41%</td>
<td>90.6%</td>
<td>96.4%</td>
</tr>
<tr>
<td>217</td>
<td>4</td>
<td>46%</td>
<td>94.5%</td>
<td>100%</td>
</tr>
<tr>
<td>366</td>
<td>8</td>
<td>43.7%</td>
<td>95.6%</td>
<td>100%</td>
</tr>
<tr>
<td>191</td>
<td>8</td>
<td>45.6%</td>
<td>96.9%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1. LENSR Accuracy: Threshold = 6/15, Euclidean Distance Metric

<table>
<thead>
<tr>
<th># Points</th>
<th># Access</th>
<th>Error &lt;= 0.5 meter</th>
<th>Error &lt;= 1.0 meter</th>
<th>Error &lt;= 1.5 meter</th>
</tr>
</thead>
<tbody>
<tr>
<td>387</td>
<td>4</td>
<td>5.2%</td>
<td>34.4%</td>
<td>51.4%</td>
</tr>
<tr>
<td>212</td>
<td>4</td>
<td>9%</td>
<td>36.3%</td>
<td>52.4%</td>
</tr>
<tr>
<td>201</td>
<td>8</td>
<td>27.4%</td>
<td>87%</td>
<td>97.5%</td>
</tr>
<tr>
<td>79</td>
<td>8</td>
<td>25.3%</td>
<td>78.5%</td>
<td>96.2%</td>
</tr>
</tbody>
</table>

Table 2. LENSR Accuracy: Threshold = 6/15, Manhattan Distance Metric

VI. Conclusions

The CPN is a viable architecture for calculating location based on RSS values. The advantages of the CPN for location estimation are manifold. The first is the inherent parallelism of RSS signature matching due to the intrinsic parallelism of the CPN architecture. Another advantage is that the actual signal distribution is recorded as it is. Complex signal deflection and attenuation calculations are avoided and hence the precision of algorithm is increased. The third advantage is the computational inexpensiveness of LENSR, an elegant CPN based pattern matching and linearization of obtained neighboring locations.

The LENSR algorithm has a higher accuracy for location estimation than other neural network, nearest neighbor, and clustering approaches reviewed in this paper. Accuracy increases, as expected, as the density of access points is increased for the radio map area of coverage.

VII. References