Cellular Positioning Using Fingerprint Matching in WCDMA Networks

Anita A. Cherian, Rex Van Olst
Department of Electrical and Information Engineering,
University of Witwatersrand, Johannesburg, South Africa
email: {anita.cherian@students.wits.ac.za, rex.vanolst@wits.ac.za}

Abstract- Accurate cellular positioning in an African context still has hurdles since it is impractical to expect all mobile users to purchase new GPS-enabled phones. This paper presents a comparison between two methods from literature, which have been adapted to the context. These methods analyze the effect of using clusters, as well as the effect of penalty factors and weights on the accuracy. The results show very acceptable levels of accuracies with accuracies of up to 12.9m in 67% of the cases and 32.6m in 95% of the cases.

Index Terms—cellular positioning, fingerprint matching, location estimation, KNN, weighted distance

I. INTRODUCTION

Cellular positioning refers to locating a cellular phone and its user by utilizing signal measurements. The need for a greater study into cellular localization was motivated by the release of the U.S. Federal Communication Commission report in 1999. This report required that all cellular network operators be able to provide location identification of mobile stations, for safety reasons, by 2001. Minimum required accuracy levels of 100m for 67% of the cases, and 300m for 95% of the cases, are required for network based location estimation. A possible solution is to incorporate GPS technology into cellular phones. However, particularly in a developing or third world nations, it is impractical and expensive to expect every cellular phone to be replaced. In addition, all four cellular operators in South Africa have implemented 3G technology and the number of subscribers is growing quickly. Thus it is essential to develop better methods of estimating the location of a mobile user in this network.

A. Common Techniques

Cell Identification uses the Base Station (BS) to which a mobile phone is connected, to identify its location. The accuracy depends on the size of the cell and can be up to 100 meters radius in urban areas, to 20 kilometers in rural areas. Thus it is used in environments where high levels of accuracies are not needed, such as restaurant enquiries.

The time taken for the signal to travel between the BS and the mobile phone is measured for the Time of Arrival technique. The corresponding distance is equal to the measured time multiplied by the speed of light. This method requires clock synchronization, which can be obtained by using more stable clocks, which in turn results in hardware changes cost increases.

The Time Difference of Arrival method requires that two BS’s be used to derive a hyperbola with constant time difference. Once again, good clock synchronization is required. The Angle of Arrival method uses the angle of the signal arriving from the mobile user. In addition to it requiring the installation of antennas arrays at the BS’s, this method yields poor accuracies in Non-Line-of-Site conditions.

B. Pattern Matching

In this paper, the pattern matching technique is studied. Pattern matching involves studying signal patterns from the BS’s to a mobile phone, to obtain “fingerprints” at each reference location. These “fingerprints”, together with their corresponding locations form the database. During the location estimation process, once the observed “fingerprint” is compared with the database, a subsequent match is made.

The primary advantage of this technique is that high accuracies can be achieved with minimal costs. In addition, it allows for flexibility since the accuracy can be improved by just improving the model. This is in contrast to geometric based technologies which require more accurate measurements to be taken, to improve the accuracy.

Fingerprinting is not directly influenced by environmental factors such as multipath propagation, as are other techniques which rely on signal strength. However, this system is dependent on the environment and any infrastructural changes must be catered for. Thus the database must be updated roughly every six months.

II. FIELD TESTS

A Sony Ericsson K810i cell phone was put into field test mode to obtain the required measurements. The Received Signal Code Power (RSCP) from the six strongest WCDMA neighbouring BS’s were measured to form the “fingerprints”. An area of 1.94km² was covered in Lynnwood, Pretoria. A separate drive test route was taken in the allocated area to measure the test samples. Although it was expected to ideally obtain measurements of all six WCDMA neighbouring BS’s in each database fingerprint, this only happened in just below half of the cases.

III. METHODOLOGY

A. K-Nearest Neighbour with Clustering

This approach requires that only the strongest neighbouring WCDMA BS be used to form the fingerprints. The k-means method is used to cluster the elements, into N/2 clusters. Thereafter, the KNN classification method determines which cluster the sample belongs to.

B. Penalty Term

The approach taken by Kemppi [1] was attempted, and then adapted to determine the location estimates. The technique implemented in [1] involves the inclusion of all the BS’s in the calculation. There may be BS’s in the
sample that do not occur in the database, and vice versa. A very small threshold value, or penalty factor, is assumed for the RSCP’s of these “missing” BS’s. This penalty factor, Q, assumes that these BS’s are located far away from the fingerprint location.

Thus the difference between the sample and the \( k^{th} \) database fingerprint can be calculated as in the following three term approach:

\[
d(k) = \text{real} - \text{sample} = \sum (f_i - g_i(k))^2 + \sum (f_j - Q)^2 + \sum (m(Q - g_m(k))^2 \quad (1)
\]

where, 
\( f_i \) refers to the RSCP of the \( i^{th} \) detected BS in the sample 
\( g_k \) represents the RSCP of the \( k^{th} \) detected BS from the sample, which is present in the \( k^{th} \) database fingerprint 
\( f_j \) represents the RSCP of the \( j^{th} \) detected BS in the sample which is not present in the \( k^{th} \) database fingerprint 
\( g_m \) refers to the RSCP of the \( m^{th} \) detected BS from the \( k^{th} \) database fingerprint, which is not present in the sample

This technique was then adapted to at two term approach:

\[
d(k) = \sum (f_i - g_i(k))^2 + \sum (f_j - Q)^2 \quad (2)
\]

The error, as well as the number of estimated locations per sample, was analyzed for varying values of \( Q \) to determine the optimal value of \( Q \).

C. Weighted distance

The number of WCDMA neighbouring BS’s in the sample, that appear in the database fingerprint is a significant factor. Thus a weight, \( w_{k} \), is calculated as

\[
w_{k} = \frac{n_{j}}{n_{s}} \quad (3)
\]

where \( n_{j} \) is the number of BS’s in the sample, that appears in the database, \( k \), and \( n_{s} \) is the total number of WCDMA neighbours that is present in the sample. This weight is then used with equations 1 and 2.

The method used by Khalaf-Allah [2] has been adapted in this approach. It calculates a weight \( w_{i} \), where

\[
w_{i} = w_{i}^{(1)} + w_{i}^{(2)} + w_{i}^{(3)} \quad (4)
\]

\( w_{i}^{(1)} ) \), \( w_{i}^{(2)}, \) and \( w_{i}^{(3)} \) are the measurement model, neighborhood degree and strongest neighbor weights respectively.

\[
w_{i}^{(1)} = \frac{1}{N} \sum \frac{1}{\sigma_{RSCP} \sqrt{2\pi}} e^{-\frac{(RSCP_{j} - RSCP_{i})^2}{2\sigma_{RSCP}^2}} \quad (5)
\]

\( M \) refers to the total number of WCDMA neighbour Base Stations detected in the sample. The number of Base Stations in the sample, which is not detected in the fingerprint, is represented by \( N \). \( Q \) represents a threshold, as in the previous section. \( \sigma_{RSCP} \) is the standard deviation of the detected signal strengths in the sample. \( RSCP_{j} \) refers to the RSCP of the \( j^{th} \) Base Station in the sample. \( RSCP_{DBj} \) is the fingerprint RSCP of the \( j^{th} \) Base Station in the sample. \( w_{i}^{(2)} \) is the weight corresponding to the strongest BS in the fingerprint. If this is not the case, then \( w_{i}^{(2)} = 0 \).

Table 1 indicates the accuracies for the various methods.

<table>
<thead>
<tr>
<th>TECHNIQUES</th>
<th>67%</th>
<th>95%</th>
<th>Mean Error</th>
<th>No. of Ave. Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering, with KNN</td>
<td>15.185</td>
<td>26.655</td>
<td>12.478</td>
<td>1</td>
</tr>
<tr>
<td>Penalty Factor: 3 Term Approach, without ( w_{i} )</td>
<td>15.541</td>
<td>34.849</td>
<td>14.719</td>
<td>1</td>
</tr>
<tr>
<td>Penalty Factor: 3 Term Approach, with ( w_{i} )</td>
<td>15.541</td>
<td>34.849</td>
<td>14.719</td>
<td>1</td>
</tr>
<tr>
<td>Penalty Factor: 2 Term Approach, without ( w_{i} )</td>
<td>12.848</td>
<td>32.609</td>
<td>11.737</td>
<td>1.094</td>
</tr>
<tr>
<td>Penalty Factor: 2 Term Approach, with ( w_{i} )</td>
<td>12.848</td>
<td>32.609</td>
<td>11.737</td>
<td>1.094</td>
</tr>
<tr>
<td>Multiple weights</td>
<td>13.187</td>
<td>33.558</td>
<td>12.423</td>
<td>2.594</td>
</tr>
</tbody>
</table>

IV. RESULTS

It was observed that for values of \( Q \) below -150, the average error of the samples stabilizes. Below this value of \( Q \), \( w_{i} \) no longer has an impact on the accuracy or the number of estimates. It must be noted that in some cases, multiple location estimates were obtained for a sample. In these cases, all the possible estimates were considered, and the mean of their errors were found for the particular sample.

V. CONCLUSION

From the results it can be concluded that although the KNN approach appears to yield good accuracies, it is sensitive to outliers as a result of using only the strongest neighbour. Hence, the best method in terms of both accuracy and number of estimates is the Penalty Factor Approach with two terms, where the database BS’s that are not common with the sample, are ignored. In addition, it was noted that it is reasonable to accept a very small value for the “missing” BS’s. More importance should thus be given to how much of the sample is matched by the database fingerprint, and not to how much of the fingerprint has been matched by the sample. This justifies the omission of these database Base Stations in the calculations.

VI. FUTURE WORK

Multiple location estimates must be further analyzed using GSM data and reduced to a single estimate. In addition, the effect of clustering is to be analyzed using the entire neighbouring WCDMA BS’s. Further samples are to be obtained and tested, which will in turn provide an indication of the effects of changes in the environment and weather on the measurements and on the location estimate accuracies.

VII. REFERENCES


Anita Cheren received her degree in BSc Electrical Engineering in 2009 from the University of Cape Town, and is currently studying towards her MSc degree at the University of Witwatersrand.