Utilizing Kriging to Generate a NLOS Error Correction Map for Network Based Mobile Positioning

Binghao Li, Chris Rizos
School of Surveying and Spatial Information System, UNSW, Australia
e-mail: binghao_li@student.unsw.edu.au Tel: 61-2-93854189; Fax: 61-2-93137493

Hyung Keun Lee
School of Avionics and Telecommunication, Hankuk Aviation University, Korea

Received: 27 November 2004 / Accepted: 12 July 2005

Abstract. Network mobilephone-based positioning experiences degradation of location accuracy due to localised non-line-of-sight (NLOS) signal propagation. This is well known to be a major source of error in network-based mobilephone positioning. NLOS error systematically causes the Mobile Station (MS) to appear further away from the base station than it actually is, thereby increasing the positioning error. One method to mitigate the effect of NLOS error is to generate a NLOS error correction map, and then use the correction map to rectify the distorted MS location. The correction map can be generated using the following procedure: (1) estimating the NLOS errors at points where the real positions can be obtained utilising other information such as the points very near BTS (Base Transceiver Station) and the intersections of streets, or the location where the measurement has been made; and (2) interpolating or extrapolating the errors to specific points that we are interested in. Assuming some reference points have been obtained, this paper utilises kriging, an estimation technique that is widely used in mining, to generate the correction map. Theoretically kriging can also be used wherever a continuous measure is made on a sample at a particular location in space or time. Using simulations with a typical dense urban environment assumption, the feature of the NLOS error variogram is analysed and different models of the variogram are compared. The correction map of NLOS error is generated using some ‘sampled’ points, and compared with the ‘true’ NLOS error map to show the efficiency of kriging.

Key words: NLOS, Positioning, Kriging

1 Introduction

Over the past decade, Mobile phone positioning techniques (PT) have received considerable attention (Rappaport TS, 1996) and great number of new value-added services have been proposed or developed, such as the location based information service, navigation assistant, resource management, gaming, emergency service etc (D’Roza and Bilchev, 2003; Wilde G, 2002). The requirements set forth by the US FCC E911 is one of the major forces to push the mobile phone PT moving forwards. It requires wireless carriers to provide precise location information within 50 to 100 meters in most cases (Http://www.fcc.gov/911/enhanced/).

The most popular and simple method is Cell ID. But the accuracy of Cell ID is dependent on the density of the BTS, and is relatively poor especially in rural areas (Dru and Saada, 2001). Another basic method is based on the received signal strength. However, since the signal propagation suffers rapid deep fading and long term fading (Lee, 1991), no model can describe the feature of signal propagation very well in many environments. The accuracy of this method is better than Cell ID (Yamamoto et al., 2001). Much attention has been focused on the signal time delay and angle of arrival measurements. No matter what kind of approach such as angle of arrival (AOA) (Sakagami et al., 1992), time of arrival (TOA) (Hashemi, 1991) or time difference of arrival (TDOA) (Drane et al., 1998) is utilized, line-of-sight (LOS) propagation is necessary for accurate location estimates. In other words, non-line-of-sight (NLOS) error is the dominant error in location estimation (Jr. JC and Stüber GL, 1998). NLOS errors are always positive, and range from a small number to thousand meters (Silventoinen and Rantalainen, 1996), depending on the propagation environment.
To protect location estimates from NLOS error corruption, many approaches have been investigated. In (Morley and Grover, 1995), an algorithm based on probability density function (pdf) model is utilized to reduce the NLOS error. However, it is very difficult to formulate the pdf, and this pdf should vary greatly with changing of the environment. A widely used idea to mitigate the NLOS error is NLOS error identification and reconstruction. The method in (Wylie and Holtzman, 1996) and (Woo SS et al., 2000) reconstruct LOS TOA measurements from a time history of LOS and NLOS TOA measurements, and assumes knowledge of the NLOS standard deviation for identifying NLOS BTS. While in (Cong and Zhuang, 2001), the NLOS BTS detection is based on TDOA residuals. The algorithm in (Wang, 2003) shows a constrained optimization method. Unfortunately, none of these methods can solve the NLOS problem well, since too many elements affect the signal propagation, and the propagation environment varies from place to place. In fact the problem could be solved using a basic data base method. In this method, the NLOS error can be directly extracted from the reference measurements at the reference points. Jayaraman et al. (2000) and Gunnarsdottir and Hole (2001) describe methods for collecting data to create a database. But the data collection and data base maintenance is quite a costly process. In order to make this work easier, a wireless signal map matching method (WSMM) is proposed (Lee and Rizos, 2003). Generally, more reference points can help to generate a database that can describe the real situation more precisely. But no matter how many reference points have been measured, the area of interest cannot be completely covered. Interpolation and extrapolation is absolutely necessary.

Now the question arises: how to efficiently estimate the NLOS error at the location (say \( x_0 \)), which is not reference point, only knowing the limited reference points (say \( x_1, \ldots, x_n \)). In this paper, we are interested in both estimation of the NLOS error and the confidence on the estimate. After a brief introduction of Geostatistics and kriging, the simulated NLOS error in a Manhattan-like urban environment is generated, and the features of NLOS error is checked using a simulation; this is followed by the different variogram models fitting. Finally, the results utilizing universal kriging (UK) are shown.

2 Geostatistics and Kriging

Geostatistics was first used by the mining industry, as high costs of drillings made the analysis of the data extremely important. The prediction of the ore grade in a mining block from observed samples at irregularly spaced locations is one of the most important problems. The basic tool in geostatistics, the variogram, is used to quantify spatial correlations between observations. The estimation procedure is called kriging after D. Krige, who and his colleagues started to apply statistical techniques to ore reserve estimation in 1950s. As there many advantages of kriging, and especially with the advent of powerful computers, application of kriging can be found in very different disciplines ranging from the classical fields mining and geology to soil science, hydrology, meteorology, environmental sciences, agriculture etc. Theoretically, kriging can also be used wherever a continuous measure is made on a sample at a particular location in space or time (Cressie, 1991; Armstrong , 1998 ). In spatial information system, it is necessary to use the limited data to describe the real nature as precisely as possible, kriging is a good candidate to choose.

In geostatistics, the geological phenomenon is described in terms of fluctuations around a fixed surface ("drift" or "trend"). The fluctuations are not error but rather fully-fledged features of the phenomenon, with a structure of their own. The observed value at each data point \( x \) is considered as the outcome, \( Z(x) \), of a random variable, \( Z(x) \). Its mean is the drift, \( m(x) \). A classical assumption in geostatistics is the second order stationarity, but in practice, a slightly weaker assumption is more widely used, that is the intrinsic hypotheses. It consists of two conditions:

- The expected value of the random variable \( Z(x) \) is constant all over the domain \( D \).
- The variance of the increment corresponding to two different locations depends only on the vector separating them

This condition can be formulated as:

\[
E[Z(x)] = \mu \quad (1)
\]

for all \( x \in D \)

\[
\frac{1}{2} \text{Var}[Z(x+h) - Z(x)] = \frac{1}{2} E[(Z(x+h) - Z(x))^2] = \gamma(h) \quad (2)
\]

where \( \gamma(h) \), called variogram, depends only on the vector \( h \) and not on the locations \( x \) and \( x+h \).

Kriging provides a solution to the problem of estimation based on the knowledge of the variogram and the above assumption. Here is the simple case that the mean is constant across the entire region of study. Unfortunately, in reality it is common that the mean is not constant. The simulation shows that a drift of NLOS error exists. Assume the mean is a function of the site coordinates:

\[
Z(x) = f_0(x)\beta_0 + f_1(x)\beta_1 + \ldots + f_p(x)\beta_p + \delta(x)
\]

where \( \beta_0, \ldots, \beta_p \) are unknown parameters; \( \delta(x) \) is intrinsic and \( E[\delta(x)]=0 \).
In matrix notation, the above expression can be written as:

\[ Z = X\beta + Y \]  

(3)

In order to deal with the drift, UK (universal kriging) is proposed (The ordinary kriging can be treated as a subset of UK when \( f_0(x) = 1, \beta_1 = \ldots = \beta_p = 0 \)). To predict \( Z(x_0) \) the UK predictor is a linear combination of values of the sample \( Z(x_i) \):

\[
\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i)
\]

where \( \lambda_i \) is the weighting factor.

For the purpose of making this predictor to be unbiased for all possible vectors \( \beta \), the following conditions need to be satisfied.

\[
E\left[ \sum_{i=1}^{n} \lambda_i Z(x_i) - Z(x_0) \right] = 0
\]

(4)

As the estimation variance is:

\[
\sigma_k^2(x_0) = \text{Var}\left[Z(x_0) - \hat{Z}(x_0)\right] = -\sum_{j=1}^{n} \sum_{i=1}^{n} \lambda_j \lambda_i \gamma(x_i - x_j) + 2 \sum_{i=1}^{n} \lambda_i \gamma(x_i - x_0)
\]

(5)

The best unbiased linear estimator is the one which minimizes \( \sigma_k^2(x_0) \) under the constraint on the sum of the coefficients in (4). Introducing the Lagrange multipliers this leads to a straightforward linear equation.

\[
\begin{bmatrix}
\Gamma & X' & \lambda
\end{bmatrix}
= \begin{bmatrix}
\gamma \\
X' & 0 & m
\end{bmatrix}
\]

(6)

where

\[
\begin{bmatrix}
\gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) \\
\vdots & \ddots & \vdots \\
\gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n)
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & f_1(x_1) & \cdots & f_p(x_1) \\
\vdots & \vdots & \ddots & \vdots \\
1 & f_1(x_n) & \cdots & f_p(x_n)
\end{bmatrix}
\]

\[
\gamma = \begin{bmatrix}
\gamma(x_0 - x_1) \\
\vdots \\
\gamma(x_0 - x_n)
\end{bmatrix}
\]

\[
f = \begin{bmatrix}
1 & f_1(x_0) & \cdots & f_p(x_0)
\end{bmatrix}
\]

So, the result is

\[
\lambda' = \left( \gamma + X(X\gamma^{-1}X)' \right)^{-1} \left( f - X\gamma^{-1}y \right) \Gamma^{-1}
\]

\[
m' = \left( f - X\gamma^{-1}y \right) \left( X\gamma^{-1}X \right)^{-1}
\]

and

\[
\sigma_k^2(x_0) = \gamma' \Gamma^{-1} \gamma - \left( f - X\gamma^{-1}y \right) \left( \Gamma^{-1} \right)' \left( f - X\gamma^{-1}y \right)
\]

(7)

Kriging is the best linear unbiased estimation (BLUE) that has the following features: (a) this estimator is a linear function of the data with weights calculated according to the specifications of unbiasedness and minimum variance. (b) The weights are determined by solving a system of linear equations with coefficients that depend only on the variogram that describes the structure of a family of functions. A major advantage of kriging is that it is more flexible than other interpolation methods. The weights are not selected on the basis of some arbitrary rule that may be applicable in some cases but not in others, but depend on how the function varies in space. Another advantage of kriging is that it provides the means to evaluate the magnitude of the estimation error. The mean square error is a useful rational measure of the reliability of the estimate; it depends only on the variogram and the location of the measurements.

![Fig. 1 Manhattan-like urban environment](image)

3 NLOS error simulation

3.1 NLOS error

As a random function, NLOS error can be modeled by a deterministic part and a random part. Basically, there are three types of methods to generate the NLOS error. In most of the papers, simplified models such as deterministic, Gaussian or other distribution models (Cong and Zhuang, 2001) are utilized. Though this method is convenient, nevertheless, it can hardly describe the real NLOS error. On the contrast, 3D ray tracing plus Poisson or Rician model (Aguado et al., 1997) can
accurately generate the NLOS error in a special environment, but it is a very complex method. It is time consuming and also costly. The chosen method in this paper is the medium accuracy model; deterministic part is generated by Dijkstra algorithm (2D only), and the random part is represented by Gaussian model. Assuming in a Manhattan-like urban environment, only 3 BTSs are arranged as shown in Figure 1. By adding NLOS error and random measurement error, the TOA measurements are generated. Multiplying TOA with the speed of light, the real propagation distance can be obtained. Compare the propagation distance and the distance between the MS and BTS, the NLOS error (plus random measurement error) can be derived. Figure 2 shows the NLOS error in Manhattan-like environment. Streets 1 to 5 are the streets of south-north direction from east to west respectively (because of the symmetric environment, the NLOS error in the streets of west-east direction are similar). Two features can be noticed: first, there is a shift of NLOS error; second, the NLOS error is continuous. It is well known that using TOA measurements to compute the MS location, the MS’s clock bias should be considered. However, the TDOA measurements can get rid of the clock bias automatically. In this paper, the object will be analyzed is the injected NLOS error rather than the NLOS error directly. The TDOA measurements can be generated by simply subtracting TOA measurement of one BTS from TOA measurement of reference BTS (here BTS1 is the reference). Two TDOA measurements representing TDOA21 and TDOA31 for each point are available. Figure 3 depicts the injected NLOS error on ideal domain and distorted domain. In the ideal domain, the injected NLOS error is more regular. The figure implies that there is a strong intrinsic relationship between the NLOS error and the location. Unfortunately, what should be dealt with is the NLOS error in the distorted domain. Since the determination of the MS’ location is contaminated by the errors, the relationship between the NLOS error and the location is weaker and this makes the approach difficult.
3.2 Hyperbolic equation solving algorithms

Once the TDOA estimates have been obtained, they are converted into range difference measurements and these measurements can be converted into nonlinear hyperbolic equations. Assuming BTS1 is the reference BTS, let \((x, y)\) be the MS location and \((X_i, Y_i)\) be the known location of the \(i\)th BTS. The TDOA measurement is

\[
R_{i,1} = c d_{i,1} = R_i - R_1 = \sqrt{(X_i - x)^2 + (Y_i - y)^2} - \sqrt{(X_1 - x)^2 + (Y_1 - y)^2}
\]  

(8)

Several algorithms have been proposed, such as Friedlander’s method (Friedlander, 1987), Taylor-Series method (Foy, 1976), Fang’s method (Fang, 1990) and Chan’s method (Chan and Ho, 1994) etc. Each method has advantages and disadvantages. For example, Chan’s method can provide exact solution; it also takes advantage of redundant measurements and further more it approaches CRLB (Cramer-Rao lower bound). However, occasionally, Chan’s method has ambiguities in the solution, and if the measurement has large errors (including NLOS error), it cannot work efficiently. Taylor-Series method is an iterative method, and the redundant measurements can be used as well. The provisional value and GDOP can significantly affect the proceeding solution and convergence is not guaranteed. However, in the condition of large errors, it provides more freedom for tuning; so it works more efficiently than other methods. The Taylor-Series method is chosen to solve the hyperbolic equation in this paper.

With a set of TDOA estimates, the method starts with a provisional value \((x_0, y_0)\) and computes the deviations of the location estimation \(dx\) and \(dy\).

\[
\begin{bmatrix}
\frac{dx}{dy}
\end{bmatrix} = \left(A^T W^{-1} A\right)^{-1} A^T W b
\]

(9)

where

\[
A = \begin{bmatrix}
\frac{X_1 - x}{R_1} & \frac{X_2 - x}{R_1} & \frac{Y_1 - y}{R_1} & \frac{Y_2 - y}{R_1} \\
\frac{X_1 - x}{R_2} & \frac{X_2 - x}{R_2} & \frac{Y_1 - y}{R_2} & \frac{Y_2 - y}{R_2}
\end{bmatrix}
\]

\[
b = \begin{bmatrix}
R_{2,1} - (R_2(x_0) - R_1(x_0)) \\
R_{M,1} - (R_M(x_0) - R_1(x_0))
\end{bmatrix}
\]

and \(W\) is the covariance matrix of the estimated TDOAs. The whole process is repeated until \(dx\) and \(dy\) are sufficiently small.
Due to the injected NLOS error and noise, the position estimates based on the TDOA do not coincide with the true position. The ideal MS position distribution and the distorted MS position distribution are shown in Figure 4.

4 Choosing a variogram model

Variogram $\gamma$ is the basic tool for the structural interpretation of phenomena as well as for estimation. It is defined in (1). Normally, $\gamma$ is not known and needs to be estimated from the TDOA measurements. There are several ways to estimate the variogram (Cressie, 1991). The classical formula is:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i,j \in \Lambda} (Z(x_i) - Z(x_j))^2$$

Most of the time, the points are irregularly spaced. In order to have more pairs, the summation $x_i - x_j = h$ has to be weakened.

$$\|x_i - x_j - h\| \leq \varepsilon \quad \text{Angle}(x_i-x_j, h) \leq \delta$$

In order to make the processing easier, the data is adjusted slightly to a grid map. Before computing the $\gamma$, the nonstationarity of the injected NLOS error should be checked. Figure 5 is an attempt to summarize the possible nonstationarity in the mean using the sample median and sample mean across rows and down columns. There appears to be a shift in the east-west direction but little or no shift in north-south direction. Computing the variogram in this direction first is a good choice.

Unfortunately, there is still a little shift in north-south direction. Figure 6(a) depicts the variogram computed using the formula above, taking the width of each bin to be 24 meters. The effect of the trend is plainly obvious, leading to a steadily increasing parabolic-like curve. How to decompose the data with shift is a classical problem. Some approaches have been proposed. One of them is to start with o.l.s. (ordinary-least-squares) estimator of $\beta$ in (3), compute a variogram estimator from the residuals, fit a variogram model, then obtain a g.l.s (general-least-squares) estimator of $\beta$ based on the fitted model, and so forth. In this paper, this iterative approach is used, although this approach suffers from a bias problem (Cressie, 1991). Figure 6(b) shows the estimated north-
south variogram of injected TDOA21 error residuals after o.l.s. The reason not to use the experimental variogram directly is because most of the experimental variograms are not admissible, as it needs to be conditionally negative definite. Only after few iterations, the result is converged. Finally, two models are chosen: exponential (11) and a spherical model (12).

\[
\begin{align*}
\gamma(0) &= 0 \\
\gamma(h) &= C0 + C \left( 1 - e^{-\frac{h}{a}} \right) \\
\gamma(0) &= 0 \\
\gamma(0 < h < a) &= C0 + C \left( \frac{3h}{2a} + \frac{1}{2} \left( \frac{h}{a} \right)^3 \right) \\
\gamma(h \geq a) &= C0 + C
\end{align*}
\]

(11) (12)

Fig. 7 The chosen variogram models: exponential model (top) and spherical model (bottom)

The parameters retained were: for exponential model, \(C0=350m^2, C=2000m^2, a=500m\); and for spherical model, \(C0=300m^2, C=1700m^2, a=800m\). These two models are shown in Figure 7. \(\gamma(0)=0\) shows a discontinuity at the origin, which is called the nugget effect. This is caused by the unknown micro scale variation. The chosen models have a sill, which means that when the distance between two points is large enough, they are independent. This means the model can be described by covariance. In UK, sometimes the covariance matrix is needed (Cressie, 1991). The model is chosen visually. Some of the automatic methods such as least squares can be used, but it is not suggested by most of the professionals (Armstrong, 1998).

\[S = \frac{\hat{Z} - Z}{\sigma_K} \]

(13)

5 Results

The variogram model should be checked before the application. The evaluation of it was done through cross-validation (Isaaks and Srivastava, 1989; Armstrong, 1998). For each measurement location \(x_i\) the values are estimated as if they were unknown. The kriging variances in (5) are also computed. Then the normalized residual can be formed:

\[S = \frac{\hat{Z} - Z}{\sigma_K} \]

It should be normally distributed with 0 mean and 1 as standard deviation \((N(0,1))\). Figure 8 shows such a distribution using exponential model. The mean, standard deviation and max of the normalized residual using different model are compared in table 1. Here, the exponential model is slightly better than spherical model.

<table>
<thead>
<tr>
<th>Tab. 1 Results of cross-validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Mean normalized residuals</td>
</tr>
<tr>
<td>Std. normalized residuals</td>
</tr>
<tr>
<td>Max normalized residuals</td>
</tr>
</tbody>
</table>

As mentioned before, in real application, obtaining the measurement of reference point is not trivial. WSMM can
aid to get some references automatically. In the simulated Manhattan-like environment, 25 points can be derived automatically. A correction map should be generated by these reference points. Figure 9 plots the injected TDOA21 NLOS error correction map. It is similar with Figure 3 (bottom). Normalized residuals (the true data are known since they are generated by simulation) can be computed using (13). The results are listed in table 2.

<table>
<thead>
<tr>
<th>Results using 25 references</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of normalized residuals</td>
<td>0.1259</td>
<td></td>
</tr>
<tr>
<td>Std. of normalized residuals</td>
<td>1.0007</td>
<td></td>
</tr>
<tr>
<td>Max of normalized residuals</td>
<td>3.6893</td>
<td></td>
</tr>
</tbody>
</table>

Using the same algorithm, the injected TDOA31 NLOS error correction map can be generated. The corrected TDOA measurements can be derived by simply subtracting NLOS error from TDOA measurements. Then Taylor-Series method is applied again, and a new position distribution can be generated. Because of the reason mentioned in previous section, the NLOS error cannot be removed completely, but a part of it should be mitigated. The new distorted position exposes better relationship with the NLOS error. The processing can be done again. After two iterations, Figure 10 illustrates the final result. It is very clear after UK, the new distorted MS position distribution is closer to the ideal distribution.

6 Conclusions

The algorithm using kriging can efficiently estimate the NLOS error with the knowledge of some references and a reasonable variogram model. Normally, kriging is used in the situation where the exact location of measurement is known. In the application to mitigate NLOS error, however, the true location is the answer we are looking for. The distorted location covers some real specifics of the NLOS error and location. It makes the application harder. There are still some questions to be answered. First of all, the real data is necessary to verify this algorithm. Secondly, how to extract NLOS error at reference points efficiently. Thirdly how many references are needed to get rid of the NLOS error to a fix percentage. Finally, how to determine the variogram model for a specific environment, although some evidences show this application is not very sensitive with variogram model. Future work is focussed on optimising the algorithm for real world application.

References


Wilde G. (2002): Making the most of legacy mobiles: examining the relationship between technology choices and location revenues, BWCS white paper.


