Analysis of Statistical Estimation for Mobile Location Estimation

Junyang Zhou, Joseph Kee-Yin Ng
Department of Computer Science
Hong Kong Baptist University
Kowloon Tong, Hong Kong
Tel: (852)-3411-7864
Email: {jyzhou, jng}@comp.hkbu.edu.hk

keywords Mobile location estimation, Statistical Estimation, Signal Information Matrix, Location Information Matrix, Signal Fluctuation

Abstract

Mobile Location Estimation is drawing considerable attention in the field of wireless communications. In this research study, we focus on the signal fluctuation problem in location estimation. We present an estimation algorithm which consider all the information we have to reduce the effect of the signal fluctuation from base stations. And we named it as Statistical Estimation. In short, Statistical Estimation is derived from the signal information and the location information of the Base Stations(BSs) and utilizes these information to estimate the location of the Mobile Station(MS). In our experiments, the Statistical Estimation algorithm can provide a more accurate estimation on the location and reduce the estimation error caused by the effect of signal fluctuation. In fact, it acts as a filter to handle the signal fluctuation problem in location estimation.

1 Introduction

Mobile location estimation is an interesting problem. There are many methods trying to solve it over the past few years. And to find out the location of a Mobile Station(MS) based on the signal strength and the base stations(BSs) is one thread to solve the mobile location problem. There are some results which follow this thread [1, 2, 3, 4, 5, 6, 7, 9, 10, 11].

Through our research and observed from the field test data, signal strength at one place do fluctuate. That is, the signal strength received from the same BS is different at two sampling time in the same place, and some places are more serious than the others. Of course, this phenomenon of fluctuating can be described by the random model, like in [1, 6, 7]. But the models only use one snapshot information to find out the MS location. If we get more than one sample snapshots at one place, but only choose one of them to find out the estimation of MS location, then we may lose some information among these snapshots.

We want to use more information about the snapshots received from the BSs to find out the estimation of MS location. Since we believe that all the information about the snapshots have some contribution for estimating the MS location, the more information we use to find out the MS location, the more accurate the estimation. In this study, we present an estimation, named by the statistical estimation, and we also give two special structures of the statistical estimation.

This research study is divided into four sections. In the following section, we will discuss the Statistical Estimation in details, then we describe the simulation results of the Statistical Estimation in section 3. And lastly in section 4, we present a summary of our research and discuss about the future work.

2 The Statistical Estimation

Given a MS location (x_0, y_0), we can obtain the received signal strengths from the surrounding BSs through the mobile station handset, and provided that we also know the locations and the transmission powers of the BSs, we can derive the signal attenuation between each of the BSs and the MS. We denote signal strength received from the surroundings BSs and the MS location at one time as a snap shot. If we take a numbers of samples of received signal strength at one place, we will have captured some snapshots information which can be useful for location estimation.

Suppose we have n snapshots at MS location in a field test. The signal information collected can be denoted by a matrix, S:

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1m_1} \\ s_{21} & s_{22} & \cdots & s_{2m_2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nm_n} \end{pmatrix}, \tag{2.1}$$

And the location information of the surrounding BSs can be denoted by matrix L:

$$L = \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1m_1} \\ l_{21} & l_{22} & \cdots & l_{2m_2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nm_n} \end{pmatrix}. \tag{2.2}$$

where

each row of S and L is one snap shot information of the MS received;

$m_1, m_2, ..., m_n$ are the number of each row, which may be not the same;

Since each time the number of the signal strength received from the BSs can be different, the information and the location matrix can be irregular. Namely, $m_1, \cdots, m_n$ can be different.

In one snapshot, we can get an estimation of the location by some algorithms, such as CG, CT, Locus , the Geometric Algorithm based on EPM (GEPM) [1] and the Iterative Algorithm based on EPM (IGEPM) [2]. If we have n snapshots, then we have n estimations of the same MS location.

2.1 Main Idea of the Statistical Estimation

We have a probability method to solve this problem. We describe the problem in detail as follows.

Suppose for a particular MS location, we have the signal information matrix, S:

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1m_1} \\ s_{21} & s_{22} & \cdots & s_{2m_2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nm_n} \end{pmatrix}, \tag{2.3}$$

and the corresponding BSs location information matrix, L:

$$L = \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1m_1} \\ l_{21} & l_{22} & \cdots & l_{2m_2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nm_n} \end{pmatrix}. \tag{2.4}$$

Given the first row of information from the two matrixes S and L. For each snapshot , that is to say, for each row, we get an estimation of the MS location. Based on the signal information, s_{11},s_{12},$\cdots$, s_{1m_1}, and the location information, l_{11},l_{12},$\cdots$, l_{1m_1}, we can get an estimation of the MS location, denoted by (x_1, y_1).

Suppose we have two random variables x,y, which are independent to each other. They are used to describe the east and north coordinates, and their exact distributions can be known or unknown to us, for example, the value of the two variables can be distributed under the normal distribution. So the estimation can be seen as a conditional expectation of the signal and location information, denoted by,

$$x_1 = E(x|x_0, y_0), \text{ and } y_1 = E(y|x_0, y_0)$$

where (x_0, y_0) is the initial information.

This is the estimation of MS location, but we just use one snapshot of the information.

We consider the second row of signal strength: s_{21},s_{22},$\cdots$, s_{2m_2}, and the corresponding location information,l_{21},l_{22},$\cdots$, l_{2m_2}. And with this information, we also can get another estimation. If the second estimation does not consider the first one, then we would have wasted some valuable information,s_{11},s_{12},$\cdots$, s_{1m_1}, l_{11},l_{12},$\cdots$, l_{1m_1}. So the second estimation should have some relationship with the first one. And it can be written as

$$x_2 = E(x|x_1, y_1, x_0, y_0), \; y_2 = E(y|x_1, y_1, x_0, y_0).$$

Since we have n snapshots, so the n-th estimation can be written as

$$x_n = E(x|x_{n-1}, y_{n-1}, \cdots, x_1, y_1, x_0, y_0), \; y_n = E(y|x_{n-1}, y_{n-1}, \cdots, x_1, y_1, x_0, y_0).$$

If we choose this approach as the estimation algorithm, then it would be better than the other estimation by incorporating statistical knowledge of previous history into our calculation.

Set $D_t = \{x_t, y_t, D_{t-1}\}$, and $D_0 = \{x_0, y_0\}$, we have a series of the estimation by the above formula:

$$\begin{aligned}
&x_1 = E(x|D_0), y_1 = E(y|D_0);\\
&x_2 = E(x|D_1), y_2 = E(y|D_1);\\
&\ldots\\
&x_n = E(x|D_{n-1}), y_n = E(y|D_{n-1}).
\end{aligned}$$

That is, $x_n = E(x|D_{n-1})$, $y_n = E(y|D_{n-1})$ and $D_n = \{x_n, y_n, D_{n-1}\}$,for $n \geq 1$, where $D_0 = \{x_0, y_0\}$.

we call the estimation of (x_n, y_n) the statistical estimation.

2.2 Good Feature of the Statistical Estimation

Suppose $E(x_n) = x_0$ and $E(y_n) = y_0$ for $n \geq 1$.

Set $Error(n) = \sqrt{(x_n - x_0)^2 + (y_n - y_0)^2}$, and we use it as the estimation measuring criteria. Since we know that x_n and y_n are random variables from their definitions, $Error(n)$ is also a random variable. So we should compare them with their expectations.

And we have $var(x_n) \leq var(x_m)$, $var(y_n) \leq var(y_m)$, for $m \leq n$.

Since $var(x_n) = var(E(x|D_{n-1})) \leq var(E(x|D_{n-2}) = var(x_{n-1})$.

And now we have a conclusion about the error criteria as Lemma1.

Lemma 1: If $E(x_n) = x_0, E(y_n) = y_0$, and $var(x_n) < +\infty, var(y_n) < +\infty$, then

$$E(Error(n)) \leq E(Error(n-1)).$$

Proof:

Since

$$Error(n) = \sqrt{(x_n - x_0)^2 + (y_n - y_0)^2},$$

so

$$Error^2(n) = (x_n - x_0)^2 + (y_n - y_0)^2,$$

and

$$\begin{aligned} E(Error^2(n)) &= E((x_n - x_0)^2 + (y_n - y_0)^2 \\ &= var(x_n) + var(y_n) \leq var(x_{n-1}) + var(y_{n-1}) = E(Error^2(n-1)). \end{aligned}$$

That is

$$E(Error^2(n)) \leq E(Error^2(n-1)).$$

On the other hand, we can get it from the definition,

$$0 < E(Error(n)) < +\infty, \text{for all } n.$$

So

$$\begin{aligned} &E(Error(n) - Error(n-1)) = \\ &\qquad E((Error^2(n) - Error^2(n-1))/(Error(n) + Error(n-1))). \end{aligned}$$

By **the Slusky Theorem**,

$$\begin{aligned} &E((Error^2(n) - Error^2(n-1))/(Error(n) + Error(n-1))) = \\ &\qquad E((Error^2(n) - Error^2(n-1)))/E(Error(n) + Error(n-1)) \end{aligned}$$

There exists a positive real number c, which satisfies

$$0 < c < E(Error(n) + Error(n-1)) < +\infty.$$

So

$$E(Error(n) - Error(n-1)) \leq E(Error^2(n) - Error^2(n-1))/c \leq 0.$$

That is,

$$E(Error(n)) \leq E(Error(n-1)).$$

Q.E.D.

So we can draw a conclusion that if n increases, then the expectation of $Error(n)$ will not increase. If we have n snapshots, then choose (x_n, y_n) where $x_n = E(x|D_{n-1}), y_n = E(y|D_{n-1})$ as the estimation will produce the best result in terms of stability.

2.3 Structure of the Statistical Estimation

Since $x_n = E(x|D_{n-1}), y_n = E(y|D_{n-1})$, and $D_n = \{D_{n-1}, x_n, y_n\}$, where $D_0 = \{x_0, y_0\}$. We can rewrite (x_n, y_n) as a function of the information of signal strength and BS location. with the information matrixes:

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1m_1} \\ s_{21} & s_{22} & \cdots & s_{2m_2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nm_n} \end{pmatrix}, \tag{2.5}$$

and

$$L = \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1m_1} \\ l_{21} & l_{22} & \cdots & l_{1m_2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{1m_n} \end{pmatrix}, \tag{2.6}$$

we can rewrite the series of estimations mentioned in the previous section as

$$x_n = f_n(x_{n-1}, y_{n-1}, \cdots, x_1, y_1, x_0, y_0; s_{n1}, \cdots, s_{nm_n}; l_{n1}, \cdots, l_{nm_n}) + \epsilon_n;$$
$$y_n = g_n(x_{n-1}, y_{n-1}, \cdots, x_1, y_1, x_0, y_0; s_{n1}, \cdots, s_{nm_n}; l_{n1}, \cdots, l_{nm_n}) + \eta_n;$$
for $n \geq 1$;

where

f_n and g_n are continuous functions in the definition field;
ϵ_n and η_n are random variables, $E(\epsilon_n) = 0$, $E(\eta_n) = 0$, $var(\epsilon_n) = \sigma_{\epsilon_n}$, $var(\eta_n) = \sigma_{\eta_n}$;
ϵ_n and η_n are independent.

Since

$$x_1 = f_1(x_0, y_0; s_{11}, \cdots, s_{1m_1}; l_{11}, \cdots, l_{1m_1}) + \epsilon_1,$$
$$y_1 = g_1(x_0, y_0; s_{11}, \cdots, s_{1m_1}; l_{11}, \cdots, l_{1m_1}) + \eta_1,$$

we have

$$x_n = f_n^*(x_0, y_0; s_{11}, \cdots, s_{1m_1}, \cdots, s_{n1}, \cdots, s_{nm_n}; l_{11}, \cdots, l_{1m_1}, \cdots, l_{n1}, \cdots, l_{nm_n}) + \epsilon_n^*,$$
$$y_n = g_n^*(x_0, y_0; s_{11}, \cdots, s_{1m_1}, \cdots, s_{n1}, \cdots, s_{nm_n}; l_{11}, \cdots, l_{1m_1}, \cdots, l_{n1}, \cdots, l_{nm_n}) + \eta_n^*,$$

where

f_n^* and g_n^* are continuous functions in the definition field;
$E(\epsilon_n^*) = 0$, $E(\eta_n^*) = 0$; $var(\epsilon_n^*) = \sigma_{\epsilon_n}^*$, $var(\eta_n^*) = \sigma_{\eta_n}^*$;
ϵ_n^* and η_n^* are independent.

For each snapshot, we can find one estimation of the MS location, denoted by (x^n, y^n), which are different from the notation above. That is

$$x^n = f^n(x_0, y_0; s_{n1}, \cdots, s_{nm_n}; l_{n1}, \cdots, l_{nm_n}) + \epsilon^n,$$

$$y^n = g^n(x_0, y_0; s_{n1}, \cdots, s_{nm_n}; l_{n1}, \cdots, l_{nm_n}) + \eta^n,$$

where

f^n and g^n are continuous functions in the definition field;
ϵ^n and η^n are random variables, $E(\epsilon^n) = 0$, $E(\eta^n) = 0$, $var(\epsilon^n) = \sigma_{\epsilon^n}$, $var(\eta^n) = \sigma_{\eta^n}$;
ϵ^n and η^n are independent.

Therefore, we have

$$x^1 = f^1(x_0, y_0; s_{11}, \cdots, s_{1m_1}; l_{11}, \cdots, l_{1m_1}) + \epsilon^1,$$
$$y^1 = g^1(x_0, y_0; s_{11}, \cdots, s_{1m_1}; l_{11}, \cdots, l_{1m_1}) + \eta^1,$$
$$\cdots$$
$$x^n = f^n(x_0, y_0; s_{n1}, \cdots, s_{nm_n}; l_{n1}, \cdots, l_{nm_n}) + \epsilon^n,$$
$$y^n = g^n(x_0, y_0; s_{n1}, \cdots, s_{nm_n}; l_{n1}, \cdots, l_{nm_n}) + \eta^n,$$

and we have some idea about the structure of the series $\{x_n, y_n\}$, which is the combination of these series. In other words, we have two threads about the combination, if we consider every snapshot has the same contribution to the estimation, then we have a common average structure, otherwise, we have a weighted average structure.

We present two structures for the Statistical Estimation: one for the common average, and one for a special weighted average structure.

The average structure

It has the same contribution to the estimation for each snapshot.

That is

$$x_n = \frac{1}{n}\sum_{i=1}^{n} x^i, y_n = \frac{1}{n}\sum_{i=1}^{n} y^i.$$

$$f_n = (f^1 + f^2 + \cdots + f^n)/n,$$
$$g_n = (g^1 + g^2 + \cdots + g^n)/n,$$
$$\epsilon_n = (\epsilon^1 + \epsilon^2 + \cdots + \epsilon^n)/n,$$
$$\eta_n = (\eta^1 + \eta^2 + \cdots + \eta^n)/n,$$

The Special weighted average structure

$$x_1 = x^1, y_1 = y^1,$$
$$x_2 = (x^2 + x_1)/2, y_2 = (y^2 + y_2)/2 \ ,$$
$$\cdots$$
$$x_n = (x^n + x_{n-1})/2, y_n = (y^n + y_{n-1})/2$$

where

$$f_1 = f^1, g_1 = g^1, \epsilon_1 = \epsilon^1, \eta_1 = \eta^1;$$
$$\cdots$$
$$f_n = (f^n + f_{n-1})/2,\ g_n = (g^n + g_{n-1})/2,\ \epsilon_n = (\epsilon^n + \epsilon_{n-1})/2,\ \eta_n = (\eta^n + \eta_{n-1})/2.$$

The average structure fits for the case that each snapshot has the same contribution to the estimation, so it puts the same weighting into each snapshot. And the special weighted structure will fit for the different contribution given by the each snapshot and consider the more recent estimation will give more important information for the estimation, and it exactly give one half of the contribution to the estimation.

3 Simulation Results

We use the field test data of HK to test the effect of the Statistical Estimation. And the data used here are same in [1] and [2].

Since the field test data are sampled in one time for one region, and we consider each snapshot will have the same contribution to the estimation, so we choose the average structure of the Statistical Estimation to compute these data. If the data received in one region are sampled in different time , we suggest to choose the weighted average structure, since the effect of information will be declined as the time pass as away.

One more paragraph describing the results as presented in Table 1.

4 Summary and Future work

In this study, we have presented a filter to handle the signal fluctuation problem, namely, the Statistical Estimation. We also give two special structures of the Statistical Estimation. Simulation

Algorithm	Average (m)	Standard Deviation
CG	335.11	319.61
CT	321.62	229.06
GEPM	285.27	309.80
IGEPM	282.68	309.05
SE of GEPM	260.26	285.64
SE of IGEPM	257.94	288.53

Table 1: Comparing results among different algorithms.

results show that the Statistical Estimation has improved the property of the MS location. It can reduce the effect of signal fluctuation to the MS location, thus, it can provide a more accurate estimation for the location service.

The Statistical Estimation can provide a more accurate estimation than the Geometric Algorithm [1] and Iterative Algorithm [2], since the Statistical Estimation uses all the information we have to find out the MS location, while the Geometric Algorithm and the Iterative Algorithm just use the information from one snapshot. Hence, the Statistical Estimation is a feasible filter to handle signal fluctuation problem.

Since all the methods and models from our research group until now just consider the 2-D situation, that is to say, the estimation we provided is just a 2-D solution. But we live in a 3-D real world, therefore, it is only natural to work on a 3-D estimation for location services. By our research, the height of the antenna is also taking an important part for the MS location in the hilly terrains in Hong Kong. So one of our future work is to extend the EPM into a 3-D model in order to provide a 3-D estimation to meet the real world situation in location estimation.

References

[1] J.Y. Zhou, Kenneth M. Chu, Joseph K. Ng, Providing Location Services within a Radio Cellular Network using Ellipse Propagation Model, Submitting to AINA 2005.

[2] J.Y.Zhou, Joseph Kee-Yin NG, An Iterative Approach to Mobile Location Estimation, technical report 2004.

[3] Joseph Kee-Yin Ng, Stephen Ka Chun Chan, and Shibin Song, A Study on the Sensitivity of the Center of Gravity Algorithm for Location Estimation, Hong Kong Baptist University Technical report, May 13, 2003, available at URL: http://www.comp.hkbu.edu.hk/tech-report/tr03014f.pdf.

[4] Kenny K.H. Kan, Stephen K,C. Chan, and Joseph K. Ng, A Dual-Channel Location Estimation System for providing Location Services based on the GPS and GSM Networks, Proceedings of The 17th International Conference on Advanced Information Networking and Applications(AINA 2003), pp. 7-12, March 27-29, 2003, Xi'an, China.

[5] Joseph K. Ng, Stephan K. Chan, And Kenny K. Kan, Location Estimation Algorithms for Providing Location Services within a Metropolitan area based on a Mobile Phone Network, Proceedings of The 5th International Workshop on Mobility Databases and Distributed Systems(MDDS 2002), pp. 710-715, September 2-6, 2002, Aix-en-Provence, France.

[6] Kenneth M. Chu, Karl R.P.H. Leung, Joseph K. Ng, and Chun H. Li, Locating Mobile Stations with Statistical Directional Propagation Model, Proceedings of the 18th International Conference on Advanced Information Networking and Applications (AINA 2004), pp. 230 — 235, March 29-31, 2004, Fukuoka, Japan

[7] Teemu Roos, Petri Myllymaki, Herry Tirri, A Statistical Modeling Approach to Location Estimation, IEEE Transactions on Mobile Computing, Vol1, No1, pp.59-69, January-March 2002.

[8] Kaveh Pahlavan, Prashant Krishnamurthy, Principles of Wireless Networks a Unified Approach, Pearson Education, Inc. 2002.

[9] P.Bahl, V.N. Padmanabhan, A. Balachandran, Enhancements to the RADAR User Location and Tracking System, Technical Report MSR-TR-00-12, Microsoft Research, Feb.2000.

[10] N.Bulusu, J.Heidemann, and D.Estrin, GPS-Less Low Cost Outdoor Localization for Very Small Devices, IEEE Personal Comm., Vol.7, no.5, pp.28-34,2000.

[11] T.S.Rappaport, J.H.Reed, B.D.Woerner, Position Location Using Wireless Communications on Highways of the Future, IEEE Comm. Magazine, Vol.34, pp.33-41, 1996.