

参赛密码 _____
(由组委会填写)



“华为杯”第十三届全国研究生 数学建模竞赛

题 目 基于无线通信基站的室内三维定位问题

摘 要:

随着无线网络和移动互联网的蓬勃发展,提供基于地理位置信息的服务(Location Based Service, 简称 LBS)已经成为最具市场前景和发展潜力的业务之一。从传统的 GPS 导航,到大众点评、微信等基于地理位置的消费信息服务和社交软件,实现其功能的基础就是要通过手机、导航仪等终端设备收发信号,来获得距离、角度等测量信息,并利用定位算法将这些测量信息转换成坐标信息。

基于无线移动通信网络的定位是以获取用户手持终端(包括手机或者平板等设备)的位置为目标。而达成这一目标的手段是通过测量无线电信号的强度、传播时间、到达角等物理指标,并将其转化成终端与基站之间的距离、角度等信息,最终利用定位算法将距离、角度等信息转化成终端的坐标信息。

本文建立了复杂环境下无线基站定位模型,体现了 NLOS 和 LOS 环境下的各个特征。首先采用 CHAN 算法,加权平方以及最小二乘法评价指标对场景进行了快速识别,实现了快速定位,然后采取基于 NLOS 环境误差估计模型下的迭代方法建立了将定位误差缩小至 XY 水平面 0.1 米以下, Z 垂直高度 3 米以下的精确模型,实现了精确定位。模型具有自适应判断场景的功能,同时,我们发现了在 NLOS 环境下定位坐标的误差呈瑞利分布。

问题 1: 模型建立:基于 LOS 和 NLOS 混合环境下利用 TOA 和基站坐标计算手持终端三维信息

该部分要求我们首先对数据包里不同环境模型下的不同基站数及不同手持终端数的场景进行建模。由于基站 LOS 或 NLOS 环境下,我们需要求出手持终端的精确坐标,然而这个坐标的求解需要考虑到用户钟差问题,以及多径效应或是周围噪声所带来的误差影响。在求解时我们利用较为经典的基于 TDOA 的 CHAN 算法进行建模,再另外加上误差模型修补迭代算法修改以提升计算精度。经过分析计算,我们建立了精确的数学模型,即便在 NLOS 较多的情况下,也可以得到距离在 3 米以内的误差,且手

持终端的水平面坐标 XY 误差在 0.1 米以内。并且，我们通过误差估计模型中一个重要参数 K 对终端所处环境进行判断，可以自适应环境。

问题 2：使用尽可能少的基站数目，实现近似最优的三维定位精度

该部分要求在给定 30 个基站的情况下使用尽可能少的基站进行较为准确的定位。在不存在 NLOS 情况下我们可以通过数学模型得知，4 个基站即可完成较为精确的手持终端三维建模。然而在存在 NLOS 情况下，我们需要具体视 NLOS 数目而定，在 case10 到 case20 中，每一种模型都有不同的 NLOS 数目及环境情况。针对这种情况，我们认为 TOA 测的距离越远，存在多径传输的可能性越大。因此我们将优先使用距离终端最近的基站进行定位，使用问题 1 建立的模型，可以保证使用尽可能少的基站数目实现近似精确的定位过程。

问题 3：移动终端的运动轨迹研究

该部分研究了移动终端的运动轨迹问题，由于移动终端运动速度相较于光速可以忽略不计。因为在每一段轨迹的计算时，我们依旧可以按照之前建模来计算当前位置，并不用考虑相对运动问题，只是在不同位置中，移动终端所处于的 NLOS 或 LOS 环境在变化，我们需要每一步重新建模以求解运动轨迹问题，因为本文建立模型十分精确，故不需要再进一步卡尔曼滤波算法等。尤其在平面二维坐标系下，我们模型的定位误差在 0.1m 以内，可以完成对移动终端的精确定位，进而描绘出准确的运动轨迹。

问题 4：平均链接度数与定位精度之间的关系

该部分针基站通信范围有限与手持终端定位结果做出了建模，要求我们在求解出区域范围内可以定位的手持终端总数，并计算平均“连接度数” λ 与定位精度之间的关系，我们在这里提出了一个“终端区域精度贡献度”这一概念，将整个区域精确度用贡献度来表示，求得区域贡献度与连接度数之间的关系。由于基站通信半径的限制，手持终端被探测到的概率被大大降低，进而影响了定位精度。我们采取的策略是将情况进行分类，探测到手持终端的数目越少，我们设计粗略定位的方法；数目越多，便采用问题 1 中的模型进行精确定位。

无线通信基站的三维定位建模

1. 问题重述

1.1 提出问题

1896 年,意大利电气工程师伽利尔摩马可尼用电磁波进行约两公里距离的无线电通讯实验获得成功,自此打开了现代无线电通信的大门。无线通信使用的灵活性极大地方便了用户。近年来无线通信,特别是移动通信,一直以惊人的速度迅猛发展,已成为带动全球经济发展的主要高科技产业之一,并对人类生活及社会发展产生了巨大的影响。

基于无线电通信基站的定位问题研究,在科研和工业界都吸引了极高的关注。一方面,定位问题与统计信号处理、最优估计理论、优化算法等诸多领域都有密切的联系,诸如数据拟合、最小二乘估计、半正定规划、流形学习等诸多数学工具都能够被用于求解上述问题。另一方面,工业界对于如何高精度地在现有通信设备上完成上述功能也表现出了浓厚的兴趣,我国除了业已广泛部署商用的北斗导航系统之外,也在积极推进基于室内室外融合定位的羲和导航系统。我们相信,基于通信基站的定位系统,将会成为羲和导航系统有力的技术手段。

求解分析基站定位相关问题的有创新性和可实现性强的算法,都将有可能被快速部署到现代商业通信网络中,带来巨大的社会和经济效益。

与有线通信不同,无线通信信号以无线电波的形式来传播,其主要有三个特点:1、传播的开放性,2、接收环境的复杂性,3、通信用户的随机移动性。从发射机到接收机这一传播过程包含多种传播方式,从简单的视距传播的直达波(即自由空间波),到遇到各种复杂地形产生的地波(即表面波),更甚至是大气层反射产生的对流层反射波、电离层波等,因此其传播路径非常复杂。其次,因为无线信号在空气中传播,其具有极度的随机性。同时,通信用户随机的移动导致的发射机和接收机的相对运动速也会影响信号的传播,因此,无线电传播模型受多重因素的制约。

无线电信号在大气中从 A 点向 B 点传播时,如果传播过程中存在一个没有遮挡的直达路径,那么,这种传播环境被称为视距传播环境(Line Of Sight propagation,简称 LOS)。这种传播环境如图 1 中的左图所示。如果在传播过程中,由于建筑物或树木的遮挡、反射、折射等物理现象,使得从 A 点到 B 点之间存在多条无线电信号的传播路径,这种环境被称为非视距传播环境(Non-Line Of Sight,简称 NLOS)。需要注意的是,在 NLOS 传播环境中,仍然可能存在着无线电波的直达路径,只不过相比于 LOS 传播环境,在 NLOS 环境下因为遮挡、反射和吸收等损耗,信号强度会在传播过程中变得较弱。

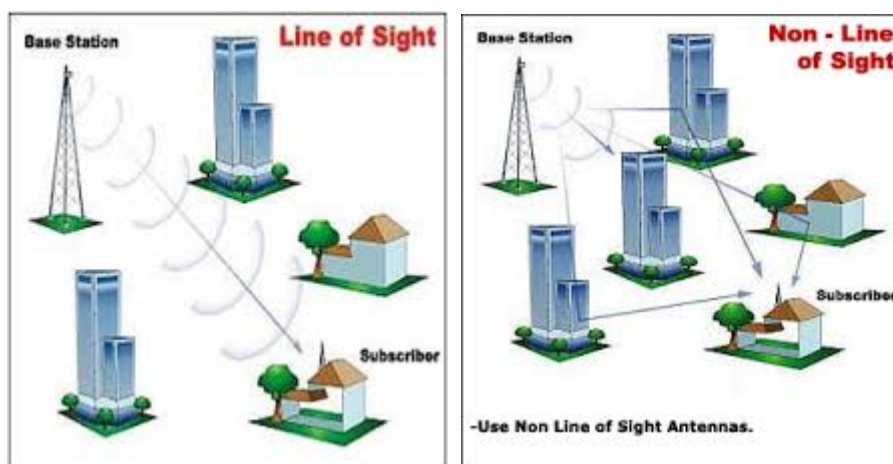


图 1-1 LOS 径与非 LOS 径示意图

当无线电信号在基站与用户手持终端之间互相传播时，就可以计算基站与手持终端之间的距离，一种常用的测量方式是记录无线电信号从手持终端发出，直到基站接收到信号为止的无线电信号传播时间，将时间乘以无线电信号的传播速度，即得到基站与终端之间沿某条路径的距离。其中，信号在基站与终端之间的传播时间，被称为无线电信号的到达时间（Time Of Arrival，简称 TOA）。

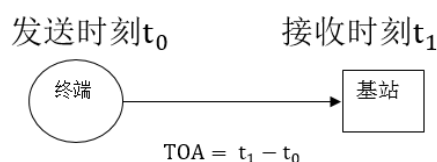


图 1-2 TOA 示意图

准确测量 TOA 所需的前提条件是基站计时与终端计时所使用的时钟是同步的。以图 1-2 为例，当基站与终端在同一个“时间坐标系”里，真实 TOA 等于接收时刻 t_1 减去发送时刻 t_0 。由于电子器件的工艺原因，基站与终端的时钟可能是不同步的存在钟差问题。

在建模过程中，以下两个因素也应该被考虑到：

（1）用基站测量的终端信号时，一个很重要因素就是基站侧接收到的信号干扰比值（SINR），也称为信噪比，定义为

$$SINR = \frac{\text{有用信号强度}}{\text{干扰信号强度} + \text{噪声信号强度}}$$

（2）室内环境下，由于反射频繁发生，会形成无线电波的多径传播（multi-path propagation），从而导致虽然距离很近接收到的信号强度却波动剧烈。

1.2 问题要求

本研究旨在挖掘、提炼和聚合实测基站数据中的各种特征，进而用数学的语言进行基站定位建模，并验证其在不同场景或不同地理位置下的适用性，具体包括：

问题一：给定 10 组 LOS 或 NLOS 传播环境下从手持终端到基站的 TOA 测量数据和所有基站的三维坐标（对应附录中编号为 case001_input.txt 到 case010_input.txt 的文件），请根据这些测量数据计算出终端的三维坐标。（请给出详细的建模分析，建模过程中建议考虑测量模型、误差分析等内容。）

问题二：给定 10 组 TOA 测量数据和所有基站的三维坐标（对应附录中编号为 case011_input.txt 到 case020_input.txt 的文件），请设计算法，使用尽可能少的基站数目，实现近似最优的三维定位精度。

问题三：给定 5 组对处于移动过程中的终端采集到的 TOA 数据（对应附录中编号为 case021_input.txt 到 case025_input.txt 的文件），请设计算法计算出终端的运动轨迹。（此时，编号为 case021_input.txt 到 case025_input.txt 的文件中，只记录一个终端的 TOA 数据，并且是这一个终端在运动轨迹中多个位置上的 TOA 数据。）

问题四：在前述 3 问中，都是假设给定区域内终端到每一个基站的距离都是可知的，但事实上，基站的通信半径是有限的，因此，只有在基站通信半径覆盖范围内的终端才有可能测到自身到基站的距离。而一个终端只有获得它与足够数目的基站之间的距离测量值，才能完成定位。假设每个基站的通信半径为 200 米（超过范围虽然有测量数据，但无效）。请根据给定的 5 组测量信息数据集（对应附录中编号为 case026_input.txt 到 case030_input.txt 的文件），设计算法寻找出可以被基站定位的所有终端。进一步，回答如下问题：每一个场景中（对应着 case026_input.txt 到 case030_input.txt 五个文件中的一个），定义终端的平均“连接度数” λ 为

$$\lambda = \frac{\text{所有可以测到终端到基站之间的连接数}}{\text{终端数}}$$

请建立模型分析连接度数 λ 与定位精度之间的关系。

2. 基本假设与符号说明

2.1 基本假设

- 不考虑介质改变，无线电传播速度始终为光速 c ；
- 接收机信噪声为高斯白噪声；
- 假设噪声 $u[k, n]$ 对于任意 k 和 n 都是独立分布的，其功率谱密度为常数。

2.2 符号说明

表 2-1 符号及其说明

符号	符号说明
T_0	手持终端发射信号时间
T_l	基站接收时间
TOA	Time of arrival
TDOA	Time Difference of Arrival
λ	平均“链接度数”
c	光速
M	基站数目
Z	手持终端三维坐标向量
x	手持终端 x 轴坐标
y	手持终端 y 轴坐标
z	手持终端 z 轴坐标
τ	手持终端到基站与基站的时间到达差值
r	基站与手持终端的几何距离
Q	TDOA 的协方差
n	噪声矢量
ξ	随机变量
σ	标准差
μ	正态分布均值
ε	误差常数

3. 问题分析

由于手持终端存在钟差问题，且 3 纳秒的钟差会影响 1 米左右的误差，因此不能忽略不计，我们在建立模型方程时需要求解四个未知量。但是由于题目中基站数目较多，很容易得到四个方程。

本题目是求解超静定方程，我们利用 CHAN 算法求解。其优点是方程简单，计算速度快。但是 CHAN 算法没有考虑 LOS 和 NLOS 的区别，因此在得出手持终端位置大致坐标后，我们需要进一步利用大致的几何距离长度与 TOA 传播时间进行比较从而求解出一个合理的精确位置解。

对于求解最少基站定位数目，我们可以利用离基站距离较近的 TOA 数据建模，依次向外增加基站定位数目，直到出现定位结果变化值小于一阈值时及可以大致确定最小定位基站数目。

运动轨迹分析主要是通过之前的基站建模将每一次的坐标记录下来，由于单点定位的运动轨迹存在轨迹误差，因此我们可以最小二乘法将轨迹拟合平滑，出一个最终轨迹值。

平均链接度的计算可以通过假定基站均为有效数据后筛选而得出，在计算平均链接度与定位误差比较中，我们可以得出平均链接度越大，定位误差越小。

4. 问题 1：基站定位模型建立

4.1 LOS 及 NLOS 环境下的定位算法

4.1.1 TD0A 定位计算方法

首先需要获取到 TD0A 的测量值，而 TD0A 参数测量值的获取主要方法是直接求两个基站信号的到达时间（time of arrival, TOA）只差。一旦获得 TD0A 参数测量值就可以采用 TD0A 定位。TD0A 定位法是通过测量点播从手持终端传播到两个基站的时间差来确定手持终端的位置，由这个传播时间差对应的距离差可以得到一条以相应的两个基站为焦点的双曲线。要想得到移动台的估计位置值，只需要测量到两组 TD0A 差值，其定位原理如图 4-1 所示。

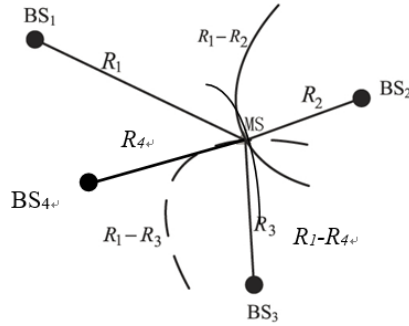


图 4-1 TD0A 定位原理图

在一个三维空间中的一组 $M \geq 5$ (在基站数目足够的情况下，暂不考虑基站数为 4 的情况，因为其定位性能差)。基站坐标分别为 $Z_i = [x_i \ y_i \ z_i]^T$, $i = 1, 2, \dots, M$ ，假设需要估计的手持终端位置为 $Z = [x \ y \ z]^T$ ，将第一个基站作为参考点，用 r_i 表示手持终端到基站 i 的测量距离， r_i^0 为 r_i 的真实值， $r_{i,1}$ 表示手持终端到参考基站 i 与参考基站 1 的距离差，则有：

$$r_{i,1} = c\tau_{i,1} = r_i - r_1, i = 2, 3, \dots, M \quad (4-1)$$

式中， $r_i = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}$, $i = 1, 2, \dots, M$ ， c 为电波传播速度， $\tau_{i,1}$ 是手持终端到基站 i 与基站 1 的时间到达差值。将上式线性化处理可得：

$$(r_{i,1} + r_1)^2 = x_i^2 + y_i^2 + z_i^2 - 2x_i x - 2y_i y - 2z_i z + x^2 + y^2 + z^2 \quad (4-2)$$

令 $K_i = x_i^2 + y_i^2 + z_i^2$, $x_{i,1} = x_i - x_1$, $y_{i,1} = y_i - y_1$, $z_{i,1} = z_i - z_1$ ，即有下式：

$$x_{i,1}x + y_{i,1}y + z_{i,1}z + r_{i+1}r_1 = \frac{1}{2}(K_i - K_1 - r_{i,1}^2) \quad (4-3)$$

将 x, y, r_i 看成未知数，那么式（3）就是一组线性方程组，只要求解出该方程组的值就可以获得手持终端的估计位置。

4.1.2 LOS 与 NLOS 环境下的误差模型

为了获得较高的定位精度，我们需要在计算中引入适当的方法削弱误差对于结果的影响，而这一切的前提是我们对于误差的模型有了足够的了解，才可以将误差模型引入计算过程中，进而提高定位精度。

在 LOS 环境下，由于在传播过程中手持终端与基站之间存在一个没有遮挡的直达路径，因而基站会直接接收到手持终端发射的电磁波，而对于其他路径传播过来的能量则远低于直接传播的电磁波能量，故一般情况下信噪比较高，系统通常会正确的采样并读取 TOA。在这样的环境下，TOA 的测量误差主要是由钟差和噪声引起。这两类误差都是典型的高斯分布的模型，属于独立同分布的误差模型。故可以将其合并为一种均值为 μ_{LOS} ，标准差为 σ_{LOS} 的高斯分布。针对这样的分析，我们结合题目中提供测试数据进行了验证：

经过对每个测试数据的仔细分析，在本次提供的测试数据中，sample_case002 的使用场景为 LOS 环境较多的场景，通过将提供的移动终端位置信息与基站位置信息进行结合，反推出每个测距过程的误差。可以根据误差大小和误差频数画出频度直方图，如图 4-2 所示。根据得到的信息对误差模型进行拟合，如图中蓝线所示，其较符合高斯分布的模型，进而对这个模型进行非参数假设检验，经过计算其符合高斯分布，故可以认为在 LOS 环境下的误差模型符合高斯分布。特别是在进行基于 TDOA 的参数求解过程中，因为对 TOA 进行求差进而得到的 TDOA 数据，因而 LOS 环境下的误差符合均值为 0，标准差为 σ_{LOS} 的高斯分布。

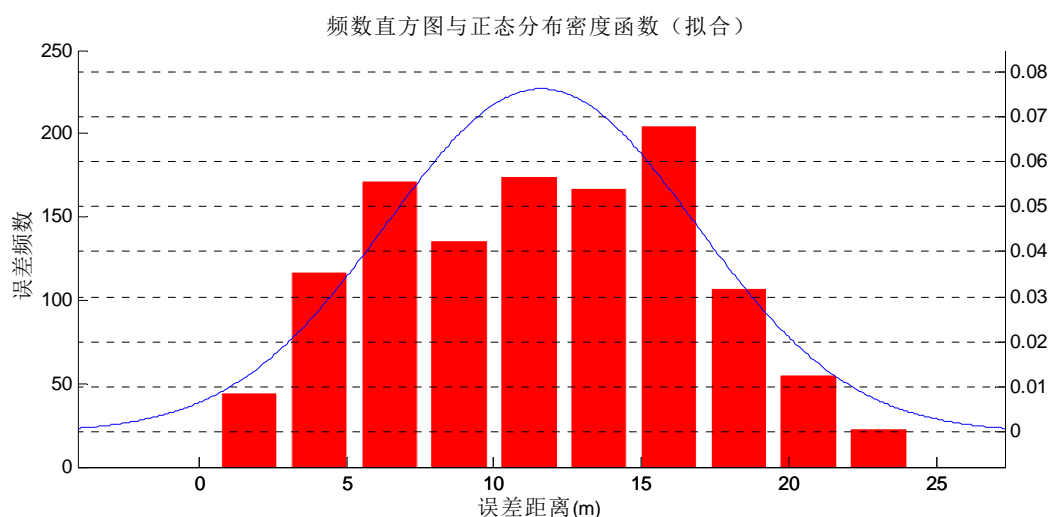


图 4-2 Sample_case002 场景中的误差频数直方图

而在 NLOS 环境下，情况变得复杂许多，其不仅包含 LOS 环境下的误差模型，同时还有其他因素造成了误差。首先，在 NLOS 环境下，信号传播的通道不只有一条，进而其各个各条路径传播的能量并不相同，通常直达路径上会产生遮挡、反射和吸收等损耗，进而到达时的信号能量较小，此时的信噪比很低，会被检测系统认为是噪声而忽略；经过墙面、物体反射到达基站的信号能量可能较大，足够基站检出，进而产生误差使 TOA 的检测值不能反映真实的基站到移动终端的距离。其次，考虑到天线的方

向以及人体运动遮挡等不稳定因素，即便是同一位置上移动终端，其在定位时检测到的信号通路也并不一定会稳定，产生很多随机误差。此时，考虑上误差可以将式(4-1)写成：

$$\tau_{i,1} = \tau_{i,1}^0 + n_{i,1} + \tau_{NLOSi,1} \quad (4-4)$$

其中， $\tau_{i,1}^0$ 是在 LOS 环境下的 TDOA 值； $n_{i,1}$ 为系统的测量误差，其服从均值为 0 的、方差为 $\sigma_{ni,1}^2$ 的高斯分布； $\tau_{NLOSi,1}$ 为由 NLOS 引起的附加时延误差，其均值为 $\mu_{NLOSi,1}$ ，方差为 $\sigma_{NLOSi,1}^2$ 。

下面将分析 NLOS 环境下的误差模型。与分析 LOS 模型类似，通过分析测试数据，发现 sample_case001 的数据中包含一定量的 LOS 与 NLOS 环境数据。进而利用测试数据提供的终端位置，算出在传播中的误差频数直方图，如图 4-3 所示。

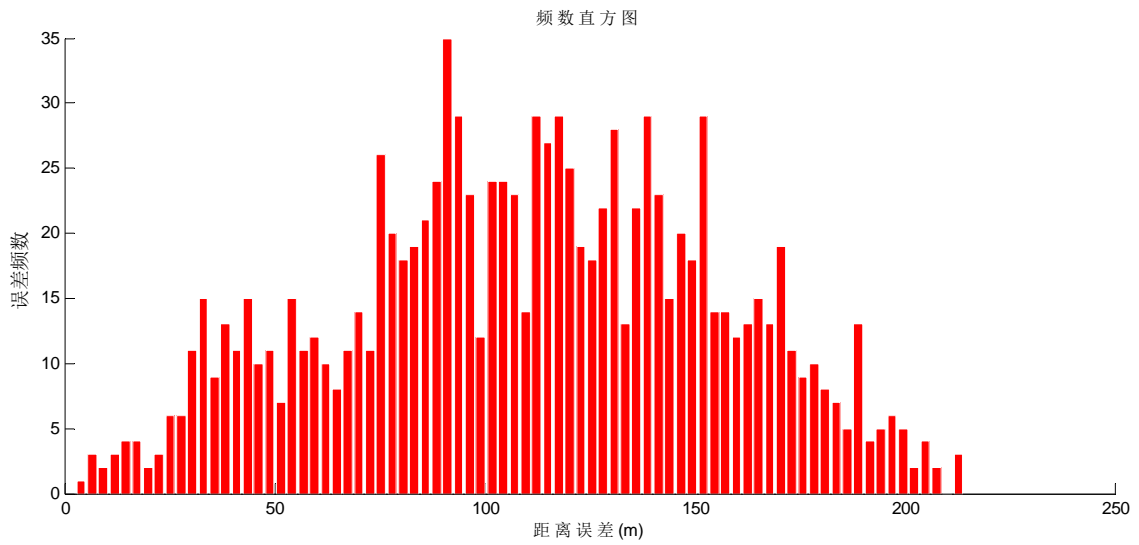


图 4-3 Sample_case001 场景中的误差频数直方图

从图中我们可以粗略的看出在 40 左右有一个峰值，而在 120 左右又出现一个峰值。针对这一模型，我们可以认为存在两种误差模型的叠加，结合之前对于 LOS 环境下的误差模型的讨论，可以认为误差较小的峰是 LOS 的高斯分布模型，而右半部分的模型则相对复杂一些。

首先，我们考虑依照之前的思路，将右半部分拟合成为一个高斯分布的误差模型。进而再后面的算法中，对测量数据减去该高斯分布的均值，使其成为一个均值为 0 的高斯分布。然而在实践中却发现很多测得的手持终端与基站的距离小于拟合出的均值，进而是这个距离变成负值，十分不合理，即对于较短的测量距离，不应存在非常大的误差。

因此，我们决定考虑误差与距离之间的相关性，并且从理论和实际经验上来看，也是终端与基站间的距离越近，其存在 NLOS 情况概率越低，而距离越远存在 NLOS 情况的概率也越高。因此尝试绘制了 Sample_case001 的误差与终端与基站的距离关系图，如图 4-4 所示。

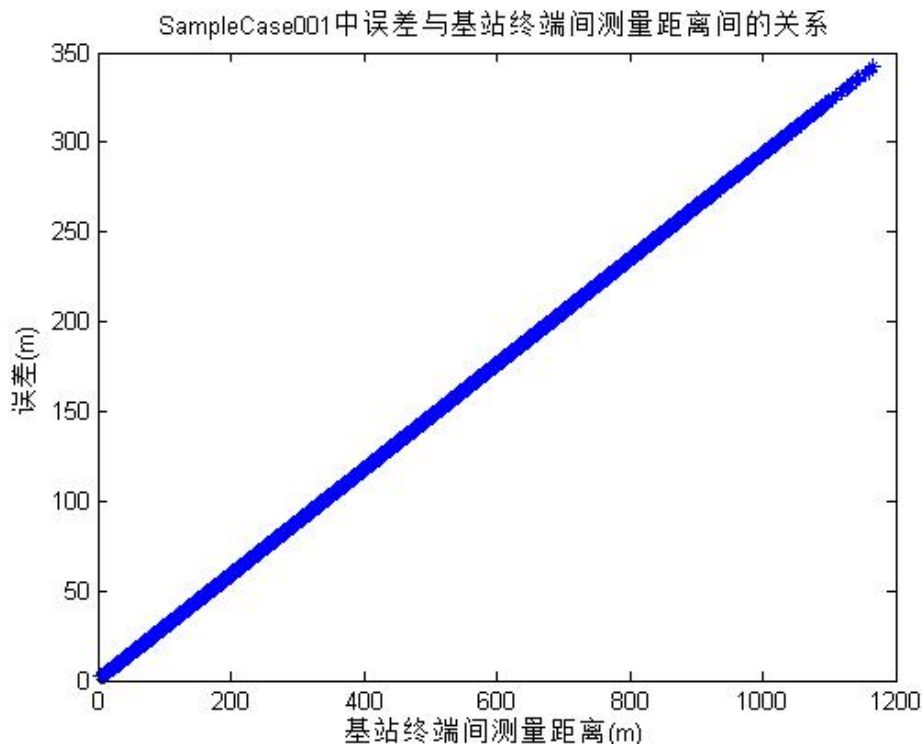


图 4-4 Sample_case001 误差与基站终端测量距离关系图

从图中我们发现，在测试场景 1 中，误差与基站误差间的测量距离间的关系非常明显，图像甚至可以近似拟合为一条过原点的直线，虽然在文献中，对于 NLOS 环境中的误差模型多认为是指数分布，在这里我们可以针对这种特殊情况拟合成为直线，对其进行拟合操作得到两者间的关系：

$$y = Kx + B \quad (4-5)$$

如针对图 4-4，我们拟合的结果为

$$y = 0.2931x + 0.3531 \quad (4-6)$$

通过这个模型可以很简单快捷地对 NLOS 环境下的误差模型进行建立与分析，对于快速去除 NLOS 环境下产生的误差具有非常重要的意义。为此，我们对剩余四组参考案例景进行验证，如图 4-5 所示，其分别为 SampleCase002-005 的误差与基站终端间测量距离的关系图。

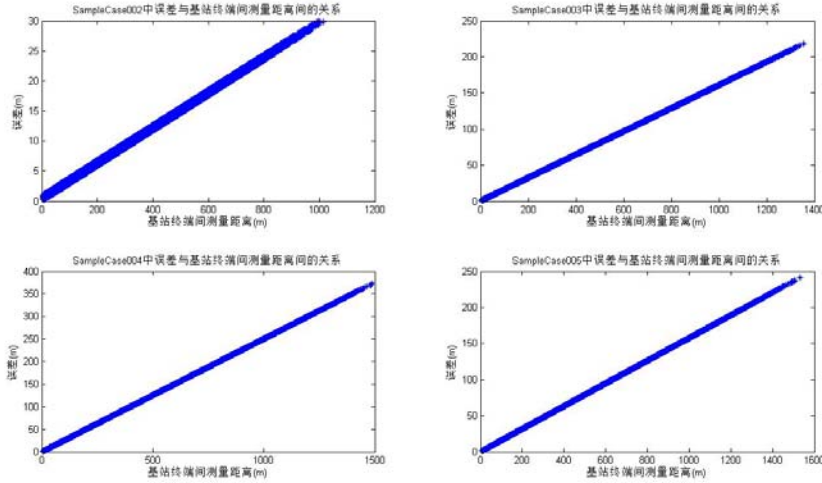


图 4-5 Sample_case002-005 误差与基站终端测量距离关系图

从图中我们可以看出，在这四个测试场景中，误差与基站终端的测量距离成非常好的线性关系，至少在提供的测试场景中是符合线性建模规律的。它们拟合非常容易，可以利用测量出的距离快速对误差进行估计，进而为消除误差提供了有效途径。

4.1.3 LOS 情况下的 CHAN 算法

根据之前在 LOS 环境下建立的误差模型，在求 TDOA 的过程中已经将误差模型转化为一个均值为 0，方差为 σ_{LOS} 的误差。令 $z_a = [z^T, r_1]$ 为未知矢量， z_a^0 为手持终端真实位置所对应 z_a 的值，假定 z_a 的元素间相互独立，则 z_a 的最大似然估计 (maximum likelihood, ML) 为：

$$z_a = \arg \min \{ (h - G_a z_a)^T \psi (h - G_a z_a) \} = (G_a^T \psi^{-1} G_a)^{-1} G_a^T \psi^{-1} h \quad (4-7)$$

式中， $\psi = E[\psi \psi^T] = c^2 B Q B$, $B = \text{diag} \{ r_1^0, r_2^0, \dots, r_M^0 \}$ ， Q 为 TDOA 的协方差，因为 B 中包含手持终端与各个基站发射机之间的距离，即 Ψ 未知，因此想要得到手持终端的位置还需要进一步近似。

假设当手持终端很远时， $r_i^0 (i = 2, 3, \dots, M)$ 与 r^0 (定义距离) 很近，所以 $B \approx r^0 I$ ，则上式可近似为：

$$z_a \approx (G_a^T Q^{-1} G_a)^{-1} G_a^T Q^{-1} h \quad (4-8)$$

当手持终端距离较近时，利用上式可以得到一个初始解来计算 B 矩阵，即 Ψ 的值，然后带入式中可获得第一次加权最小二乘估计值。其中：

$$h = \frac{1}{2} \begin{bmatrix} r_{2,1}^2 - K_2 + K_1 \\ r_{3,1}^2 - K_3 + K_1 \\ \dots \\ r_{M,1}^2 - K_M + K_1 \end{bmatrix}$$

$$G = \begin{bmatrix} x_{2,1} & y_{2,1} & z_{2,1} & r_{2,1} \\ x_{3,1} & y_{3,1} & z_{3,1} & r_{3,1} \\ \dots & & & \\ x_{M,1} & y_{M,1} & z_{M,1} & r_{M,1} \end{bmatrix}, z_a^0 = \begin{bmatrix} x \\ y \\ z \\ r_1^0 \end{bmatrix}$$

$$Q = E(nn^T) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}, n = \begin{bmatrix} n_{2,1} \\ n_{3,1} \\ \dots \\ n_{M,1} \end{bmatrix}$$

其中, n 为噪声矢量, 近似正态分布。

接着利用上面得到的第一次估计值和附加变量 r_1 等已知条件, 重新构造一组误差方程组进行第二次估计, 得到第二次估计位置为:

$$z_a \approx (G_a^T B^{-1} G_a^T Q^{-1} G_a B^{-1} G_a)^{-1} (G_a^T B^{-1} G_a^T Q^{-1} G_a G_a B^{-1} G_a) h \quad (4-9)$$

$$\text{式中, } h = \begin{bmatrix} (Z_{a,1} - x_1)^2 \\ (Z_{a,2} - y_1)^2 \\ Z_{a,1}^2 \end{bmatrix}, G_a = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, Z_a = \begin{bmatrix} (x - x_1)^2 \\ (y - y_1)^2 \\ (z - z_1)^2 \end{bmatrix},$$

$$B = \begin{bmatrix} Z_a(1) - x_1 & 0 & 0 & 0 \\ 0 & Z_a(2) - y_1 & 0 & 0 \\ 0 & 0 & Z_a(3) - z_1 & 0 \\ 0 & 0 & 0 & Z_a(4) \end{bmatrix}$$

通过上述两次最小二乘法可知手持终端的最终位置为:

$$Z = \pm \sqrt{Z_a} + \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \quad (4-10)$$

以上两次 WLS 估算都是基于测量噪声服从零均值的高斯分布前提下, 只有在这前提下才能得到较准确的手持终端估计值, 因为该算法的加权矩阵已经包含系统的测量误差, 这在一定程度上消除了系统误差带来的影响。并且因为在一定程度上消除了系统误差带来的影响。并且因为系统误差是零均值的, 所以结果是无偏估计的。所以如果没有上述前提条件, 那么获得的定位估计结果误差会很大。

NLOS 环境下获得的 TDOA 测量值误差除了系统误差外还有附加延时引入的误差, 因此它比在 LOS 环境下得到的 TDOA 测量误差更大。并且 NLOS 误差均值不是为零的, 而 CHAN 算法中的加权矩阵只考虑系统误差, 无法抵消附加延时带来的误差影响, 所以在这种情况下得到的定位结果误差很大。需要对算法进行改进, 从而能在 NLOS 环境下获得较好的性能。

4.1.4 NLOS 情况下优化的 CHAN 算法

在 NLOS 环境下，由于 TOA 测量存在较大误差，无法直接使用如 4.1.3 章节实现的 CHAN 算法进行精确计算，但是可以利用 CHAN 算法得到一个参考解。统计在当前参考解的情况下的误差与测量距离的关系，如图 4-6 所示是对 case006 求解的过程中，误差与测量距离的关系图，根据 4.1.2 章节拟合的 NLOS 线性误差模型进行拟合，得到一个当前场景下误差与测量距离之间的线性关系式，初步估计出误差模型。

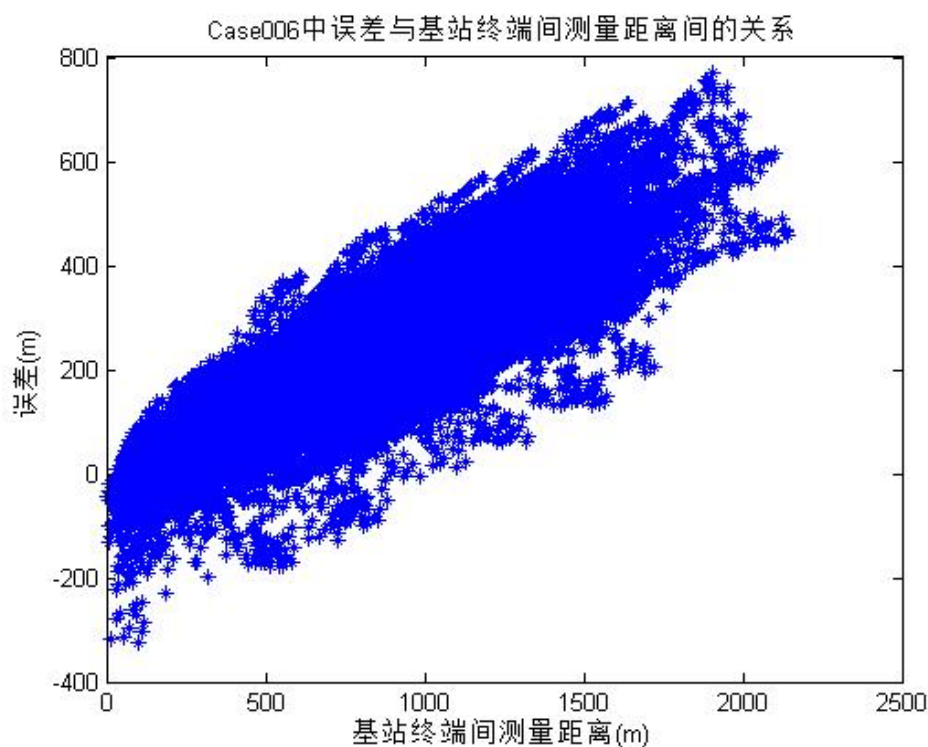


图 4-6 Case006 第一次求解后误差与测量距离的关系

将得到的初步误差模型代入测量数据中，对由基站 TOA 得到的距离信息进行修正。使用修正过的数据重复上述过程，迭代数次后可以得到精度较高的定位信息和误差估计模型。最终的误差估计模型如图 4-7 所示。

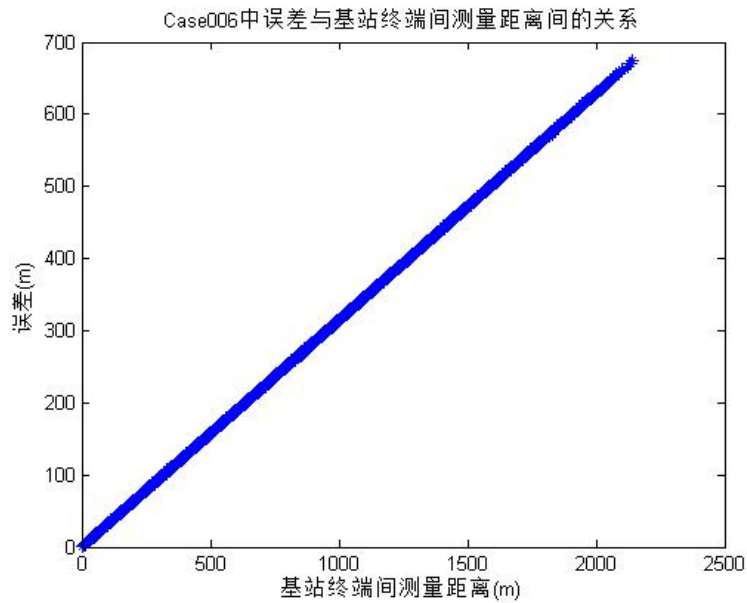


图 4-7 Case006 完成求解后误差与测量距离的关系
整个过程的流程如图 4-8 所示。

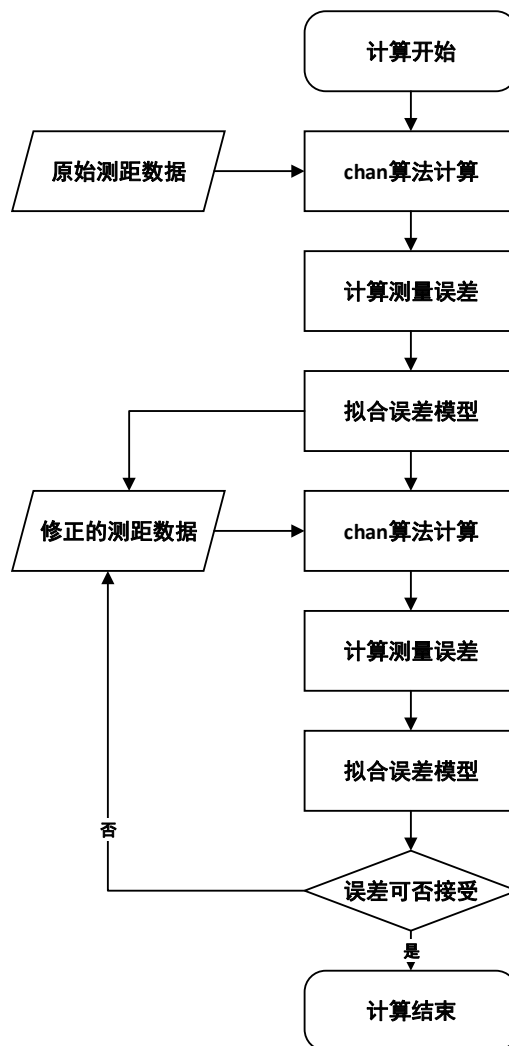


图 4-8 针对 NLOS 环境下优化 CHAN 算法流程图

4.2 手持终端三维坐标计算

通过上一小节分析，我们可以建立较为完善的基于 CHAN 的基站定位模型算法，将其通过 MATLAB 程序予以实现，处理题目中提供的真实环境下手持终端数据 case_input.txt 文件，将其输出得到结果。这里我们只展示 case1, 5, 10 的结果分布图，其他就不再罗列。

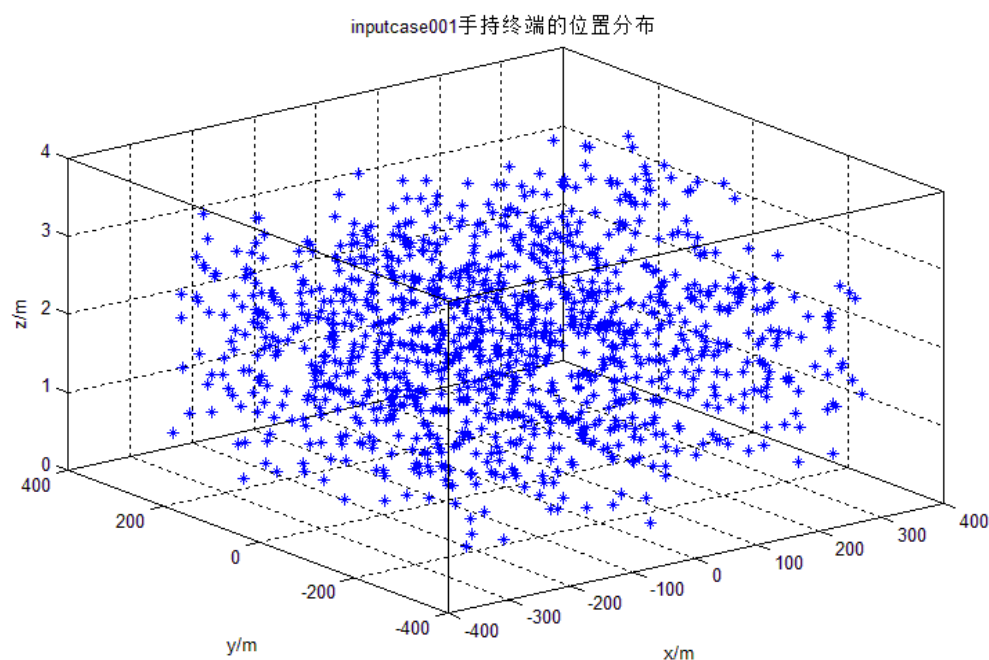


图 4-9 input_case_001 手持终端的位置分布图

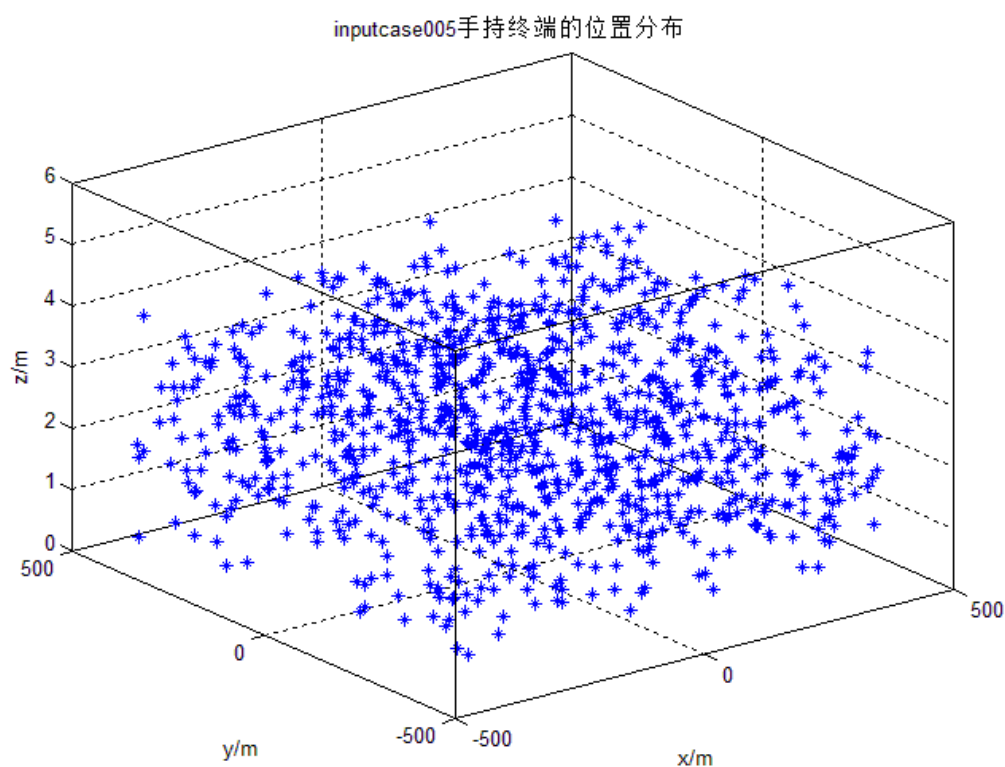


图 4-10 input_case_005 手持终端的位置分布图

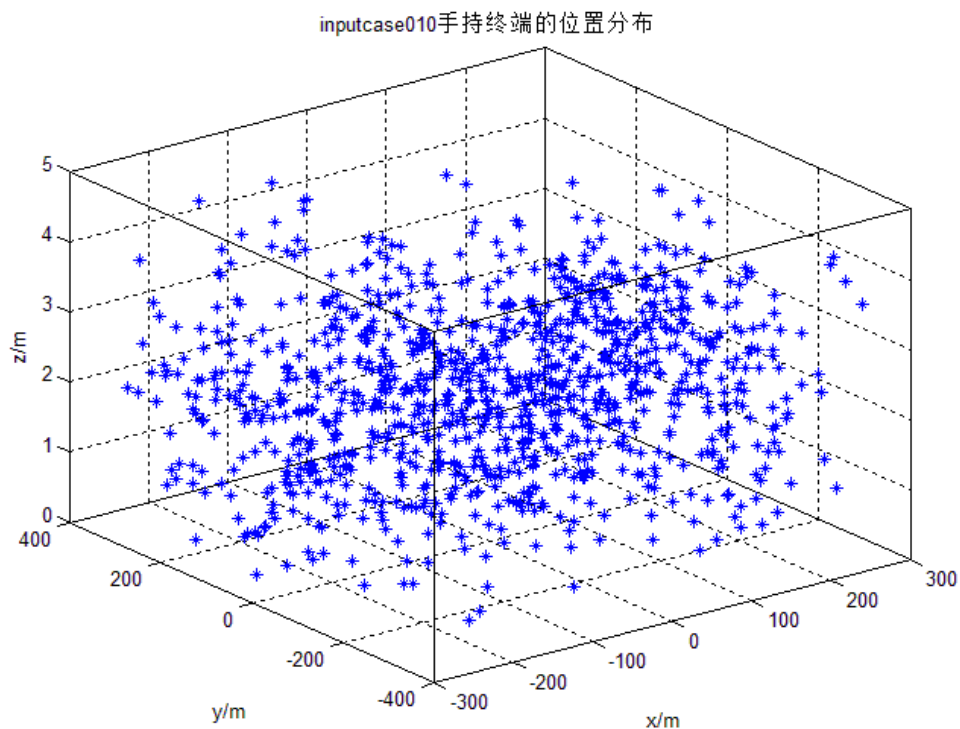


图 4-11 input_case_010 手持终端的位置分布图

我们将十个 case 结果输出成 txt 供专家检查。下面，我们利用算例的信息来分析本模型的准确性及可行性。

4.3 误差评价及分析

我们利用本模型算法将 sample 中的算例进行处理得出一组求解后的手持终端坐标，将此坐标与真实坐标同建立在一个三维场景中，如下图所示。

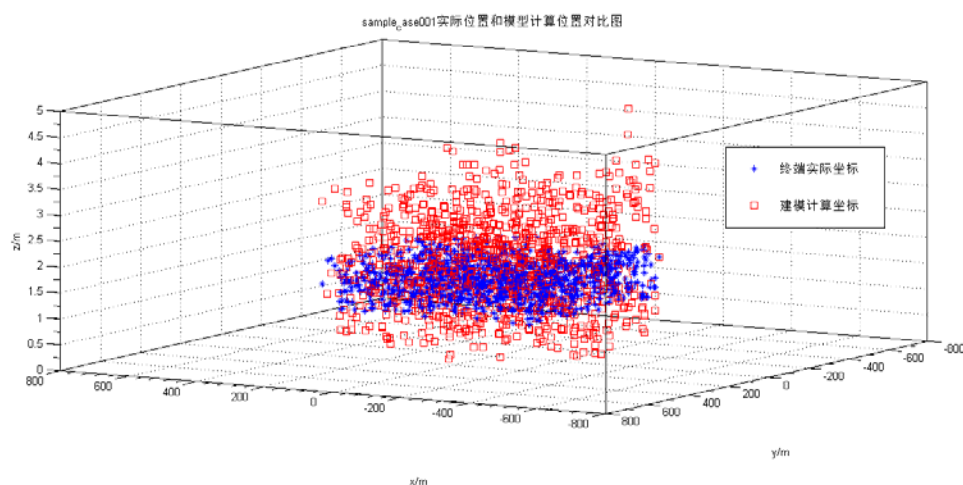


图 4-12 sample_case001 实际位置与模型计算对比图

上图中，蓝色图点代表实际坐标位置，红色图点为本模型下计算出来的坐标位置，我们可以发现计算值十分接近真实坐标。

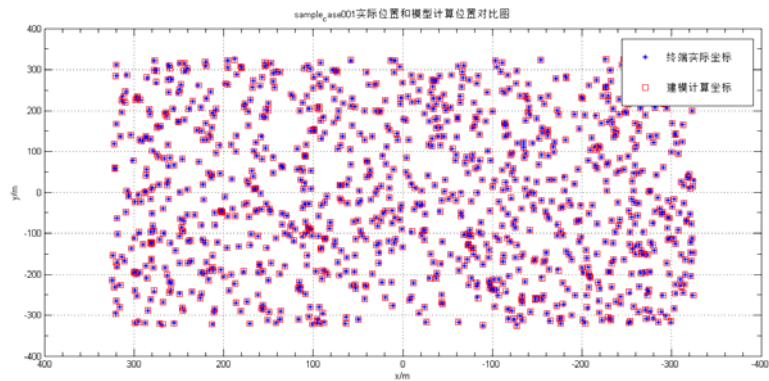


图 4-13 sample_case001 实际水平位置与模型计算对比图

上图表示在三维图俯视情况下的建模坐标与真实坐标差距，我们可以发现，模型误差十分小，平面差距在 0.1m 以内甚至是厘米级别的误差。

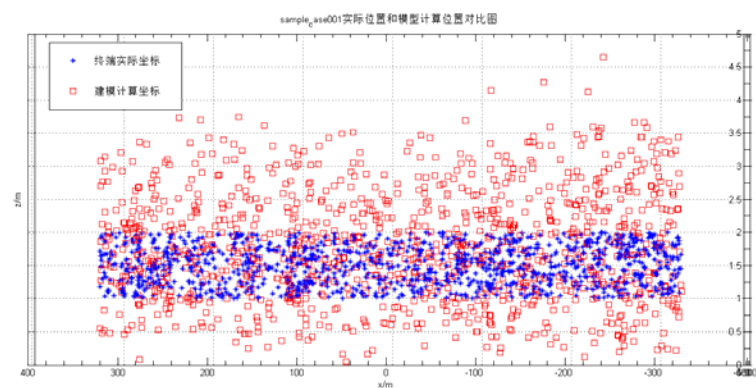


图 4-14 sample_case001 实际垂直高度与模型计算对比图

上图将高度坐标提取出来，我们可以发现，本模型计算的高度存在一定的误差，最大差值可以达到 3 米左右。是造成误差的主要源。

现在我们对误差进行分析:首先对测量点的距离误差特性进行分析，作出不同终端的误差变化曲线，如图所示，没有发现误差变化有和规律。

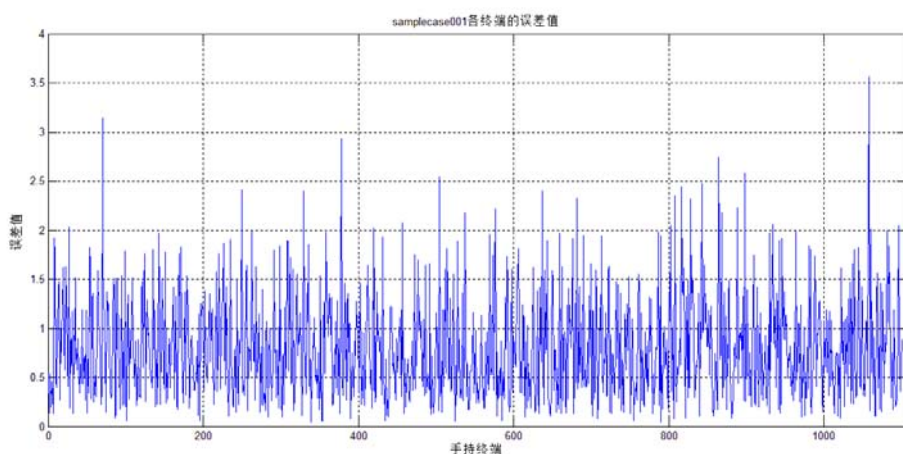


图 4-15 sample_case001 误差统计图

进一步，我们将误差按照最小值 0 和最大值 $\max([0, \max])$ 区间 100 等分，统计误差频率，作出误差频率曲线，如图所示：

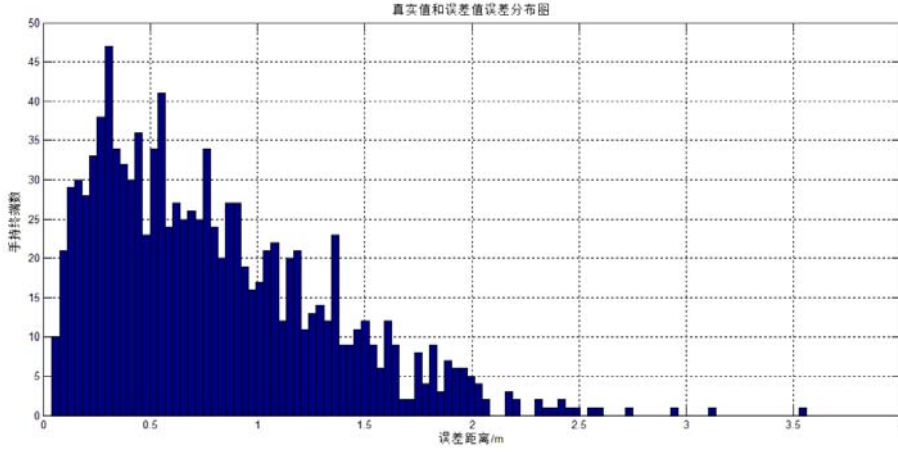


图 4-16 sample_case001 误差统计图

从图 4-16，我们发现误差分布类似于瑞丽分布，为了验证这一结论，我们首先对误差分布的瑞丽分布参数进行极大似然估计。

瑞丽分布的概率密度函数为：

$$f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), x \geq 0 \quad (4-11)$$

则似然函数为：

$$L(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i) = \frac{\prod_{i=1}^n x_i}{\sigma^{2n}} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}\right) \quad (4-12)$$

取对数：

$$\ln L = \sum_{i=1}^n \ln x_i - 2n \ln \sigma - \frac{\sum_{i=1}^n x_i^2}{2\sigma^2} \quad (4-13)$$

再对 σ 求偏导：

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{2}{\sigma} + \frac{\sum_{i=1}^n x_i^2}{\sigma^3} \quad (4-14)$$

令偏导数 $\frac{\partial \ln L}{\partial \sigma} = 0$ ，得到 σ 的极大似然估计为：

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{2n}} \quad (4-15)$$

则可以做出瑞丽分布的概率密度函数，如图所示，与测量值的误差频率曲线较为接近。

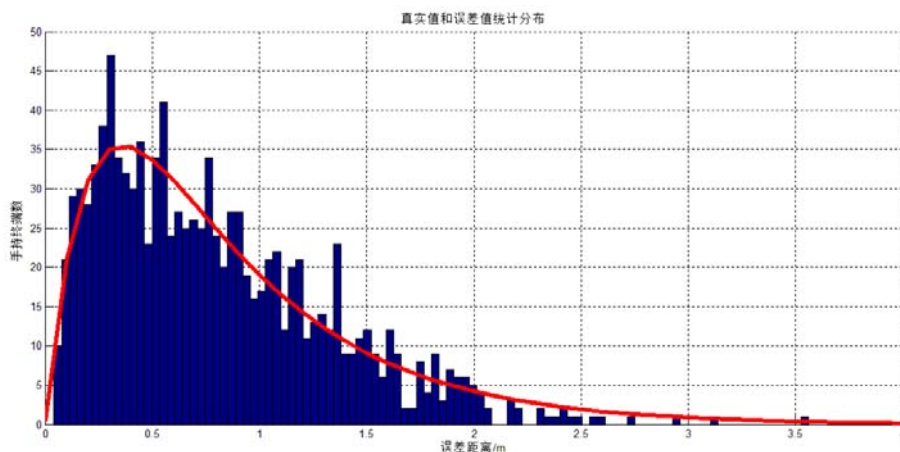


图 4-17 sample_case001 误差统计瑞丽分布拟合

进一步，我们进行 χ^2 拟合优度检验，确定测量误差统计包络是否符合瑞丽分布。我们作如下假设：

H_0 ：包络分布为瑞丽分布， H_1 ：包络不是瑞丽分布

样本区间为 $[0, \max]$ 的 100 等分，区间数 $s=100$ ，设 A_i 表示事件 $(a_i < X < a_{i+1})$ ，统计各区间样本数为 N_i ，样本总数为 n ，我们取统计量 χ^2 ：

$$\chi^2 = \sum_{i=1}^r \frac{(N_i - np_i)^2}{np_i} \quad (4-16)$$

其中， p_i 为测量值 Z_i 服从 H_0 假设分布时，事件 A_i 的概率：

$$p_i = \int_{a_i}^{a_{i+1}} f(x) dx \quad (4-17)$$

对于给定的显著性水平 α ，由 χ^2 拟合优度检验法得：

当 $\chi^2 < \chi_{\alpha}^2(s-r-1)$ 时， H_0 成立：

当 $\chi^2 \geq \chi_{\alpha}^2(s-r-1)$ 时， H_1 成立：

在计算概率 p_i 时， σ 由极大似然估计得到，所以， $r=1$. 在显著性水平的 $\alpha=0.05$ 情况下，各场景测量结果中抽取样本进行计算，所得结果如表所示，均满足瑞丽分布。因此这个我们在现实中对测量值有一个更准确的评估，三维定位测量的位置距离真实的误差满足瑞丽分布。

表 4-1 分布拟合实验结果

Sample_case001	Sample_case002	Sample_case003	Sample_case004	Sample_case005
H0 成立	H0 成立	H0 成立	H0 成立	H0 成立

通过试题提供的测试用例我们可以比较本模型的客观性及误差，我们所建立的模型误差均存在瑞丽分布曲线。在信息的坐标图中我们可以发现本模型建立所计算出来的平面信息 (x, y) 坐标误差值均在 0.1 米以内甚至是厘米级的误差！然而在比较高度误差的时候， z 的误差稍大，在图表中我们可以发现其高度误差均在 3 米以内。这个误差原因也是因为基站分布较为平面化，其空间分布规律只能是在地平面附近，导致基站坐标的 z 值太过于集中，HDOP 值不好，所以计算出来的手持终端高度值误差较大也是因为这个客观原因的问题。在模型建立中不易消除。

4.4 模型自适应性和普适性

从 case001 到 case025 的案例计算下来，我们的基于 NLOS 误差估计模型下的迭代算法能够精确计算所有 case 下的手持终端坐标，且不管是终端移动还是不移动，因此我们的算法具有普适性。

同时，在当前条件下，我们可以针对 case001 到 case025 直接运用 4.1.4 节的迭代法，而且我们的 K 值随案例的改变都能收敛于某一值，说明我们的模型具有自适应环境的特性，即自适应性。case001 到 case025 的最终 K 收敛值见下表

表 4-2 各个案例下的 K 收敛值

案例	Case001	Case002	Case003	Case004	Case005
K 收敛值	0.2647	0.2050	0.2096	0.1572	0.0462
案例	Case006	Case007	Case008	Case009	Case010
K 收敛值	0.3144	0.2483	0.2915	0.0584	0.0059
案例	Case011	Case012	Case013	Case014	Case015
K 收敛值	0.0436	0.0795	0.2140	0.0612	0.0618
案例	Case016	Case017	Case018	Case019	Case020
K 收敛值	0.2008	0.0898	0.2803	0.2656	0.4040
案例	Case021	Case022	Case023	Case024	Case025
K 收敛值	0.3998	0.1356	0.3841	0.3395	0.4292

NLOS 环境下的误差模型估计 $y=Kx+B$ ，从模型中我们可以看到误差 y 与实际距离 x 成正比，其中 K 为正值，那么 K 越大，表示测得的误差就越大。而造成误差的主要来源是钟差和 NLOS 环境下的多径传输。因为我们采用的是基于 TDOA 算法的 chan 算法，钟差带来的误差可以极大地减小，因此误差来源主要是 NLOS 环境下的多径传输。那么 K 越大说明存在的多径传输情况就越多，NLOS 环境越明显，从而导致误差越大。因此，我们可以根据 K 值的大小来判断手持终端的所处室内环境的 NLOS 情况。我们假定：

表 4-3 K 值与环境对应表

K 值	0~0.1	0.1~0.2	0.2~0.3	0.3 以上
环境	空旷	较空旷	较拥挤	拥挤

由以上总结表格我们可以将各个案例室内环境为：

表 4-4 各案例与环境对应表

环境	空旷	较空旷	较拥挤	拥挤
案例	Case5,9,10,11, 12,14,15,17	Case4,22	Case1,2,3,7,8, 13,16,18,19	Case6,20,21, 23,24,25

5. 问题2：最少基站数目求解精确定位值

5.1 场景特征

在本模型中，我们要求，基站在尽可能少的情况下提供手持终端的较精确坐标值。通过以上 TDOA 模型分析，我们知道四个基站方程即可确定一个手持终端三维坐标信息，然而这个坐标的准确性一般很低，其中不仅因为四个基站中有可能包括 NLOS 情况，且环境噪声带来的误差也无法忽略。因此，下文我们介绍一种总结思考出来的算法思路。

5.1.1 基站数目确定思路图

下图为关于本题最少基站数目的思路路线图。

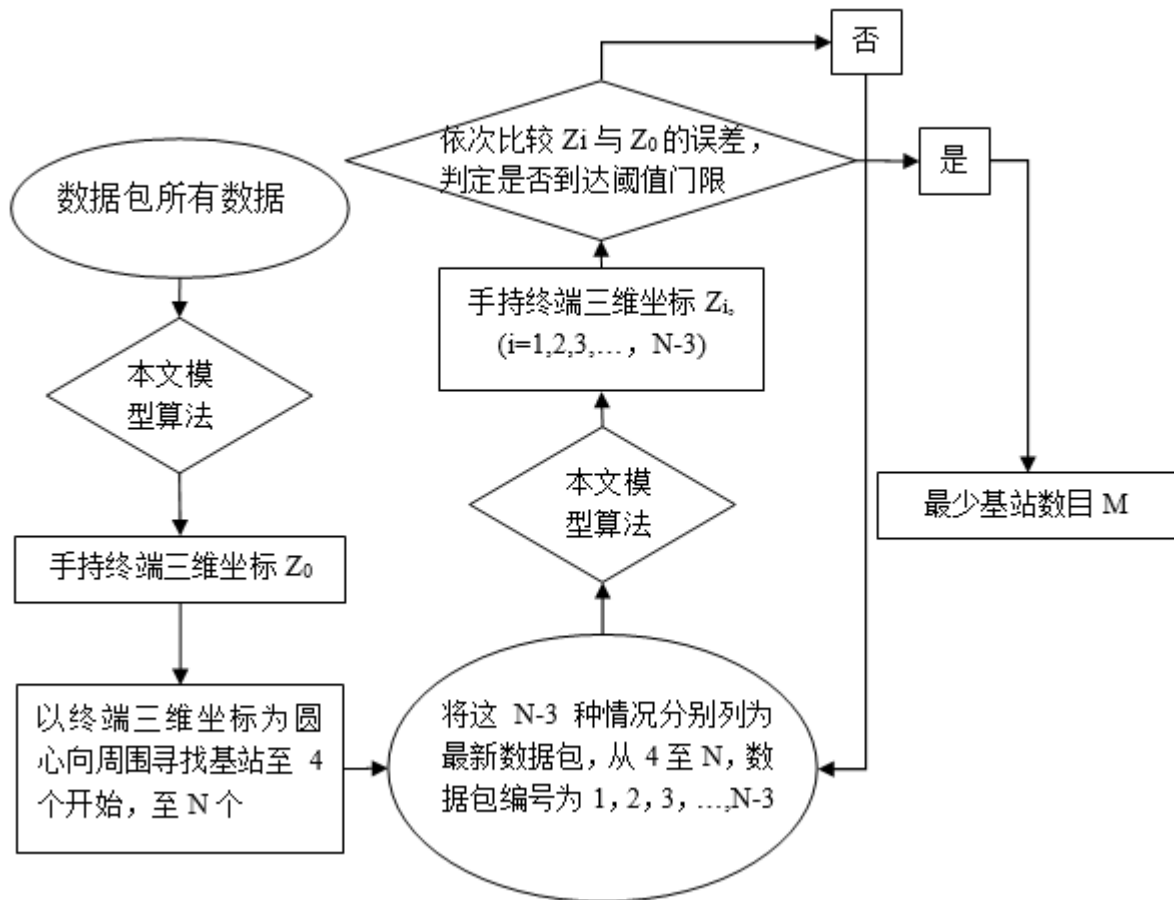


图 5-1 基站数目确定流程图

首先我们将数据包里的数据使用第一问的数学模型建模处理, 此时得到一个手持终端的精确解。记录为 Z_0 , 以精确解坐标 Z_0 为原点, 半径做球面, 半径依次变大, 会逐渐开始包含基站。在这里我们假设距离手持终端近的基站存在 NLOS 可能性较小, 但是距离精确解最近的四个基站不存在 NLOS 问题, 也可能由于接收机噪声等问题造成较大误差, 因此才在精确解附近开始寻找基站, 将基站包含数目从 4 开始记录, 依次为 4 个、5 个、6 个...N 个, 共 $N-3$ 组数据, 将数据从小到大依次重新建模计算手持终端位置, 并于精确解进行比较, 设定一个门限阈值, 在这里我们取高于误差一个数量级的当量作为阈值 (1 米)。当定位范围在精确值为球心, 半径为 1 米球面以内时即可以认为定位结果近似最优三维定位精度。此时记录基站数目, 则该基站数目即为最少定位数目。

5.2 最少基站数统计

在这里我们以 case_011 举例说明: 在 30 个基站分布于周围, 1100 个手持终端的情况下, 我们将每个终端的定位误差以及所需的最少基站数目归纳为下图 5-2。

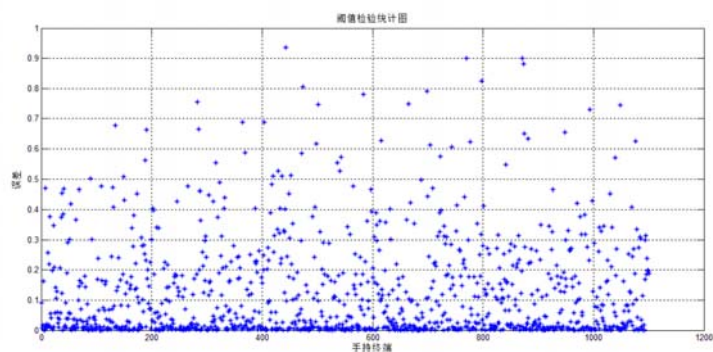


图 5-2 阈值判定图

图 5-2 为最少基站定位后的位置与精确位置误差图，可以发现每个定位精度均小于所设置阈值 1 米，所以符合阈值设置。

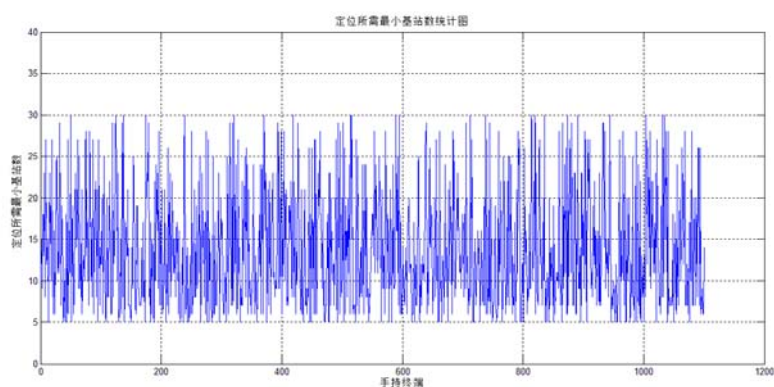


图 5-3 定位最小基站数目统计图

由图 5-3 我们可以发现所需最小基站数值从 5 到 30 不等，也就是说有些终端由于处于地理位置好可以仅通过 5 个基站即可以较精确的定位，而有些用户由于地理位置问题一个基站都不能少。

图 5-4、图 5-5 和图 5-6 为 case011,case015,case020 利用尽可能少的基站数所得的近似最优的三维定位的结果：

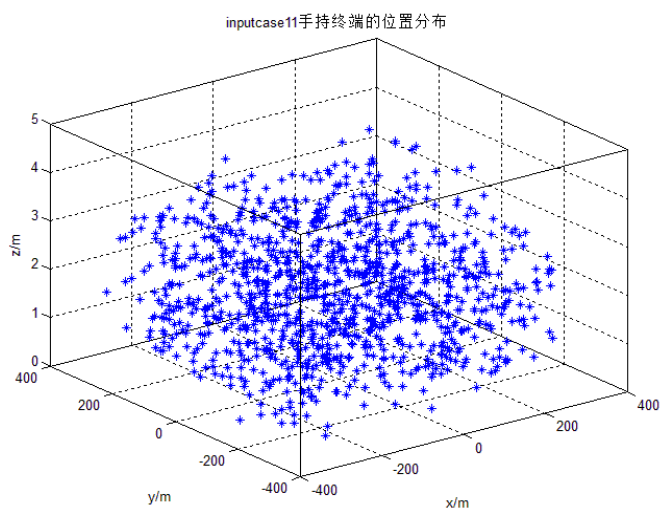


图 5-4 case011 近似最优三维定位坐标

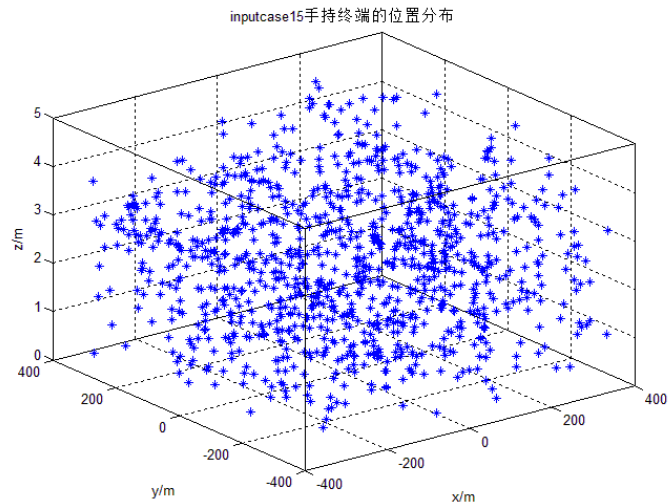


图 5-5 case015 近似最优三维定位坐标

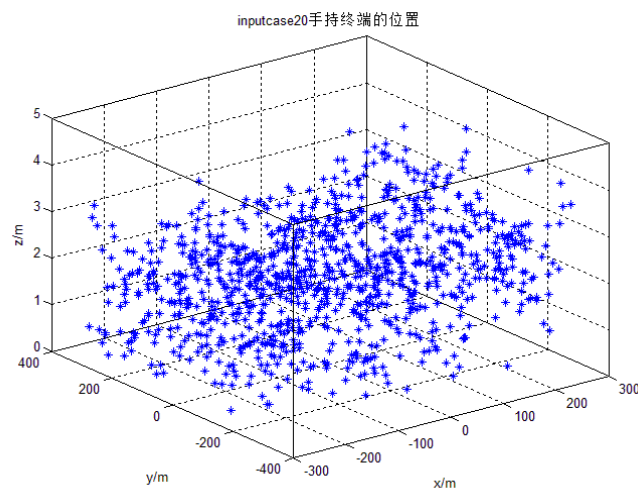


图 5-6 case020 近似最优三维定位坐标

6. 问题 3：终端运动轨迹分析

6.1 用户运动轨迹分析

在考虑用户运动轨迹的过程中，由于我们计算的误差值很小，尤其是在二维平面的情况下，平面 XY 的定位误差是厘米级别的范围，所以我们可以采取简单的拟合方法直接对画出的轨迹点进行拟合。

计算出各组的手持终端位置并输出在一个 `txt` 文件里即可。

6.2 运动轨迹图

五组运动轨迹图片如下所示：

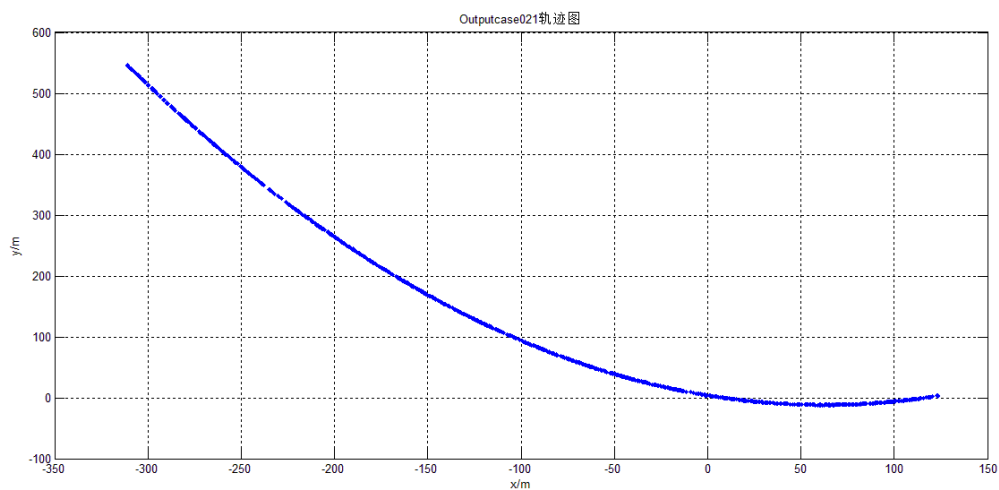


图 6-1 case_021 运动轨迹图

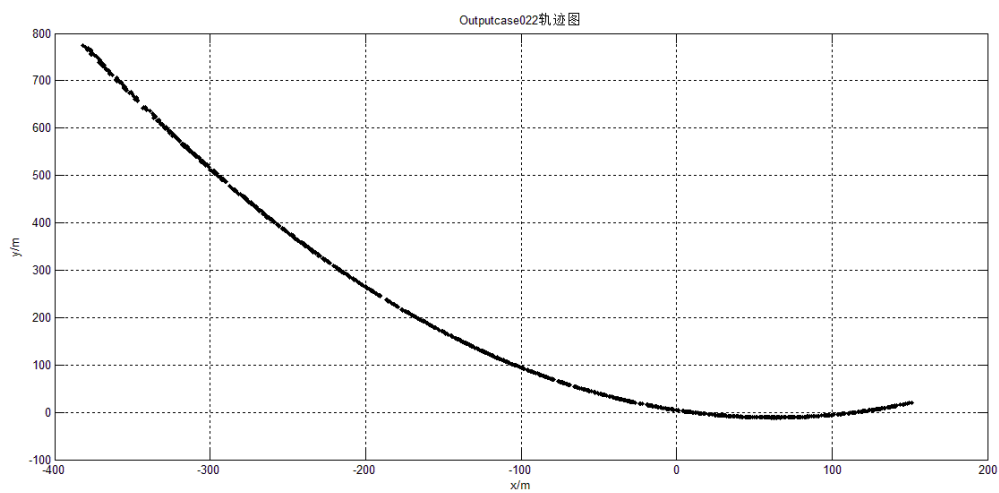


图 6-2 case_022 运动轨迹图

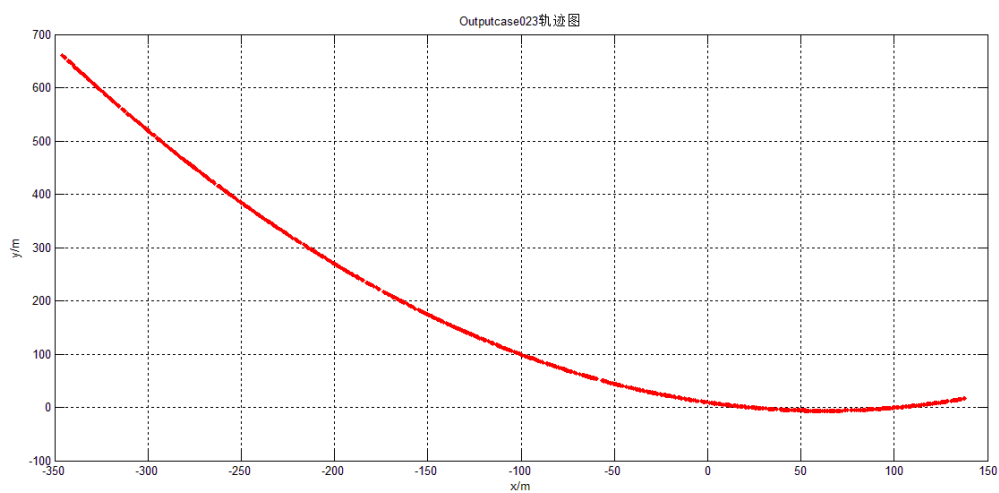


图 6-3 case_023 运动轨迹图

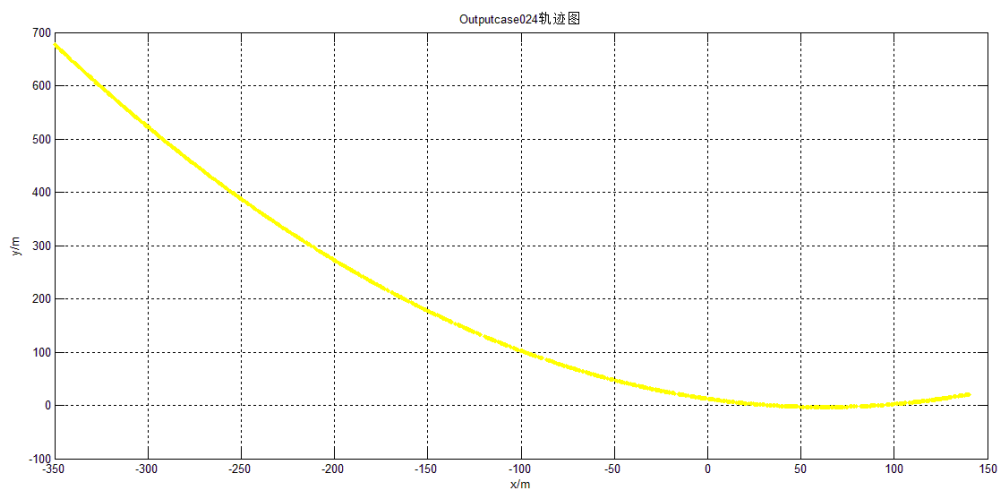


图 6-4 case_024 运动轨迹图

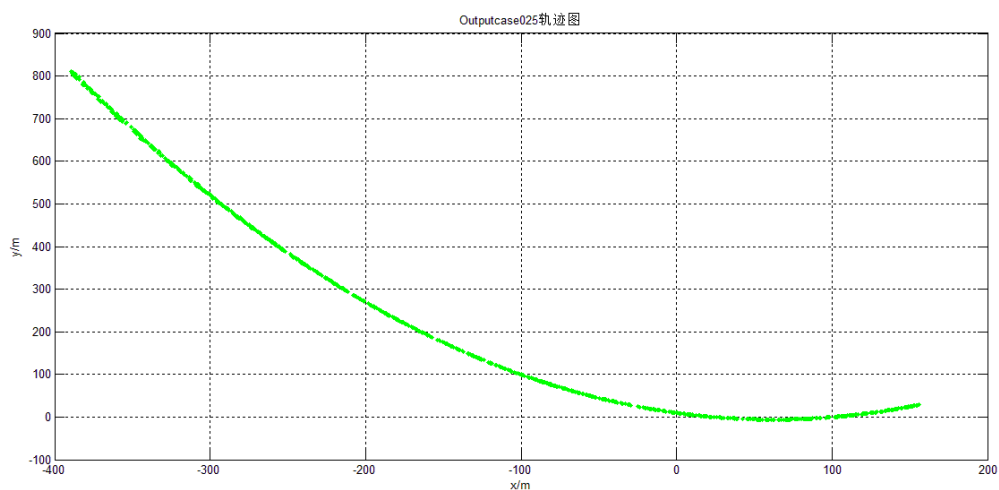


图 6-5 case_025 运动轨迹图

7. 问题 4：平均链接度与定位误差关系

7.1 有效基站选取

由于基站有效半径只有 200 米，因此我们需要寻找终端距离 200 米以内的基站，且由于算法要求，我们至少需要 4 个有效基站以上的已知条件才能精确求解手持终端的三维坐标信息。但是，考虑到在基站少于 4 个的情况下，虽然无法求出精确位置，但是仍有部分信息可以估计出终端的大概位置，因此针对此问题的思路如下：

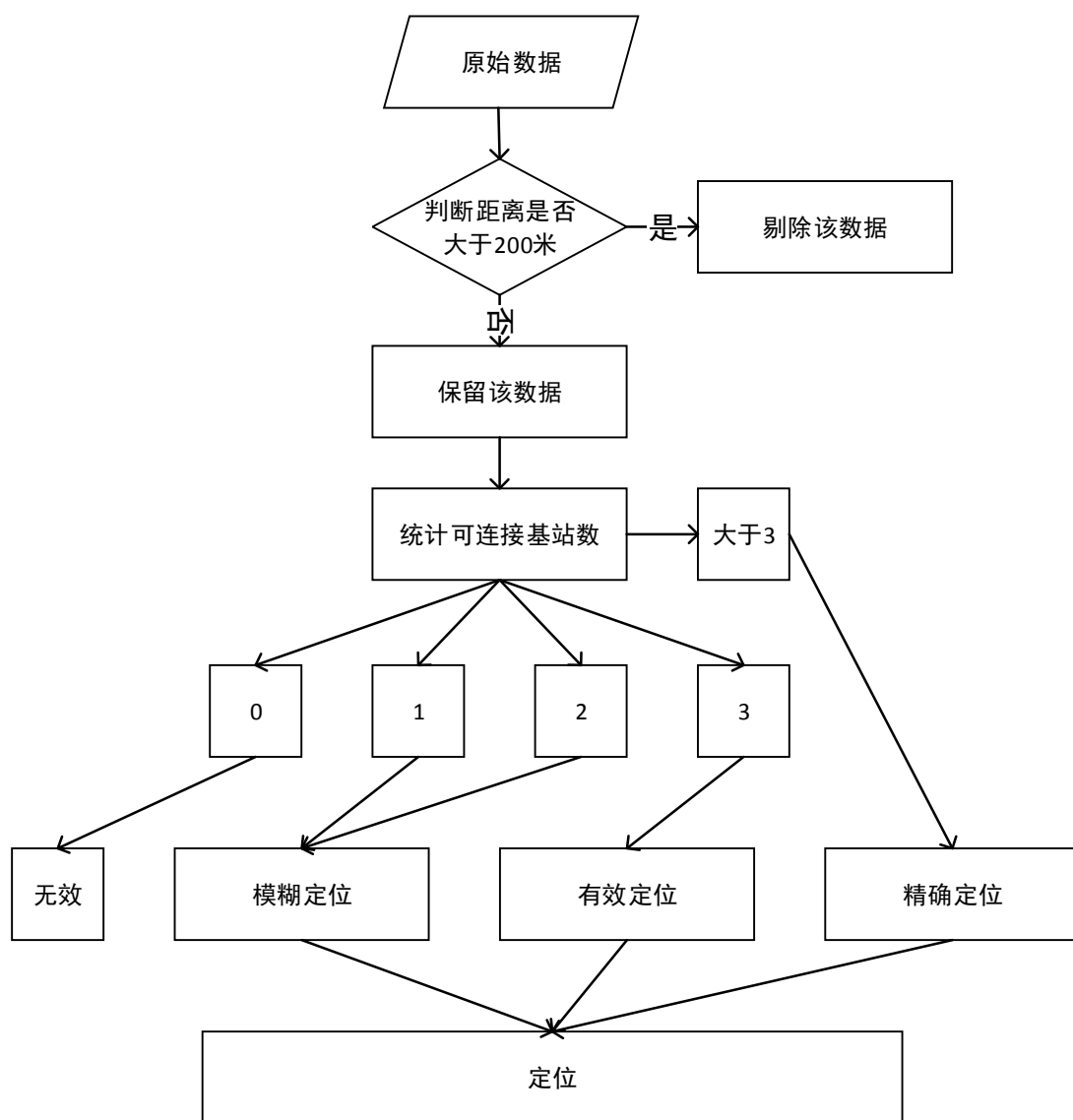


图 7-1 针对第四问的思考流程图

首先我们判断哪些连接是有效连接，并计算出所有假定范围内的有效基站数目，其值可能为 0，1，2，3，4 及以上，并依照可连接基站数目分为了无效定位、模糊定位、有效定位和精确定位四类分开讨论。并定义了**终端区域精度贡献度（后简称精度贡献率）**这一统计概念，精度即精度是测量值与真值的接近程度，每一个终端的定位精度共同构成整个了整个区域的定位精度，故每个终端都应该对这个区域有一定的贡献，利用终端区域精度贡献度这个概念来反应终端对于区域在精度方面的贡献程度。进而整个区域的定位精度可以由所有终端的精度贡献度求和进行表征。

7.2 终端区域精度贡献度

（1）无效定位

在有 0 个有效基站数目的情况下，相当于是终端不在任何基站有效范围内，故基站不存在接收信号，所以不会有任何有效定位结果。将其归类为无效定位，故该终端对于整个区域的精度没有贡献，故认为可连接数为 0 的终端的**精度贡献度为 0**。

(2) 精确定位

由先验模型我们可以得知存在四个基站及以上有效信息时，我们可以通过数学模型求得手持终端的精确三维信息，在本章节之前我们也做了许多分析，对它的细节这里不做赘述。在这样数据匮乏的情况下，我们可以认为有足够的基站数据可以得到相当准确的定位信息，故其对整个区域的贡献率最大，这里认为可连接数大于等于 4 的终端的精度贡献度 5。

(3) 有效定位

当有效基站数目为 3 个时，我们定义一下这种状态称为有效定位状态。虽然三个基站的三组方程是一个欠定方程，我们可以求解出无数组终端三维坐标信息，但是从考虑本题目中所要解决的实际问题出发，我们是要解决移动终端在无线基站中的定位问题，移动终端一般不会出现在太空中，或是地壳下，我们可以针对具体的实际问题增加一个约束方程，比如在基站定位的问题中，我们假定移动终端的高度 z 值为 3 个基站高度的平均值 $z = ave(z_a + z_b + z_c)$ ，这样虽然不能准确估计出手持终端的具体高度，但就常见的情况而言，人的活动范围有限，其误差并不会特别大，而且可以顺利求解出手持终端的三维信息。这种考虑方法也是出于我们对于实际使用中的考虑，用户在使用地图等定位应用时更关注用户的平面位置信息，即 (x, y) 坐标值。但是在某些特殊情况下，如图 7-2 所示的包裹了重叠区域等情况，这样三个基站提供的数据并没有比两个基站更多，这便退化成了有效连接数为 2 的情况（这里仅仅是为了说明会有退化风险，并没有完全列举出所有退化的可能）。

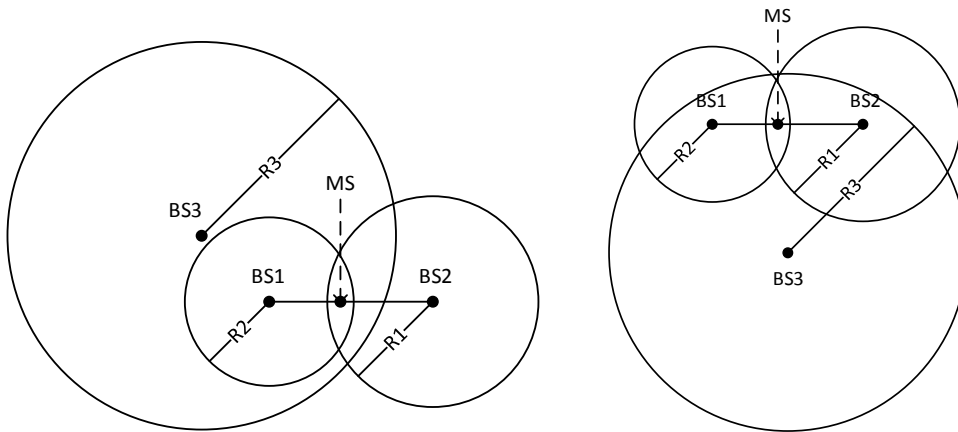


图 7-2 三基站退化至两基站图

在大部分情况下，能连接三个有效基站可以满足大多数用户的实际需求，故对于区域的精度贡献也较高，这里认为有效可连接数为 3 的终端的精度贡献率为 4，但是因为其存在退化的风险，故降低至 3。

(4) 模糊定位

1) 基站数目为 2 个：这时两基站在空间中会有部分区域交集，利用手持终端高度 z 值等于 2 个基站的平均值这一约束条件，我们可以确定一个交平面，取交平面的几何中心作为基站定位的最终值 Z ，如图 7-3 所示。对这样的定位模型进行精度分析计算

可以得出，其沿两基站的系统误差为 $\frac{R_1 + R_2 - D_{BS1-BS2}}{2}$ ，而垂直于该方向的系统误差

达到了 $\frac{\sqrt{(3R_1 - R_2 + D_{BS1-BS2})(R_1 + R_2 - D_{BS1-BS2})}}{2}$ ，故其精度并不高，且在某些特殊情况

下，会出现如图 7-4 所示的情况。在这两种情况下，两个基站所提供的信息量并没有比一个基站大很多，会退化成为有效连接为 1 的情况。在这样的情况下，我们可以认为

它对于区域的精度贡献度为 2，又因为有退化风险，所以最终降为 1.5。

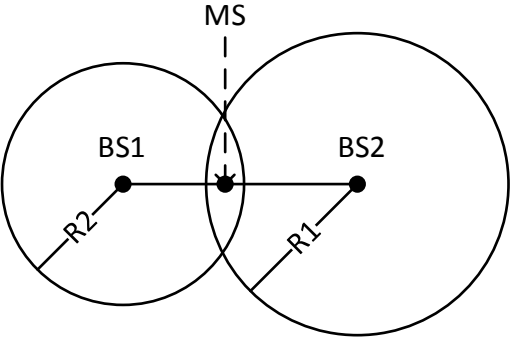


图 7-3 两基站定位示意图

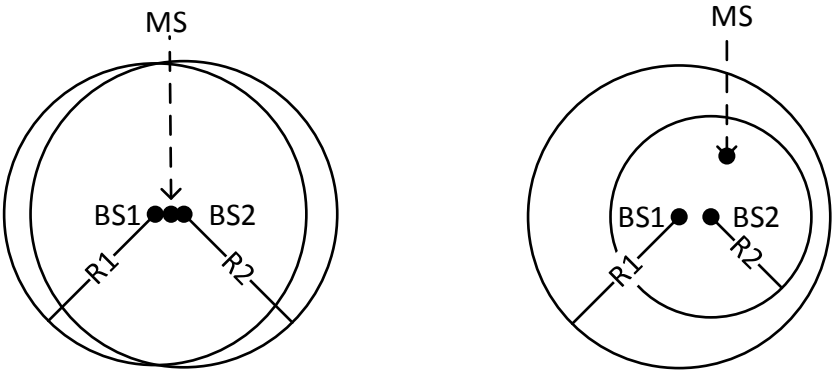


图 7-4 两基站退化为一基站示意图

2) 基站数目为 1 个:当我们只有 1 个有效基站数目时,此时的模糊定位算法即假定手持终端的三维坐标为基站三维坐标。当然此时的定位误差一定会较大,甚至会达到测量半径 R 。故将有效连接数为 1 的终端精度贡献率设为 1。

模糊定位方法是一种基于较少已知信息的定位方法,其特点是可以通过少量信息进行移动终端的三维定位,缺点是由于简单的信息所带来的较大的误差。

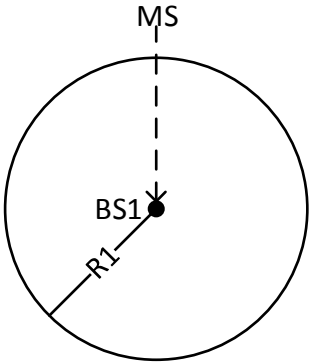


图 7-5 一基站定位图

本题中所要求测试的 case26-case30 数据按照以上建模理论得到的结果如表 7-1 所示。

表 7-1 链接度数与精度贡献度关系

算例	平均连接 度数	可连接 数为 0	可 连 接 数为 1	可 连 接 数为 2	可 连 接 数为 3	可 连 接 数大于 4	各终端精度 贡献度之和
Case26	1.335455	490	184	202	94	130	1419
Case27	5.6275	169	137	69	86	739	4193.5
Case28	2.047692	475	185	184	157	299	2427
Case29	5.775	277	105	133	81	804	4567.5
Case30	1.097	420	225	227	101	27	1003.5

通过对结果的分析，我们可以看出，整体区域的精度和平均连接度数是成正相关的关系。并利用计算出的坐标数据进行了简单的比对，可以证明这个理论可以在一定程度上反应精度与平均连接度数之间的关系。但是因为场景数较少，对其进行拟合和误差分析数据量不够，相信在大数据时代，有足够的对这种理论思想进行分析和验证。

8. 模型的评价

本文建立了复杂环境下无线基站定位模型，体现了 NLOS 和 LOS 环境下的各个特征。利用这一模型，采用加权平方以及最小二乘法评价指标对场景进行了快速识别，然后采取基于 NLOS 模型下的迭代方法将模型误差缩小至平面 0.1 米以下，垂直高度 3 米以下的精确模型，且有自适应判断场景的功能。如果将瑞利分布曲线存在的误差进行求解，我们可以得到更为精确的高度计算值。

9. 参考文献

- [1]. 陆音, 王宝全, 丘颢玮. CHAN 算法在 LOS 和 NLOS 环境下的定位研究[J]. 计算机技术与发展, 2015(9):61-65.
- [2]. 徐彤阳. NLOS 误差模型下的无线传感网定位方法与仿真[J]. 计算机工程与设计, 2013, 34(8):2680-2684.
- [3]. 万瑶. 无线电定位技术的研究[D]. 中北大学, 2013.
- [4]. 徐英凯. 蜂窝网无线定位方法研究[D]. 兰州理工大学, 2013.
- [5]. 姚光乐. 基于无线网络的定位系统的设计与实现[D]. 成都理工大学, 2009.
- [6]. 赵春晖, 吴舟. 基于 NLOS 环境下的 TOA/AOA 混合定位算法[J]. 哈尔滨工业大学学报, 2007, 39(1):85-88.
- [7]. 谢钢. GPS 原理与接收机设计[M]. 电子工业出版社, 2009.
- [8]. 章坚武, 唐兵, 秦峰. Chan 定位算法在三维空间定位中的应用[J]. 计算机仿真, 2009, 26(1):323-326.
- [9]. 段凯宇, 张力军. 一种在 NLOS 环境下提高精度的 TDOA 定位方法[J]. 南京邮电大学学报(自然科学版), 2005, 25(5):15-19.
- [10]. 李静, 刘琚. 用卡尔曼滤波器消除 TOA 中 NLOS 误差的三种方法[J]. 通信学报, 2005, 26(1):130-135.
- [11]. 雷仁大. 多天线短距离三维定位的研究[D]. 大连工业大学, 2012.
- [12]. 田孝华, 廖桂生, 王洪洋. 利用单基站对三维运动目标定位与跟踪[J]. 电子与信息学报, 2004, 26(9):1364-1370.
- [13]. 龚福祥, 王庆, 张小国. NLOS 环境下无线通信网络中的 TDOA/AOA 混合定位算法[J]. 东南大学学报(自然科学版), 2010, 40(5):905-910.
- [14]. Guvenc I, Chong C C. A Survey on TOA Based Wireless Localization and NLOS Mitigation Techniques[J]. Communications Surveys & Tutorials IEEE, 2009, 11(3):107-124.
- [15]. Zhang S, Gao S, Wang G, et al. Robust NLOS Error Mitigation Method for TOA-Based Localization via Second-Order Cone Relaxation[J]. IEEE Communications Letters, 2015:1-1.
- [16]. Sieskul B T, Zheng F, Kaiser T. A Hybrid SS-ToA Wireless NLoS Geolocation Based on Path Attenuation: Mobile Position Estimation[C]// IEEE Conference on Wireless Communications & NETWORKING Conference. IEEE Press, 2009:2744-2749.