

東海大學資訊工程學系研究所

碩士論文

Department of Computer Science

Tunghai University

指導教授：石志雄博士

Advisor : Chihhsiong Shih, Ph. D.

使用最小偏差法及蜂群法之無線室內定位精度改進探討

與比較

The Improvement of Indoor Localization Precision
through Partial Least Square(PLS) and Swarm(PSO)

Methods

研究生：梁昭隆

Graduate Student : Chao-Lung Liang

中華民國一〇五年七月

July, 2016

東海大學資訊工程學系研究所

碩士論文

Department of Computer Science

Tunghai University

指導教授：石志雄博士

Advisor : Chihhsiong Shih, Ph. D.

使用最小偏差法及蜂群法之無線室內定位精度改進探討
與比較

The Improvement of Indoor Localization Precision
through Partial Least Square(PLS) and Swarm(PSO)
Methods

研究生：梁昭隆

Graduate Student : Chao-Lung Liang

中華民國一〇五年七月

July, 2016

東海大學碩士學位論文考試審定書

東海大學資訊工程學系 研究所

研究生 梁 昭 隆 所提之論文

使用最小偏差法及蜂群法之無線室內定位精度

改進探討與比較

經本委員會審查，符合碩士學位論文標準。

學位考試委員會

召集人

孔崇旭

簽章

委

員

孔崇旭

熊桴安

石志雄

指導教授

石志雄

簽章

中華民國 105 年 7 月 1 日

摘要

近年來室內定位技術經歷了一個快速發展的時期，許多的方法被提出並使用，不論是到達時間定位 (Time of Arrival, TOA)、到達時間差定位法(Time Difference of Arrival, TDOA)、接收訊號角度法 (Arrival of Angle, AOA)、超音波定位(Ultra Sonic)、接收信號強度指標(Received Signal Strength Indicator, RSSI)……都是使用無線訊號進行室內定位時，經常使用的方法。地區特徵指紋定位法，是指收集各定位點特徵指紋建立指紋資料庫，在收到定位請求後進行比對，判斷定位結果。我們的指紋資料庫建立方式，是由收集到的 RSSI 值建立高斯模型，在收到定位請求後，計算在指紋資料庫中機率最大的定位點。使用單一高斯模型求取機率十分簡便，但與原始數據的分佈機率難以符合，使用高斯混合模型，則能降低誤差值。每一組參與混合的高斯模型都包含平均值、標準差、權重三個屬性，若使用暴力演算法計算由 n 個高斯模型組合的混合模型，每個屬性 m 等分，時間複雜度為 $O(m^3n)$ ，以 PSO 演算法求取近似解，時間複雜度為 $O(n)$ ，可以在合理的時間內，求出低誤差的近似解。然而 PSO 演算法有易陷入局部極值、過早收斂的缺點。因此我們稍加改良，在收斂至個體與局部最佳解產生黏著時，保留局部最佳解，同時給予與它黏著的個體新的屬性，經改良後避開了陷入局部極值的狀況，對原始數據分佈機率的誤差量總和是原本的 0.1~0.5 倍。在實測以高斯混合模型進行指紋定位後，為了能取得更精準的定位結果，也嘗試了使用最小偏差法(partial least squares, PLS)將定位資料分群的作法建立指紋資料庫，並使用基因演算法部分最小偏差法(genetic algorithm-based partial least squares, GA-PLS)，減少定位計算時的取樣點，以提高定位時的計算效率，在實測中使用 PLS 做為定位分群的定位精準度高達 91% 以上，平均誤差量也降到 15.5 公分，相較於單一高斯模型的定位 44% 精準度及 94 公分的平均誤差量，以及混合高斯的定位 65.15% 精準度及 75.5 公分平均誤差量，有明顯的改善。

關鍵字：室內定位，指紋定位法，高斯混合模型，PSO 演算法，RSSI，最小平方法

目錄

摘要.....	1
第一章 研究動機.....	1
第二章 文獻探討.....	2
2.1 相關背景介紹.....	2
2.2 相關研究.....	5
2.3 蜂群演算法(PSO).....	5
2.4 高斯混合模型(Gaussian Mixture Model).....	6
2.5 最小偏差法(Partial Least Squares).....	6
2.6 基因演算法(Genetic Algorithm).....	8
第三章 研究方法.....	9
3.1 Zigbee 無線訊號.....	9
3.2 室內定位系統.....	10
3.2.1 系統架構.....	10
3.2.2 離線訓練.....	10
3.2.3 線上定位.....	11
3.3 建立高斯混合模型.....	11
3.3 PSO 演算法.....	12
3.4 基因演算法.....	16
3.4 PLS 與 GA-PLS.....	18
第四章 實作.....	20
4.1 實驗設備.....	20
4.2 實驗場地.....	20
4.3 實驗結果.....	21
4.3.1 PSO 改良前後混合模型比較.....	21
4.3.2 高斯與混合高斯模型定位結果比較.....	23
4.3.3 PLS 分群.....	23

4.3.4 PLS 定位結果比較	26
4.3.5 PLS 訓練資料分群與定位權重比較.....	28
4.3.6 PLS 訓練資料數量對定位影響	34
4.4 結果討論.....	34
第五章 結論.....	37
參考文獻.....	39

圖目錄

圖 2.1 到達時間法(TIME OF ARRIVAL, TOA).....	2
圖 2.2 接收訊號角度法(ARRIVAL OF ANGLE, AOA).....	3
圖 3.1 將三筆同一定位點不同感測節點強度訊號合併為一筆訊號波形	11
圖 3.2 單一高斯模型	12
圖 3.3 高斯混合模型	12
圖 3.4 PSO 演算法流程圖.....	13
圖 3.5 改良後 PSO 演算法流程圖.....	15
圖 3.6 (A)使用 PSO 演算法逼近結果, (B)使用改良後 PSO 演算法逼近結果.....	16
圖 3.7 基因演算法(GENETIC ALGORITHM, GA)流程圖.....	17
圖 4.1 實驗場地 1 示意圖	20
圖 4.2 實驗場地 2 示意圖	21
圖 4.3 改良前 PSO 差異值.....	22
圖 4.4 改良後 PSO 演算法差異值.....	22
圖 4.5 PLS 分群資料(GA 前).....	24
圖 4.6 PLS 分群資料(GA 後).....	25
圖 4.7 PLS 分群資料(每點 10 組長度 180).....	26
圖 4.8 分割組數與定位準確率	30
圖 4.9 分割組數與平均誤差量	30
圖 4.10 使用權重定位準確率比較.....	32
圖 4.11 使用權重平均誤差量比較	32

表目錄

表 2.1 相關研究比較.....	5
表 2.2 SIMPLS ALGORITHM	8
表 4.1 實驗設備簡介.....	20
表 4.1 實驗場地 2 定位命中次數統計	28
表 4.2 分割數量與定位結果比較	29
表 4.3 加入權重影響後，分割數量與定位結果比較	31
表 4.4 各種權重測試比較	33
表 4.5 降低訓練資料量對定位結果影響比較	34

第一章 研究動機

近年來，由於無線技術的發展，以及可攜式設備的普及，適地性服務(Location-Based Service, LBS)的應用也成為大家關注的焦點，其中 LBS 最基本的功能就是定位服務，全球定位系統(Global Positioning System, GPS)[20]即是一個為人們所熟知的戶外定位系統，它可以為絕大部分的戶外地區提供準確的定位，GPS 系統擁有許多優點：覆蓋面積廣、快速、省時、高效率。然而，對於由建築物遮蔽的室內環境，如封閉的辦公大樓或是半開放式的校園空間及工廠，GPS 系統依賴的衛星訊號會被嚴重削弱，失去參考價值，因此無法進行準確的定位。

目前室內定位感測技術發展，常見的有可見光、超音波、地球磁場、慣性定位、Zigbee、紅外線與 RFID 等。以現階段而言，藉由無線傳輸技術發展的室內定位是最多人使用的，因為智慧型手機上相當普遍存在 Wi-Fi 與藍牙等各種無線傳輸元件，這也使得基於無線訊號的室內定位應用能夠在現階段，更容易被實現。

因為室內環境擺設複雜，所以室內定位系統在實作上總是會產生不小的偏差，而為了降低誤差提高定位精準度，就要對無線訊號的強度擷取方式、資料庫建置、定位演算法等方面進行改善。

第二章 文獻探討

2.1 相關背景介紹

近年來，無線感測網路(Wireless Sensor Network, WSN)技術逐漸發展成熟後 [21][30]，室內定位系統再度地被提及實現之可行性，許多可應用在室內空間的定位技術已相繼被發表，且可直接實作於室內定位，下面將簡單說明目前已被提出的常用無線網路室內定位技術及討論其優缺點：

一、到達時間法

到達時間法(Time of Arrival, TOA)也稱為 TOF(Time of Flight)，此定位方法是利用目標物發送的訊號到達感測裝置的傳遞時間，來估算感測標籤與感測節點之間的距離[29]。原理如圖 2.1 所示，感測節點 $i(i=1,2,3\dots)$ 送出訊號，訊號由感測標籤(TAG)接收。根據時間差，可得知訊號在空氣中傳播的時間 $T = (t_i - t_0)$ ，推算出感測節點 i 與感測標籤的距離 $r_i = T * C$ (傳播速度)，再由 $r_i^2 = (X_i - X_m)^2 + (Y_i - Y_m)^2$ 算出 Tag 所在位置 (X_m, Y_m) 。

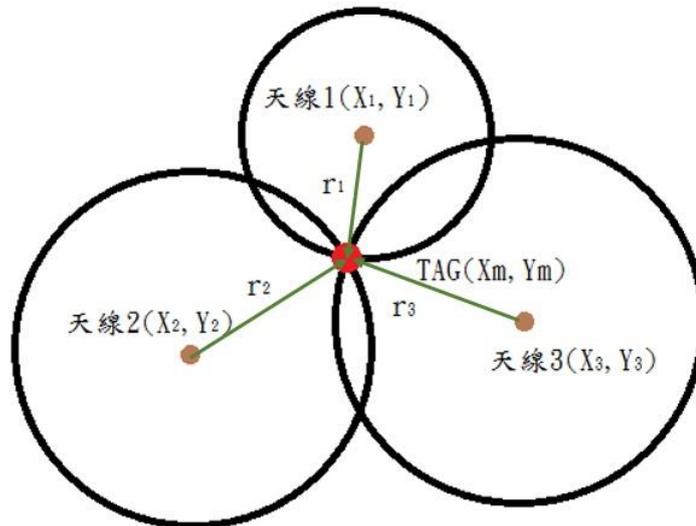


圖 2.1 到達時間法(Time of Arrival, TOA)

Klee 等學者的研究[32]中提出，到達時間法要求感測節點以及感測標籤的時間軸須精確地同步，因為訊號的傳播速度非常快(3×10^8 m/sec)，所以即使是 1 微

秒的時間誤差，也會導致數百公尺的定位距離誤差，因此 TOA 對裝置的硬體需求較高，花費成本相對來說也較大。

二、到達時間差法

到達時間差法(Time Difference of Arrival, TDOA)為到達時間法的延伸技術[24]，不同的地方是，感測標籤會發射兩種傳遞速率不同的訊號，感測節點根據這兩種訊號到達的時間差，計算兩者之間的距離，再去推算出感測標籤所在位置，雖然 TDOA 對時間的精確度也有要求，但相對於 TOA 來說，因為使用的是同一個節點中的訊號時間差，所以並不需要所有裝置都做精確的時間同步，對裝置的硬體需求就比較低，能降低建置成本。

三、接收訊號角度法

接收訊號角度法(Arrival of Angle, AOA)[7]，只需要利用兩個或更多感測節點接收到的訊號入射角度，就能判斷定位標籤的方位。工作原理是利用具方向性的天線(Directional Antenna)所量測的訊息推測目標位置。如圖 2.2 所示，定位標籤(TAG)，A 與 B 則是事先架設好的具方向性的天線，根據感測節點接收到的訊號角度 θ_A 與 θ_B ，與已知的兩天線距離 d_{AB} ，代入三角定理計算距離 d_A 、 d_B 即可求得定位標籤所在位置。

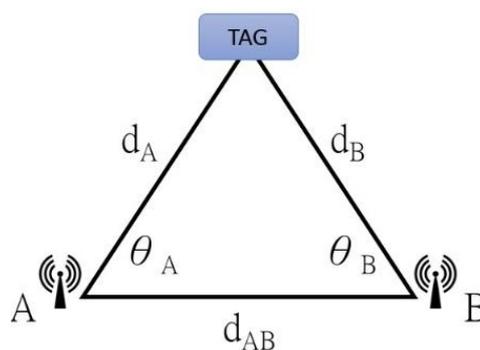


圖 2.2 接收訊號角度法(Arrival of Angle, AOA)

AOA 的最大優勢在於不需要與時間同步，但在障礙物較多的環境會發生多路徑(Multi-path)效應，所謂多路徑效應是指一個訊號經由兩個或兩個以上的不同

路徑到達一個天線，主要是因為在天線的周圍有反射平面的物體存在，例如金屬板或者水面等，這將會造成較大的噪聲(noise)導致觀測量產生偏移(biased)，這也是 AOA 主要的誤差來源，另外接收端天線本身的角度解析度有其極限，當定位標籤距離感測節點過遠時，角度解析度的誤差也會影響定位準確度。

四、接收信號強度指標

接收信號強度指標(Received Signal Strength Indicator, RSSI)[9][25]，是使用接收訊號的強度來計算距離，定位方式與 TOA 非常相似，它透過訊號強度和已知的頻道衰弱模型來估計參考點與定位點的距離，根據多個參考點得出的距離值，就可以得到定位點的可能位置。與 TOA、TDOA 相比，完全不需要時間同步，但是和 AOA 一樣容易受到多路徑效應影響。

Larranaga 等人在研究[16]中提出，讓參考點間互相測量各鄰近節點的 RSSI 值，透過量測這些參考點 RSSI 值的改變，定位點可以依照 RSSI 值的變化幅度來計算，透過這種修正方式，可以降低多路徑效應的影響，遇到環境改變或訊號強度減弱的狀況，依舊可以取得不錯的定位精度。

五、訊號特徵比對法

訊號特徵比對法(Signal Pattern Matching)，又稱為特徵指紋法，又因為必須事先裝設感測節點，然後針對該每一個定位點做測量，取得該地點的訊號樣本，也被稱為場測法[26]，其工作原理是針對各定位點採集訊號樣本建立指紋資料庫，收到定位請求時，將資料與指紋資料庫中比對，判斷定位標籤可能在的位置。

此定位方式簡易，成本也不高，因此許多研究都基於此種方式發展不同演算法，以提高定位精度，此種方式缺點是，當定位區域改動或增加了場測時所沒有的干擾源，則指紋資料庫內的訊號樣本就會失去參考價值，需要以演算法或是重新場測來更新資料庫內樣本。

本實驗就是基於訊號特徵比對法，架設 Zigbee 感測節點，收集各定位點訊號樣本後，使用高斯模型建構指紋資料庫，進行定位運算，並使用高斯混合模型

及最小偏差法等方式進行定位精準度的改善。

2.2 相關研究

許多相關研究已經在室內定位的領域中做出了一定成果，但是大部分的解決方案都在於降低噪音，使用多演算法或配合路徑運算來降低誤差，下面列出一些相關研究的比較。

系統/解決方案	無線技術	演算法	誤差	準確	複雜度	成本
Microsoft RADAR [26][27]	WLAN,RSS	KNN,Viterbi-like	3~5m	90%	中	低
Horus[22][23]	WLAN,RSS	Probabilistic method	2m	90%	中	低
DIT[19][28]	WLAN,RSS	MLP,SVM	3m	90%	中	低
SpotON[15]	主動 RFID RSS	Ad-Hoc lateration	N/A	N/A	中	低
LANDMARC[18]	主動 RFID RSS	KNN	<1m	50%	中	低
GSM 指紋[33]	GSM cellular network(RSS)	Weighted kNN	<10m	80%	中	中
張凡生(2014)	N/A	Novel Fingerprinting Mechanisms(NFM)	1.2m	N/A	高	中
林嘉慶(2015)	WiFi RSSI	kNN、K 分群、粒子群優化	1m	N/A	中	低
葉宇珊(2014)	N/A	Weighted kNN,差分修正演算法	1.9m	N/A	中	低
梁昭隆(2016)	ZigBee RSSI	基因演算法部分最小偏差法	16cm	91%	中	低

表 2.1 相關研究比較

2.3 蜂群演算法(PSO)

PSO 是一種模仿自然界蜜蜂覓食行為進行問題求解之方法，該法為具有群體智慧的仿生演算法，其最大優點在所需參數設定少以及搜尋範圍廣，它利用偵查點分散的方式進行大範圍的搜尋，並判斷是當前解是否為搜尋區域內遇到的最佳解，如此反覆搜尋數次，以求得最佳近似解。

PSO 的主要特點是不需要獲取問題的特殊信息，只需要對問題進行計算，並

對計算結果比較優劣，通過各個偵查個體的對區域內進行尋求最佳的行為，最終展現出整個群體中的全體最佳值。

2.4 高斯混合模型(Gaussian Mixture Model)

高斯混合模型 (Gaussian mixture model, 簡稱 GMM) 是單一高斯機率密度函數的延伸，它是由多組高斯模型乘上各自的權重後進行相加，因此 GMM 能夠比單一高斯模型更加平滑地貼近統計資料中的機率分佈。

$$\text{高斯模型數學式如右所示 } g(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

高斯混合模型數學式則針對每一個加入混合的高斯模型乘上權重 w ，數學式如右所示 $p(x) = \sum_{i=1}^n w_i g(x, \mu_i, \sigma_i)$ ，其中的 n 代表加入混合的高斯模型數量，理論上來說，混合的高斯模型數量越多，則取得的高斯混合模型將會與真實機率越貼近。

2.5 最小偏差法(Partial Least Squares)

最小偏差法(Partial Least Squares, PLS)是一種多元統計分析的方法，由瑞典統計學家 Herman Wold (1975)提出，然後由他的兒子 Svante Wold 發展，是一種統計學上分析數據的方法，相較於其他迴歸分析的方法，更適合解決有多重共線性的問題，多重共線性是指多變量線性迴歸中，變量之間由於存在高度相關而使迴歸估計不準確的現象。

PLS 與主成份分析(principal components analysis, PCA)方法很相似，兩者皆是以縮減資料維度為目的，主要不同的地方在於，PCA 目的是要提取隱藏在矩陣 X 中的相關信息，然後用於預測變數 Y 的值，只選出了變數 X 中影響最大的變異，但當一些有用變數的相關性很小時，選取主成份時就很容易忽略，使得預測模型的可靠性下降，PLS 則透過建立新的解釋變數，這種新的解釋變數稱作潛在變數(Latent Variables)，先將 X 轉換成潛在變數，再引入反應變數 Y ，同時從 X 、

Y 中尋找影響較大的變數建立預測模型[6]。因此，PLS 最大的目的在藉由縮減資料維度，從較少的解釋變數中獲取最多資訊來預測反應變數。

PLS 在模型的建立上，融合了多元線性迴歸、主成分分析及典型相關分析等重要的統計技術，實現了多種數據分析的應用。一般的多元線性迴歸(Multiple Linear Regression, MLR)用在配適多個變數，然而當變數個數過多(變數個數大於觀察的樣本數, i.e. P 大 n 小)，MLR 就會過度配適(over-fitting)，模型預測新的變數時效果就會較差，相較之下 PLS 利用潛在變數找出最大變異的解釋變數，使預測模型最佳，因此 PLS 也被稱為 projection to latent structure (Garthwaite, 1994)。

PLS 是基於這樣的假設，矩陣 X 和它的標籤向量 y 透過線性變換產生潛在變數向量 T 和 U，P 和 G 是各自的負荷矩陣和向量，E 與 F 為殘差矩陣，可以寫為下列算式。[11]

$$X = TP^T + E$$

$$y = UG^T + F$$

透過 PLS 建立模型疊代過程，得出由權重向量 w、c 組成的權重矩陣 W、C 使得

$$[cov(t, u)]^2 = [cov(Xw, yc)]^2 = \max_{|a|=|b|=1} [cov(Xa, yb)]^2$$

$cov(t, u) = t^T u$ 表示得分向量 t 和 u 之間的協方差。最大化協方差的潛在變數向量，相當於最大化在同一空間的辨別。

傳統的 PLS 必須經過反覆的疊代運算以及資料的壓縮來求得正確的權重向量與分數向量，是比較耗時的。因此，在 1993 年，De Jong, S. 介紹了一個直接由原始資料取得分數向量 t 的方法，稱為 SIMPLS[10]，本文實驗中所使用的也是這個較為簡潔的方式，演算法如下表 2.2。

SIMPLS Algorithm

1	$S = X_0^T Y_0$
2	for $i = 1$ to k
3	if $i = 1$, $[u, s, v] = \text{svd}(S)$
4	if $i > 1$, $[u, s, v] = \text{svd}(S - P_{i-1}(P_{i-1}^T P_{i-1})^{-1} P_{i-1}^T S)$
5	$r_i = u(:, 1)$
6	$t_i = X_0 r_i$
7	$p_i = X_0^T t_i / (t_i^T t_i)$
8	end
9	$B_{\text{PLS}} = R_k (T_k^T T_k)^{-1} T_k^T Y_0$

表 2.2 SIMPLS Algorithm

首先將矩陣 X 與向量 y 減去它們的平均值後，成為標準化的 X_0 及 Y_0 ，經過第一步算出 S ，接著進行遞迴運算，第一輪先將 S 進行奇異值分解(SVD)，取出第一個左奇異向量 r ，透過 r 計算求得 t 及 p ，將 p 組成的矩陣回到第四步驟中進行奇異值分解，最後可得到 $B_{\text{PLS}} = R_k (T_k^T T_k)^{-1} T_k^T Y_0$ 。

2.6 基因演算法(Genetic Algorithm)

基因演算法(Genetic Algorithm, GA)是計算機科學人工智慧領域中用於解決最佳化的一種搜索啟發式演算法，它借鑒了達爾文的進化論和孟德爾的遺傳學說，這些生物現象包括遺傳、突變、自然選擇以及交配等。其本質是一種高效、並行、全局搜索的方法，能在搜索過程中自動獲取和積累有關搜索空間的知識，並自適應的控制搜索過程以求得最優解。遺傳演算法廣泛應用在生物信息學、系統發生學、計算科學、工程學、經濟學、化學、製造、數學、物理、藥物測量學和其他領域中[8]。

第三章 研究方法

本實驗是要實現在一個區域內進行室內定位，利用訊號特徵比對法，收集 Zigbee 感測節點的定位訊號樣本，顯示感測節點在移動後的位置。

3.1 Zigbee 無線訊號

利用無線訊號進行定位，最常見的是利用基於現有的無線區域網路(Wireless Local Area Network, WLAN)架構來做討論與研究[12][13]，因為可以在不增加額外的硬體配置，只需要開發適當演算法，即可提供定位功能，但基於 802.11 的標準所設計的定位裝置通常甚為耗電，對於長期的追蹤定位，恐怕不是一個優秀的設計方案，因此有研究改用 Zigbee 802.15.4 標準傳送定位資料[17]。

此外，根據現有文獻的結論，基於訊號強度定位系統，其精度完全依賴在感測節點的個數，對於高精度定位需求的情況下，建置大量感測節點則必須付出許多的成本。因此，ZigBee 技術所擁有的低成本架構、低功率損耗、佈建容易…等優點，成為我們挑選的解決方案。

關於 ZigBee 技術，是由 ZigBee 聯盟與 IEEE 共同制定關於低功耗、低資料傳輸率、低成本之標準，爾後以『ZigBee』來表示此項標準。ZigBee 技術被期望於提供低成本、設備間連線之低功耗，這樣的優勢將可提供只需以裝備電池的方式，便可使裝置維持數月甚而超過一年的使用壽命。

ZigBee 技術還可以實現於 Mesh 網路架構，且可提供的較多的連接節點數，當無線設備符合於 ZigBee 的無線傳輸標準時，通常其傳輸範圍可達 10-75 公尺，傳輸距離將依賴於電波傳輸的環境狀況以及各種不同應用情境之無線電波發射功率，同時，ZigBee 工作於不需申請的 ISM (Industrial, Scientific and Medical, ISM) 頻段，即適用於全球的 2.45GHz、915MHz 以及 868MHz。資料傳輸率分別為 250Kbps(2.45GHz)、40Kbps(915MHz)與 20Kbps(868MHz)，整個封包堆疊格式則由 IEEE 與 ZigBee 聯盟共同制定，IEEE 802.15.4 著重在較底層的封包格式制

定，其分別為實體層(Physical layer)以及連結層(Link layer)，ZigBee 聯盟則著重於制定網路層(Network layer)、傳輸層(Transport layer)、交談層(Session layer)、表現層(Presentation layer)以及應用層(Application layer) 之封包格式。

3.2 室內定位系統

3.2.1 系統架構

本文提出的室內定位策略是以接收信號強度指標(Received Signal Strength Indicator, RSSI)與訊號特徵比對法(Signal Pattern Matching)為基礎所建立的。

訊號特徵比對法又分為離線訓練和線上定位兩個階段，離線訓練階段以及線上定位階段，離線訓練階段的任務是收集定位點資訊，建立資料庫模型，以作為線上定位階段比對樣式之用。

3.2.2 離線訓練

離線定位階段，首先在定位區域內劃分 N 個訓練點，並於每個訓練點以定位標籤量測感測節點發送的訊號 RSSI 值，最後將訓練點的座標位置，RSSI 樣本資訊，儲存為訊號資料庫。

在完成資料收集後，要針對資料內容進行處理，排除差異過大的雜訊，統計各點 RSSI 強度分布機率，依照統計資料對各點建立高斯模型訊號指紋特徵資料庫、以 PSO 演算法建立高斯混合模型訊號指紋特徵資料庫，或是將資料合併為波形並分組，進行 PLS 分群處理建立 PLS 訊號指紋特徵資料庫。

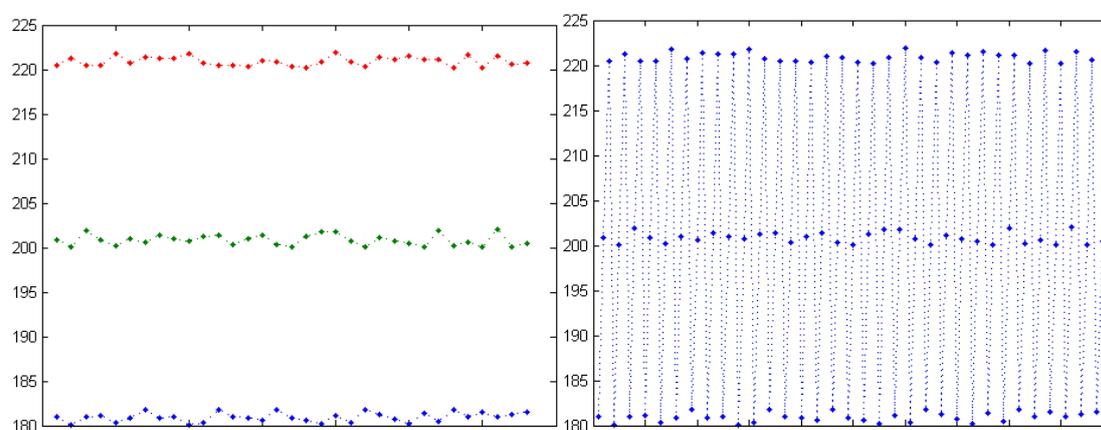


圖 3.1 將三筆同一定位點不同感測節點強度訊號合併為一筆訊號波形

3.2.3 線上定位

線上定位階段主要是推估使用者於已建置指紋特徵資料庫之室內定位區域所在位置，其運作原理是收集線上定位階段的訊號樣本，連同定位請求傳送給資料庫主機，與資料庫中離線訓練階段於各個訓練點所採集的訊號特徵比對，找出其中最相似訊號特徵的訓練點，以推估使用者所在位置。

3.3 建立高斯混合模型

建立高斯模型與高斯混合模型是為了在之後定位階段，讀取訊號強度時，可對應出一個機率密度值，定位系統即可依據這個機率值來判斷定位點的可能性。

建立高斯混合模型，主要目的是能更加貼近原始分布機率，每一組高斯混合模型都是由數組高斯模型所組成，每一組高斯模型都有平均值與標準差兩種屬性，而高斯混合模型相較於高斯模型，則是多了一項權重屬性，也就是說將多個高斯模型依照分配的權重相加，就能建構出高斯混合模型，平均值、標準差和權重，分別影響分布的中心位置、分佈的幅度和影響程度。

下圖 3.2 為其中一個定位點，針對其中一個感測節點 RSSI 統計所建立的單一高斯模型，圖 3.3 則為同一個定位點及感測節點所建立的 4 組高斯混合模型，後方長條圖為原始分布機率，此組數據中，單一高斯模型機率分布誤差為 0.5168，而高斯混合模型則為 0.1043，兩者的差距高達四倍以上。

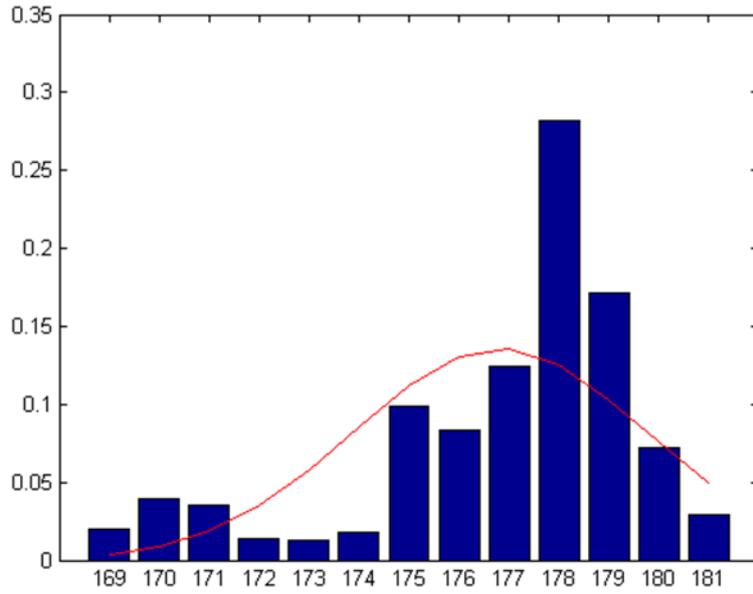


圖 3.2 單一高斯模型

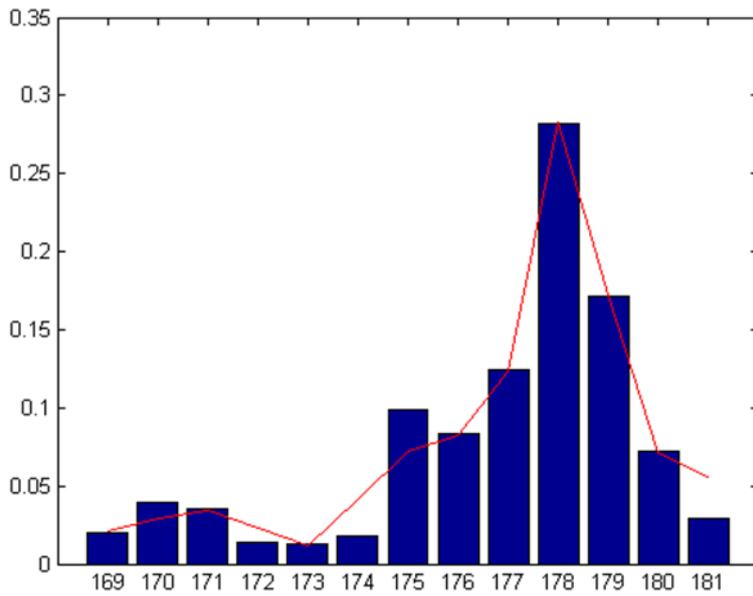


圖 3.3 高斯混合模型

3.3 PSO 演算法

在建立高斯混合模型時，因為考慮到場測階段不需要及時運算，只需要建立符合原始機率的混合模型，嘗試過使用暴力演算法來找尋最佳解，但是經過實測，

一組高斯混合模型需要數十小時才能得出結果，故而改選用近似演算法來求取近似解，在這個階段，我們選用了 PSO 演算法。

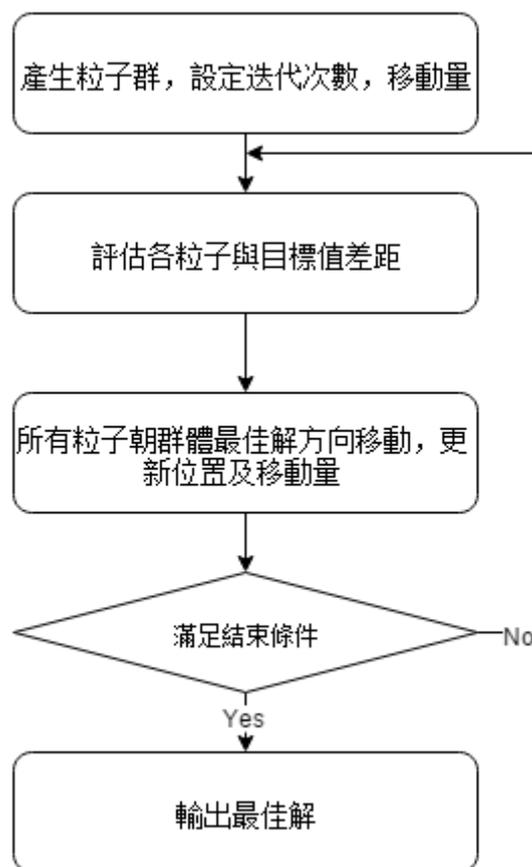


圖 3.4 PSO 演算法流程圖

PSO 演算法其流程圖如圖 3.4 所示。然而，PSO 演算法有易陷入局部極值、過早收斂的缺點。PSO 的早熟收斂通常是因為群體喪失了多樣性，大部分個體集中在一小範圍之內，如果小範圍內並不包含最佳解，演算結果就很難跳離這個範圍，因為這時個體的移動量已經不具有擴大搜索範圍的可能。

因此我們對 PSO 演算法稍加改良，在收斂至個體與局部最佳解產生黏著時，

保留局部最佳解，同時給予與它黏著的個體新的屬性，避免陷入局部極值的處境。以此改良後的 PSO 演算法，我們可以比原 PSO 演算法算出更接近原始數據的分佈機率的高斯混合模型，原 PSO 演算法與改良後的 PSO 演算法，兩者計算出的高斯混合模型與原始數據分佈機率的誤差量總和相比，前者是後者的 2~10 倍，由此可以看出，改良後的 PSO 演算法有效的避開陷入局部極值的狀況。

我們改良的部分是上圖 3.4 PSO 演算法流程圖中，“所有粒子朝群體最佳解方向移動，更新位置及移動量”的步驟，經過改良後的流程圖如下圖 3.5 所示。

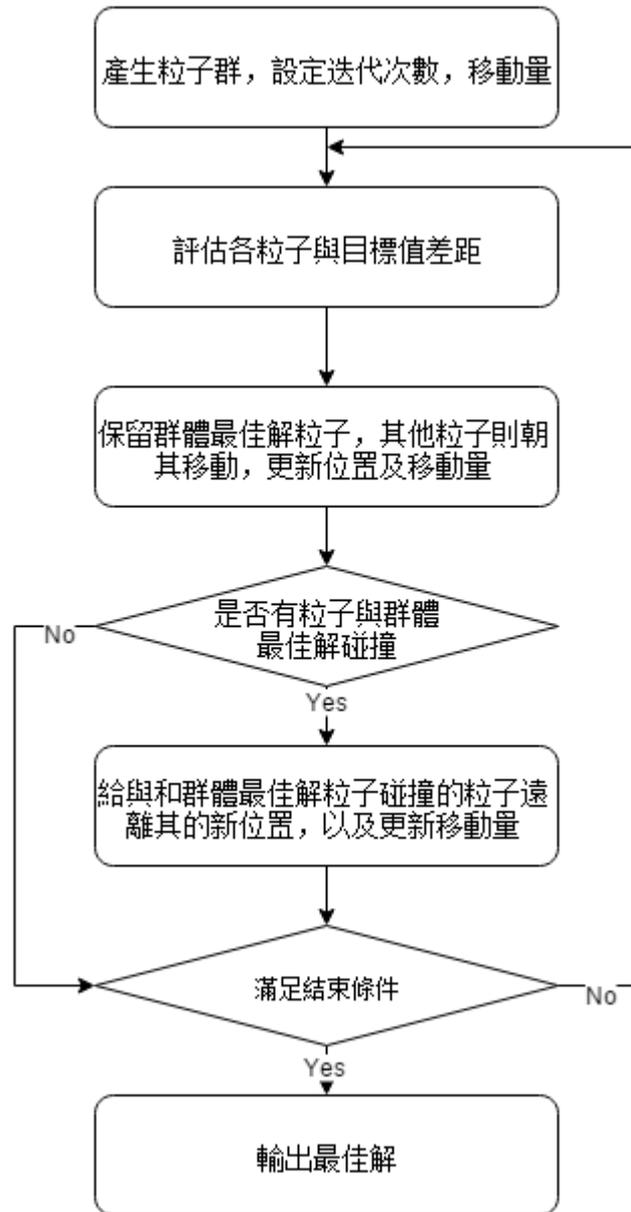


圖 3.5 改良後 PSO 演算法流程圖

在改良後，我們建立了數組線型，分別用改良前和改良後的 PSO 演算法，去求取逼近的線形，結果如下圖 3.6 所示，藍線為預設線型，紅色為演算法逼近線型。

可以看出改良前因為陷入了局部極值，在右半邊的線型無法逼近，而改良後的 PSO 演算法，避開了這種狀況，幾乎和預設線型完全重合。

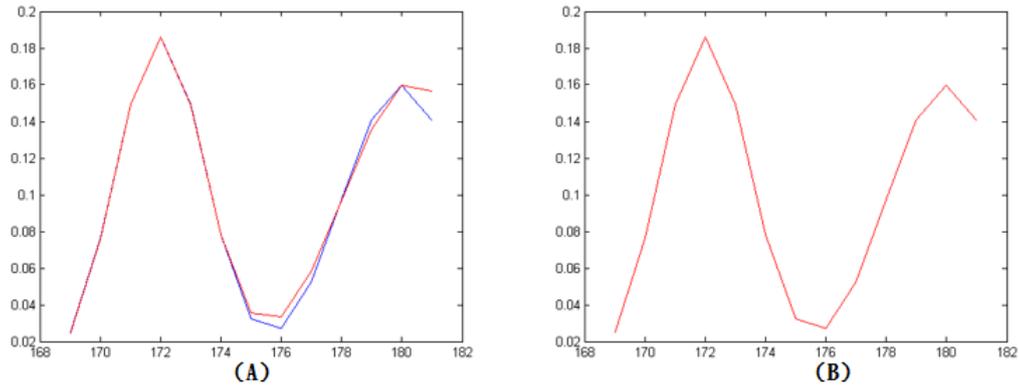


圖 3.6 (A)使用 PSO 演算法逼近結果，(B)使用改良後 PSO 演算法逼近結果
 分別使用改良前和改良後的 PSO 演算法去做 100 次與隨機模擬線形的吻合程度比較，模擬線形為 3 組隨機高斯混合，PSO 演算法則是使用 4 組高斯混合模型以及 50 搜尋次數，得到的結果，改良前 PSO 演算法，誤差總和最小為 0.0046，最大為 6.3434，平均誤差為 0.2007，改良後 PSO 演算法，誤差總和最小為 0.00025，最大為 0.7793，平均誤差為 0.1349。

3.4 基因演算法

基因演算法(Genetic Algorithm，GA)流程圖如圖 3.7 所示。

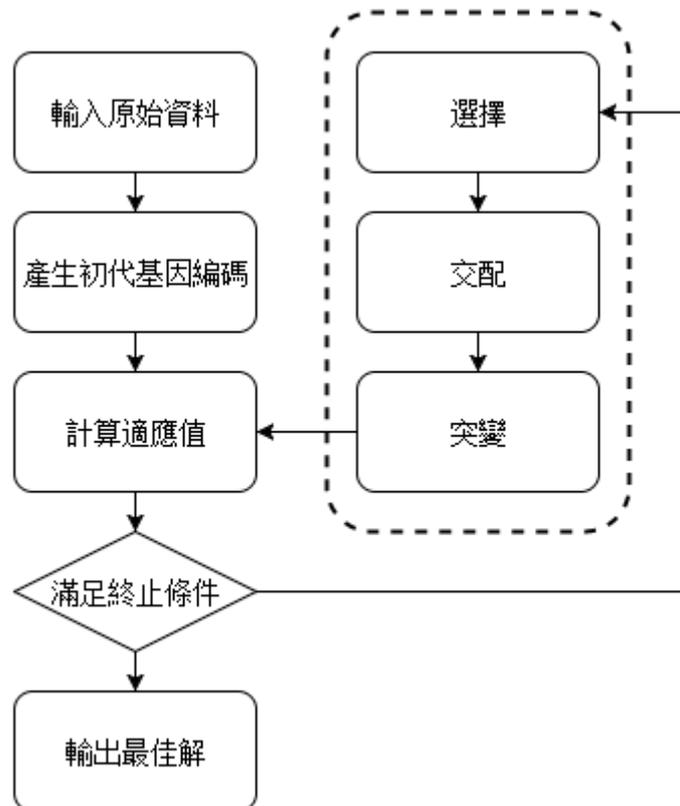


圖 3.7 基因演算法(Genetic Algorithm, GA)流程圖

基因演算法一開始會隨機生成數個長度等同原始訓練資料長度的二進制字串，稱為第零代(generation)或是初代編碼，其中的每一組二元字串被稱為一個染色體(chromosome)，每組編碼皆代表一個可行的解(可能包含最佳解)；對於一個染色體，所有被設置為 1 的資料會被抽出並包含到一個子集裡，反之則不包括。此外定義一適應函數(fitness function)計算該親代所有染色體對此「環境」適性程度，以達遺傳學說適者生存的法則，適應函數在不同的應用會對應到不同的計算方式，而基因的改變主要透過三種機制，挑選機制、交配機制、突變機制。

挑選機制主要觀念為「留下好的染色體，排除不好的染色體」，挑選機制其實就是探討如何從群體(樣本空間)挑選出個體(樣本)的取樣方式，被挑選的個體就是親代，可經由遺傳運算來產生子代，本文的實驗中使用 Holland 提出的輪盤式(roulette wheel selection)選擇，由計算完的適應值做排序，淘汰較差的 50%，餘下前 50%者，適應值越高的，優先被挑選的機率越高。

交配機制是將兩個挑選出來的親代染色體經過合併產生子代，讓子代含有雙親的部分特性。目的是希望子代能夠組合出含有更高適應度的染色體，然而有可能子代只遺傳到雙親的缺點，所以交配並不保證一定可以產生出更好的子代，不過有了適應值挑選的機制，較差的子代會被淘汰，而優良的子代則會因為被挑選的機率較大，而可以繼續繁衍出下一代的子代。

突變機制會導致染色體做出隨意的變化，常見的變化是改變染色體上的一個基因。突變的作用會引導基因演算法進入未曾尋找過的基因架構，將新的基因導入群體中。但是過多的突變將會導致基因演算法變成隨機演算法，因為過多的突變會破壞原來的基因結構，造成子代與親代之間相似的特徵值減少；而過少的突變則會造成有用的基因沒被發覺，陷入區域的最佳解的困境。因此 基因演算法將突變視為次要的遺傳運算子，以較低的機率來反轉子代的某一個位元 (0 變 1 或 1 變 0)。

3.4 PLS 與 GA-PLS

建立 PLS 與 GA-PLS 指紋特徵資料庫，本實驗使用的方法是把定位點鐘收集到的各感測節點強度，以穿插的方式組合成一個連續波形，通過 PLS 降維成 5 維，挑選其中兩個差異最大的維度做為分群標準。

而 GA-PLS 則是多了一個以基因演算法挑選部分樣本為參考值，進行 PLS 降維分群，一樣也是挑選其中兩個差異最大的維度做為分群標準，在經過基因演算法挑選樣本後，可以有效減少在線上定位階段時參與計算的資料數量，以達到提高效能的目的。

其中基因演算法的適應函數定義如下：

$$F(d) = \sum_c \sum_{s \in c} \frac{1}{K_c} * SHC(s) * SW(s) \quad (1)$$

每個染色體透過適應函數的計算，會產生一個 PC 圖，其中 K_c 定義為在 PC 圖中某一類樣本相同類別的個數，以 K-nearest neighbor 演算法(KNN)用來計算距離

該樣本最近的同類別的樣本個數有幾個，也就是式(1)當中的(Sample Hit Count) ($0 \leq SHC(s) \leq K_c$)， $SW(s)$ 表該樣本的權重，定義如式(2)、(3)所示。權重起始分配為 100，均分至每個類別，再均分至每個類別包含的樣本個數。

$$CW(c) = \frac{CW(c)}{\sum_c CW(c)} * 100 \quad (2)$$

$$SW(s) = \frac{SW(s)}{\sum_{s \in c} SW(s)} * CW(c) \quad (3)$$

透過往後連續子代的權重值的重新分配，遺傳算法的適應函數能提高聚焦在一些比較難分類出的類別上；考慮 SHR(Sample Hit Rate)式(4)，計算在當一代中所有 PCA 圖中某一樣本 SHC 的平均值，CHR(Class Hit Rate)與 SHR 類似，定義在當一代屬同一類別所有樣本 SHR 的平均值，如式(5)所示。類別和樣本的權重會在每次新的子代更新一次，根據式(6)、(7)，再於式(2)、(3)重新做權重的比例分配。權重值變化對於提高基因演算法成功運作至關重要，從而改變適應函數的評分，降低陷入局部收斂的問題。

$$SHR_g(s) = \frac{1}{\emptyset} \sum_{i=1}^{\emptyset} \frac{SHC_i(s)}{K_c} , \forall s \in c \quad (4)$$

$$CHR_g(c) = \frac{1}{\emptyset} \sum_{s \in c} SHR_g(s) , \forall s \in c \quad (5)$$

$$CW_{g+1}(c) = CW_g(c) + (1 - CHR_g(c)) \quad (6)$$

$$SW_{g+1}(s) = SW_g(s) + (1 - SHR_g(s)) \quad (7)$$

第四章 實作

4.1 實驗設備

本實驗使用之設備符合 ZigBee 聯盟之規範，其利用無線網路 802.15.4 的標準所制定，其特色為低耗電及低傳輸速率。設備包含一個連接 PC 的接收器、多個感測節點，及一個定位標籤，其功能列表如下：

連接 PC 的接收器	接收定位標籤傳送的 RSSI 值，用以建立指紋資料庫或計算定位點機率。
感測節點	發送感測訊號。
定位標籤	接收感測節點傳送的訊號，並將其 RSSI 值傳送給接收器。

表 4.1 實驗設備簡介

4.2 實驗場地

本次實驗分兩組，第一組進行高斯模型與混合高斯模型的比較，實驗場地如下圖 4.1 為一約寬十公尺，長十二公尺之電腦教室，內部包含兩個白板，數十台電腦、電腦桌及座椅(黑色實心區域)。

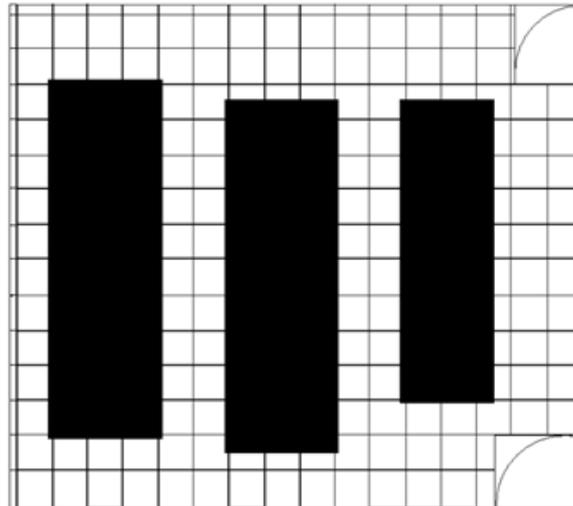


圖 4.1 實驗場地 1 示意圖

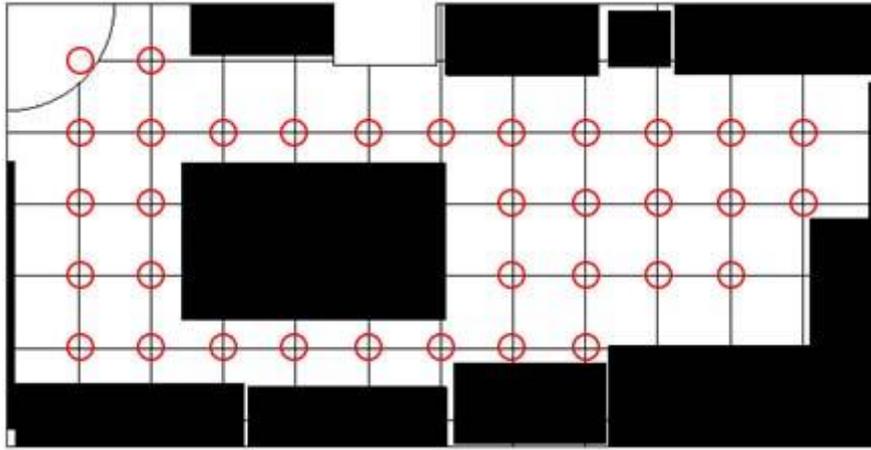


圖 4.2 實驗場地 2 示意圖

第二組進行高斯模型與 PLS 模型的比較，實驗場地如上圖 4.2 為一約寬四公尺，長約七公尺之研究室，內部包含兩個白板，四台電腦、五張桌子、數個櫃子及座椅(黑色實心區域)，定位點 34 個(紅色圓圈)。

在設定好感測節點的位置之後(本次實驗，兩組場地均為使用三個感測節點)，選定所有需要定位的位置後，開始進行場測階段，紀錄共定位點的 RSSI 值，每個點紀錄一千筆資料。

4.3 實驗結果

4.3.1 PSO 改良前後混合模型比較

RSSI 值收集完成後，針對每個定位點計算其分布機率，並分別建立單一高斯模型與高斯混合模型。

我們挑選多組定位資料樣本，分別以改良前與改良後的 PSO 演算法，進行了 0~50 組混合模型，迭代次數 0~50 次的測試，發現混合模型數在 4 組以上，以及迭代次數 40 代左右時，與原始數據的機率分布差異值，不再出現大幅變動，所以我們的指紋資料庫建立選用了 4 組高斯混合模型，並使用改良後的 PSO 演算法進行 50 迭代次數的運算，下圖 4.3, 4.4 分別為改良前與改良後 PSO 演算法對其中一組樣本作 0~50 組混合模型，迭代次數 0~50 次的測試結果。

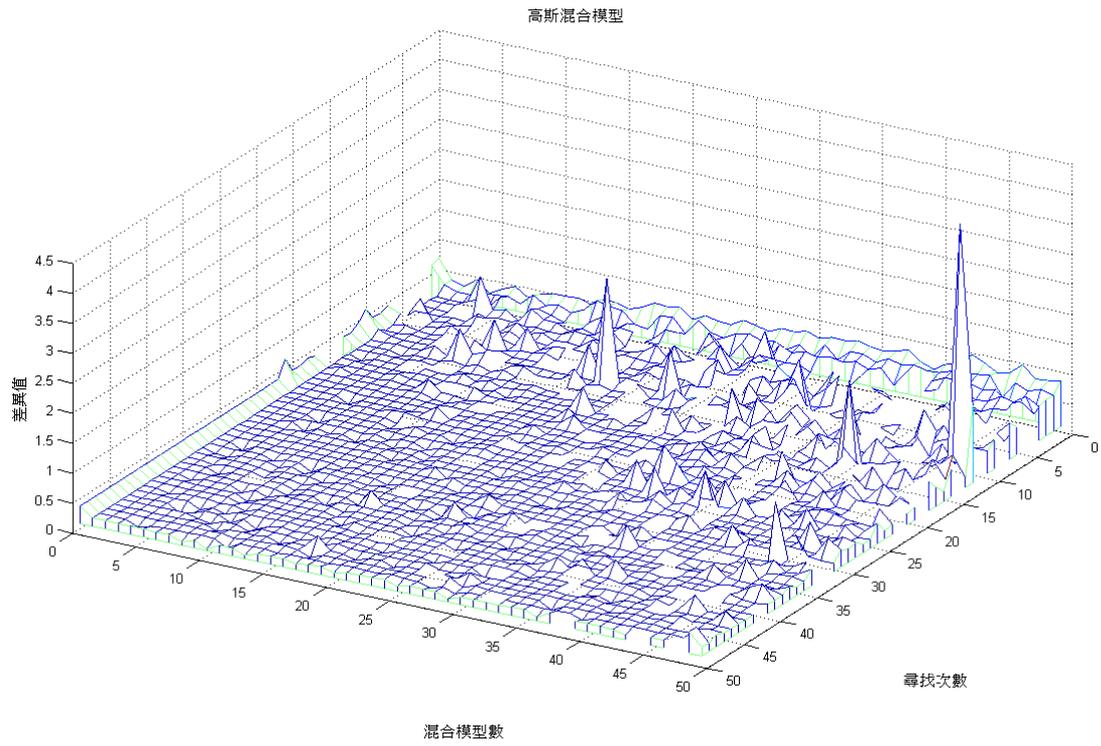


圖 4.3 改良前 PSO 差異值

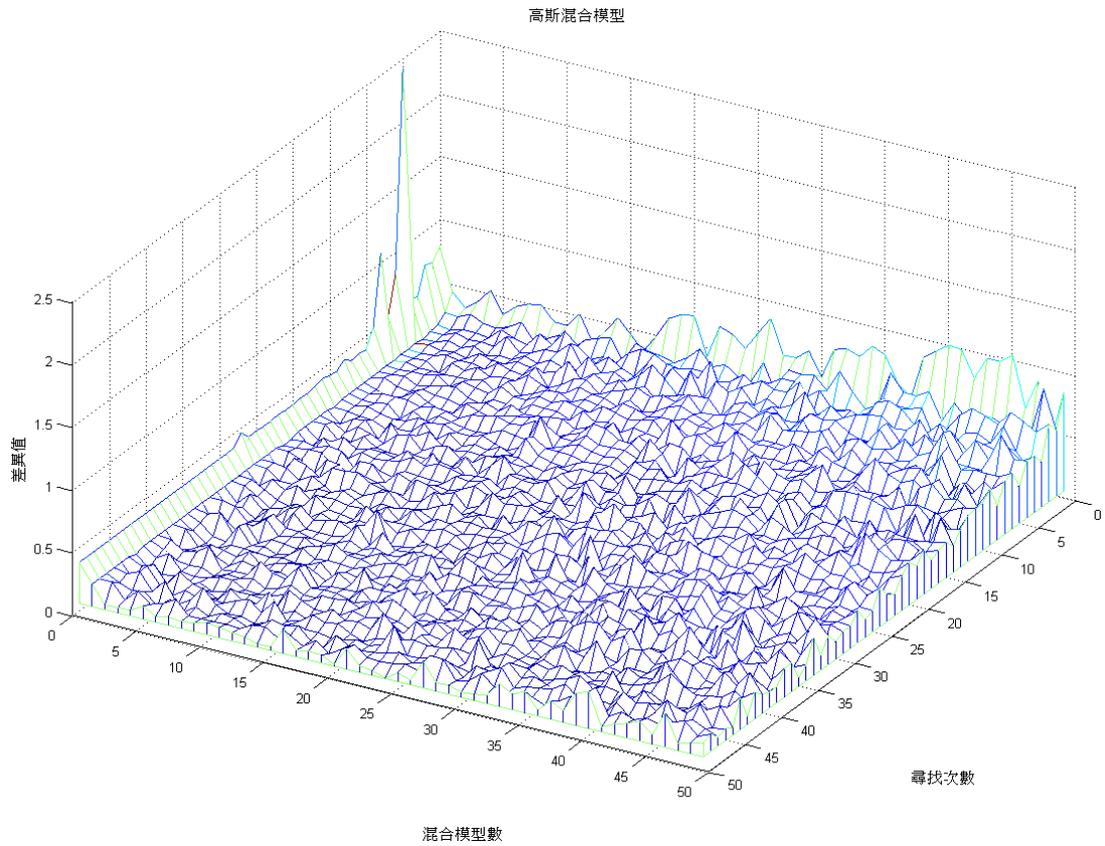


圖 4.4 改良後 PSO 演算法差異值

4.3.2 高斯與混合高斯模型定位結果比較

在實驗場地 1 完成指紋資料庫建立之後，我們對所測量的 97 個定位點進行定位實測，測量方式為，定位標籤分別收集三組感測節點各 10 筆 RSSI 強度，計算其 RSSI 平均值後，與指紋資料庫中資料進行比對，在機率累積的過程中，我們使用的是累乘的方式而計算總機率，最後再以最大概率法則(Maximum likelihood, ML)，來求出機率最大的定位結果。

我們首先使用單一高斯模型的指紋資料庫進行定位，在使用單一模型的情況下，定位點與實際位置的誤差總量高達 99 公尺以上，平均誤差約為 102 公分，而使用高斯混合模型的指紋資料庫進行定位，定位點與實際位置的誤差總量也高達 95 公尺，平均誤差約為 98 公分，其結果並不讓人滿意。

因此我們將 97 個定位點，縮減為 36 個定位點，拉大每個定位點的間距後，再次使用兩種模型的資料庫進行定位實測，使用單一高斯模型，定位點與實際位置的誤差總量降為 7.5 公尺，平均誤差約為 21 公分，而使用高斯混合模型的指紋資料庫進行定位，定位點與實際位置的誤差總量降為 6.8 公尺，平均誤差約為 19 公分，相對於單一模型雖然稍有改進，但是並不明顯。

4.3.3 PLS 分群

以 PLS 建立分群指紋資料庫時，我們先採用了 5 個點的資料進行測試，每個點包含三個感測節點強度各 600 筆，將此 1800 筆資料以穿插方式接續成一連續波形，再平均切成每段長度 600 的波形做為一組，並進行 GA 挑選。

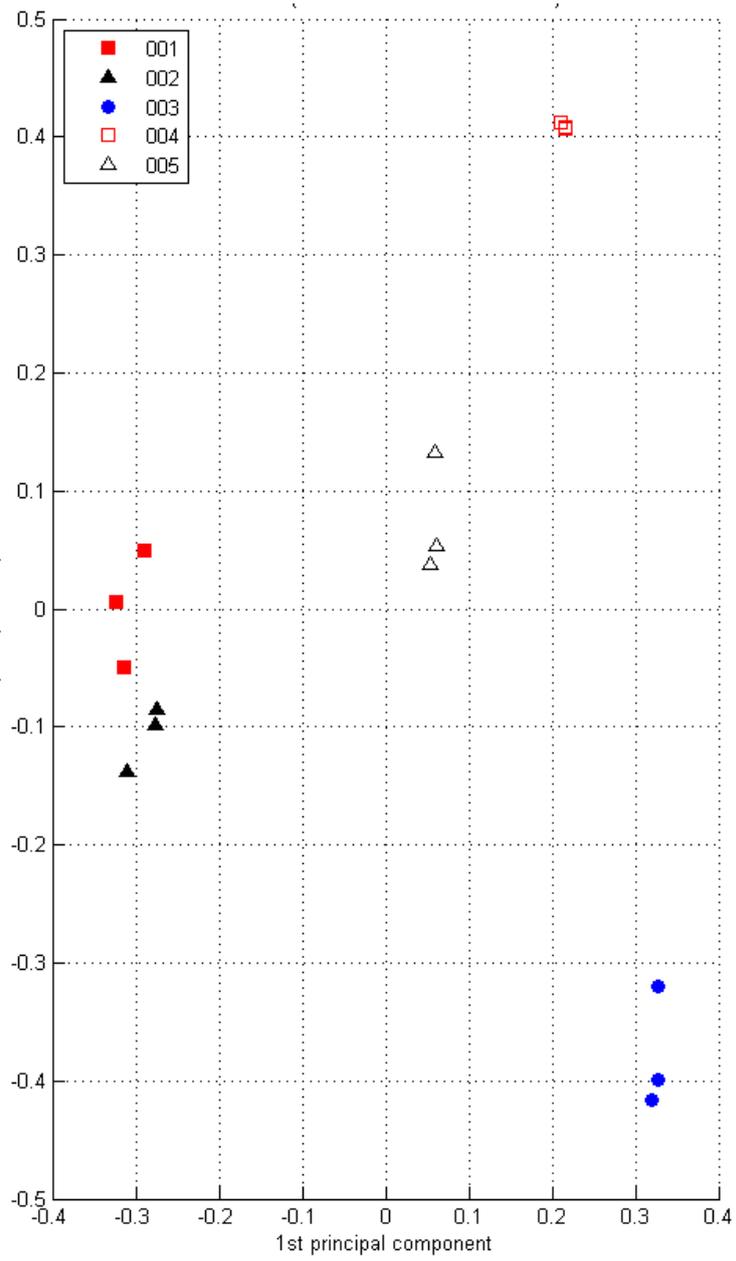


圖 4.5 PLS 分群資料(GA 前)

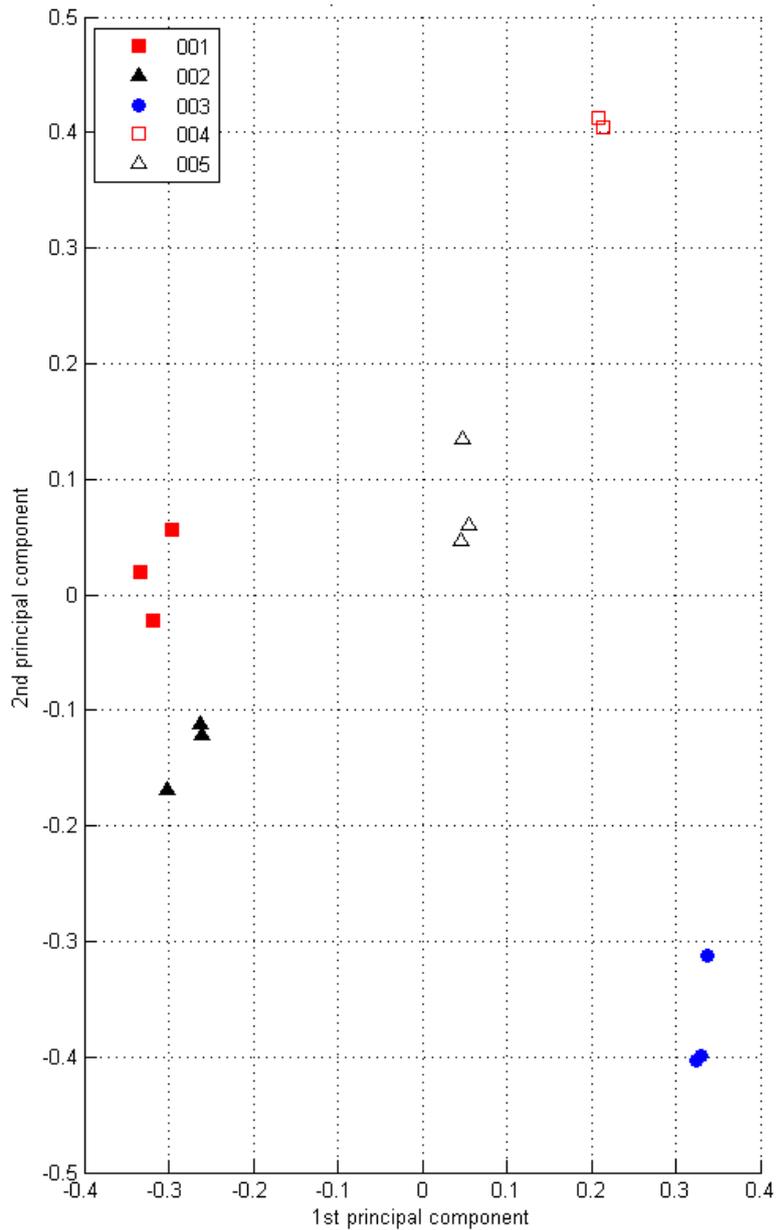


圖 4.6 PLS 分群資料(GA 後)

由上兩張圖可以看出，在點數較少的情況下，使用基因演算法挑選前後，並沒有大幅變動，但是挑選進行資料比對的資料量從，5 個點每點 3 組每組 600 筆，共 9000 筆資訊降為，5 個點每點 3 組每組 286 筆，共 4290 筆減少超過 50%。

經過測試後，決定將每筆資料分為長度 180 的波形 10 組，進行分群計算，分群結果如下圖 4.7。

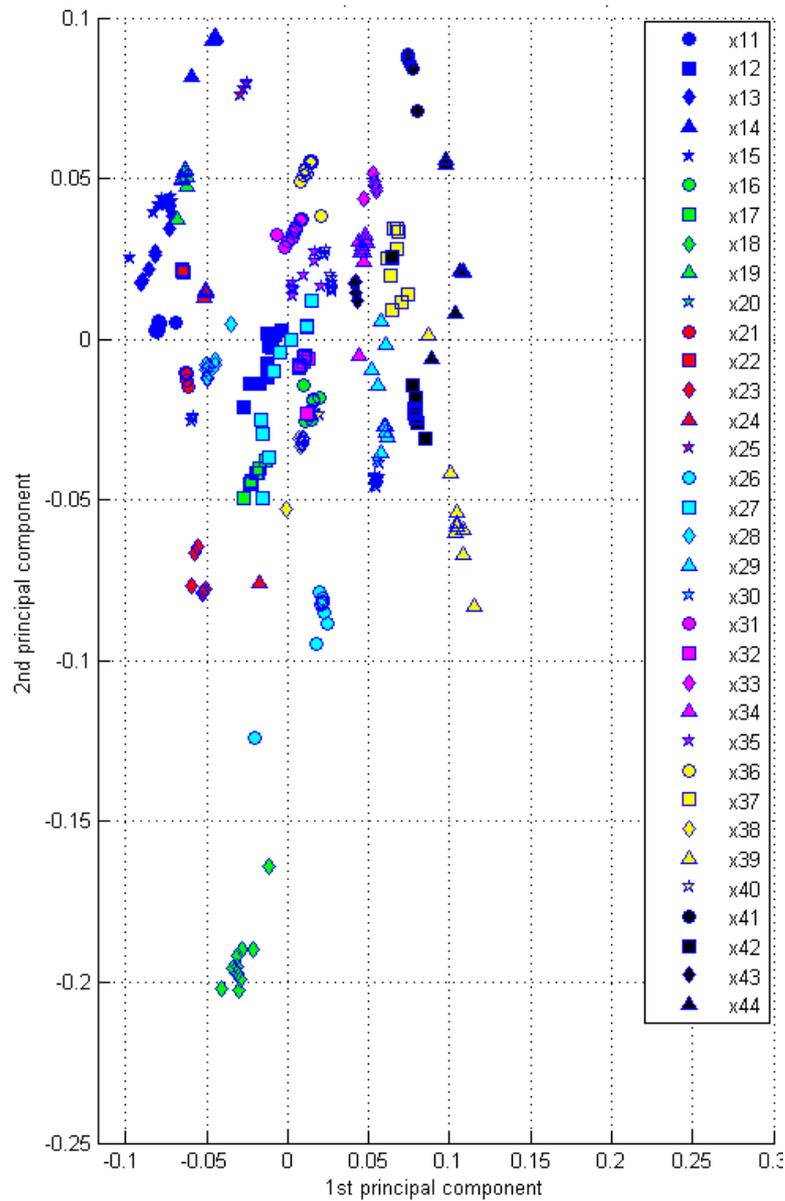


圖 4.7 PLS 分群資料(每點 10 組長度 180)

4.3.4 PLS 定位結果比較

在實驗場地 2 完成指紋資料庫建立之後，我們對所測量的 34 個定位點進行定位實測，測量方式為，定位標籤分別收集三組感測節點各 10 筆 RSSI 強度，將其 RSSI 值混合為一段波形，並視為連續波形重複，直到波形長度與建立指紋資料庫的樣本長度相同後進行比對。

比對方式為將定位資料與指紋資料庫中訓練的分群資料進行 PLS 運算，統計最靠近的前 10 個鄰近點中屬於哪個分群的數量最多，則判斷為定位結果，如果有數量相同的情況，則去除最遠的鄰近點，再次比較所屬分群之數量。

其中每個點最近距離為 60 公分，對每個點進行 40 次定位，將定位結果紀錄後統計，只有三個點定位正確機率低於 50%，分別是 0%，32.5%，50%，而另外 29 個點定位正確機率皆在 90% 以上，平均誤差值為 15.5 公分。

而同樣的資料用單一高斯模型進行定位，其結果只有 15 個點的資料有成功定位，且都是 40 次定位都命中，其他 19 個點定位失敗的，命中率都是 0，定位正確機率為 44%，平均誤差值為 94.4 公分。

使用同樣的資料用 PSO 演算法組建混合高斯模型進行定位，有 10 個點完全定位失敗，3 個點定位成功機率不到 90%，平均準確率為 65.15%，平均誤差值為 75.5 公分，三者定位結果如表 4.1。

定位點編號	PLS 定位		單一高斯模型		混合高斯模型	
	準確次數	機率	準確次數	機率	準確次數	機率
x11	13	32.50%	40	100.00%	40	100.00%
x12	37	92.50%	40	100.00%	40	100.00%
x13	40	100.00%	0	0.00%	38	95.00%
x14	40	100.00%	0	0.00%	40	100.00%
x15	39	97.50%	40	100.00%	40	100.00%
x16	34	85.00%	0	0.00%	0	0.00%
x17	38	95.00%	0	0.00%	0	0.00%
x18	40	100.00%	40	100.00%	40	100.00%
x19	40	100.00%	0	0.00%	0	0.00%
x20	40	100.00%	0	0.00%	0	0.00%
x21	40	100.00%	0	0.00%	40	100.00%
x22	40	100.00%	0	0.00%	0	0.00%
x23	39	97.50%	0	0.00%	40	100.00%
x24	40	100.00%	0	0.00%	0	0.00%
x25	40	100.00%	0	0.00%	40	100.00%
x26	40	100.00%	40	100.00%	40	100.00%

x27	0	0.00%	40	100.00%	0	0.00%
x28	40	100.00%	40	100.00%	40	100.00%
x29	40	100.00%	0	0.00%	40	100.00%
x30	40	100.00%	40	100.00%	40	100.00%
x31	40	100.00%	40	100.00%	40	100.00%
x32	30	75.00%	40	100.00%	30	75.00%
x33	40	100.00%	0	0.00%	0	0.00%
x34	37	92.50%	40	100.00%	40	100.00%
x35	20	50.00%	40	100.00%	40	100.00%
x36	40	100.00%	40	100.00%	40	100.00%
x37	40	100.00%	0	0.00%	17	42.50%
x38	39	97.50%	40	100.00%	40	100.00%
x39	40	100.00%	0	0.00%	40	100.00%
x40	40	100.00%	0	0.00%	38	95.00%
x41	40	100.00%	0	0.00%	40	100.00%
x42	40	100.00%	0	0.00%	0	0.00%
x43	39	97.50%	0	0.00%	3	7.50%
x44	40	100.00%	40	100.00%	0	0.00%
平均準確率		91.54%		44.12%		65.15%

表 4.1 實驗場地 2 定位命中次數統計

4.3.5 PLS 訓練資料分群與定位權重比較

接下來實驗對 PLS 定位訓練資料的分群數量，以及定位時對 PLS 指紋資料庫樣本比對依照相似度給予權重影響，以同樣的訓練及定位資料進行比較。

原本實驗中，將定位資料分為 10 段，也就是說原本 34 個定位點，三個感測節點強度各 600 筆，總共 61200 筆的資料，被我們分為 34 群，每群包含 10 組數據，每組數據長度為 180。

實驗中對分段數量進行改變，從不分切的 1 段長度 1800，到切成長度 90 的 20 段，進行了定位精準度與誤差量的比較，得到的數據資料如下表 4.2，根據數據繪製了圖 4.8 及 4.9。

分割份數	定位準確率(%)	平均誤差量(cm)
1	91.84	16.95
2	93.60	28.92
3	89.78	23.06
4	92.72	19.04
5	91.84	17.49
6	92.35	15.66
7	89.85	21.56
8	92.50	13.91
9	90.96	15.97
10	91.54	15.48
11	91.76	14.72
12	91.25	15.90
13	90.44	16.32
14	89.26	19.69
15	91.18	14.74
16	90.59	16.34
17	91.25	15.10
18	91.25	15.31
19	91.18	15.38
20	91.47	15.72

表 4.2 分割數量與定位結果比較

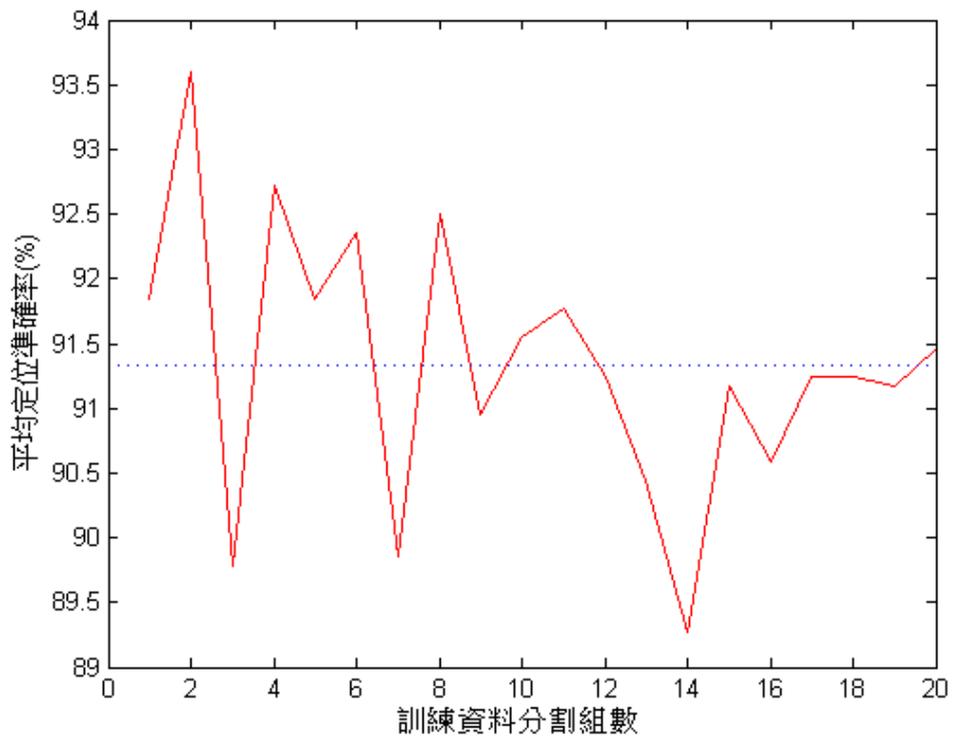


圖 4.8 分割組數與定位準確率

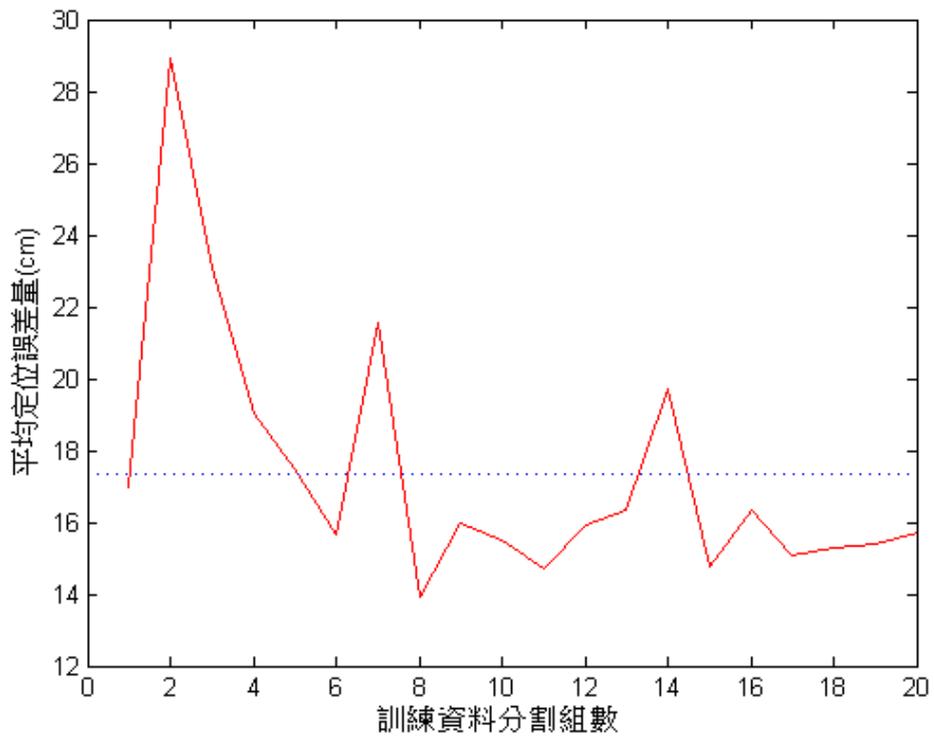


圖 4.9 分割組數與平均誤差量

數據顯示，訓練資料的分割量對定位的準確度關聯性不大，但是確能夠有效降低誤差量，會造成這種結果的原因可能是，定位失敗的時候，定位請求提供的強度值偏離所屬群組略遠，而偏離的位置正屬於多個群組都很靠近的區域，導致判斷相似度時被判給錯誤的群組，但是在分段數量提升時，即使判斷到錯誤的群組，也會更容易判斷給與正確定位點實際位置接近的定位點，因為這些定位點訊號強度與正確位置的差異比起遠離正確位置的差異小，這樣即使定位錯誤，也會定位到更接近正確位置的定位點，從而降低定位的平均誤差量。

比較過分割數量的影響後，接著對定位時判斷鄰近群組的方式加上權重分配，使得在 PLS 分布中越靠近的點，對最後定位結果影響越大得到的數據資料如下表 4.3。

分割份數	定位準確率(%)	平均誤差量(cm)
1	91.84	16.95
2	91.54	18.60
3	95.44	18.46
4	92.57	16.78
5	92.06	17.51
6	93.24	13.85
7	92.35	16.59
8	92.35	14.24
9	91.25	16.05
10	91.47	15.64
11	92.28	13.68
12	92.43	13.86
13	91.47	15.14
14	93.16	12.72
15	91.99	14.10
16	91.54	15.08
17	92.35	13.93
18	92.06	14.06
19	91.91	14.54
20	91.99	14.32

表 4.3 加入權重影響後，分割數量與定位結果比較

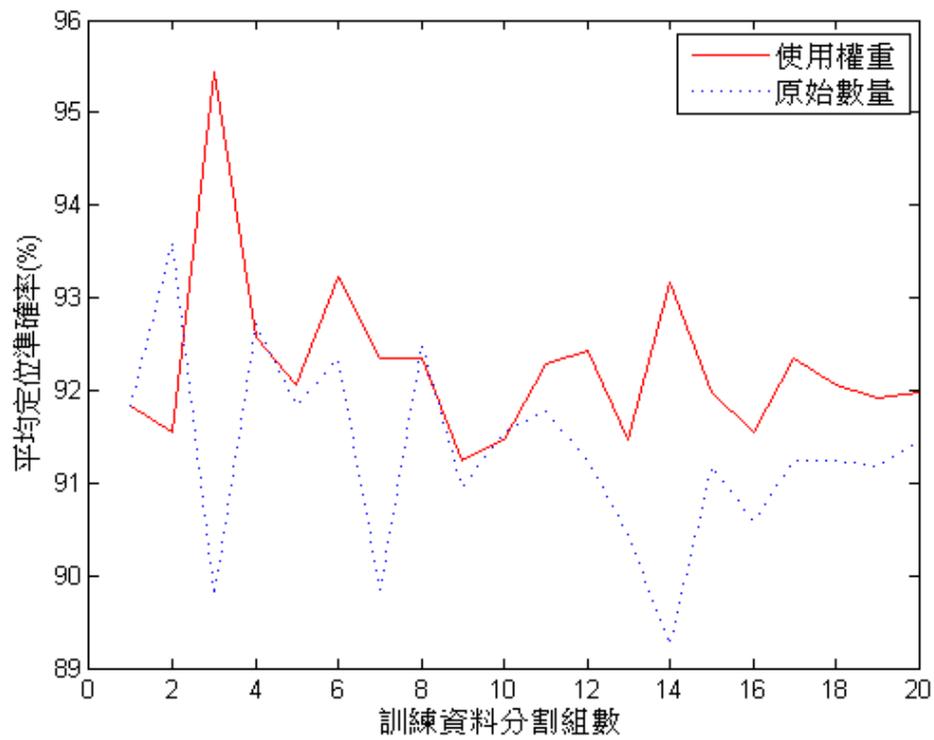


圖 4.10 使用權重定位準確率比較

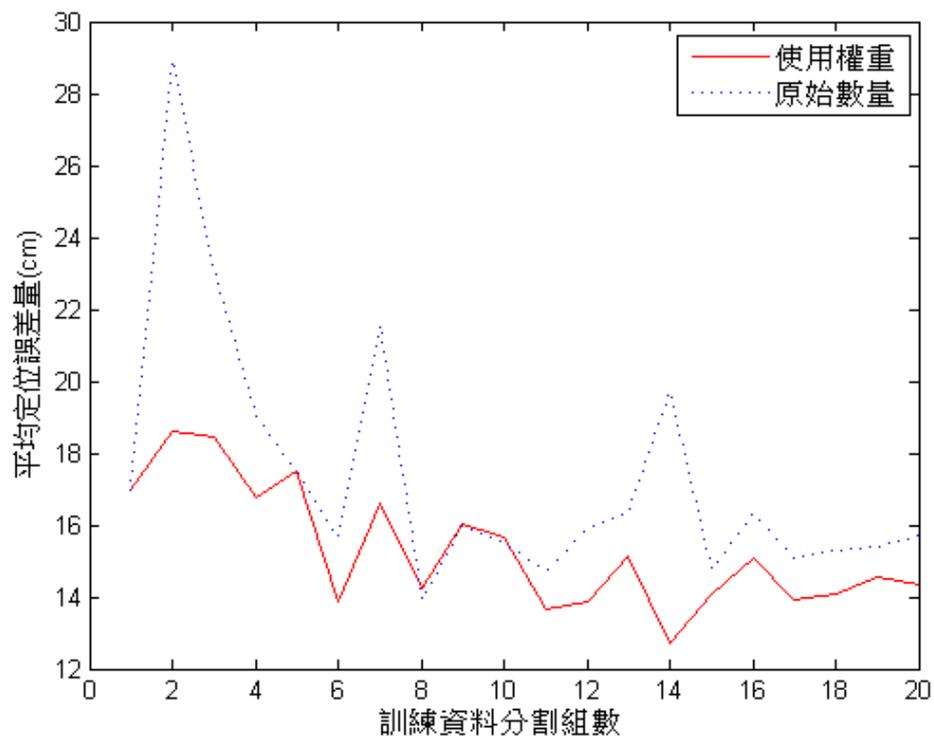


圖 4.11 使用權重平均誤差量比較

這裡使用的權重分配是比較簡易的，將依照分割數量配給權重，例如分割成 10 份，則 PLS 分群中與定位資料最靠近的 10 個點，將由近到遠分別獲得 11 到 1 的影響值，而獲得影響值總和最高的定位點為定位結果。

從圖 4.10 與圖 4.11 中，可以明顯看出使用權重和使用原始數量方式的差異，加上了權重影響後，定位的準確率大部分都是優於原始數量，而平均誤差量也因為準確率的提升而降低。

接著又對權重影響方式做了幾種測試，實驗結果如下表 4.4。

分割份數	累積次數		倒數		平方倒數		修正倒數 1		修正倒數 2	
	準確率	誤差								
1	91.84%	16.95	91.84%	16.95	91.84%	16.95	91.84%	16.95	91.84%	16.95
2	91.54%	18.6	91.54%	18.60	91.54%	18.60	91.54%	18.60	91.54%	18.60
3	95.44%	18.46	94.19%	14.01	94.19%	14.01	89.63%	22.94	95.44%	18.46
4	92.57%	16.78	91.91%	18.06	90.81%	20.06	91.99%	17.56	92.57%	16.78
5	92.06%	17.51	94.04%	13.54	95.74%	10.14	91.84%	17.43	92.06%	17.51
6	93.24%	13.85	93.90%	12.82	96.84%	7.75	92.13%	15.57	93.24%	13.85
7	92.35%	16.59	92.57%	18.39	91.40%	21.55	90.07%	20.67	92.35%	16.53
8	92.35%	14.24	93.16%	12.66	92.65%	15.79	92.43%	13.96	92.35%	14.24
9	91.25%	16.05	93.75%	11.80	95.66%	9.76	91.54%	15.13	91.25%	16.05
10	91.47%	15.64	93.68%	12.01	93.09%	14.24	91.25%	15.55	91.47%	15.64
11	92.28%	13.68	94.78%	10.03	92.79%	17.01	91.91%	14.43	92.28%	13.68
12	92.43%	13.86	95.00%	9.32	92.87%	20.01	91.18%	15.49	92.43%	13.86
13	91.47%	15.14	94.19%	10.61	91.18%	22.22	90.88%	16.02	91.47%	15.14
14	93.16%	12.72	95.51%	6.76	94.71%	9.53	90.07%	18.18	93.16%	12.72
15	91.99%	14.10	95.00%	9.25	93.53%	16.64	91.40%	14.58	91.99%	14.10
16	91.54%	15.08	95.66%	7.97	92.28%	17.58	90.81%	15.88	91.54%	15.08
17	92.35%	13.93	95.29%	9.04	92.87%	16.77	91.40%	15.21	92.35%	13.93
18	92.06%	14.06	95.29%	8.83	91.69%	17.30	91.32%	14.83	92.06%	14.06
19	91.91%	14.54	94.63%	10.08	91.69%	17.16	91.47%	14.73	91.91%	14.54
20	91.99%	14.32	95.37%	8.75	94.26%	11.51	91.40%	15.41	91.99%	14.32

表 4.4 各種權重測試比較

累積次數是前面實驗比較中使用的方式，也就是依據鄰近程度，給予比更遠的鄰居多 1 的影響值，例如判斷 10 個鄰居，則由近到遠給予 10 到 1 的影響權

重，倒數則是分別給予 1/1 到 1/10 的影響權重，平方倒數則是 1/1 到 1/100，修正倒數則是倒數加上一定值後再除以另一數的結果。

由上表中可以看出，使用倒數方式的定位準確率及平均誤差量在這次實驗數據中，優於原本的線性提升，而其他的方式則無明顯改善。

4.3.6 PLS 訓練資料數量對定位影響

對於參與訓練的資料筆數，進行了刪減的測試，以測試是否可以降低訓練資料收集量，而保持一定的定位精準度，於是將原本參與訓練的 600 筆資料，分別刪減為 500，400，300，200，100，然後進行 40 次定位來比較，實驗結果如下表 4.5。

訓練資料	累計次數法		權重分配法	
	定位準確率	平均誤差量(cm)	定位準確率	平均誤差量(cm)
100	72.72%	58.35	75.51%	42.49
200	84.93%	31.14	85.37%	29.06
300	86.03%	27.28	86.91%	25.93
400	90.88%	17.03	91.54%	14.95
500	91.10%	18.16	92.21%	14.02
600	91.54%	15.48	91.47%	15.64

表 4.5 降低訓練資料量對定位結果影響比較

從上表得知，訓練資料量在 400 筆以下，對定位準確率影響較大，提升到 400 筆以上後，對準確率的影響則不大。

4.4 結果討論

在實驗場地 1 中，單一高斯模型的指紋資料庫進行定位，定位點與實際位置的誤差總量高達 99 公尺以上，平均誤差約為 102 公分，而使用高斯混合模型的指紋資料庫進行定位，定位點與實際位置的誤差總量也高達 95 公尺，平均誤差約為 98 公分。

將 97 個定位點，縮減為 36 個定位點，拉大每個定位點的間距後，再次進行定位實測，使用單一高斯模型，定位點與實際位置的誤差總量降為 7.5 公尺，平

均誤差約為 21 公分，而使用高斯混合模型的指紋資料庫進行定位，定位點誤差總量降為 6.8 公尺，平均誤差約為 19 公分，相對於單一模型雖然稍有改進，但是並不明顯。

在實驗場地 2 中，只有 34 個定位點，但是密度較實驗場地 1 的時候更為密集，使用單一高斯模型的指紋資料庫進行定位，定位點與實際位置的平均誤差約為 94.4 公分，使用同樣的資料用 PSO 演算法組建混合高斯模型進行定位，平均誤差值為 75.5 公分，而使用 PLS 分群的指紋資料庫進行定位，定位點與實際位置的平均誤差僅約為 15.5 公分，相對之下有明顯的進步。

而在實驗場地 2 中，單一高斯模型的指紋資料庫進行定位，能成功定位的定位點，命中率都是 100%，定位失敗的定位點，命中率都是 0%，是一個值得注意的現象，因為這表示只要能定位成功，就一定準確，但只要一失敗，就全部都失敗，不存在於修正命中區域的狀況。

再第一次實驗中雖然在建立高斯混合模型的部分，成功的使用改良後的 PSO 演算法建立了逼近原始分布機率的模型，但是在定位階段，比較使用單一高斯模型與高斯混合模型的指紋資料庫進行定位的部分，並沒有產生顯著差異，經過討論後，其原因可能有以下幾種：原始訊號噪音過多，定位點間隔太密，感測節點數量不足，區域內有大量干擾物導致嚴重的多路徑效應，收集樣本相關。

其中原始訊號噪音過多，區域內有大量干擾物導致嚴重的多路徑效應，這兩者在真實環境中是難以避免的，但是可以使用演算法或其他方式對訊號做修正，譬如利用參考點間互相測量各鄰近節點的 RSSI 值，透過量測參考點 RSSI 值的改變，修正定位階段測得的 RSSI 資訊，降低多路徑效應的影響，及訊號強度減弱的影響。

定位點間隔太密則是因為距離差異不大，RSSI 強度差異過小，在原始資料庫中稍微差 1~2 的強度，可能就判斷到鄰近的定位點上，解決方式則有拉開定位點間距，以及增設感測節點，同時總機率的計算方式可能不能單純的使用累乘方

式，以避免因些微差距導致總機率變動過大。

而收集樣本相關則是，指紋法的定位成功與否，其指紋資料庫的建置方式影響重大，可能是收集的樣本數量不足，或是取得的樣本誤差太大不具有參考性，又或者因為環境變動導致樣本失效等，能改善的方式大概就是，提高設備的精準度，增加樣本的收集數，針對環境改變對 RSSI 樣本進行修正運算。

而將訊號由數值訊號轉為波形訊號，並以 PLS 進行分群，計算定位點在分群中的鄰近節點，統計鄰居節點的群組數量比例，達到了大幅提高定位精準度的結果是一大突破，可能的原因也許是因為將三組 RSSI 同時進行判斷，而不是分別計算出現機率，因為干擾產生時，同時影響了三組 RSSI，在繪製波形時能明顯的記錄下來，只做訊號出現機率統計時，短暫變化的訊號強度在整體樣本中所佔比例太低，容易被判斷到其他定位點去。

這次實驗中，PLS 分群後針對鄰居節點的遠近進行影響權重的分配，僅只有做最簡單的測試，如果針對權重分配方式進行更深入的分析，也許還能進一步的提高定位精準度，並降低定位誤差量，而訓練資料分群的數量，也是一個可以研究的方向，合理的分群數量，才能夠兼顧定位精準度與運算成本，分群數量過少，精準度不足，分群數量增加則能降低定位誤差量，但過多的分群數量，將會大幅提高定位時在 PLS 分群後尋找鄰居點的運算成本，是需要注意的地方。

第五章 結論

近年來，由於無線技術的發展，以及可攜式設備的普及，定位系統也成為大家重視的應用服務項目，全球定位系統即是一個為人們所熟知的戶外定位系統，它可以為絕大部分的戶外地區提供準確的定位，GPS 系統擁有許多優點：覆蓋面積廣、快速、省時、高效率。然而，對於由建築物遮蔽的室內環境，如封閉的辦公大樓或是半開放式的校園空間及工廠，GPS 系統依賴的衛星訊號會被嚴重削弱，失去參考價值，因此無法進行準確的定位。

因為室內環境擺設複雜，所以室內定位系統在實作上總是會產生不小的偏差，而為了降低誤差提高定位精準度，就要對無線訊號的強度擷取方式、資料庫建置、定位演算法等方面進行改善，在第一次實驗中，使用改良後的 PSO 演算法建立了逼近原始分布機率的高斯混合模型，在定位階段分別使用單一高斯模型與高斯混合模型的指紋資料庫，使用高斯混合模型指紋資料庫的定位精準度相較於使用單一高斯模型指紋資料庫略有提升，同時也降低了 4% 的平均定位誤差量。

為了進一步提升指紋定位精確度，提出一新穎的資料探勘及訓練方法。將訊號由數值訊號轉為波形訊號，以 PLS 進行分群，同時在離線訓練階段使用基因演算法來減少參與計算的資料量，透過計算定位點在分群中的鄰近節點所屬群組及間隔距離，統計鄰居節點所屬的群組數量，挑選所佔比例最高的群組來判斷定位點，這種定位方法大幅提高了定位的精準度，相較於單一高斯模型指紋資料庫定位準確率只有 44.12%，且平均定位誤差量高達 94.4 公分，以及混合高斯模型指紋資料庫定位準確率 65.15%，平均定位誤差量為 75.5 公分，使用 PLS 分群的判斷方式準確率高達 91.54%，正確命中定位點的次數是使用單一高斯模型指紋資料庫的兩倍以上，而平均定位誤差量也降到了 15.5 公分，比起用單一高斯模型指紋資料庫定位的結果，平均定位誤差量大幅降低 79 公分，也比混合高斯模型指紋資料庫平均定位誤差量低了 60 公分。

實驗結果正確達到預期目標，不僅定位準確率比單一高斯模型定位有雙倍之

提升，對比混合高斯模型定位也有接近 1.4 倍提升，同時對定位誤差也有大幅改善，此為本論文主要貢獻。對現今最盛行之物聯網應用佔有重要之角色。

參考文獻

- [1] 王兆慶, 黃國勝, “以 Wireless Sensor Network 作定位之實作”, 全國碩博士論文, 七月, 2006.
- [2] 李維平, 李元傑, 謝明勳, “以群中心策略改良人工蜂群演算法”, 資訊管理學報, Vo1. 21, no. 1, pp. 25-44, 2014.
- [3] 陳大鈞, “Measure Point selection Algorithms for Wireless Indoor Positioning Systems”, 國立臺灣大學資訊管理研究所碩士論文, 七月, 2004.
- [4] 陳嶽東, “以訊號強度為基礎利用高斯混合模型之 IEEE802.11 無線區域網路定位演算法”, 國立成功大學資訊工程研究所碩士論文, 六月, 2005.
- [5] 駱俊, 馬盡文, “高斯混合模型的遺傳分基融合算法”, 信號處理, Vo1. 21, no. 4A, pp. 395-398, 八月, 2005.
- [6] 韓佩君, “PLS2 演算法之比較”, 國立高雄大學統計學研究所碩士論文, 七月, 2010.
- [7] 顏啟森, 溫志宏, “無線感測網路之感測器定位系統實作”, 全國碩博士論文, 七月, 2005.
- [8] B. K. Lavine, et al., Raman Spectroscopy and Genetic Algorithms for the Classification of Wood Types, *Journal of Spectroscop*, 55(8):960-966, 2001.
- [9] Chang-Yi Hsu, “Implement of Personnel Tracking Wireless Sensor Networks.” National Cheng Kung University, 2008
- [10] De Jong, S. (1993). SIMPLS : an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18:251–263.
- [11] Dong-Yan Huang, Zhengchen Zhang, and Shuzhi Sam Ge, “Speaker state classification based on fusion of asymmetric simple partial least squares (simpls) and support vector machines,” *Computer Speech & Language*, vol. 28, no. 2, pp. 392–419, 2014.
- [12] F. Duvallet and A.D. Tews, "WiFi Position Estimation in Industrial Environments Using

- Gaussian Processes," Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems, pp. 2216-2221, Sept. 2008.
- [13] I-En Liao, and Kuo-Fong Kao, "Enhancing the Accuracy of WLAN-based Location Determination System using Predicted Orientation Information," Information Sciences, Vol.178, pp.1049-1068, 2008.
- [14] J.-G. Park et al., "Growing an organic indoor location system," in Proc. 8th Int. Conf. MobiSys, Appl., Serv., Jun. 2010, pp. 271–284.
- [15] J. Hightower, R. Want, and G. Borriello, "SpotON: An indoor 3D location sensing technology based on RF signal strength," Univ. Washington, Seattle, Tech. Rep. UW CSE 2000–02-02, Feb. 2000.
- [16] Janire Larranaga, Leire Muguira, Juan-Manuel Lopez-Garde, Juan-Ignacio Vazquez, "An Environment Adaptive ZigBee-based Indoor Positioning Algorithm," International Symposium on International Conference on Indoor Positioning and Indoor Navigation, pp.1-8, Sept. 2010
- [17] Jianwei Niu, Bowei Wang, Lei Shu, Duong, T.Q. , Yuanfang Chen, "ZIL: An Energy-Efficient Indoor Localization System Using ZigBee Radio to Detect WiFi Fingerprints," IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 33, NO. 7, JULY 2015.
- [18] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor location sensing using active RFID," Wireless Netw., vol. 10, no. 6, pp. 701–710, Nov. 2004.
- [19] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless LANs," Comput. Netw., vol. 47, pp. 825–845,
- [20] Michalski, A., Czajewski, J., "The accuracy of the global positioning systems," IEEE Instrumentation & Measurement Magazine, Volume 7, Issue 1, Mar. 2004.
- [21] Min Li Hunang and Sin-Chong Park, "A WLAN and ZigBee Coexistence Mechanism for Wearable Health Monitoring System," International Symposium on Communications and Information Technology, pp. 555-559, Sept. 2009.
- [22] M. Youssef, A. Agrawala, and A. Udaya Shankar, "WLAN location determination via clustering

- and probability distributions,” IEEE Int. Conf. Pervasive Comput. Commun., Mar. 2003, pp. 143–151.
- [23] M. Youssef and A. K. Agrawala, “Handling samples correlation in the Horus system,” IEEE INFOCOM 2004, Hong Kong, vol. 2, pp. 1023–1031, Mar. 2004.
- [24] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, “Locating the Nodes: cooperative localization in wireless sensor networks,” IEEE SIGNAL PROCESSING MAGAZINE, vol. 22, no. 4, pp. 54-69, July 2005.
- [25] Pablo Corral, Eladio Peña, Ricardo Garcia, V. Almenar, A. C de C. Lima, “Distance Estimation System based on ZigBee,” Proceedings of the Computational Science and Engineering Workshops, pp.405-411, July 2008.
- [26] P. Bahl, and V. N. Padmanabhan, “RADAR: An In-Building RF-based User Location and Tracking System,” IEEE Computer and Communications Societies, vol. 2, pp. 775-784, 2000.
- [27] P. Bahl and V. N. Padmanabhan, “Enhancements to the RADAR user location and tracking system,” Microsoft Corp., Tech. Rep. MSR-TR-2000–12, Feb. 2000.
- [28] R. Battiti, T. L. Nhat, and A. Villani, “Location-aware computing: A neural network model for determining location in wireless LANs,” Tech. Rep. DIT-02–0083, 2002.
- [29] S. Venkatraman, J. Caffery, and H. R. You, “A Novel TOA Location Algorithm Using Los Range Estimation for NLos Environments,” IEEE Trans. on Vehicular Technology, vol. 53, no.5, pp. 1515-1524, 2004.
- [30] Tai-Wei Lu, “Outdoor Localization Algorithm in Wireless Sensor Networks,” Yuan Ze University, 2009.
- [32] U. Klee, T. Gehrig, and J. McDonough, “Kalman filters for time delay of arrival-based source localization,” EURASIP Journal on Applied Signal Processing, vol. 2006, no. 12378, pp. 1-15, Oct 2006.
- [33] V. Otsason, A. Varshavsky, A. LaMarca, and E. de Lara, “Accurate GSM indoor localization,” UbiComp 2005, Lecture Notes Computer Science, Springer-Verlag, vol. 3660, pp. 141–158, 2005.

- [34] Y. Wu, J. Hu, and Z. Chen, "Radio Map Filter for Sensor Network Indoor Localization Systems."
2007 5th IEEE International Conference, vol.1 23-27, pp.63-68, June 2007.