



結合影像和慣性元件之 定位技術與系統

版 別：第 1.0 版

總 頁 數： 頁 （含封面及附件）

文件編號：



Abstract

在早期的社會裡，傳統常見的定位技術主要以全球定位系統(Global Positioning System, GPS)或雷達(Radio Detection And Ranging, RADAR)為主，其中，GPS 以廣泛、全球覆蓋率高、可移動定位和高精度著稱；而 RADAR 以高靈敏度與高可靠性為主要優點。在現代，室外定位由於全球定位技術的發展已趨近成熟，然而，室內定位因室內環境為一複雜空間，無線或衛星訊號在傳播的過程中難免受到室內建築結構、天候以及訊號強度等影響，造成室內精確定位的困難。基於此，本文針對當前室內定位技術的研究現況進行探討，並整理了近年來的相關研究。大致上，取得室內定位資訊上的方法可分為使用純影像、純慣性元件與影像結合慣性元件等三大類，其中又以影像結合慣性元件定位較為精確。

1. 介紹

在現代社會裡，全球定位技術的發展越趨成熟之時，我們的生活基本上受到全球定位系統、輔助全球定位系統(Assisted Global Positioning System, AGPS)或無線網路等全面追蹤。突飛猛進的定位技術將所有相關應用如 Wi-Fi、藍牙、GPS 以及 AGPS 整合到了一片小小的晶片之中，以達到全方位定位與節省成本、功耗等目的。然而，室內環境是一個複雜的空間，無線或衛星訊號在傳播的過程中難免受到室內建築結構的影響，造成室內精確定位的困難。

有鑑於此，室內環境的移動物體追蹤與定位在近年來有相當多的研究與應用，以往發表過的系統通常是量測待定位目標物體與外部感測器如攝影機或雷射掃描器等相對運動來達到追蹤或定位的目的。其中，又以利用視覺影像分析為基礎的定位技術最被廣泛探討，主因為基於視覺分析的定位系統精確度比較高。然而，電腦視覺分析的演算法往往複雜較高，因而需要付出較高的計算量；另外，移動物體超出感測器的監控範圍或是量測目標物被遮蔽，都會進一步影響影像式定位系統的功效，因



此也限制了影像式定位系統的應用層面。為了解決視覺上定位的限制，許多研究在欲定位的目標物上加裝慣性量測元件(Inertial Measurement Unit, IMU)等內部原件以輔助外部感測器的定位。慣性量測元件主要包含了加速度儀(Accelerometer)與陀螺儀(Gyroscope)，可用以量測加速度和角速度。這些資訊經過初始位置設定以及資料整合後，可推論得到載具(目標物)位置、速度和姿態的趨近資料。由於慣性量測元件不會有影像式定位的問題，可被妥善用來彌補影像式定位系統的不足；相反地，以慣性量測為主的定位系統會因偏差(Bias)和飄移(Drift)而造成最後定位的誤差，這部分則可用利用影像式定位系統的高精確度特性來補強。當前，有許多研究探討如何利用外部感測器與內部感測器的互補性來強化室內定位系統的效能，其中，利用“擴展型卡爾曼濾波器(Extended Kalman Filter, EKF)”等類的技術來模擬統計式運動模型，以達到穩定的資訊整合，最近被廣泛地討論並已得到初步的研究成果。以下我們分別對當前的定位技術進行探討並比較。

2. 相關研究

關於定位種類，可分為室外定位與室內定位，室外定位不外乎就是經由GPS和RADAR來實現，而室內定位大致上可分為以下三種方法，純影像、純慣性元件和影像結合慣性元件等三種方法，其目的都是要定位使用者的位置或姿態，在此我們將分別介紹以下5種方法的室內室外方法相關研究並比較各種方法的優劣。

2.1 全球定位系統

首先，我們先介紹目前最常見的室外定位方法，全球定位系統 GPS，目前導航軟體絕大部分都是依賴 GPS 來實現定位功能[1]。

GPS 系統的前身為美軍研製的一種子午儀衛星定位系統，1958 年研製，1964 年正式投入使用。該系統用 5 到 6 顆衛星組成的星網工作，每天最多繞過地球 13 次，並且無法給出高度資訊，在定位精度方面也不盡如人意。然而，子午儀系統使得研發部門對衛星定位取得了初步的經驗，並驗證了由衛星系統進行定位的可行性。由於衛星定位顯示出在導航方面的巨大優越性及子午儀系統存在對潛艇和艦船導航方面的巨大缺陷。美國海陸空三軍及民用部門都感到迫切需要一種新的衛星導航系統。為此，美國海軍研究實驗室(NRL)提出了名為 Tinmation 的用 12 到 18 顆衛星組成 10000km 高度的全



球定位網計劃，並於 67 年、69 年和 74 年各發射了一顆試驗衛星，在這些衛星上初步試驗了原子鐘計時系統，這是 GPS 系統精確定位的基礎。而美國空軍則提出了 621-B 的以每星群 4 到 5 顆衛星組成 3 至 4 個星群的計劃，這些衛星中除 1 顆採用同步軌道外其餘的都使用周期為 24h 的傾斜軌道。該計劃以偽隨機碼（PRN）為基礎傳播衛星測距信號，其強大的功能，當信號密度低於環境噪聲的 1% 時也能將其檢測出來。偽隨機碼的成功運用是 GPS 系統得以取得成功的一個重要基礎。海軍的計劃主要用於為艦船提供低動態的二維定位，空軍的計劃能供提供高動態服務，然而系統過於複雜。由於同時研製兩個系統會造成巨大的費用，而且這裡兩個計劃都是為了提供全球定位而設計的，所以 1973 年美國國防部將二者合二為一，並由國防部牽頭的衛星導航定位聯合計劃局（JPO）領導，還將辦事機構設立在洛杉磯的空軍太空處。該機構成員眾多，包括美國陸軍、海軍、海軍陸戰隊、交通部、國防製圖局、北約和澳大利亞的代表。

GPS 系統主要由空間星座部分、地面監控部分和用戶設備部分組成，其中空間星座部分，GPS 衛星星座由 24 顆衛星組成，其中 21 顆為工作衛星，3 顆為備用衛星。24 顆衛星均勻分布在 6 個軌道平面上，即每個軌道面上有 4 顆衛星。衛星軌道面相對於地球赤道面的軌道傾角為 55° ，各軌道平面的升交點的赤經相差 60° ，一個軌道平面上的衛星比西邊相鄰軌道平面上的相應衛星升交角距超前 30° 。這種布局的目的是保證在全球任何地點、任何時刻至少可以觀測到 4 顆衛星；地面監控部分主要由 1 個主控站（Master Control Station, MCS）、4 個地面天線站（Ground Antenna）和 6 個監測站（Monitor Station）組成，主控站是整個地面監控系統的管理中心和技術中心，天線站的作用是把主控站計算得到的衛星星歷、導航電文等資訊注入到相應的衛星，故注入站同時也是監測站，監測站的主要作用是採集 GPS 衛星數據和當地的環境數據，然後發送給主控站；用戶設備主要是 GPS 接收機，作用是從 GPS 衛星收到信號並利用傳來的資訊計算用戶的三維位置及時間。

再者，為了使精確度提昇，科學界發展另一種技術，稱為差分全球定位系統（Differential GPS, DGPS）[2]。亦即利用附近的已知參考座標點，來修正 GPS 的誤差。再把這個即時誤差值加入本身座標運算的考慮，便可獲得更精確的值。



主要特點：

- 全天候，不易受任何天氣的影響
- 全球覆蓋（高達 98%）
- 三維定點定速定時高精度
- 快速、省時、高效率
- 應用廣泛、多功能
- 可移動定位。

2.2 雷達

其工作原理是將電磁能量以定向方式發射至空間之中，藉由接收空間內存在物體所反射之電波，可以計算出該物體之方向，高度及速度，並且可以探測物體的形狀。以地面為目標的雷達可以用於探測地面的精確形狀。

而早期雷達是由於二戰期間當時英國和德國交戰時，英國急需一種能探測空中金屬物體的雷達（技術）能在反空襲戰中幫助搜尋德國飛機。二戰期間，雷達就已經出現了地對空、空對地（搜索）轟炸、空對空（截擊）火控、敵我識別功能的雷達技術；當代雷達同時多功能的能力使得戰場指揮員在各種不同的搜索與跟蹤模式下對目標進行掃描，並對干擾誤差進行自動修正，而且大多數的控制功能是在系統內部完成的。

早期的雷達天線是固定的、無方向的陣列，只有距離信息。天線在一定的時間間隔內發射射頻脈衝，將接收到的回波放大，並在示波器的 CRT 上顯示，產生一個與目標位置對應的水平線，供雷達操作員識別目標的大致距離，但由於當時所用的射頻電波頻率較低，為了有效地發射和接收射頻信號，雷達系統需要一個很大的天線，這種天線不能遷移或者改變方向，而且只能探測到大目標，且距離信息的精度也很低[3]，如圖(1)



圖(1). 雷達站

主要特點：

- 同時多功能
- 傳感器融合
- 高靈敏度
- 隱身
- 反隱身
- 雷達 ECCM(增強感應，使其干擾成功率降低)
- 自動目標識別
- 高可靠性

2.3 純影像的定位

關於室內定位，首先，我們先介紹以純影像來實現定位的方法，所謂純影像就是以單純的影像來實現定位，因此方法不結合慣行元件，在此，我們介紹[4]所提出的 MoVIPS(Mobile Visual Indoor



Positioning System)，它主要使用了距離估測法和影像辨識技術來獲得更精確的定位，作者使用智慧型手機來實現定位，因為智慧型手機硬體規格日趨成熟，其所提供的相機更是扶搖直上，固使用智慧型手機就可以拍攝出高解析度的影像，以提供更多的特徵資訊，使拍攝影像與資料庫的匹配更準確，而這裡的資料庫包含了多張影像和影像拍攝時所處的位置資訊，且每張影像進行特徵提取。使用者拍攝影像後，擷取影像的特徵，接著與資料庫中的影像進行匹配，找出使用者拍攝影像時所處的位置，最後使用 MoVIPS 使定位更精確。

要以影像實現定位，不外乎就是要比對影像，而比對影像，就必須搜尋特徵點，搜尋特徵點早期較著名的有尺度不變特徵轉換(Scale invariant feature transform, SIFT)，而作者在這裡使用加速穩健特徵(Speed Up Robust Features, SURF)，簡單地說 SURF 是 SIFT 的改良，因 SIFT 計算速度較慢，故 SURF 就是要改進其計算速度，使其更有效率，以下先分別介紹 SIFT 以及 SURF。

SIFT 是一個尺度不變的特徵點擷取與描述演算法，其獨特的特徵點描述法可以不受旋轉、縮放、位移、雜訊與亮度的差異所影響，這樣的特徵點描述法因此非常適合應用於影像間的特徵點匹配。在特徵點的擷取上，SIFT 使用(1)高斯金字塔(Gaussian pyramid)與(2)高斯差分(Difference-of-Gaussian, DOG)的方式來取樣各個不同尺度的候選特徵點，以達到尺度不變的目的。在特徵點的描述上，SIFT 計算(1)影像的梯度強度與方向，(2)統計特徵方向的直方圖，(3)找出特徵點的主要方向，最後(4)構成特徵向量，以達到旋轉不變的特性。至於在影像間的特徵點匹配上，各類特徵向量相似性的量測法都可適用，以 [5] 為例，作者直接利用單純的歐式距離(Euclidean distance) 相似性量測法來計算特徵點之間的匹配。下面我們說明 SIFT 演算法的詳細步驟。

SIFT 演算法主要可分為四個步驟，尺度空間極值偵測、特徵點定位與篩選、計算特徵點方向性以及特徵點描述，以下說明各個步驟的細節。

(1) 尺度空間極值偵測

Koenderink(1984)和 Lindeberg(1994)認為一張經過高斯函數處理後的模糊影像最能代表不同的尺度空間(scale-space)，如圖影像經過放大縮小後的情形，而為了使特徵點有尺度不變的特性，利用了高斯金字塔與高斯差分，如圖(2)，將尺度空間中可能出現的特徵

點找出來。

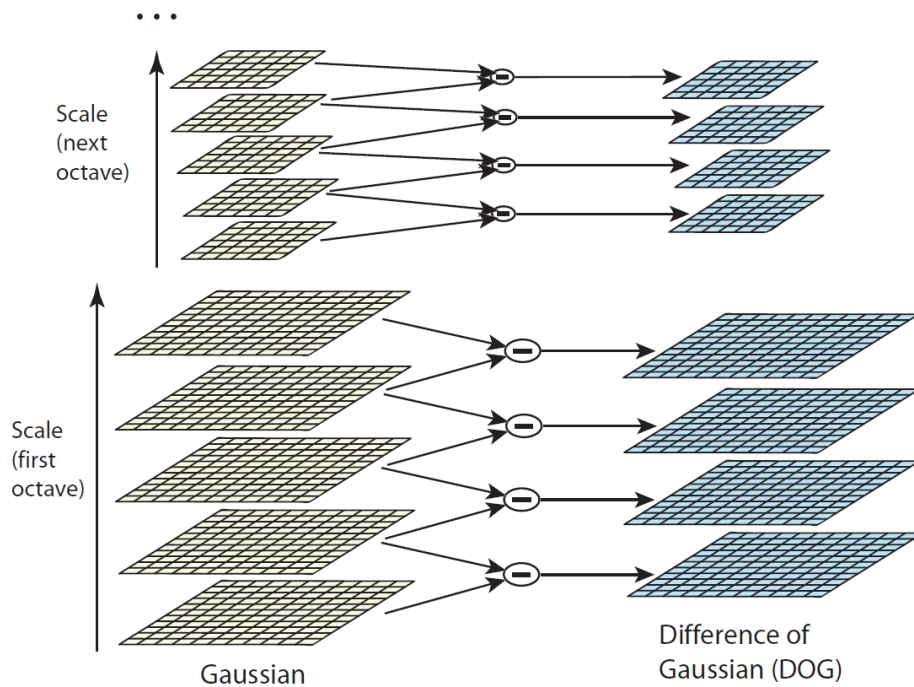
首先定義 σ 為一尺度空間的尺度， σ 也代表高斯函數的標準差，如公式(1)，並將原始影像與高斯函數做摺積(Convolution)運算，將影像模糊化，如公式(2)，其中 $I(x, y)$ 為原始影像， $L(x, y, \sigma)$ 為一尺度空間，也就是高斯模糊後的影像。

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp(-(x^2 + y^2)/2\sigma^2) \quad (1)$$

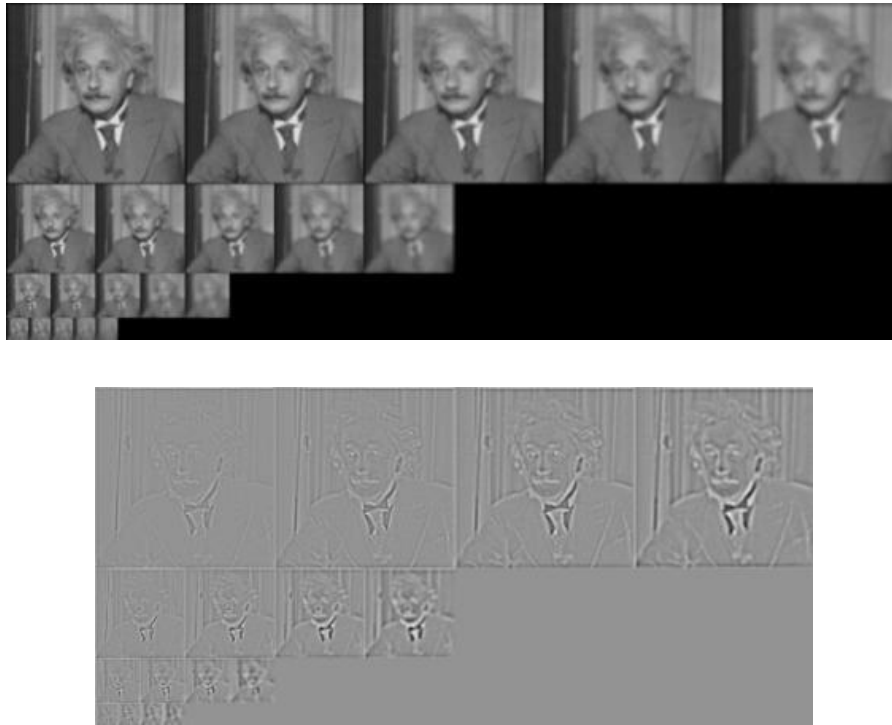
$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2)$$

每張模糊影像之間相隔 k 倍的 σ ，組成高斯金字塔的第一層(octave)，再將影像降採樣(down-sampled)後，開始第二層的高斯金字塔，此時第二層的第一張影像尺度為 2σ ，以此類推形成高斯金字塔，之後再將每層中的影像兩兩相減，進而形成高斯差分，如公式(3)，其中 $D(x, y, \sigma)$ 為高斯差分後的影像，實際的高斯金字塔與高斯差分結果如圖(3)。

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (3)$$

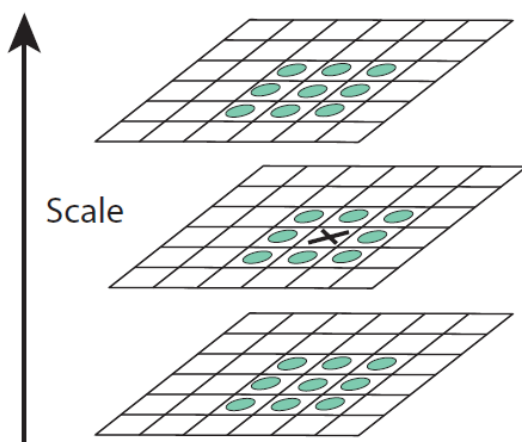


圖(2). 高斯金字塔與高斯差分示意圖

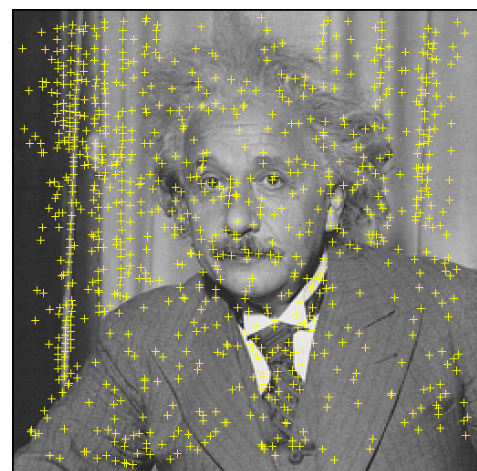


圖(3). 上圖為高斯金字塔，下圖為高斯差分的實際結果

取得高斯差分影像後，即可進行極值的偵測。所謂的極值(extrema)就是每一組高斯差分影像中，若某點的像素值為本身以及前後尺度共 26 個鄰近點中的區域最大值或區域最小值，如圖(4)，則稱此點為極值點，也就是此影像候選的特徵點。將所有高斯差分的影像進行極值偵測，就可以找出所有候選的特徵點，並決定此特徵點的最佳觀察尺度，如圖(5)。



圖(4). 某點與 26 個鄰近點比較



圖(5). 所有候選的特徵點

(2) 特徵點定位與篩選

經過尺度空間的極值偵測所找出的特徵點，並非全部都穩定，所以必須將一些不穩定的候選特徵點刪除，在這個步驟中將會把低對比度和邊緣的點去除。首先，將公式(3)泰勒展開得公式(4)，

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x} \quad (4)$$

其中， $\mathbf{x}$ 為極值點 $(x, y, \sigma)^T$ ， D 為高斯差分後的結果。將公式(4)對 $\mathbf{x}$ 偏導數，並令偏導數等於零，可得極值點的偏移量 $\hat{\mathbf{x}}$ ，如公式(5)，這個值可以看作是區域極值的實際位置，

$$\hat{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2} \frac{\partial D}{\partial \mathbf{x}} \quad (5)$$

然後再把此偏移量代入公式(4)，若 $|D(\hat{\mathbf{x}})|$ 小於 0.03，就表示此點為低對比度的特徵點，並將此候選特徵點刪除。

接下來要將邊緣的點刪除，利用類似 Harris 的角落偵測，首先計算出海森矩陣(Hessian matrix)，如公式(6)，其中 D_{xx} 表示高斯差分對 x 偏導數兩次； D_{yy} 表示對 y 偏導數兩次，而 D_{xy} 則表示高斯差分對 x 偏導數後，再對 y 偏導數，

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (6)$$

這裡定義 α 為海森矩陣最大的特徵值， β 為最小的特徵值，所以海森矩陣的跡(trace)和行列式(determinant)可以如公式(7)和公式(8)表示，然後令 $\alpha = r\beta$ ，可以推導出公式(9)，

$$\text{Tr}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta \quad (7)$$

$$\text{Det}(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \quad (8)$$

$$\frac{\text{Tr}(\mathbf{H})^2}{\text{Det}(\mathbf{H})} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \quad (9)$$

因此得以下判斷式，若不等式不成立，則表示此候選的特徵點為邊緣的點，並將此候選特徵點刪除，此時的 r 設為 10。

$$\frac{\text{Tr}(\mathbf{H})^2}{\text{Det}(\mathbf{H})} < \frac{(r+1)^2}{r} \quad (10)$$

(3) 計算特徵點方向性

這個步驟的主要目的，就是要將影像轉到與特徵點相同的方向，此時相對應的特徵點就會在相同的方向描述特徵向量，因此就可以使特徵不會受到旋轉而影像。

首先，以特徵點的位置為中心，在此時特徵點所在尺度的高斯模糊影像中取一個區塊，計算區塊內所有像素的梯度強度與方向，計算如公式(11)和公式(12)，其中 $m(x,y)$ 代表像素點 (x,y) 上的梯度強度； $\theta(x,y)$ 代表像素點 (x,y) 上的梯度方向。

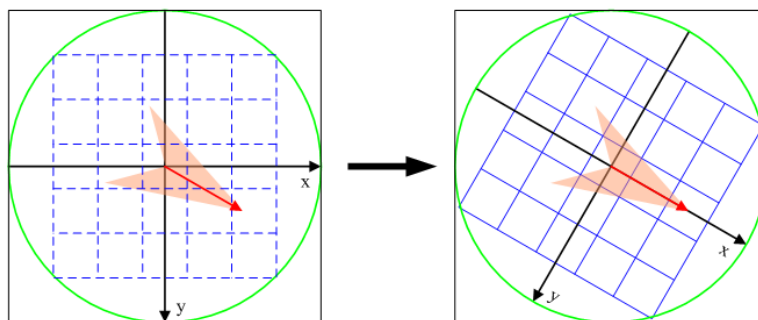
$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \quad (11)$$

$$\theta(x,y) = \tan^{-1}((L(x+1,y) - L(x-1,y)) / (L(x,y+1) - L(x,y-1))) \quad (12)$$

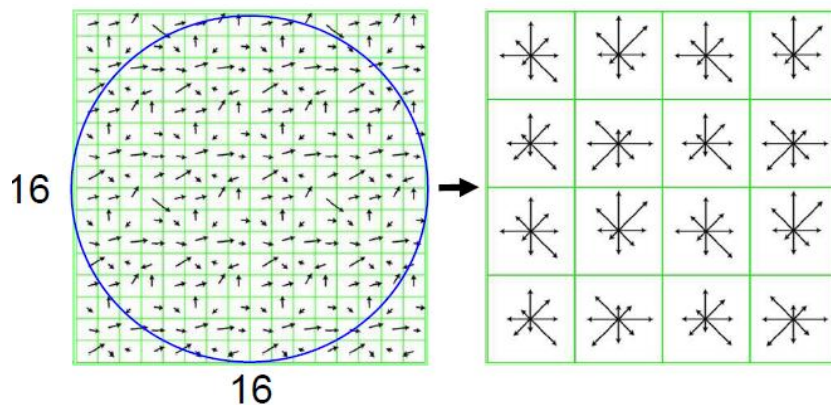
求出梯度的強度及方向後，將所有像素點的梯度強度及方向統計成一個以 10 度為間隔的方向直方圖，而其總和最高的方向，即為此特徵點的主要方向。

(4) 特徵點描述

求出每個特徵點的方向後，接下來就是描述特徵點的特徵向量，首先旋轉影像的主軸，使其和特徵點的主方向相同後，如圖(6)，同樣以特徵點為中心，取一 16x16 大小的區塊，並加入一個尺度為 0.5σ 的高斯函數作為權重，再以每 4x4 大小為一個小區塊，共切割成 16 個小區塊，接著統計每個小區塊內每個像素的梯度強度與方向，每個小區塊有 8 個方向，代表有 8 個特徵向量，因此總共有 $4 \times 4 \times 8 = 128$ 個特徵向量，如圖(7)，最後將特徵向量正規化，以避免亮度差異造成的影響。

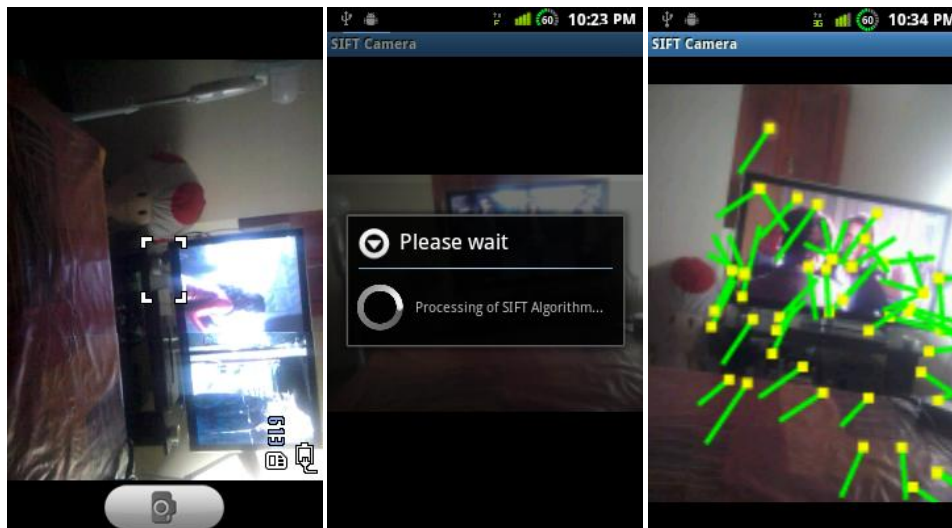


圖(6) 史影像與特徵點方向相同



圖(7). SIFT 特徵點描述

為了應用在智慧型手機上的室內定位，可以將 SIFT 實作在 Android 作業系統上，如圖(8)，這麼一來就可以直接將智慧型手機所拍攝的影像進行特徵的擷取[6]。



圖(8). SIFT 實作在 Android 作業系統上

以上介紹了 SIFT 演算法，接下來介紹作者所使用的 SURF，SURF 簡單地說就是 SIFT 的改良，其計算速度比 SIFT 更快且有效率。而 SURF 演算法主要分為特徵點偵測與特徵點描述，特徵點的偵測與 SIFT 相同都是用來找出區域特徵點的坐標，不同的是 SURF 結合了盒子濾波器(box filter)與積分圖像(Integral image)的方法，縮短影像與高斯濾波器進行摺積運算所花費的時間。而特徵點的描述也與 SIFT 不同，SURF 是利用特徵點鄰近區域內的 Haar 小波響應(Haar wavelet responses)分佈，描述出特徵點的特徵向量，以下簡單地介紹 SURF 的原理[6]。

(1) 特徵點偵測

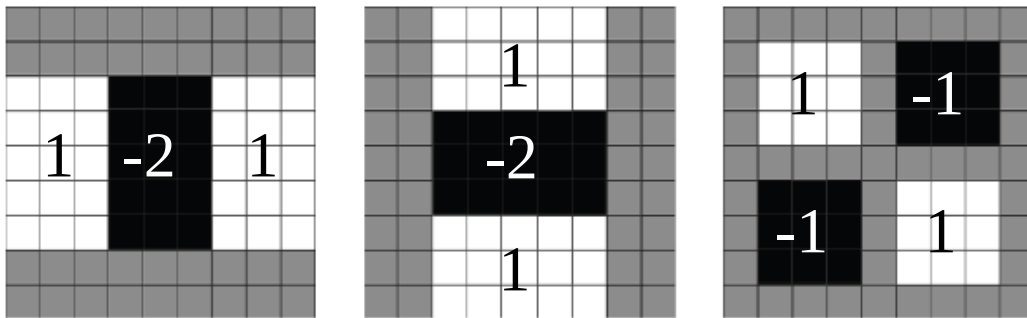
特徵點偵測是根據影像 Hessian 矩陣的概念，找尋 Hessian 矩陣行列式(determinant)的極值，極值的位置即是特徵點的座標。Hessian 矩陣的定義如公式(13)，

$$H(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix} \quad (13)$$

其中 $\mathbf{x}$ 代表影像上的某點 (x, y) ， σ 代表尺度， $L_{xx}(\mathbf{x}, \sigma)$ 為影像上的 $\mathbf{x}$ 與高斯二階導數(Gaussian second order derivative) $\frac{\partial^2}{\partial x^2} g(\sigma)$ 的摺積，同理定義 $L_{yy}(\mathbf{x}, \sigma)$ 與 $L_{xy}(\mathbf{x}, \sigma)$ 。SURF 使用盒子濾波器搭配積分圖像的方法，改善 Hessian 矩陣在計算上費時的缺點。而使用盒子濾波器是因為高斯二階導數的效果與使用的盒子濾波器近似，如圖(9)，盒子濾波器以 D_{xx} 、 D_{yy} 與 D_{xy} 近似原本 Hessian 矩陣中的 L_{xx} 、 L_{yy} 與 L_{xy} ，所以可以求出 Hessian 矩陣行列式的近似值，如公式(14)。

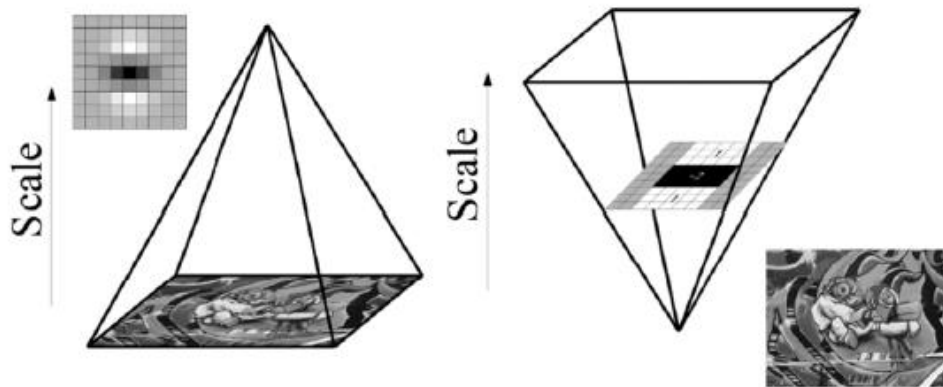
$$\det(H_{\text{approx}}) = D_{xx}D_{yy} - (wD_{xy})^2 \quad (14)$$

其中 w 為使方程式平衡的權重係數， w 設為 0.9。



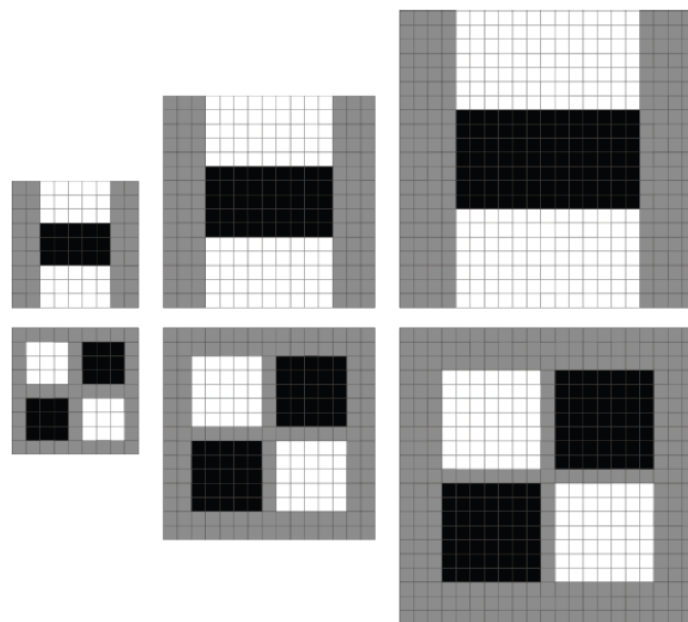
圖(9). 盒子濾波器，由左至右分別為 x 方向、y 方向與 xy 方向

SURF 利用盒子濾波器建立尺度空間，而 SURF 建立尺度空間的方式也與 SIFT 有所不同，SIFT 是將原始影像不斷的進行高斯模糊與降採樣，形成高斯金字塔，進一步得到高斯差分，而與 SURF 不同之處在於 SIFT 所使用的高斯濾波器大小不變，只改變影像的大小，但 SURF 則是影像大小不變，改變盒子濾波器的大小，如圖(10)。



圖(10). 尺度空間建立方式，左為 SIFT，右為 SURF

再來，為了保證盒子濾波器有中心點，必須使邊長永遠是奇數，9×9，15×15，21×21，27×27 等，如圖(11)。

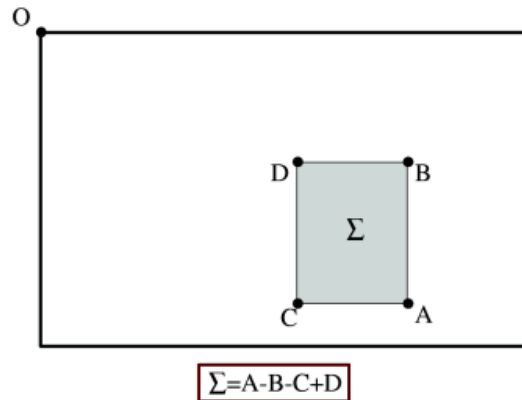


圖(11). 濾波器的大小隨著尺度的不同而改變

因為盒子濾波器每個區塊內的權重都一樣，所以使用積分圖像的方法來搭配盒子濾波，目的就是為了加快濾波的速度，因為積分圖像可以快速地計算出某一個區域內像素值的總和，積分圖像的定義如公式(15)，

$$I_{\Sigma}(\mathbf{x}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad (15)$$

其中 $\mathbf{x}$ 代表影像 I 的某點位置 (x, y) ，而當需要計算出影像上任一區域內像素值的總和，只需要三個加(減)法的運算就可以達成，如圖(12)



圖(12). 計算任一區域內像素值的總和，只需要三個加(減)法的運算

經過盒子濾波器偵測出的特徵點數量眾多，必須加以過濾，在 SURF 中稱之為特徵點的定位。首先，在鄰近 $3 \times 3 \times 3$ 的空間中進行非極大值抑制(non-maximum suppression)，目的是找出極大值，接著將 Hessian 矩陣行列式的極大值在影像尺度空間中做內插，以便更精準地定位特徵點。

(2) 特徵點描述

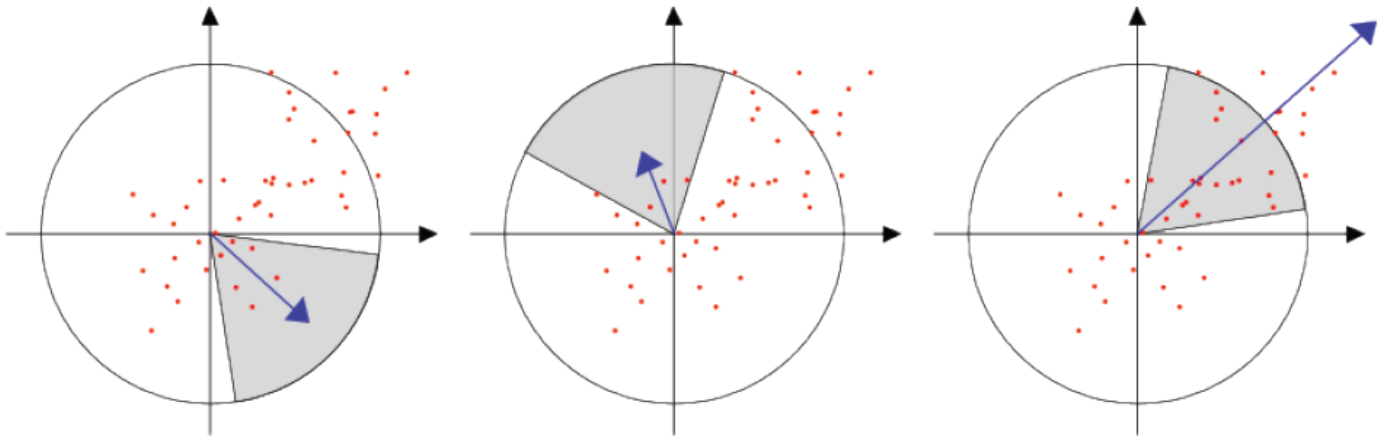
SURF 與 SIFT 一樣為了使特徵點具有旋轉不變性，必須找出特徵點的主方向，與 SIFT 不同的地方是，SIFT 是取一個以特徵點為中心區塊，直接計算出區塊內所有像素的梯度強度與方向，在統計成方向值方圖，進而找出主要的方向；而 SURF 是利用 Haar 小波濾波器來找出區塊內所有像素的梯度強度與方向，再使用 Haar 小波響應來找出特徵點的主方向。

Haar 小波濾波器可以使用積分圖像來加速計算，因為 Haar 小波是一個簡單的濾波器，如圖(13)，它可以用來找出 x 和 y 方向上的梯度強度與方向。



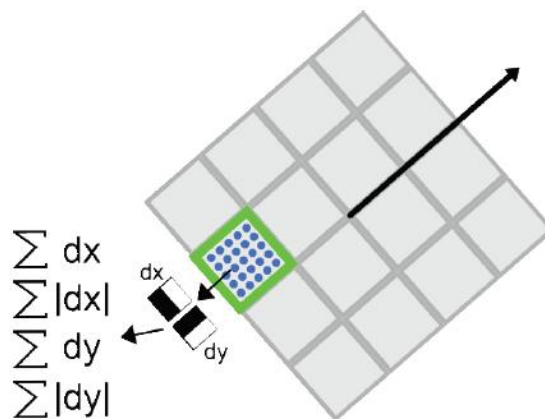
圖(13). Haar 小波濾波器，左為 x 方向，右為 y 方向

再來，以 6σ 為半徑的圓來計算 x 和 y 方向上的 Haar 小波響應，然後採用一個大小 $\frac{\pi}{3}$ 的扇形視窗，計算視窗內水平和垂直響應的總和，扇形視窗滑動整個圓，找出響應總和最大的即為特徵點的主方向，如圖(14)。



圖(14). 特徵點的方向分配

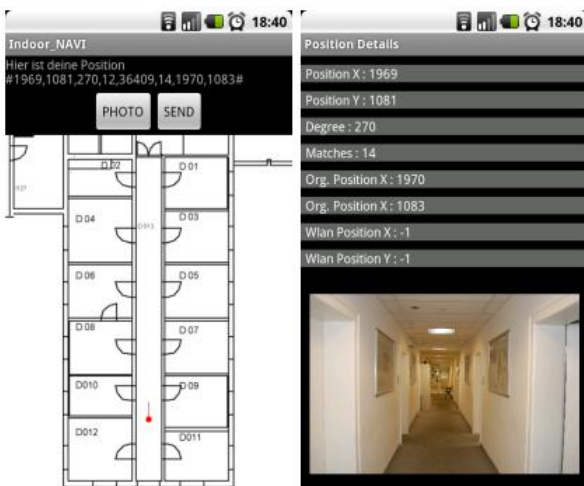
找出特徵點的主方向後，接著描述特徵點的特徵向量，首先取一個以特徵點為中心的區塊，區塊大小為 20σ ，而區塊的方向就是特徵點的主方向。然後將此區塊分成 16 個小區塊，每個小區塊內將會有 4 維的特徵向量，特徵向量表示為 $\mathbf{v} = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$ ，其中 $\sum d_x$ 與 $\sum d_y$ 為 Haar 小波響應的總和，因此，整個特徵點的特徵向量總共有 $4 \times 4 \times 4 = 64$ 個維度，如圖(15)，最後再將特徵向量進行正規化，使其不受亮度差異所影響。



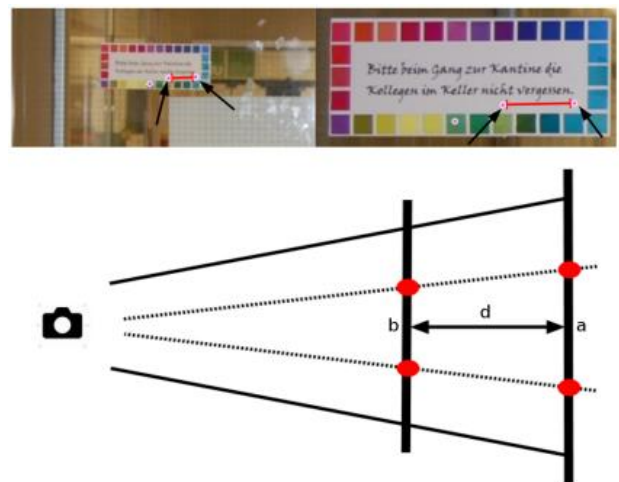
圖(15). 特徵點的描述，藍點為小波響應的取樣點

以上介紹了 SIFT 與 SURF，並比較彼此的差異，而作者使用 SURF 來擷取影像上的特徵，因為 SURF 的計算速度比 SIFT 快，再來解說作者提出的 MoVIPS 方法。

MoVIPS 提供了照片模式、平均值影片模式和投票影片模式。首先，在照片模式，於一個場景內，先取得多張影像用來建立場景資料庫，這些在資料庫內的影像都事先得知其位置資訊，在照片模式下，手機鏡頭所拍攝的高解析度影像，上傳至伺服器，並提取 SURF(Speed Up Robust Features)特徵描述符，且在先前事先建立好的資料庫中搜尋最相近的那張影像並將其結果返回顯示在手機上，如圖(16)，右圖為攝影機拍攝影像，可得知位置、角度、匹配點數等等資訊，左圖為此影像在平面圖中的位置。



圖(16). 手機螢幕畫面



圖(17). 距離估測演算法

而圖(17)為距離估測演算法，估算距離的比率，以用來衡量視點與圖像之間的距離，可用來提高定位的精確度。

平均值影片模式先拍攝連續的影像序列，上傳至伺服器，再提取特徵並與資料庫影像比對，並返回所有幀估測位置的平均值。

投票影片模式也必須先拍攝連續的影像序列，並上傳至伺服器，一樣提取特徵並比對提取位置，不同的是，這裡是從資料庫中提取返回最多匹配數圖像的位置，並將結果顯示在手機上。

最後，我們可以由表(1)所示，三種模式的累積分布函數位置的誤差與雷達相比，其中都以照片模式因其擁有高解析度圖像且無動態模糊，故其誤差最小，而平均值影片模式，因其動態模糊與低解



析度的關係，造成誤差較大。

System	25%[m]	50%[m]	75%[m]
Photo Mode (Stationary)	0.28	0.68	1.25
Video Mode (Averaging)	2.39	3.88	4.4
Video Mode (Voting)	1.0	2.85	4.4
RADAR	1.92	2.94	4.69

表(1). 各種模式下累積分布函數位置誤差.

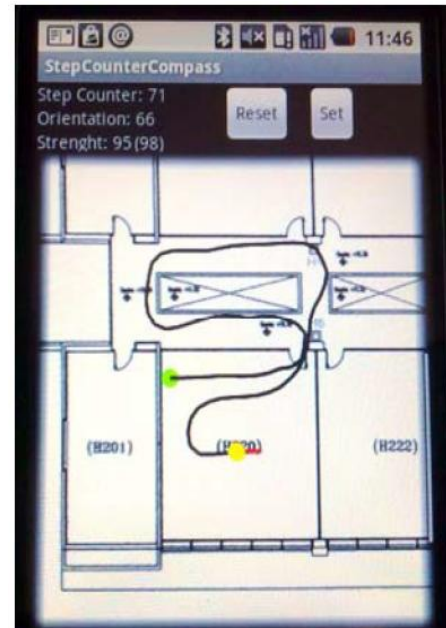
2.4 純慣性元件的定位

這裡主要是以 Serra[7] 等人所提到的方法，利用加速度計和陀螺儀來實現慣性導航系統 (Inertial Navigation Systems, INS)，加速度計量測物體的加速度，經過二次積分計算出物體的位移，陀螺儀量測物體的三軸角速度，經過一次積分計算出物體依三軸旋轉偏移的角度，再經由導航演算法，推算出移動物體在室內的位置。而現代的智慧型手機內都含有加速度計與陀螺儀，故在此將 INS 應用於智慧型手機上。

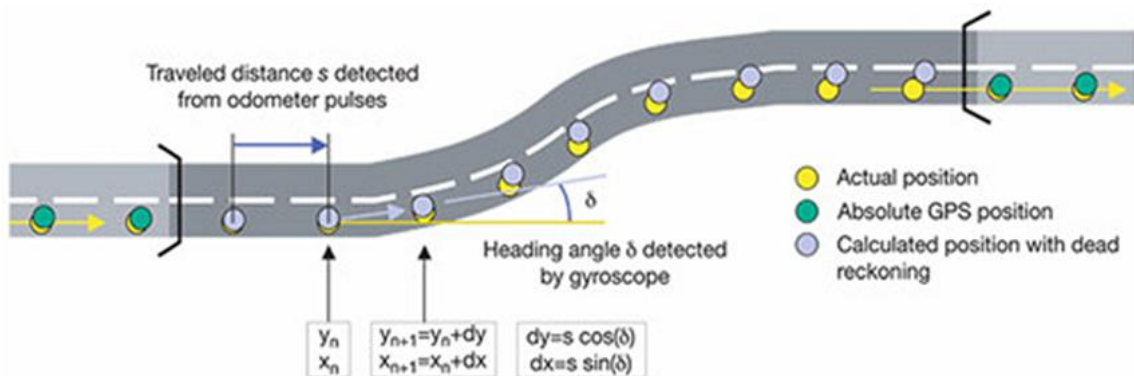
如圖(18)，作者先以智慧型手機掃描二維資料矩陣，此二維資料矩陣提供建築物的平面地圖資訊和二維資料矩陣的位置資訊，再來，如圖(19)，綠色的點是以二維資料矩陣位置資訊作為導航演算法的初始位置，當使用者移動時，依加速度計提供的位移資訊和陀螺儀提供的角度資訊，進而不斷推算下一步的位置，其中黑色的線為行進軌跡，黃色的點為終點位置。以圖(20)說明導航演算法，圖中， x_n 和 y_n 為初始位置，經由初始位置 x_n 和 y_n 和 dx 和 dy 計算出的下一時間點的位置 x_{n+1} 和 y_{n+1} ，而 dx 、 dy 可經由加速度計計算其中的位移 s 和陀螺儀計算出角度 δ ，以達到位置推算，不過加速度計與陀螺儀會隨著時間不斷地累積誤差，進而導致位置越來越偏移，所以必須不斷地進行校正，以確保INS的精確度和可靠性。



圖(18). 使用者讀取二維資料矩陣和初始位置



圖(19). 螢幕上顯示預測路徑



圖(20). 導航演算法概念

因拍攝二維資料矩陣時，只能知道二維資料矩陣的位置，無法得知使用者的確切位置，為了使初始位置更加準確，作者提出了使用plane-to-plane homography的方法，以下將一步一步推算說明，如公式(16)，物體在三維空間中表示為 $(X, Y, Z, 1)^T$ ，經由一攝影機投影矩陣 P 投影回二維影像平面表示為 $(x, y, 1)^T$ ，而投影矩陣 P 可由攝影機內部參數和外部參數計算出來，表示為公式(17)，而plane-to-plane homography的作法如公式(18)，令 $Z=0$ ，使其定義為三維中的XY平面，所以公式(17)少了對Z軸的旋轉，改寫為公式(19)。



$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = [P] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (16)$$

λ ：比例參數

(x, y) ：影像平面座標

P ：攝影機投影矩陣

(X, Y, Z) ：影像轉換到三維空間的座標

$$P = K[R|T] = K[r1 \ r2 \ r3 \ t] \quad (17)$$

P ：攝影機投影矩陣

K ：攝影機內部參數

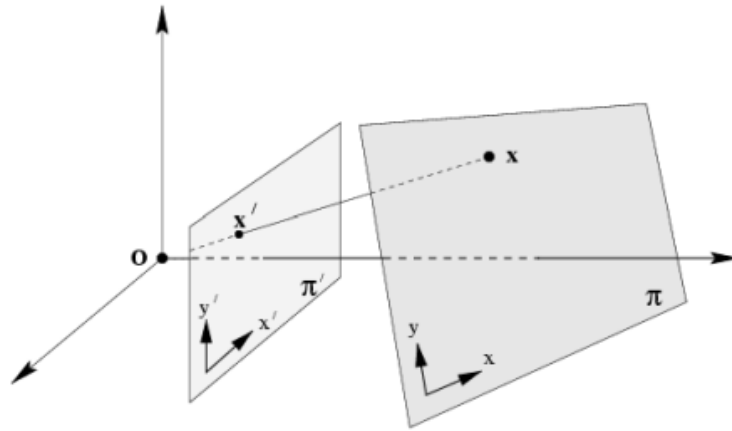
R ：攝影機旋轉矩陣

T ：攝影機平移矩陣

$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = [P] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (18)$$

$$H = [K][r1 \ r2 \ t] = \begin{bmatrix} h11 & h12 & h13 \\ h21 & h22 & h23 \\ h31 & h32 & h33 \end{bmatrix} \quad (19)$$

經過上面的推導後，可由圖(21)表示為二維平面對應到影像平面的關係， x 表示物體在三維空間中令 $z=0$ 二維平面 π 的點，而 x' 表示物體投影在影像平面 π' 的點，以數學式可表示為公式(20)， P_i 表示為三維空間中令 $Z=0$ 二維平面的點，經由一矩陣 H 投影在影像平面點 P_i' 。



圖(21) 二維平面對應到影像平面

$$P_i' = HP_i \quad (20)$$

P_i' ：由 P_i 投影在影像平面的點

H ：投影矩陣

P_i ：三維空間中令 $Z=0$ 二維平面的點

我們可由公式(19)、(20)推導出公式(21)、(22)，再來將二維資料矩陣的四個角點

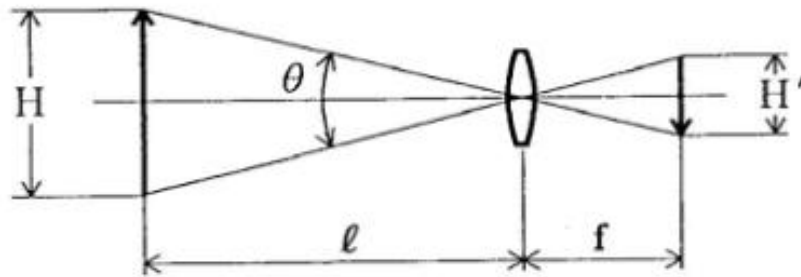
$P1'(x1,y1)$ ， $P2'(x2,y2)$ ， $P3'(x3,y3)$ 與 $P4'(x4,y4)$ 和投影影像上的四個座標點

$P1(x1,y1)$ ， $P2(x2,y2)$ ， $P3(x3,y3)$ 與 $P4(x4,y4)$ 帶入公式(21)和(22)解出矩陣 H ，經由 QR 分解，分解出 $[r1 \ r2 \ t]$ 和內部參數 $[K]$ ，再由 $[K]$ 可得出焦距 f 。

$$xi' = \frac{h11xi+h12yi+h13}{h31xi+h32yi+h33} \quad (21)$$

$$yi' = \frac{h21xi+h22yi+h23}{h31xi+h32yi+h33} \quad (22)$$

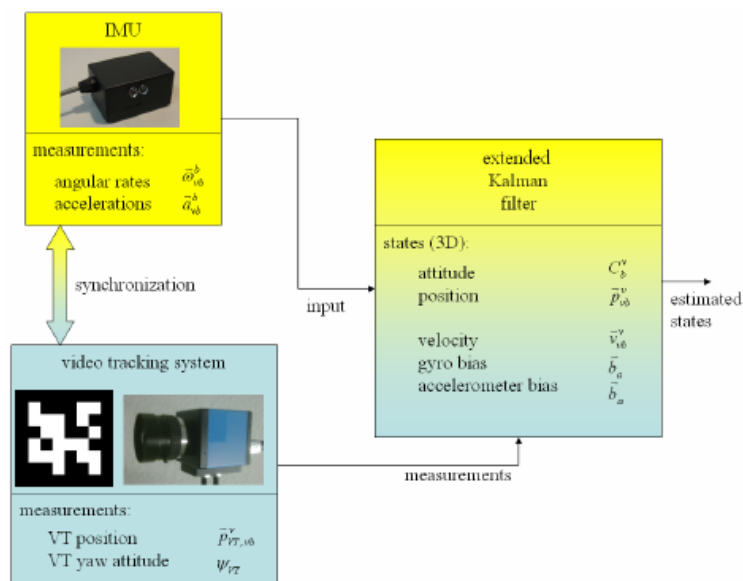
再來，如圖(22)， H 為三維空間中二維資料矩陣的大小， H' 為影像上二維資料矩陣的大小， f 為攝影機焦距， ℓ 為二維資料矩陣與使用者之間的距離，最後，我們可經由三角關係求得二維資料矩陣與使用者的距離以提高初始位置定位的精確度。



圖(22) 計算二維資料矩陣與使用者之間的距離

影像結合慣性元件

Hartmann[8]發表了一個對室內物體位置估測的系統，其系統是利用到一個微機電慣性量測元件 (MEMS IMU) 及一個基於標誌(marker-based)的視訊追蹤系統(Video Tracking System)，並且利用擴展型的卡爾曼濾波器將兩個互補的感測器整合起來，以達到追蹤定位的效果，如圖(23)，以下先介紹作者使用的慣性元件、視訊追蹤系以及測試的設置。



圖(23). 系統利用擴展型卡爾曼濾波器整合 IMU 及視訊追蹤系統

(1) IMU

Hartmann 是選用 Analog Devices[9]的 IMU (型號 ADIS16350)，並將其安裝於移動物體上，量測出移動物體的角速度及加速度。IMU 訊號的採樣頻率是 400Hz，陀螺儀和加速度儀分別有 $0.4^{\circ}/s$ 和 $0.05m/s^2$ 的標準偏差。

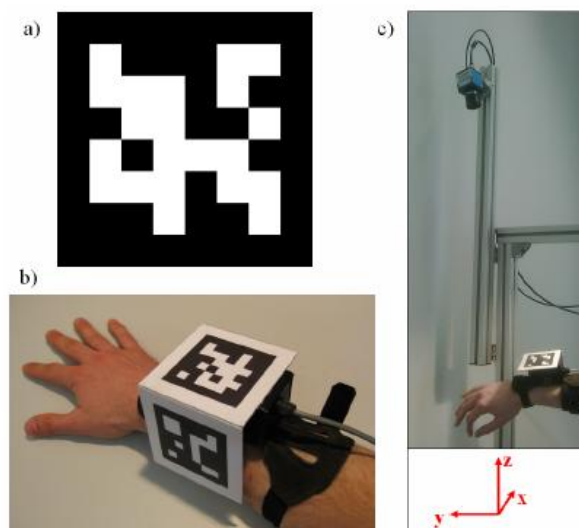
(2) 視訊追蹤系統

為了獲得移動物體的絕對位置，以補償 IMU 因為偏差和飄移而造成的累積誤差，Hartmann 使用基於標記的視訊追蹤系統。視訊追蹤系統是使用一個或多個外部攝影機，選用 The Imaging Source [10]的攝影機(型號 DBK21BF04)，來觀察物體的移動，利用[11]的視訊追蹤(Video Tracking, VT)演算法量測一個基準標記，如圖(23).a)在定義座標系上的位置和姿態。假設已知這個標記的尺寸，並且標記是非對稱的，如此一來，就可以利用一台攝影機來估測一個貼有標記的物體在 3D 空間的位置與姿態。

(3) 測試設置

在這裡是以追蹤手的移動來當作測試的情況，所以這種情況下必須將 IMU 及標記安裝在手上，如圖圖(23).b)，而攝影機的架設及攝影機的座標系如圖(23).c)。校正好的攝影機利用 VT 演算法將物體移動的位置及姿態記錄在 PC 上，典型視訊追蹤在這種測試設置下能達到平均每秒 5-10 的量測值，且位置的精度範圍是毫米，值得一提的是，由於攝影機大多是光軸平行於 Z 軸的情況下拍攝，因此垂直於攝影機光軸的姿態量測比較不可靠。

至於 IMU 與視訊追蹤系統是利用觸發來進行同步，而攝影機的曝光時間是 1/1000s，因此每 1ms 觸發一次。



圖(24). a)視訊追蹤的標記，b)IMU 和標記安裝在手上，c)測試環境使用一台攝影機

在 3D 空間中的物體追蹤，需要適當的採樣頻率才能得到可靠的估計，但由於偏差與飄移會產生累積誤差，所以微機電的 IMU 沒辦法長時間的估計位置，另一方面，視訊追蹤系統會被可視範圍所限制。因此，利用擴展型的卡爾曼濾波器整合這兩種感測器所得到的資訊，就可以增加精確度和採樣頻率，以下將介紹作者所使用的導航模型與擴展型卡爾曼濾波器的模型。

(1) 導航模型

在這裡導航模型可描述成兩個座標系，一個是 IMU 的座標系 b-frame，另一個則是視訊追蹤系統的座標系 v-frame，在這個導航模型裡，我們必須要將 IMU 量測到的加速度及角速度轉換成移動物體在 v-frame 的姿態、位置以及速度。

而 IMU 量測到的尤拉角(roll ϕ , pitch θ , yaw ψ)可以組成一個矩陣來表示 b-frame 與 v-frame 之間的旋轉關係，旋轉矩陣如下

$$C_b^v = \begin{pmatrix} c\psi c\theta & c\psi s\theta s\phi - s\psi c\phi & c\psi s\theta c\phi + s\psi s\phi \\ s\psi c\theta & s\psi s\theta s\phi + c\psi c\phi & s\psi s\theta c\phi - c\psi s\phi \\ -s\theta & c\theta s\phi & c\theta c\phi \end{pmatrix} \quad (23)$$

上式的 s 表示 *sine*， c 表示 *cosine*，而尤拉角與 b-frame 量測到的角速度 $\vec{\omega}_{vb}^b$ 之間的微分關係如下

$$\dot{\phi} = (\omega_{vb,y}^b \sin\phi + \omega_{vb,z}^b \cos\phi) \tan\theta + \omega_{vb,x}^b \quad (24)$$

$$\dot{\theta} = \omega_{vb,y}^b \cos\phi - \omega_{vb,z}^b \sin\phi \quad (25)$$

$$\dot{\psi} = (\omega_{vb,y}^b \sin\phi + \omega_{vb,z}^b \cos\phi) \sec\theta \quad (26)$$

我們正確的估計出姿態角之後，就可以利用式(23)將 b-frame 的加速度 $\vec{a}_{vb}^b$ 轉換到 v-frame 的加速度 $\vec{a}_{vb}^v$ ，再加上重力 $\vec{g}^v$ 的影響，所以動態速度可以表示如下

$$\vec{v}_{vb}^v = \vec{a}_{vb}^v + \vec{g}^v = C_b^v \vec{a}_{vb}^b + \vec{g}^v \quad (27)$$

最後位置的微分方程式如下

$$\dot{\vec{p}}_{vb}^v = \vec{v}_{vb}^v \quad (28)$$

(2) 擴展型卡爾曼濾波器

卡爾曼濾波器被廣泛的運用在導航上的問題，在此追蹤的應用中要利用 EKF 估計 15 個狀態，如下

根據式(27)，所有的狀態向量包括：

- 表示姿態的尤拉角 $\vec{\Psi}(\phi, \theta, \psi)$
- 位置 $\vec{p}_{vb}^v(x, y, z)$
- 速度 $\vec{v}_{vb}^v(v_x, v_y, v_z)$
- 角速度的偏差 $\vec{b}_\omega(b_{\omega x}, b_{\omega y}, b_{\omega z})$
- 加速度的偏差 $\vec{b}_a(b_{ax}, b_{ay}, b_{az})$

$$\vec{x}_k = \left(\vec{\Psi}_k^T \quad \vec{p}_{vb,k}^{vT} \quad \vec{v}_{vb,k}^{vT} \quad \vec{b}_{\omega,k}^T \quad \vec{b}_{a,k}^T \right)^T \quad (29)$$

從 IMU 量測到的加速度 $\vec{a}_{vb,IMU}^b$ 和角速度 $\vec{\omega}_{vb,IMU}^b$ 被當作濾波器的輸入，如下

$$\vec{u}_k = \left(\vec{\omega}_{vb,IMU,k}^{bT} \quad \vec{a}_{vb,IMU,k}^{bT} \right)^T \quad (30)$$

IMU 的量測值通常會有偏差 $\vec{b}$ 和雜訊 $\vec{w}$ ，在卡爾曼濾波器會其考慮在模型裡，因此，感測器的偏差及雜訊表示如下

$$\vec{\omega}_{vb,IMU}^b = \vec{\omega}_{vb}^b + \vec{b}_\omega + \vec{w}_\omega \quad (31)$$

$$\vec{a}_{vb,IMU}^b = \vec{a}_{vb}^b + \vec{b}_a + \vec{w}_a \quad (32)$$

在這裡擴展型卡爾曼濾波器是用離散的微分方程來表示此導航模型的狀態方程式及量測方程式，如下

$$\vec{x}_{k+1} = \vec{f}(\vec{x}_k, \vec{u}_k, \vec{w}_k) \quad (33)$$

$$\vec{z}_k = \vec{h}(\vec{x}_k, \vec{v}_k) \quad (34)$$

根據上一節的微分方程，可以推導出上述的非線性方程式

$$\vec{f}(\vec{x}_k, \vec{u}_k, \vec{w}_k) = \begin{pmatrix} \vec{\Psi} + \vec{f}_{EulerUpdate}(\vec{x}, \vec{u}, \vec{w}) \\ \vec{p}_{vb}^v + \vec{v}_{vb}^v \Delta T + \frac{1}{2} \Delta T^2 (C_b^v \vec{a}_{vb}^b + \vec{g}^v) \\ \vec{v}_{vb}^v + \Delta T (C_b^v \vec{a}_{vb}^b + \vec{g}^v) \\ \vec{b}_\omega \\ \vec{b}_a \end{pmatrix} \quad (35)$$

上式中的 $\vec{a}_{vb}^b$ 用來取代式(32)，包含加速度的量測值、偏差和雜訊，而尤拉角的更新方程式如下

$$\vec{f}_{EulerUpdate}(\vec{x}, \vec{u}, \vec{w}) = \begin{pmatrix} (\omega_{vb,y}^b s\phi + \omega_{vb,z}^b c\phi) \tan\theta + \omega_{vb,x}^b \\ \omega_{vb,y}^b c\phi - \omega_{vb,z}^b s\phi \\ (\omega_{vb,y}^b s\phi + \omega_{vb,z}^b c\phi) \sec\theta \end{pmatrix} \quad (36)$$

上式中的 $\vec{\omega}_{vb}^b$ 用來取代式(31)，包含角速度的量速值、偏差和雜訊。

視訊追蹤系統可以直接測量位置及姿態，所以量測方程式可以表示成

$$\vec{h}(\vec{x}, \vec{v}) = \begin{pmatrix} \psi \\ \vec{p}_{vb}^v \end{pmatrix} + \vec{v} \quad (37)$$

這裡的 $\vec{v}$ 式視訊追蹤系統量測到的雜訊，而在式(37)中只有偏航角 yaw 被考慮，因為視訊系統是從上面往下拍移動物體，所以量測到的旋轉角 roll 與俯仰角 pitch 是不可靠的。

根據式(35)和式(36)，可以計算出 EKF 所需的偏導數矩陣，如下

$$\Phi_k = \frac{\partial \vec{f}}{\partial \vec{x}}(\vec{x}_k, \vec{u}_k, 0) \quad (38)$$

$$W_k = \frac{\partial \vec{f}}{\partial \vec{w}}(\vec{x}_k, \vec{u}_k, 0) \quad (39)$$

$$H_k = \frac{\partial \vec{h}}{\partial \vec{x}}(\vec{x}_k, 0) \quad (40)$$

$$V_k = \frac{\partial \vec{h}}{\partial \vec{v}}(\vec{x}_k, 0) \quad (41)$$

這些矩陣將會用在擴展型卡爾曼濾波器的預測的步驟

$$\vec{x}_{k+1}^- = \vec{f}(\vec{x}_k^+, \vec{u}_k, 0) \quad (42)$$

$$P_{k+1}^- = \Phi_k P_k^+ \Phi_k^T + W_k Q_k W_k^T \quad (43)$$

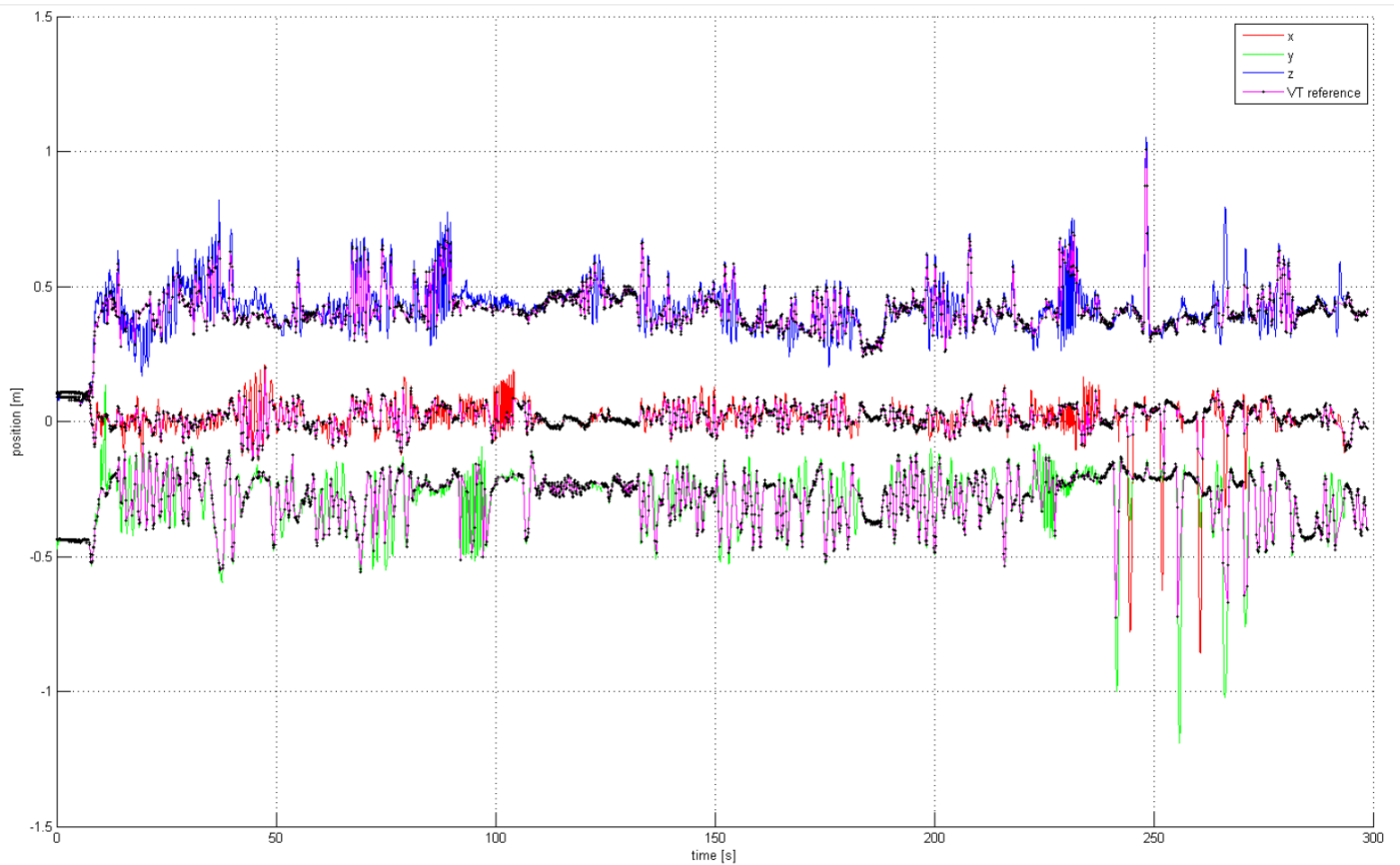
和校正的步驟中

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + V_k R_k V_k^T)^{-1} \quad (44)$$

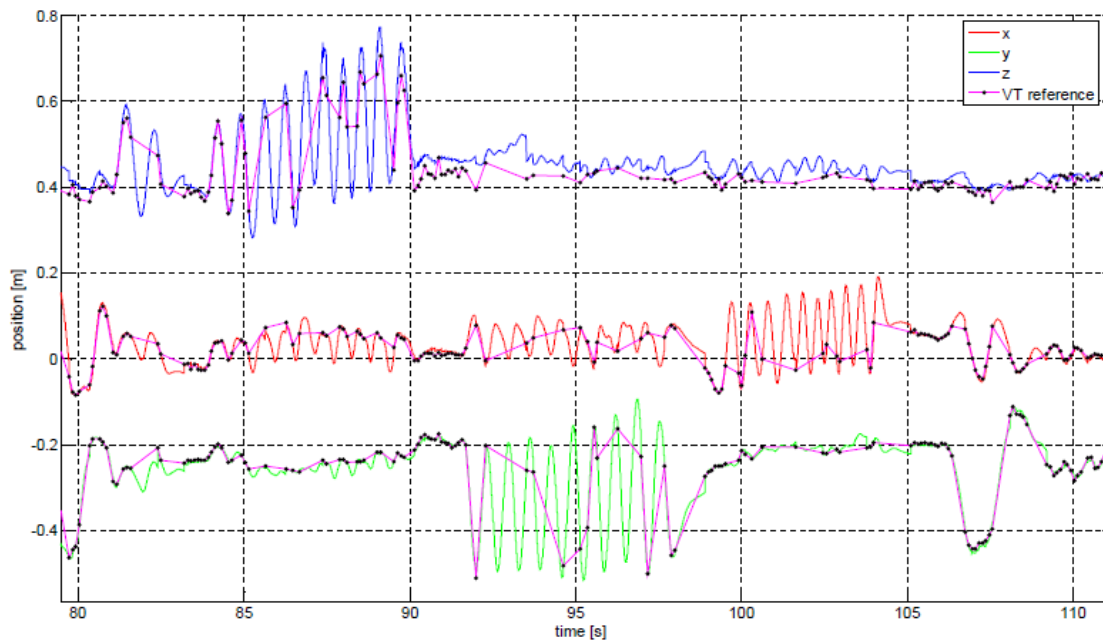
$$\vec{x}_k^+ = \vec{x}_k^- + K_k (\vec{z}_k - \vec{h}(\vec{x}_k, 0)) \quad (45)$$

$$P_k^+ = (I - K_k H_k) P_k^- \quad (46)$$

Hartmann 的實驗結果，是將 EKF 做出來的結果與視訊系統的結果作比較，如圖(25)，將圖(25)其中一段放大來看，如圖(26)，可以明顯的觀察到其中有幾段視訊系統的結果，會因為物體移動的速度過快而導致資訊遺失的現象。

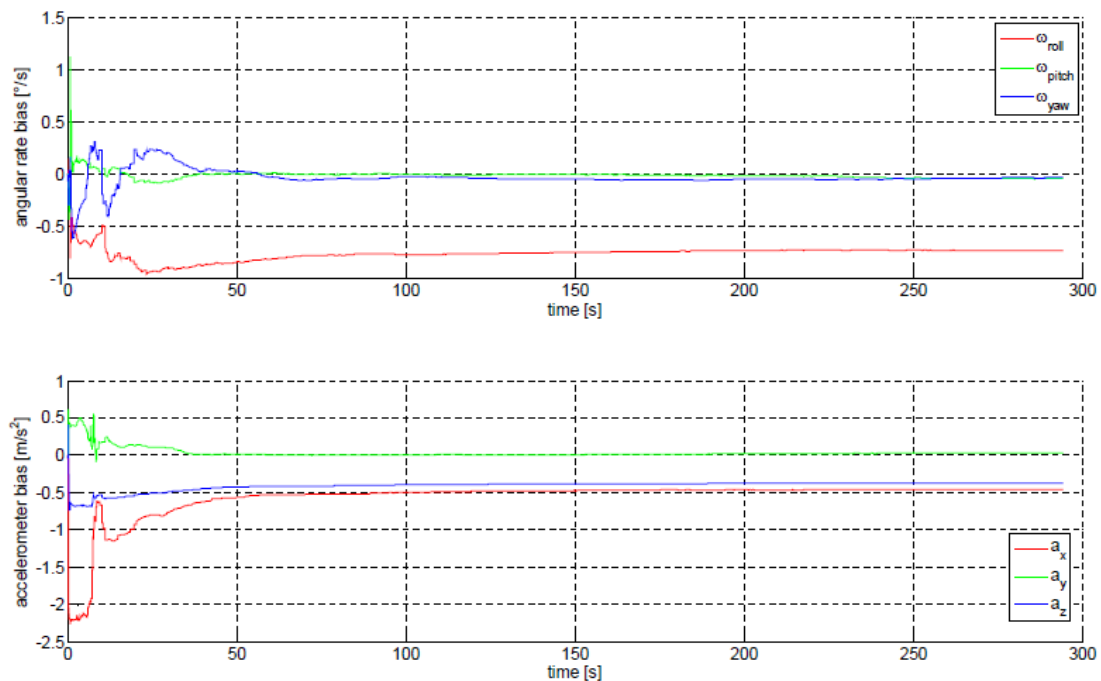


圖(25). Hartmann 的實驗結果



圖(26). EKF 位置估測結果與 VT 系統實際量測位置的結果

而 EKF 在角速度與加速度偏差的估測，如圖(27)，也可以明顯的觀察到偏差的估測經過大約 50 秒後會開始穩定。可以看得出 EKF 的整合的確可以改善 IMU 和視訊追蹤系統彼此的缺點。



圖(27). EKF 偏差的估測



資料名稱： 結合影像和慣性元件之定位技術與系統

版別：第 1.0 版

test sequence	Position - mean deviation to VT [m]		
	x	y	z
1	0.0063	0.0075	0.0167
2	0.0075	0.0082	0.0137
3	0.0069	0.0067	0.0163
4	0.0072	0.0064	0.0130
5	0.0096	0.0099	0.0221
6	0.0078	0.0076	0.0168
overall	0.0075	0.0077	0.0164

test sequence	Position - maximal deviation to VT [m]		
	x	y	z
1	0.0747	0.1054	0.1434
2	0.1123	0.0846	0.1066
3	0.0964	0.1167	0.0890
4	0.0884	0.1390	0.0811
5	0.1408	0.1778	0.1489
6	0.1540	0.1093	0.0998
overall	0.1540	0.1778	0.1489

test sequence	VT measurements		
	sampling rate ¹⁾	outage max ²⁾	#outages >1s ³⁾
1	6.01	2.42	26
2	5.86	2.23	18
3	6.47	1.62	9
4	6.35	2.05	7
5	4.42	3.19	49
6	5.83	1.94	20
overall	5.8229	2.2405	21.5
¹⁾ average sampling rate in test sequence [Hz]			
²⁾ longest outage in test sequence [s]			
³⁾ number of outages >1s in test sequence			

表(2). EKF 與視訊追蹤系統的比較結果

3. 結論

室內定位因建築環境使得衛星或雷達的定位訊號衰減，其無法由全球定位系統和雷達來實現，所以我們研讀了許多室內定位技術並且在以上做了大致的介紹，包括純影像定位、純慣性元件定位和影



像結合慣性元件定位等方式，而在這之中，又以影像結合慣性元件定位的精確度最為準確，因其將影像與慣性元件利用卡爾曼濾波器整合，如實驗結果所示，可提供較小的定位誤差。

4. 參考文獻

- [1] <http://zh.wikipedia.org/wiki/GPS>
- [2] http://en.wikipedia.org/wiki/Differential_GPS
- [3] <http://en.wikipedia.org/wiki/Radar>
- [4] Martin Werner and Moritz Kessel, “Indoor Position Using Smartphone Camera,” *In Position Location and Navigation (IPIN)*, 2011 IEEE.
- [5] D. Lowe., “Distinctive image features from scale-invariant keypoints,” *IJCV*, 60(2):91 – 110, 2004.
- [6] JiDuL, 2012, [Android] Utiliser et appliquer l’algorithme SIFT sur une image avec Android. Available:
<http://www.jidul.com/wp/2011/07/15/android-utiliser-et-appliquer-lalgorithme-sift-sur-android/>
- [7] H. Bay, T. Tuytelaars, and L. Van Gool., “Surf: Speeded up robust features.” *EuropeanConference on Computer Vision*, 1:404–417, 2006.
- [8] Alberto Serra, Tiziana Dessi, Davide Carbon, Vlad Popescu, Luigi Atzori “Inertial Navigation Systems for User-Centric Indoor Applications,” *NEM proceedings*, 2010.
- [9] B. Hartmann, N. Link, and G. F. Trommer, “Indoor 3d position estimation using low-cost inertial sensors and marker-based video- tracking,” *In Position Location and Navigation Symposium (PLANS)*, 2010. *IEEE/ION*, May 2010, pp. 319 – 326.
- [10] Analog Devices, Inc. All Rights Reserved, 1995–2012. Available:
<http://www.analog.com>.



資料名稱： 結合影像和慣性元件之定位技術與系統

版別：第 1.0 版

-
- [11] The Imaging Source Europe GmbH, 1991-2012. Available:
<http://www.theimagingsource.com>.
- [12] H. Kato and M. Billinghurst, "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," *IWAR '99: Proceedings of the 2nd IEEE and AC International Workshop on Augmented Reality*, p. 85, 1999.