



國立台灣科技大學
資訊工程研究所

碩士論文

3D 點雲與 SIFT 特徵點室內影像定位之研究

3D Point Cloud and SIFT Descriptor Indoor Localization
Research

研 究 生：陳致良
學 號：M9915057

指 導 教 授：項天瑞博士

中華民國一百零一年七月十日

3D 點雲與 SIFT 特徵點室內影像定位之研究

學生：陳致良

指導教授：項天瑞博士

國立台灣科技大學資訊工程研究所

摘要

本篇論文探討如何利用單張彩色影像來重建出三維人臉模型。我們的方法是利用彩色影像來取代灰階值去建立張量模型 (tensor model)，在人臉資料庫中，是用典型相關分析 (Canonical correlation analysis) 是來建立彩色影像與深度資訊的對應關係，一旦建立好屬於各自的對應關係後，在重建人臉的過程中，就只需要單張彩色影像即可利用典型相關分析的對應關係來推算出正確的深度資訊。實驗中，我們的方法可以在不同的光線環境跟人臉角度得到不錯的效果。

3D Point Cloud and SIFT Descriptor Indoor Localization Research

Student: ZZhi-Liang Chen

Advisor: Dr. Tien-Ruey Hsiang

Submitted to Department of Computer Science and Information
Engineering

College of Electrical Engineering and Computer Science
National Taiwan University of Science and Technology

ABSTRACT

This paper develops a tensor-based 3D face reconstruction approach from a single color image. Instead of the grayscale image, we also consider additional color factors in constructing the tensor model. Canonical correlation analysis is applied to establish the relationship between the color image and the depth information in the face database. During the face reconstruction, given a single color face image, the depth estimation is computed from the CCA-based mapping between the tensor models. Experimental results show our approach is better suited under different lighting conditions and poses.

誌謝

本論文能夠完成，首先要感謝的是指導教授項天瑞老師。老師嚴謹的治學態度，讓我不但在學術研究上學習到更謹慎的思考，也在日常生活上獲益良多。

感謝實驗室的同學們，實驗室的生活有苦有樂，有你們才讓我能撐得下這三年漫長的時間。感謝建群、松翰學長給我的指導。感謝益偉、崇峰、承志、誠儀、慶豪學長以及恩緯、盈樽、嘉駿、世寬學長給我的指導。感謝實驗室的學弟妹們，訓哲、立昂、致良、青緯、庭耀、宗博、雅筑，常常幫了我不少忙。更要感謝我的同梯們，貴彥、薇穎、冠佑，我們同甘共苦，一起奮鬥，特別是在弄計畫的過程中，冠佑幫助了我很多，有你們在我才能走到這一步。也感謝好朋友志強跟我一起修正文法與用詞。

最後我要把我最深的感謝留給我的家人。謝謝我的爸媽跟兩位姊姊，你們讓我沒有經濟壓力地讀完這這碩士學位，也常常給我很多鼓勵，今天我終於拿到這個學位，終於可以讓你們放下心上的一塊大石頭了。

內文目錄

論文指導教授推薦書	i
考試委員審定書	ii
摘要	iii
Abstract	iv
誌謝	v
內文目錄	vi
圖目錄	viii
表目錄	ix
1 相關研究介紹	1
1.1 室內定位的相關研究	1
1.2 SLAM 在建置 3D 環境的應用	2
1.3 特徵點的比對與匹配	3
1.3.1 SIFT 尺度不變特徵向量	4
1.3.2 隨機抽樣一致演算法 (RANSAC)	7
2 虛擬環境建置與定位方法描述	9
2.1 方法大綱	9
2.2 利用 3D 點雲重建當時定位環境	9
2.2.1 重建初步模擬環境	11
2.2.2 將點雲環境限制範圍符合真實環境重建	12
2.3 虛擬照相機設置及影像資料庫建置	13
2.3.1 均勻分布設置虛擬攝影機	14

2.3.2	虛擬照相機成像原理	15
2.3.3	利用計算深度來優化攝影機角度	17
2.3.4	儲存虛擬相機影像建立資料庫	19
2.4	虛擬影像定位	20
2.4.1	導入虛擬相機坐標位置作參考	20
2.4.2	尋找特徵點並找出最多的特徵點投票選出位置	20
2.4.3	利用虛擬相機位置來定位	22
3	定位實驗方法比較分析	25
3.1	實驗目的	25
3.2	特徵點固定環境下定位結果分析	25
3.2.1	特徵點固定環境下實驗方法	25
3.2.2	分析不同因素影響特徵點固定環境下實驗結果	26
3.3	一般室內環境定位實驗	29
3.3.1	室內環境定位方法說明	29
3.3.2	定位數據結果分析比較	30
	Bibliography	33

圖 目 錄

3.1 控制環境實驗參數設定	26
3.2 室內定位環境大小分布	30

表 目 錄

1.1	glFrustum 矩陣圖示說明	5
1.2	特徵點向量方為與強度示意直方圖	6
2.1	3D 點雲環境座內定位整體流程圖	10
2.2	初步建置好的點雲環境	12
2.3	將點雲給予界限範圍	13
2.4	調整點雲坐標軸角度方法	13
2.5	在點雲上設置虛擬照相機位置	14
2.6	glFrustum 矩陣圖示說明	15
2.7	根據物體距離鏡頭遠近來調整方位	18
2.8	內差法補強前後的差異圖	19
2.9	特徵點比較差異圖	21
2.10	相機與特徵點的夾角示意圖	22
2.11	定位點夾角與長度向量關係示意圖	23
3.1	依照同心圓的覆蓋算出每個半徑內的每張圖的特徵點平均數，左 控制環境 1 右。控制環境 2	27
3.2	特徵點數量趨勢圖	27
3.3	間距定位平均誤差趨勢圖	28
3.4	定位環境成果	29
3.5	室內定位環境與待定位照片分布位置	31
3.6	室內環境定位成果	32

第 1 章 相關研究介紹

室內定位研究大部分是利用訊號接收作三角定位，但是需要設備上的支援，如：紅外線、超聲波、Wifi 等設備，訊號的不穩定也會影響定位的誤差。之後有研究把影像特徵點當作訊號作定位研究，影像定位可以減少訊號不穩定導致的定位誤差，但是照片取樣分布不均及相機角度上的限制，使得室內定位可定位的範圍受到很大的限制。我們方法利用照片組建出 3D 點雲環境，在 3D 環境內拍攝虛擬相片作定位，彌補訊號不穩定的情形及增加可定位的範圍。在相關研究的章節中我們分成三個階段來說明使用的研究技術。首先基於現今室內定位的研究作探討，再來說明本文方法中 3D 環境技術的相關研究，最後說明如何在照片中找到特徵點及特徵點批倍的方法。

1.1 室內定位的相關研究

現在室內定位的技術多用於訊號接收定位，需要設備上的支援，下面介紹幾種常用的技術：

紅外線技術：紅外線室內定位技術定位的原理是，紅外線 IR 標識的紅外射線，通過安裝在室內的光學傳感器接收進行定位。雖然紅外線具有相對較高的室內定位的精準度，但是由於光線不能穿過障礙物，當標識放在口袋裡或者有牆壁及其他遮擋時就不能正常工作。因此，紅外線只適合短距離傳播，而且容易被房間內的燈光所干擾，在定位上有一定的限制。

超聲波技術：超聲波測距主要採用反射式測距法，通過三角定位等算法確定物體的位置，即發射超聲波並接收由被測物產生的回波，根據回波與發射波的時間差計算出待測距離。當同時有 3 個或 3 個以上不在同一直線上的應答器做出回應時，可以根據相關計算確定出被測物體所在的二維坐標系下的位置。超聲波定位整體定位精度較高，結構簡單，但超聲波需要大量的底層硬體設施投資，成本太高。

藍牙技術：藍牙技術通過測量訊號強度進行定位。這是一種短距離低功耗的無線傳輸技術，在室內安裝適當的藍牙裝置，配置成多用戶的連接模式，並

保證此藍牙設備始是這個區域內的主要設備，就可以獲得用戶的位置訊息。藍牙技術主要應用於小範圍定位，例如大廳或倉庫。藍牙室內定位技術最大的優點是設備體積小、易於作在 PDA、PC 以及手機中，因此很容易推廣普及。但不足在於藍牙器材和設備的價格比較昂貴，而且對於復雜的空間環境，藍牙系統的穩定性稍差，受信號干擾大。

Wi-Fi 技術：無線區域網路 (WLAN) 是一種訊息平台，可以在廣泛的應用領域內實現複雜的大範圍定位、監測和追蹤任務 [10]，而網路節點自身定位是大多數應用的基礎和前提。當前比較流行的 Wi-Fi 定位是 IEEE802.11 無線網路標準的定位解決方案。易於安裝，只需要很少的 Access Point(AP) 設備，也能採用相同的底層無線網絡架構，定位精準度高。但是，如果定位的測算僅僅依賴於哪個 AP 點最近，而不是依賴合成的訊號強度，那麼在樓層定位上很容易出錯。目前，它應用於小範圍的室內定位，成本較低。但無論是用於室內還是室外定位，Wi-Fi 收發器都只能覆蓋半徑 90 公尺以內的區域，而且很容易受到其他信號的干擾，從而影響其精度，定位器的能耗也較高。

ZigBee 技術：ZigBee 是一種新興的短距離、低速率無線網絡技術，可以用於室內定位。它有自己的無線電標準，在數千個微小的感測器之間互相傳送訊息以實現定位。這些感測器只需要很少的能量，以接力的方式通過無線電波將數據從一個感測器傳到另一個感測器，所以它們的傳送效率非常高。ZigBee 最顯著的特點是它的低功耗和低成本。

除了以上提及的定位技術，還有基於電腦視覺、圖像分析、磁場定位等。圖像分析定位就是我們這次所研究的目標，目的可以減低設備上的使用成本，也包含更多關於整體環境的資訊。

1.2 SLAM 在建置 3D 環境的應用

上個章節介紹一些室內定位的方法，接下來的章節將介紹 Simultaneous Localization and Mapping(SLAM) 的方法，以及相關的研究。SLAM 是利用拍攝路徑的連串影像，針對這些影像的相對關係來製作出一條可被看見的路徑。由 Harris 和 Pike 在 1987 年所提 [1] 的方法，利用連串的影像拼湊出一個影像路徑的地圖，當初的這項研究為最早 SLAM 的起源。這方法可以秀出路徑上特定的 3D

特徵，但是卻無法找出這些 3D 特徵點的相對關係。這方法因為特徵點之間沒有相對關係，導致找出的路徑與實際影像路徑有所差異，且所做出的路徑無法封閉，也就是說根據這條路徑沒辦法走回原點。在之後的方法，找出路徑上特徵點的相對關係以及做出一條封閉的路徑，就成了改進 SLAM 的研究目標。

在之後的研究，像是 [2]、[3] 與 [4]，都朝找出特徵點對應關係，這個方向做加強。這些研究的共同點都是以 Extended Kalman Filter(EKF) 來做路徑分析，EKF 方法主要是說，由現在的狀態去預測出未來會發生的情況，利用最大似然估算法 (Maximum Likelihood) 來找出最有可能發生的情況，運用在 SLAM 上面則是改善了特徵點與特徵點相互對應的情況，藉由特徵點分布找出最有可能的連通關係，形成一個封閉的路徑。由於科技的進步加快運算的速度，在 [5] 與 [6] 的研究把 EKF SLAM 技術用在一般室內環境之中，也經實驗證實 FKF 能有效改善 SLAM 預估出來的路徑所產生的誤差。這對 SLAM 研究上有新的發展方向，也有其他研究把機率模型應用在 SLAM 改善路徑誤差，像是 [7] 的研究。

在現今的研究中，有很多的研究利用 SLAM 來建置 3D 模型，SLAM 在應用上的好處在於增加影像之間匹配關係，以及可用在一般室內環境中。

1.3 特徵點的比對與匹配

建置 3D 模型中，我們的方法利用 Graph SLAM 方法來減少相片之間匹配的錯誤，在上個章節介紹了 SLAM 的相關研究後，在這個章節我們將說明如何利用照片找出特徵點，及特徵點匹配的相關研究。找出圖形特徵點及匹配特徵點，目的在不同的照片中拍攝到相同的景物，利用同樣的景物找出兩個照片相對應的特徵點，再利用特徵點求出這些相片的相對位置，決定如何將照片放置在 3D 環境中。基於這項目的，先利用 SIFT 找出特徵點，再用 RANSAC 找出特徵點匹配的關係，使得照片能夠在正確的位置重合。下面的章節將介紹 SIFT 與 RANSAC 的方法。

1.3.1 SIFT 尺度不變特徵向量

取出相片中特徵點的方法，一開始由 lowe [8] 所提出，稱為尺度不變特徵向量 (SIFT)。後來有人針對 SIFT 做演算法的加速 (SURF) [9] 以及增強找出特徵點數量的方法 (ASIFT) [10]。SURF 的方法雖然快速，但是簡化許多計算流程，影響找出圖片當中特徵點的數量，如果套用在我們的方法中，會導致之後特徵點無法做到有效的匹配。ASIFT 是將照片可能會旋轉、放大或縮小的情況考慮進去，將這些照片模擬不同變化，再根據不同變化的照片中找出匹配的特徵點。此舉會導致運算量的增加，考慮到龐大的運算量導致時間的增加，我們還是使用一般的 SIFT 方法找出特徵點。

尺度不變特徵向量 (SIFT)，主要分成四個階段：

1. 區域空間極值分布
2. 特徵點定位與篩選
3. 特徵點方向分配
4. 特徵點描述向量建立

第一階段 (1) 區域空間極值篩選，先利用不同尺度間的高斯金字塔選擇區域中的最大極值，其高斯分布式子如下：

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp(-(x^2 + y^2)/2\sigma^2) \quad (1.1)$$

不同尺度的高斯分布利用摺積 (Convolution) 將影像模糊化。 $I(x, y)$ 代表原始影像， $G(x, y, \sigma)$ 代表高斯函數：

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1.2)$$

再利用每組影像相鄰的高絲模糊影像進行高斯差分 (Difference-Of-Gaussian)，目的用於在集合內 4 組高斯差分影像中找出極值，式子如 (1.3) 所示：

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (1.3)$$

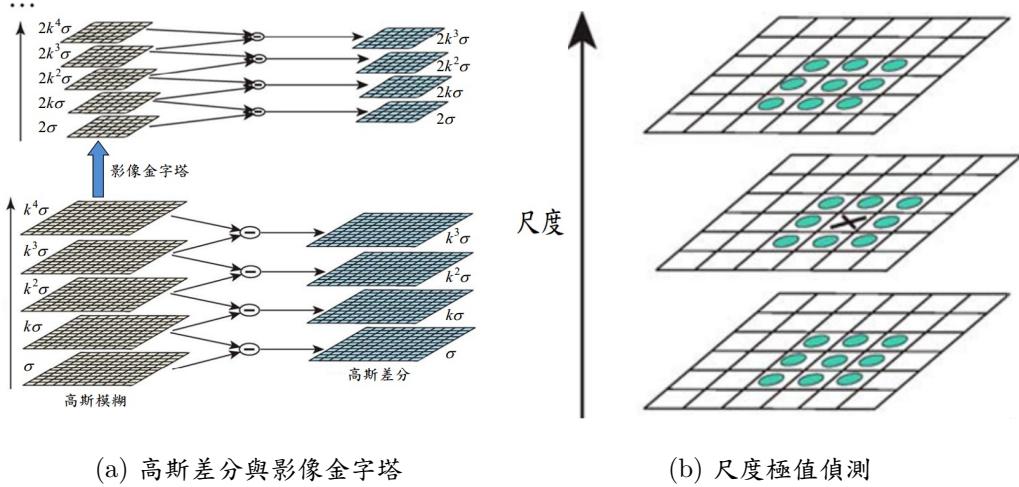


圖 1.1: glFrustrum 矩陣圖示說明

在此 k 為高斯模糊的尺度比值，設為 $\sqrt{2}$ ，若某個像素的極值為 26 個相鄰的像素中最大或最小的話，則此像素的位址即為區域極值的所在。

第二階段 (2) 特徵點定位篩選，其主要的目的在於找出真正有用的特徵點，在此特徵點的精準度必須要達到次像素的精度。有的特徵點其極值為低對比度的點，這時候這些低對比度的特徵點就會不予採用，剩下的特徵點即可為下一階段所使用。作法首先將 (1.3) 利用泰勒展開得到 (1.4):

$$D(x) = D + \frac{\delta D^T}{\delta X} X + \frac{1}{2} X^T \frac{\delta^2 D}{\delta X^2} X \quad (1.4)$$

式中 X 為極值 $(x, y, \delta)^T$ ， D 為高斯差分後的結果，再將 (1.4) 對 X 作偏微分可得 \vec{X} 算出 X 為極值點的偏移量。

$$\vec{X} = -\frac{\delta^2 D}{\delta X^2}^{-1} \frac{\delta D}{\delta X} \quad (1.5)$$

若是 $|\vec{X}| >= 0.5$ ，或是 $\sigma > k/2$ ，表示此區域極值點較靠近相鄰的點位，則需要再將此點移至相鄰的極值再經 (1.5) 計算後得到最佳的位置。若將 \vec{X} 帶入 (1.4) 中，可得 (1.6) 我們所用來篩選的式子:

$$D(\vec{X}) = D + \frac{1}{2} \frac{\sigma D^T}{\sigma X} \vec{X} \quad (1.6)$$

利用 (1.6) 將求出的絕對值與其他絕對值相比，可將對比度小的特徵點刪除以達到過濾的效果。

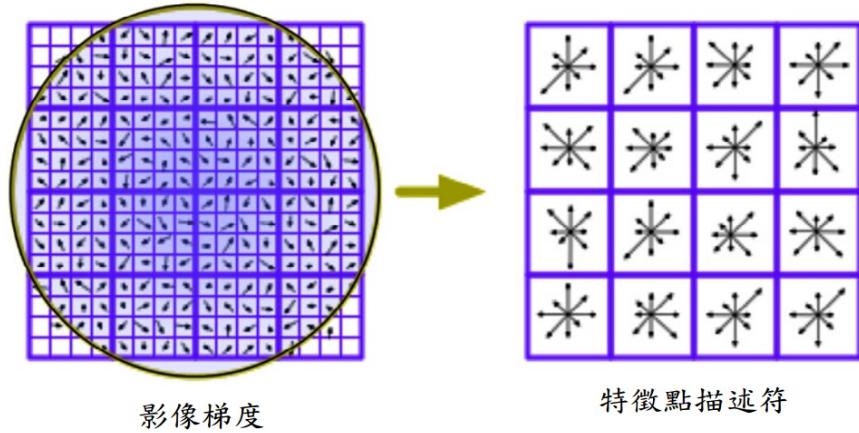


圖 1.2: 特徵點向量方為與強度示意直方圖

第三階段 (3) 特徵點方向分配，目的在於當對比的圖片有旋轉或者是尺度上的變化，相同的特徵點為了保有相同方向的特性，必須賦予每個特徵點一組特定的方向。其做法則是利用統計的方式，將所有的梯度值以角度每 10 個單位做方位直方圖記錄，並且記錄每個梯度的強度，以 (1.7)(1.8) 表示：

$$\theta(x, y) = \tan^{-1}\left(\frac{\delta L}{\delta y} / \frac{\delta L}{\delta x}\right) \quad (1.7)$$

$$m(x, y) = \sqrt{\left(\frac{\delta L}{\delta x}\right)^2 + \left(\frac{\delta L}{\delta y}\right)^2} \quad (1.8)$$

第四階段 (4) 徵點描述向量建立，最後一個階段為提供特徵點作為依據。做法上先將影像旋轉，使其特徵點向量與畫面主方巷一致之後，再以特徵點為中心，作出一個 16×16 的視窗，並加一個尺度為 0.5σ 的高斯函數為權重。把每個區塊區分成 4×4 的大小，切割成 16 個區塊，在每個區塊統計出梯度方位 $\theta(x, y)$ 以及強度值 $m(x, y)$ ，而後分別每個區塊有 8 個區間，代表 8 個不同的方向，如 1.2 所表示。

每個特徵點有 16 個方位的直方圖，每個方位直方圖內有 8 個梯度強度值，因此共有 $16 \times 8 = 128$ 個特徵值，這些特徵點則為我們所需要的描述向量。

1.3.2 隨機抽樣一致演算法 (RANSAC)

在獲得照片 2D 特徵點的資訊後，利用這些照片與 Kinect 的深度資訊做結合，建置 3D 環境。但是找出的特徵點當中，包含深度錯誤的資訊，為了將這些錯誤的資訊過濾掉，利用 RANSAC(RANdom Sample Consensus) 將不必要的深度特徵刪除，使得作出的轉移矩陣能將 3D 影像投影至正確的位置。RANSAC [11] 由 1981 年被提出，大多運用在電腦視覺上，用來預測數學模型有可能分布的狀況，透過迭代的方法使得預測的精準度增加。

RANSAC 的流程分為下面幾個步驟：

1. 設定隨機挑選的 inlier 數量 n ，以及迴圈的重複次數 k
2. 從所有的匹配中隨機挑選 n 個作為 inliers，數量至少能夠計算出轉移矩陣
3. 以 Levenberg-Marquardt 演算法從 inliers 計算出投影矩陣 R_{in}
4. 把 inliner 之外的匹配特徵帶入 R_{in} ，計算投影之後匹配特徵的距離，把距離小於設定門檻值得關配都加入 inliers
5. 紀錄所有 inliers，並比較其總數是否大於現在的最大值，如果是則更新最大值
6. 重複執行 k 次步驟 2. 到步驟 5.，最後留下的最大值就是過濾好的特徵點，再利用 Levenberg-Marquardt 虛這些過濾好的批被去計算出最佳轉移矩陣 R_{best}

至於 n 和 k 如何決定，可以假設每個特徵匹配為良好的機率設 w ，因此一開始隨機挑選的 n 個匹配的機率為 W^n ，其中涉某個錯誤的機率為 $1 - W^n$ ，而只要有一個是錯的，就找不到最佳的轉移矩陣，定義 p 為重複執行 k 次之後，至少選到一次 n 個都為良好匹配的機率即為 $1 - p$ ，可表示為：

$$1 - p = (1 - w^n)^k \quad (1.9)$$

$$p = 1 - (1 - w^n)^k \quad (1.10)$$

$$k = \frac{\log(1 - p)}{\log(1 - w^n)} \quad (1.11)$$

從 (1.10) 可看出，若希望 p 要盡可能大，則 k 要夠大，然而 k 越大表示所需計算的時間也越長，因此要先假設合理的 w 值，定義一個至少足以計算轉移矩陣的 n 值，來衡量 p 值與 k 值 (1.11)。

第 2 章 虛擬環境建置與定位方法描述

為了提升定位成功的覆蓋率與精準度，我們方法在於改進影像定位取樣的不足，與增加相片角度取樣的範圍。傳統影像定位所拍出的相片只能片段的取得環境特徵，導致部分環境範圍定位準確，但沒有被拍照的範圍會定位誤差過大。我們的作法利用點雲模擬當時環境，藉由取得環境利用虛擬相機拍攝相片，將虛擬相片集合製作成資料庫，與待定位照片比較求出定位點。此舉可增加相片取樣的不足與拍攝角度的變化性。

2.1 方法大綱

整套虛擬影像室內定位作法我們分成三大步驟來敘述：(1). 利用 3D 點雲重建當時定位環境 (2). 虛擬照相機設置及影像資料庫建置 (3). 虛擬影像定位。利用 Kinect 紅外深度攝影機，取得深度資訊後幫助我們做出 3D 的點雲環境。透過 3D 的點雲環境，再根據點雲中的坐標系來作格狀均勻分布 (Grid Permutation)，藉由有規律的分布來決定虛擬照相機的位置。因為每個照相機間距相同，代表觀察到的區域都有固定範圍，藉由角度上的調整就可以取得比一般 2D 影像包涵更廣的視角空間與更大的覆蓋範圍，取出更好的照片，以減低之後定位所造成的誤差。圖?? 為整體 3D 點雲環境座內定位的過程，這些流程會在之後章節逐一解釋步驟與這些步驟的目的。

2.2 利用 3D 點雲重建當時定位環境

要重建環境，先利用 3D 點雲建立初步的模擬資料，再經過調整逼近原來定位環境。藉由這些 3D 點雲，作出虛擬照片建置影像定位所需的資料庫。在這章節首先介紹怎麼重建當時的定位環境，3D 環境的重建分成下列五個步驟：

1. 取得 Kinect 照片
2. 將偵測到照片中的特徵點作隨機抽樣一致演算法 (RANSAC)

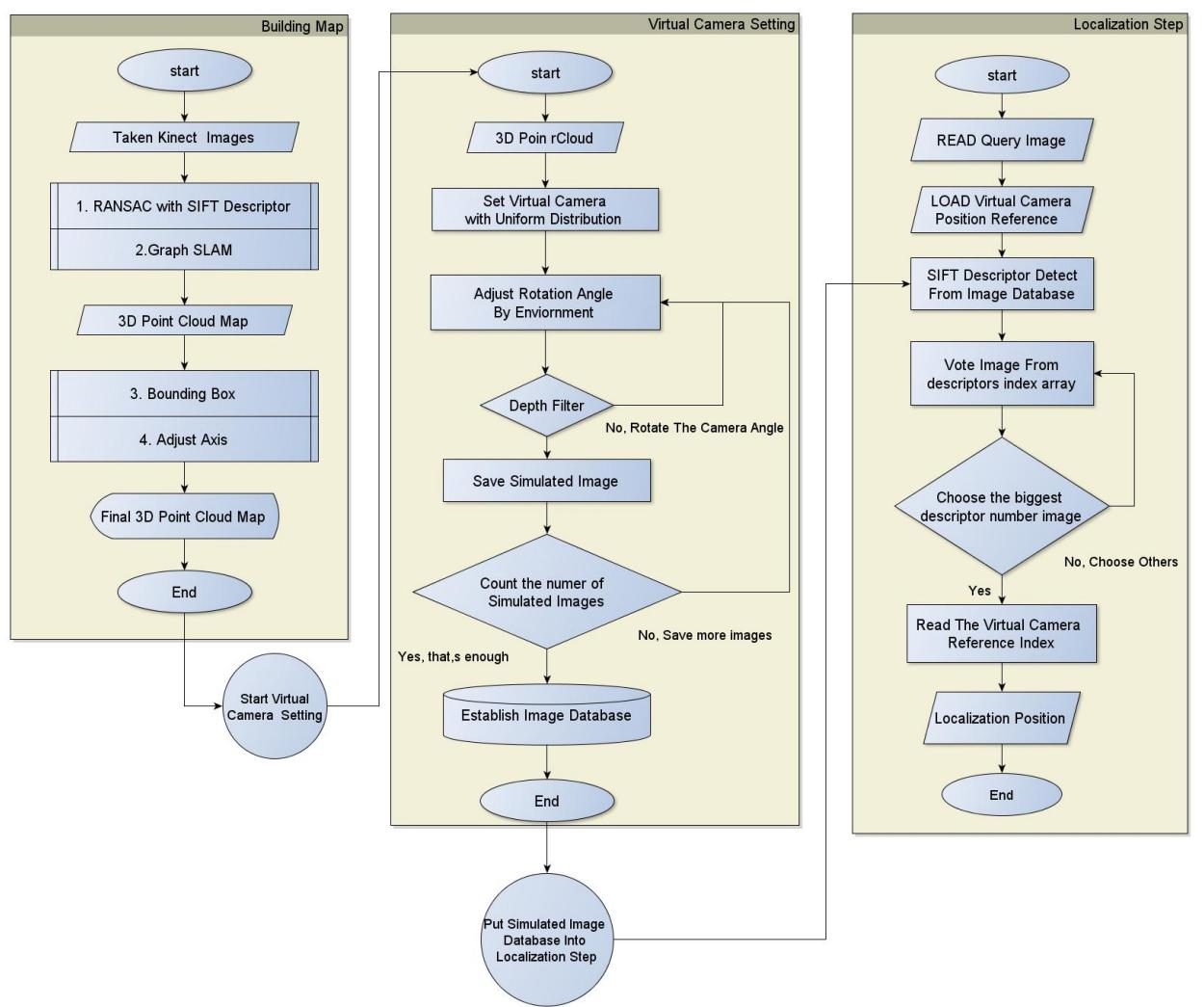


圖 2.1: 3D 點雲環境座內定位整體流程圖

3. 將組好的點雲作 Graph SLAM
4. 設置點雲涵蓋範圍的界線作界限範圍 (Bounding Box)
5. 調整坐標系

前面三個步驟主要是建置初步的環境，當環境完成模擬後，再透過剩下的兩個步驟對環境進行微調與限制。避免在四周看不到任何景物的地方，製造出沒有使用價值的虛擬相片。接下來將如何製作 3D 點雲環境分成兩大部分做說明。

2.2.1 重建初步模擬環境

一開始先用 Kinect 在環境中拍攝照片，照片取得的作法類似之前 [12] 的方法，在環境上對每個景物做一連串的拍攝，且每張拍攝出來的照片都需要有一些相似之處，環繞整個拍攝環境確保照片的組成具有連續性。當一個環境包含的景物越多代表所包含的特徵點越多，擁有豐富的特徵點數量在做 RANSAC 之後會有更好的重和效果。像之前在相關研究的章節所述，一個密集的點群所找到的平面會越接近真實點群的表現，所做出的轉移矩陣 (Transformation Matrix) 也會越精準，在進行重合時，不會有影像疊影或者是破碎導致點雲中空、物體歪斜扭曲的現象產生。連續的環繞拍攝是為了確保每一張影像相對位置關係沒有錯誤。RANSAC 最怕沒有順序的影像排列，順序不正確就無法找出影像的相對位置，也就是說根據轉移矩陣所做出的點雲位置可能會和實際景物在環境中的位置相差甚遠。因此拍攝照片在點雲建置的步驟中是影響最大的因素，其中可能光線的不足或是玻璃的反射等一些外在的因素都會導致之後在建置點雲的困難，所以在作拍攝時最好都避免這些不利的因素。

拍攝完照片組之後，我們要利用這些照片依據 3D 位置拼貼出點雲環境。如何重和這些照片拼貼出點雲環境，需要利用到尺度不變特徵向量 (Scale Invariant Feature Transform) 在從這些照片中取得特徵點的位置，將這些特徵點的位置求取 RANSAC，使得每一張影像都能夠在 3D 坐標系中與正確的位置中重合。作法將每張圖片找出來的特徵點作配對，求出來配對關係後，再將這些配對關係作最小平方法 (Least Square Error) 求出想要的平面，根據不同平面求出轉移矩陣 (Transformation Matrix)，最後可以得到之前照片影像位置的絕對關係，我們稱為 Global Pose，Global Pose 包含照片在三維座標以及照片在當時拍攝的角度。之



圖 2.2: 初步建置好的點雲環境

後 2D 影像的定位都需要以它作每張影像定位的起始原點，在其他地方像是點雲之後需要調整亦或是找出界限範圍，都需要利用 Global Pose 的起始原點當做參考，在我們的研究中，我們是使用虛擬相機的 Global Pose，至於虛擬相機的 Global Pose 的設置方式會在之後詳加說明。有了 Global Pose 的位置後，最後我們利用 Graph SLAM 將點雲作最後的調整。

Graph SLAM 描述，記得加完整版上去。

2.2.2 將點雲環境限制範圍符合真實環境重建

前三個步驟將點雲環境完成之後，在我們需要在環境中制定界線範圍 (Bounding Box)，用照相機的 Global Pose 調整角度。目的是為了讓虛擬照相機能夠被擺放在限定的範圍內，而不會有照相機放置在點雲環境之外，因而無法產生出有效的虛擬影像作定位。具體的做法如下，當我們讀入整個點雲之後，找在點雲中最大及最小的 X 及 Y 座標，知道了這四個座標，可以求出位於 Bounding Box 的頂點座標。最後座標相減求出的長度即為 Bounding Box 的長寬。有了這些長度之後，就可以知道整個點雲環境的長寬距離為多少，在之後可防止虛擬照相機坐落在散布點雲以外的位置。

制定點雲環境的界限範圍之後，做出來的點雲可能會因為之前 Kinect 攝影機的 Global Pose 歪斜分布而使得點雲也會有歪斜的狀況產生，這時候就必須將點雲作角度上的調整。這個步驟是為了之後的虛擬相機在照相時不會因點雲角度歪斜而使得照出來的角度與真實相機的角度差異過大，減少許多對應的特徵點，導

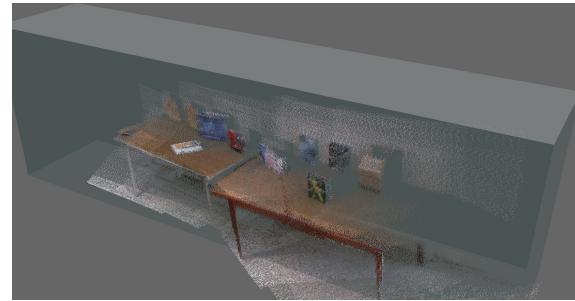
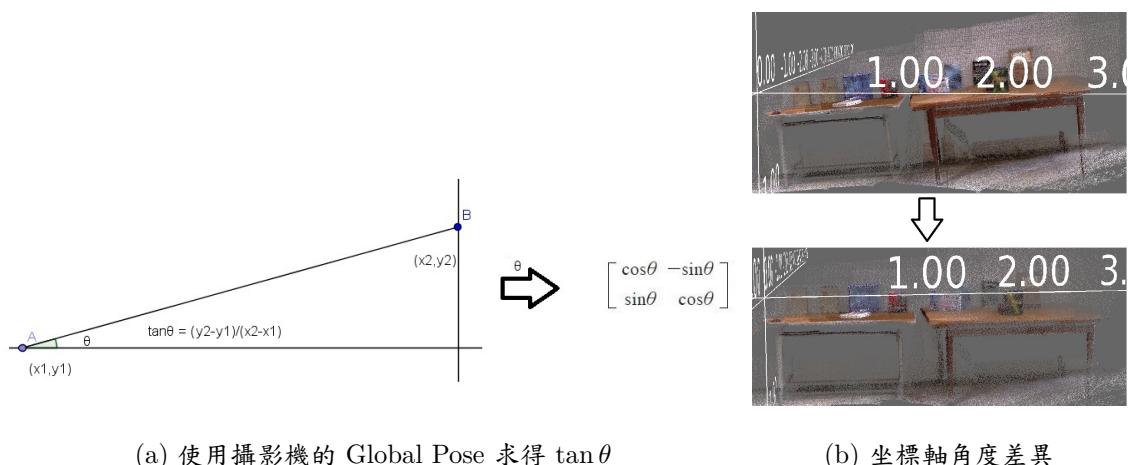


圖 2.3: 將點雲給予界限範圍



(a) 使用攝影機的 Global Pose 求得 $\tan \theta$

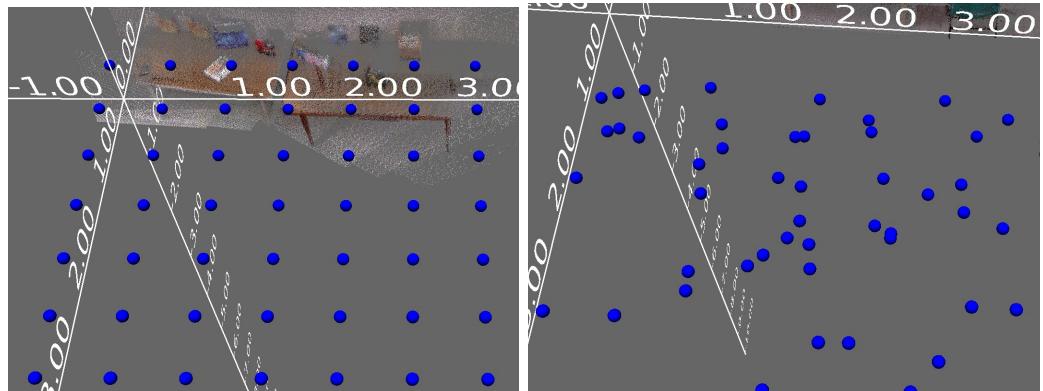
(b) 坐標軸角度差異

圖 2.4: 調整點雲坐標軸角度方法

致定位的誤差產生。我們的做法是看出點雲分布的情況是往哪個方向歪斜，利用 Global Pose 來算出兩個 Kinect 攝影機角度的 $\tan \theta$ ，求出 θ 之後，將點雲帶入求出來的 θ 旋轉矩陣旋轉至與座標軸平行的角度。調整完角度後，點雲的界線範圍與旋轉角度都與座標軸方向一致，就可以準備虛擬相機的準備工作。

2.3 虛擬照相機設置及影像資料庫建置

在建置完環境之後，接下來利用這個章節來描述如何決定虛擬照相機的位置、角度以及虛擬照相機成像的原理。與一般 2D 影像定位方法不同的是，傳統的 2D 影像定位內的影像資料庫，大部分都是利用隨機位置取得影像資料，而在我們的作法是先利用格狀分布設置虛擬照相機的位置，再利用隨機分布的角度來決定照相機拍攝的角度。依照這樣的作法，我們能取得比一般影像定位更多的環境資訊，



(a) 格狀分布虛擬相機位置

(b) 均勻分布虛擬相機位置

圖 2.5: 在點雲上設置虛擬照相機位置

而這些資訊都是利用點雲所產生的，不必再額外人工存取 kinect 攝影機的影像資料，我們所要輸入的資料只需要點雲就可以了。之後我們將分成四個部分描述虛擬照相機的設置：

1. 均勻分布設置虛擬攝影機
2. 虛擬照相機成像原理
3. 根據深度來調整攝影機角度
4. 儲存虛擬照相機圖片

2.3.1 均勻分布設置虛擬攝影機

上一個章節中，我們完成了實驗環境的建置，也就是點雲環境的資料。在這個章節中為了環境內每個景物都有充分拍攝而取得足夠的特徵點，將虛擬相機位置設置成格狀分布，在每個區塊上設置一個虛擬照相機。格狀分布的好處在於能夠減少相機集中在某處的情形發生，以圖來說，格狀分布會均勻分布在環境內，但隨機分布卻過於集中在圓圈處，再之後實驗會分別進行定位比較。格狀分布作法依據環境而有所改變，為了希望環境內建置 50 部以上的虛擬相機，我們會將環境的長度分成 8 個等分、寬度分成 7 個等分，這樣每個等分都會有一樣的距離間隔，完成 56 部虛擬相機擺設的位置。接下來隨機分布照相機角度，完成虛擬照相機布置的工作。隨機分布照相機角度比一般 2D 影像定位所建置的資料庫比較有

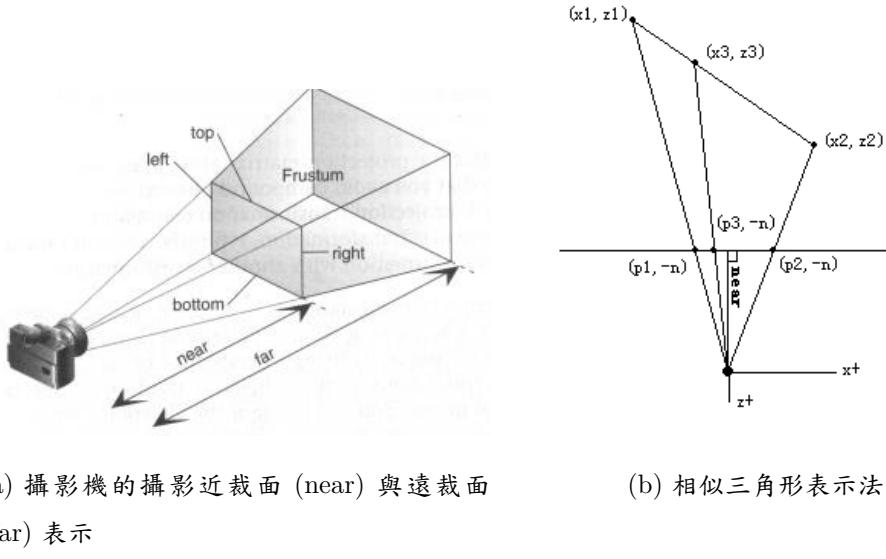


圖 2.6: glFrustum 矩陣圖示說明

更寬廣的角度。一般影像資料庫可能只針對特定區域的特徵點作取樣，而導致特定區域內的影像定位效果非常好，但在其他區域卻沒有足夠的影像特徵點資料，使得定位誤差範圍過大，透過均勻分布不會有部分景物或場景沒有被拍攝到，也可以增加定位的覆蓋率。

2.3.2 虛擬照相機成像原理

當虛擬相機位置固定之後，接下來利用虛擬相機拍攝照片，透過攝影機的影像角錐來模擬相機的成像，取出角錐內範圍的 3D 點雲，模擬照相機所照出的照片。透過 OpenGL 的坐標系，先將視野調整到虛擬照相機的位置，再利用 OpenGL 中 glFrustum 這個矩陣取得相機影像角錐，這個矩陣目的在於模擬相機光線經過透鏡成像，矩陣表示法如下：

$$glFrustum = \begin{pmatrix} \frac{2near}{right-left} & 0 & \frac{right+left}{right-left} & 0 \\ 0 & \frac{2near}{top-bottom} & \frac{top+bottom}{top-bottom} & 0 \\ 0 & 0 & -\frac{far+near}{far-near} & -\frac{2far\times near}{far-near} \\ 0 & 0 & -1 & 0 \end{pmatrix} \quad (2.1)$$

將座標轉為齊次座標後，利用攝影機的攝影近裁面 (near) 與遠裁面 (far) 的相似三角形來推出這個矩陣，這個矩陣會將角錐內的景像投影到深度在 $[-1, 1]$ 之間。這部分攝影機的焦距設定，以及解析度都參照 Kinect 紅外深度攝影機的參數設定。要解釋如何求出 glfrustrum 矩陣，需要用到兩項條件來說明：(1) 證明 $\frac{1}{z}$ 為線性關係，(2) 將 (1) 所求出的公式帶入投影座標求出矩陣關係式。

(1) 證明 $\frac{1}{z}$ 為線性關係：

根據圖2.6(b)所示，由相似關係三角形得出的關係式：

$$p = \frac{-n}{z} \times x \quad (2.2)$$

而我們知道直線關係視為 $y = ax + b$ ，將式 (1-2) 帶入直線關係式得出

$$p = \frac{n}{a} \left(\frac{z}{b} - 1 \right) \quad (2.3)$$

利用線性關係， $p_3 = tp_2 + (1-t)p_1$ ，帶入其中得出：

$$\frac{n}{a} \left(\frac{b}{z_3} - 1 \right) = t \frac{n}{a} \left(\frac{b}{z_1} - 1 \right) + (1-t) \frac{n}{a} \left(\frac{b}{z_2} - 1 \right) \quad (2.4)$$

化簡後得出：

$$\frac{1}{z_3} = t \frac{1}{z_1} + (1-t) \frac{1}{z_2} \quad (2.5)$$

(2) 將 (1) 所求出的公式帶入投影座標求出矩陣關係式：

在最後一個步驟中，要把之前求出的關係式都帶到投影座標內，我們假設 (x, y, z, w) 為攝影機座標， (x', y', z', w') 為投影座標，而 (P_x, P_y, P_z, P_w) 為攝影角錐內的座標，以圖2.6(a)為例，t=top, l=left, r=right, b=bottom，對於 x,y 其中關係式可以寫為：

$$x' = \frac{-nx}{z} \quad y' = \frac{-ny}{z} \quad (2.6)$$

將其縮放到可視範圍 $[-1, 1]$ 之間，得出：

$$\frac{1 - P_x}{1 - (-1)} = \frac{r - x'}{r - l} \quad \frac{1 - P_y}{1 - (-1)} = \frac{t - y'}{t - b} \quad (2.7)$$

化簡後得出：

$$P_x = \frac{2x'}{r-l} - \frac{r+l}{r-l} \quad P_y = \frac{2y'}{t-b} - \frac{t+b}{t-b} \quad (2.8)$$

帶入 x', y' ：

$$P_x = \frac{2n}{r-l} \left(-\frac{x}{z} \right) - \frac{r+l}{r-l} \quad P_y = \frac{2n}{t-b} \left(-\frac{y}{z} \right) - \frac{t+b}{t-b} \quad (2.9)$$

已知 P_z 與 $\frac{1}{z}$ 呈現性關係，設 $P_z = \frac{a}{z} + b$ ，求 a 與 b 。已知兩點 $(-n, -1), (-f, 1)$ ，所以：

$$a = \frac{2nf}{f-n} \quad b = \frac{f+n}{f-n} \quad (2.10)$$

$$P_z = \frac{2nf}{f-n} \left(\frac{1}{z} \right) + \frac{f+n}{f-n} \quad (2.11)$$

把 P_x, P_y 與 P_z 轉成齊次坐標系得出：

$$\begin{cases} -zP_x = \frac{2n}{r-l}x + \frac{r-l}{r+l}z \\ -zP_x = \frac{2n}{t-b}x + \frac{t+b}{t-b}z \\ -zP_z = -\frac{2nf}{f-n} - \frac{f+n}{f-n}z \\ w = -z \end{cases} \quad (2.12)$$

上述為 glFrustum 矩陣式子所推導的過程，我們將攝影角錐內的點雲投影成平面，存取虛擬影像。當存取完虛擬影像後，我們會判斷相機位置是否會太逼近虛擬環境內的景物，太靠近點雲邊界。這時候我們利用影像內的平均深度，判斷相機角度的選擇是否適當，這部分會在之後的章節作說明。

2.3.3 利用計算深度來優化攝影機角度

當虛擬照相機的圖片擷取出來後，因為拍照的相機深度過淺，而導致拍攝的景物無法辨識，這時候我們利用深度過濾的機制來將照相機取得角度作過濾。一般深度 buffer 分為 z-buffer 與 w-buffer 兩種，先從兩種不同的深度分辨方式作探討：

首先作關於深度的計算，利用四維座標軸 (x, y, z, w) 表示三維座標軸 (x', y', z') 的點，以圖2.6(a)為例， $t=top$, $l=left$, $r=right$, $b=bottom$ ，空間關係



圖 2.7: 根據物體距離鏡頭遠近來調整方位

的表示法為：

$$\begin{cases} x' = x/w \\ y' = y/w \\ z' = z/w \end{cases} \quad (2.13)$$

根據圖2.6(a)的示意圖表示， $Z_n = near$ 面的 z 範圍， $Z_f = far$ 面 z 範圍， $w = \frac{2 \times Z_n}{right-left}$ ， $Q = \frac{Z_f}{Z_f - Z_n}$ 所以由 z 座標求得 w 縮放的比例，式子可以寫為：

$$w = \frac{Q \times Z_n}{(Q - Z)} \quad (2.14)$$

z-buffer 是保存經過 glFrustum 投影變換後的 z 坐標，投影後物體會產生近大遠小的效果，所以距離眼睛比較近的地方，z 坐標的分辨率比較大，而遠處的分辨率則比較小。換句話說，投影後的 z 坐標在其值得分布上，對於景物對眼睛的物理距離變化來說，不是線性變化的（即非均勻分佈），這樣的一個好處是近處的物體得到了較高的深度辨識，但是遠處物體的深度判斷可能會出錯。

w-buffer 保存的是經過投影變換後的齊次坐標系中的 w 坐標，而 w 坐標通常跟世界坐標系中的 z 坐標成正比，所以變換到投影空間中之後，其值依然是線性分佈的，這樣無論遠處還是近處的物體，都有相同的深度分辨率，這是它的優點，當然，缺點就是不能用較高的深度分辨率來表現近處的物體。

針對兩種不同的深度 Buffer 比較，因為我們的做法是來判別景物是否距離鏡頭過近，所以在深度判斷上是採用 z-Buffer 的作法，當我們判斷鏡頭與物體距離

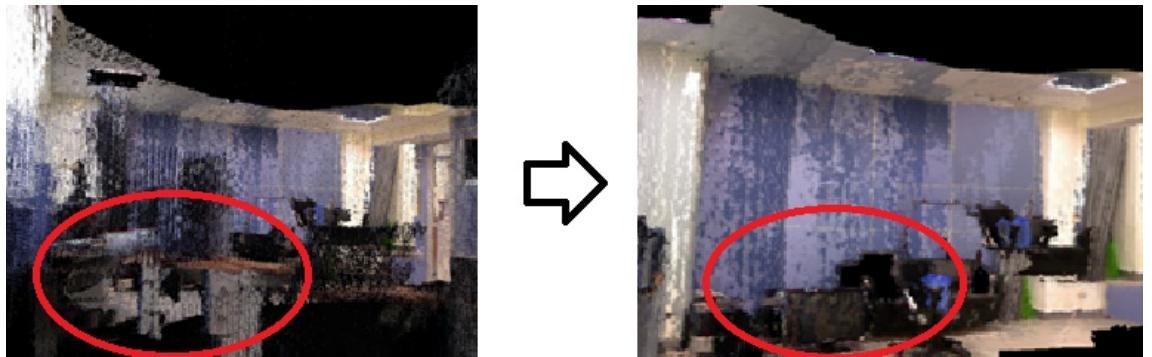


圖 2.8: 內差法補強前後的差異圖

實際深度小於 80 公分時，我們會將照相機鏡頭角度轉向 180 度，也就是正後方來重新拍攝。

2.3.4 儲存虛擬相機影像建立資料庫

當已經決定取好的照片之後，利用虛擬照相機將取出的照片來儲存至影像資料庫，來進行接下來定位的前置作業。虛擬影像儲存是透過虛擬照相機鏡頭裡的每一個像素寫入相片裡頭，主要做法如下。當從 z-buffer 讀出來的深度錯誤時，代表這個像素對應在點雲上是一個黑點或者是說根本沒有點雲的資訊，則以黑色為代表，當深度沒影錯誤時，則代表它具有實際點雲的資料，我們找出點雲對應點的顏色資訊，寫入圖檔裡，這樣即可完成初步的虛擬相片。根據上述的方法，還會遇到透視的問題，就是說原本不應該出現的景物因為深度有誤差，而原本在障礙物之後的物體卻跑在障礙物之前，像是穿透障礙物一樣，例如圖2.8原本不該出現桌子的地方，因為發生了透視的現象而出現了桌子。改進方法為根據周圍的深度來做內插補強。

虛擬影像的資料量因環境而變，主要根據 Global Pose 在每個位置上取出相隔 120 度的兩個不同角度的相片，在一般情況下環境中取出 50 點的 Global Pose，所以總共會有 100 張的虛擬相片。藉由這些虛擬相片，我們取得了環境所在內的不同位置與不同角度的資料，比起一般的影像定位資料多出了更豐富的特徵點資訊。之後的實驗可以比較出來，在不同位置以及距離特徵點的遠近對定位會帶來什麼樣的影響。到了最後定位的流程，將介紹虛擬影像的定位方法。

2.4 虛擬影像定位

在定位的流程中，利用讀取待定位的圖片，根據定位照片的特徵點找出最合適的虛擬影像，參照虛擬影像所在的相機的位置來定位。可以節省利用三角定位 (Triangulation) 的時間，這種根據之前不同的影像資料庫的建置方法，可以增加許多以前傳統影像所定位不到的地方，增加定位的覆蓋率，關於這種覆蓋率的數據比較，在之後的實驗分析會有詳細的數據可以佐證定位覆蓋率的改善。

在定位的程序上，主要會分成 3 個階段：

1. 導入虛擬相機坐標位置作參考
2. 尋找特徵點並找出最多的特徵點投票選出位置
3. 利用虛擬照相機位置來定位

2.4.1 導入虛擬相機坐標位置作參考

在定位之前，先輸入待定位的照片以及虛擬照相機的位置。虛擬照相機的位置記錄檔格式包含每個虛擬相機的 X, Y, Z 座標以及每個攝影機的角度位置，當流程步驟做到特徵點定位時，就會需要參考到虛擬相機的位置。

2.4.2 尋找特徵點並找出最多的特徵點投票選出位置

在前置作業完成之後，接下來利用所有資料庫中的照片進行比對，並將每張照片所擁有找到與被定位照片相同的特徵點數量記錄下來。在這裡使用的的方法為尺度不變特徵向量 (Scale Invariant Feature Transform)，簡稱為 SIFT，我們利用 SIFT 找出與相片中相同的特徵點，並將找出的特徵點的數量給記錄下來。關於 SIFT 的作法在之前已經有了相關的敘述，所以可以得知當存在虛擬相片中特徵點的數量越多，代表與所要定位的照片有越密切的關係，當我們在所有照片中選出來擁有最多特徵點數量的虛擬照片時，我們參考這個拍攝虛擬照片的相機的編號，根據編號找出相機所在的位置，再利用這個虛擬相機的 pose 當作初步所在定位的定位位置。



(a) 根據虛擬相片找出的特徵點



(b) 根據實際相片找出的特徵點

圖 2.9: 特徵點比較差異圖

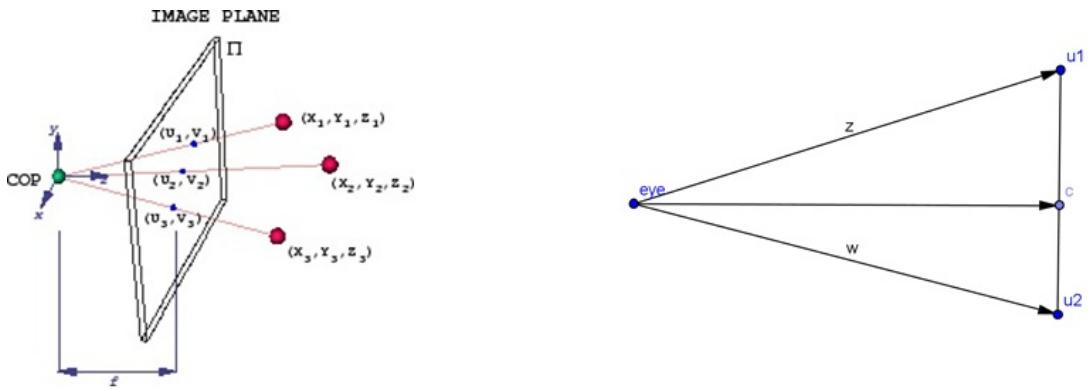


圖 2.10: 相機與特徵點的夾角示意圖

特徵點的分布跟環境景物的分布有密切的關係，在我們的作法上藉由虛擬照片找到距離特徵景物遠的待定位照片，卻可以比一般的照片找出更多的特徵點。利用虛擬照片我們可以有效的找出更多的環境特徵，在之後的定位上不管是覆蓋率或是精準度都有一定程度的提升。

2.4.3 利用虛擬相機位置來定位

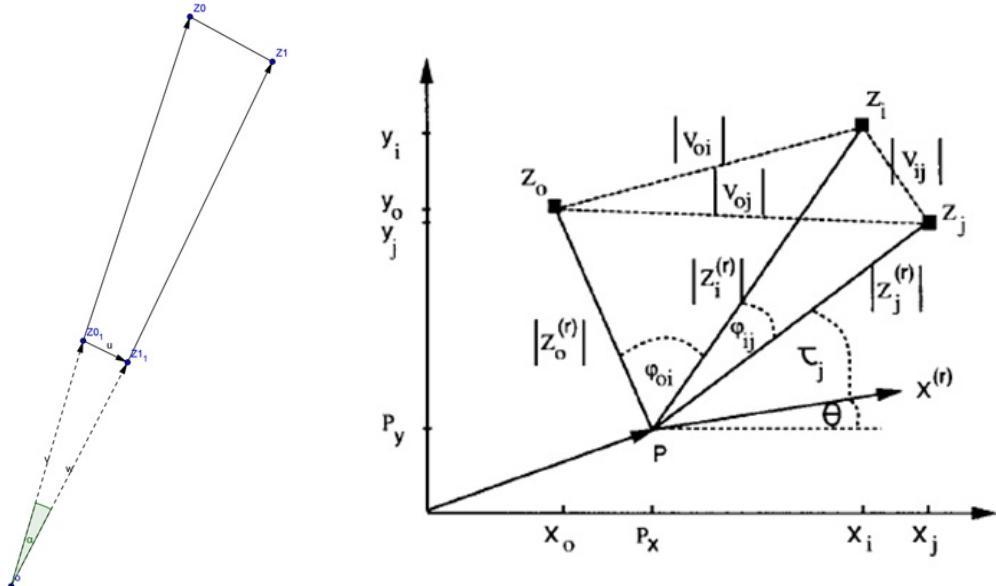
最後定位我們利用虛擬相機的位置來當作定位的參考位置，在利用虛擬相機的位置來定位與一般影像定位不同的地方在於三角定位的使用。在這裡我們先解釋一般三角定位的流程：

1. 找出特徵點對於鏡頭的夾角
2. 利用夾角帶入餘弦定理 (Cosine Law) 求出特徵點所距離鏡頭位置的長度
3. 利用已知的長度求出待定位圖片的位置

找出特徵點對於鏡頭圖片的夾角

在 2.10 當中我們要先求得 \vec{z} 與 \vec{w} 的長度，當我們知道 \vec{U}_1 與 \vec{U}_2 之後，帶入下列求解的算式：

$$\begin{cases} \vec{z} = (u_{1x} - c_x)\vec{u} + (v_{1y} - c_y)\vec{v} + \vec{f}d \\ \vec{w} = (u_{2x} - c_x)\vec{u} + (v_{2y} - c_y)\vec{v} + \vec{f}d \end{cases} \quad (2.15)$$



(a) 角度 α 對於夾角 \vec{z} 與 \vec{w}
示意圖 (b) 利用夾角帶入餘弦定理 (Cosine Law) 求出特徵點所距
離鏡頭圖片位置的長度

圖 2.11: 定位點夾角與長度向量關係示意圖

在這之中， f 為焦距向量， d 為深度。

我們得到 $|z|$ 與 $|w|$ 的長度，在圖2.10我們知道 $|u|$ 的長度之後，再帶入餘弦定理求得角度 α 的夾角：

$$|\vec{u}|^2 = |\vec{z}|^2 + |\vec{w}|^2 - 2zw\cos\alpha \quad (2.16)$$

利用夾角帶入餘弦定理 (Cosine Law) 求出特徵點所距離鏡頭圖片位置的長度

由圖2.11 我們可以知道 φ_{oi} 與 φ_{ij} 也知道 $|V_{oi}|$, $|V_{oj}|$ 與 $|V_{ij}|$ 的長度，藉由餘弦定理可以推出下列算式：

$$\begin{cases} |V_{oi}|^2 = |Z_o^{(r)}|^2 + |Z_i^{(r)}|^2 - 2|Z_o^{(r)}||Z_i^{(r)}|\varphi_{io} \\ |V_{oj}|^2 = |Z_o^{(r)}|^2 + |Z_j^{(r)}|^2 - 2|Z_o^{(r)}||Z_j^{(r)}|\varphi_{jo} \\ |V_{ij}|^2 = |Z_i^{(r)}|^2 + |Z_j^{(r)}|^2 - 2|Z_i^{(r)}||Z_j^{(r)}|\varphi_{ij} \end{cases} \quad (2.17)$$

其中 (r) 代表從定位點 P 所觀測出的位置與視角。

當我們解出 $|Z_o|, |Z_i|$ 以及 $|Z_j|$ 之後，根據圖上的座標表示法，我們最後帶入

式子 (2.7) 中求解

利用已知的長度求出待定位圖片的位置

$$\left\{ \begin{array}{l} |Z_o^{(r)}| = (x_o - p_x)^2 + (y_o - p_y)^2 \\ |Z_i^{(r)}| = (x_i - p_x)^2 + (y_i - p_y)^2 \\ |Z_j^{(r)}| = (x_j - p_x)^2 + (y_j - p_y)^2 \end{array} \right. \quad (2.18)$$

我們利用上述式子整理可得出下列式子：

$$\left\{ \begin{array}{l} |Z_o^{(r)}|^2 - |Z_i^{(r)}|^2 = X_o^2 - X_i^2 + 2p_x(x_i - x_o) + y_o^2 - y_i^2 + 2p_y(y_i - y_o) \\ |Z_o^{(r)}|^2 - |Z_j^{(r)}|^2 = X_o^2 - X_j^2 + 2p_x(x_j - x_o) + y_o^2 - y_j^2 + 2p_y(y_j - y_o) \end{array} \right. \quad (2.19)$$

在依照 (2.7) 式子兩兩相減，可得出六項聯立方程組，(2.8) 為其中的兩項，在式子當中我們求出 p_x 及 p_y 則為我們想要定位之座標。當然所有的特徵點會超過 3 點以上，這些將些的餘弦等式利用最小平方法求解，得出我們想要的定位結果。

上面為傳統 2D 平面影像根據特徵點的定位流程，我們根據虛擬影像也可以與待定位照片根據特徵點定位。但是虛擬影像為 3D 投影回 2D 的平面影像，在座標空間表示會面臨到投影所產生的誤差，再者點雲所見出的環境深度因為 Kinect 深度攝影機本身偵測的深度也會產生誤差，由虛擬影像跟平面影像特徵點利用式子求解比平面影像定位求解來的誤差更大。根據這點，我們利用虛擬相機的位置來當參考的定位點，當我們找出最多特徵點的虛擬照片後，我們還是利用虛擬影像作三角定位，當所求的定位點與虛擬相機絕對距離超過 70 公分，我們就利用虛擬相機位置做最後定位點，否則則三角定位的位置則為最後定位完成的結果。

會這樣做的理由基於每個相機的 x 軸距離與 y 軸距離是 50 公分，為均勻分布，所以假定最大的平均定位誤差就為 $\sqrt{x^2 + y^2} = 70.7$ 公分，當定位位置與相機距離超過平均誤差距離，我們會捨棄三角定位的結果，改以最多特徵點的虛擬相機位置為最終的成果。

第 3 章 定位實驗方法比較分析

3.1 實驗目的

為了改善室內定位的覆蓋率以及定位的精準度，我們利用建置虛擬相片的方法來增加照片的範圍及廣度。有了更多的取樣範圍，藉由實驗來跟以前傳統的 SIFT 方法做比較。我們分成兩個實驗環境來說明方法所改善的定位數據：(1) 可以控制的實驗環境、(2) 一般室內定位的環境。

首先製造一個可以控制特徵點數量的環境，在這個環境中我們驗證每個固定距離內根據 SIFT 所涵蓋的特徵點數量作比較，增加與物體的固定距離算出每個距離中的平均定位誤差，再算出定位誤差範圍的覆蓋率與傳統的照片影像定位作比較。在可以控制的定位環境下我們根據這些實驗方法說明改善的成果，再把方法放建築一般實際的室內環境中作比較，最後呈現出改善的平均定位誤差與增加環境所能定位的覆蓋率。為了完成這些實驗，所用到的設備為 Intel Core I5 2.0GHz 的 CPU 與 8 GB 的 RAM，顯卡為了能夠使用 CUDA 平行運算加速，所採用的是 Nvidia 的顯示晶片。

3.2 特徵點固定環境下定位結果分析

3.2.1 特徵點固定環境下實驗方法

建立可以控制的實驗環境主要目的為在可以控制的特徵點環境下與一般平面影像定位做比較，藉由實驗驗證出有更好的定位覆蓋率與更小的平均定位誤差。在根據有限景物數量的環境下，將待定位的照片依距離增加，生成出格狀的位置的照片定位點，每個定位點前後都有固定的間距距離，利用這些待定位照片，分別用三種方法比較實驗結果。

一開始建置實驗環境，我們將這些定位照片分布在 4 公尺 X 5 公尺的環境大小內，而景物的分布在 1.5 公尺 X 0.7 公尺的大小範圍內，Kinect 攝影機放置在距離景物 0.5 公尺前的地方，利用 Kinect 攝影機拍照取得深度照片與待定位的照

片。如圖3.1所示，在兩個控制環境的實驗中，利用不同的定位環境、不同的間距以及不同待定位圖片的數量來做實驗比較。

表 3.1: 控制環境實驗參數設定

實驗設定:	控制環境 1	控制環境 2
待定位照片數量:	45 張	54 張
間距寬度:	0.53m	0.5m
間距長度:	0.68m	0.5m

3.2.2 分析不同因素影響特徵點固定環境下實驗結果

為了證明在可以控制的環境下有比較好的定位結果，將實驗分成三個部分作討論：

1. 特徵點數量分布分析
2. 間距定位誤差分布分析
3. 定位精準度覆蓋率以及平均定位誤差

特徵點數量分布對定位結果影響探討

這個實驗目的是要看出根據景物的距離增加與特徵點數量的關係圖，實驗的作法是先在待定位圖片的第一排中間設為圓心，如圖 3.1所示。根據這個圓心將圓的半徑以 0.5 公尺的長度增加，依據每個圓之間所能覆蓋的特徵點位置做特徵點數量的總和並算出平均值，藉由圖中的趨勢看出特徵點數量的變化。

如圖3.2所示，橫軸表示與景物之間的距離增加，縱軸為特徵點的數量，三條線分別代表 (1).2D 隨機排列的攝影機影像位置，(2).3D 隨機排列的虛擬攝影機影像位置與 (3).3D 格狀排列的虛擬攝影機影像位置，一開始距離越近根據 2D 影像所找出的影像特徵點也越多，符合實際的情況，但之後的趨勢分布呈現卻發現距離越遠平面影像所找出的特徵點減少許多，表示說 2D 平面影像因為受到攝影機位置取樣的關係都侷限在景物附近，導致距離越遠卻沒有好的匹配影像做比對，

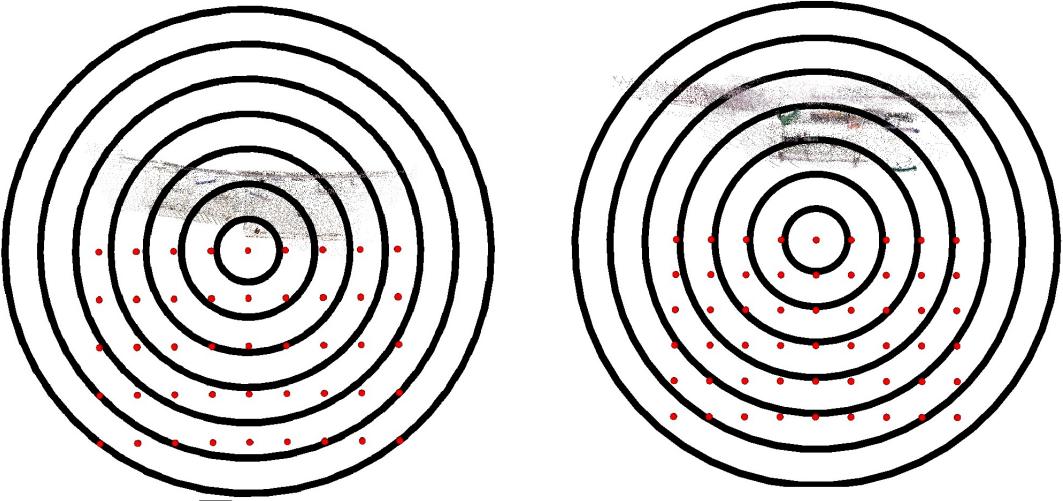
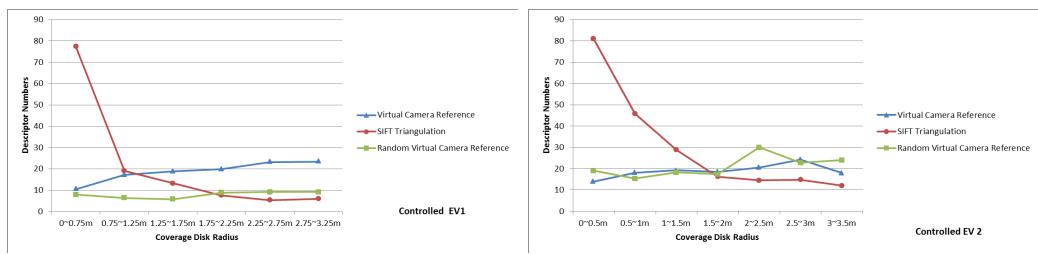


圖 3.1: 依照同心圓的覆蓋算出每個半徑內的每張圖的特徵點平均數，左. 控制環境 1 右. 控制環境 2



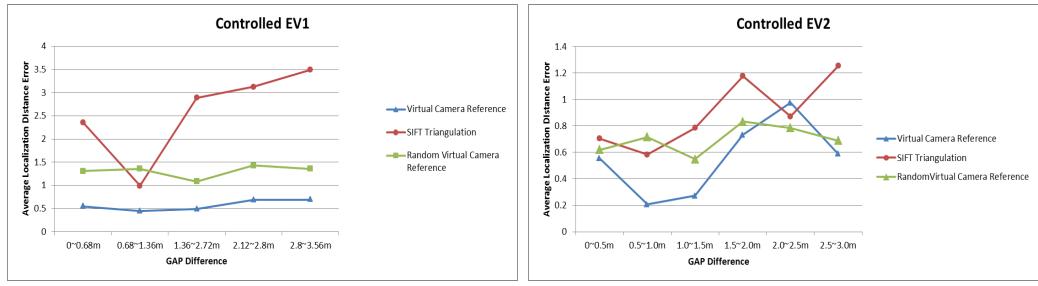
(a) 控制環境 1. 特徵點平均數量比較 (b) 控制環境 2. 特徵點平均數量比較

圖 3.2: 特徵點數量趨勢圖

但是 3D 虛擬影像的好處在於可以分布在整個環境區域中模擬平面影像所照出的照片，所以特徵點的數量不會因距離增加而有太大的改變，對於之後的定位精準度也不會因為距離增加而導致定位誤差有明顯擴大的影響。

間距定位誤差分布比較

以圖3.3可看出，以每一橫排的待定位圖片的平均誤差分布，根據每一個不同的間距，虛擬相機圖片定位比一般相機所做出的影像定位誤差都有改善，尤其以格狀分布的 3D 虛擬影像改善最為明顯。根據我們的方法，我們想要模擬在不同的角度及位置產生 3D 虛擬影像，比起一般的傳統影像有更多可以做特徵點匹配的相片可供使用。傳統的平面影像定位所做出的資料庫中，對於景物比較遠的照



(a) 控制環境 1. 間距定位平均誤差比較 (b) 控制環境 2. 間距定位平均誤差比較

圖 3.3: 間距定位平均誤差趨勢圖

片並沒有資訊提供，只能利用拘限在景物較近的照片可供定位，但是少許的特徵點使得定位誤差更加放大，所以照圖中趨勢來看，距離一增加，定位誤差就會加大。但是在 3D 虛擬影像不會因為距離的增加，導致定位誤差增加。除了 3D 虛擬影像可以增加更多可以被匹配的相片以外，好的虛擬相機分布，也可以產生更多的特徵點可以被匹配。以隨機分布與格狀分布來說，隨機分布的 3D 虛擬相片雖然有改善，但沒有比格狀分布的虛擬相片改善來的明顯。在我們的方法中，我們定位會參考照相機所在的位置，所以定位的位置都會在相機位置的附近，隨機分布可能會造成某些區域的相機分布過於集中，某些相機卻又過於分散的情況發生。所以在隨機虛擬相機分布的定位其實就跟平面相機分布的定位分布差不多，差別在於相片角度會避開障礙物，但不會均勻分布在環境中；格狀平均分布的攝影機位置，就會均勻地分布在環境內，而對於影像特徵點匹配上比隨機分布來的更有幫助。

定位精準度覆蓋率以及平均定位誤差改善情況

研究發現與景物的距離越遠在 3D 虛擬影像定位並不會影響誤差，整體來看，對定位的覆蓋率也有一定的提升。根據圖 3.4(a) 所示，在格狀分布的 3D 虛擬相機定位在誤差 0.5 公尺左右有超過 60% 的覆蓋率，但隨機分布的虛擬相機以及 2D 平面影像定位坐落在 20%~30% 左右上下，表示相機的分布影響定位結果的好壞。這也是我們想讓照相機格狀分布平均的原因，再回到圖 3.2(a) 來看，在 1.75 公尺以後的隨機虛擬相片所找出特徵點平均數量比起格狀分布虛擬相片所找出的特徵點數量少上許多，而在圖 3.3(a) 來看，每個間距的定位平均誤差，格狀分布都比隨機分布的誤差好上許多。所以當以誤差在 0.6 公尺以內的覆蓋率來說，隨機分布

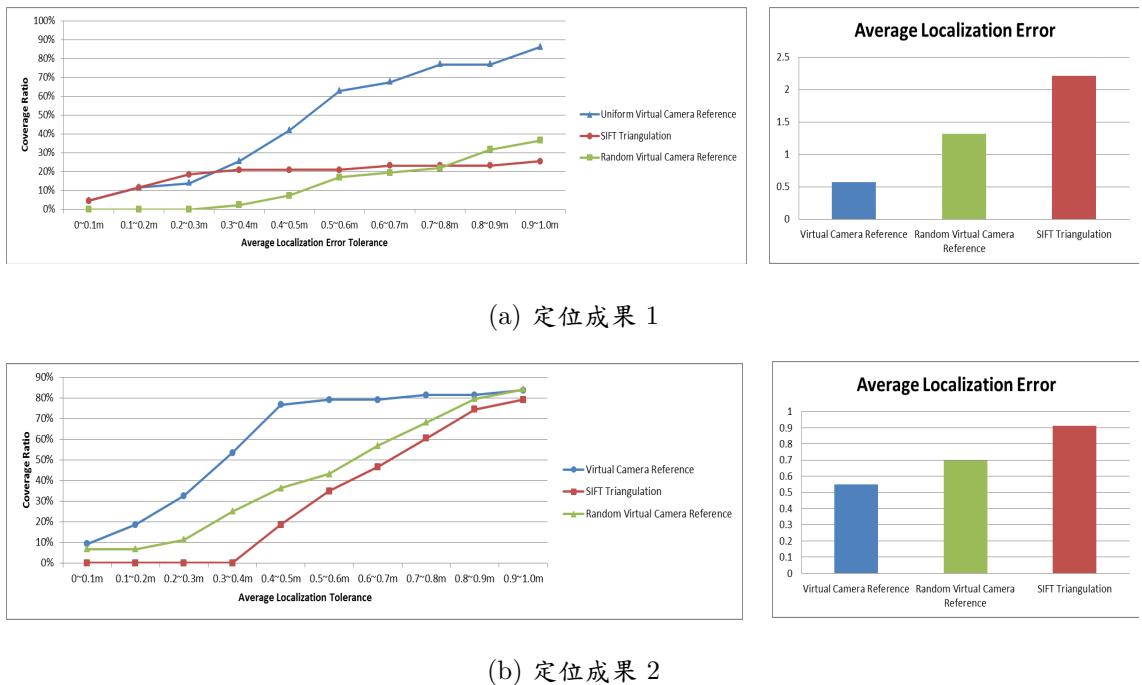


圖 3.4: 定位環境成果

的虛擬相片並沒有產生比較好的定位覆蓋率，但以格狀分布的虛擬相片定位覆蓋率與 2D 影像定位覆蓋率相比，結果好上許多。這說明虛擬相機分布的位置，影響了虛擬相片的品質，平均定位誤差平均定位的誤差以格狀的虛擬影像為物誤差為最低，2D 影像定位的誤差為最高，改善了整體定位的平均誤差。

3.3 一般室內環境定位實驗

在控制環境下，改善了定位的覆蓋率與精準度，在這章節將在一般室內環境下進行定位測試。我們將室內定位環境分成三種情境：(1) 居家客廳，(2) 居家廚房與 (3) 居家房間，分別在這三種環境下進行定位實驗。在一般室內定位主要進行定位覆蓋率與定位平均誤差測試。

3.3.1 室內環境定位方法說明

在室內定位的情況下，拍攝待定位照片的方法依據環境情況而定。待定位照片拍攝方法是依照人能夠活動的範圍作依據，在這些區域進行格點分布拍攝，如

圖3.5(c)與圖3.5(d)所示。每張待定位圖片的間距距離為 0.5 公尺，照片數量根據環境大小而定，平均在 30 張上下。虛擬照片依據間距距離，取出不同數量的虛擬相機。因為考量不同環境的景物分布，每組相機位置分別拍攝兩種不同的角度，最後根據虛擬相機拍攝出的照片作影像定位。除了與 2D 影像定位方法做比較之外，分別測試在不同虛擬照片資料數量的定位情形。

表 3.2: 室內定位環境大小分布

實驗設定:	客廳	廚房
待定位照片數量:	35 張	30 張
間距距離:	0.5m	0.5m
環境長度:	3.8m	4.1m
環境寬度:	4.1m	3.2m

表 2.2 記錄了不同室內定位環境的設定，分別在不同環境下進行實驗，每個環境的虛擬相機位置均採用格狀分布。圖3.6 記錄當時點雲環境的建置以及待定位照片的位置分布。點雲的建置是利用 Kinect 環繞室內環境四周所拍攝，再利用這些拍攝的圖片，當作平面 2D 影像定位所需的影像資料庫。接下來分別以不同虛擬相機照片的數量與平面 2D 影像做實驗比較分析。

3.3.2 定位數據結果分析比較

在室內定位的環境下，特徵點分布的數量，以及環境內觀測物的不同對定位結果增加許多變動因素。為了使定位結果能夠量化比較，我們將實驗結果分成 (1). 定位精準度覆蓋率，與 (2). 平均定位誤差兩個指標來分析成果好壞。在圖3.6之中來看，橫坐標表示定位誤差的容忍範圍，縱座標代表在這個誤差範圍下的定位成功率，以三種不同虛擬照片的數量與傳統 2D 影像照片做實驗比較。

以覆蓋率來看，100 張虛擬照片的定位結果為最好，代表取得越多點雲環境的資料，定位的改善越明顯。因為在每個相機位置上，我們取出兩張虛擬照片，所以在環境中總共有 50 個虛擬相機位置作比較，分別與 40 個與 35 個相機位置相比，有更多定位參考的依據。則平面 2D 影像因為只有一部分的環境參考依據，所以在定位覆蓋率與平均誤差，都比 100 張與 80 張虛擬照片定位成果來的差。以3.6(a)來看，誤差範圍在 0.5m 0.6m 之間的覆蓋率，以 100 張虛擬照片覆蓋率最

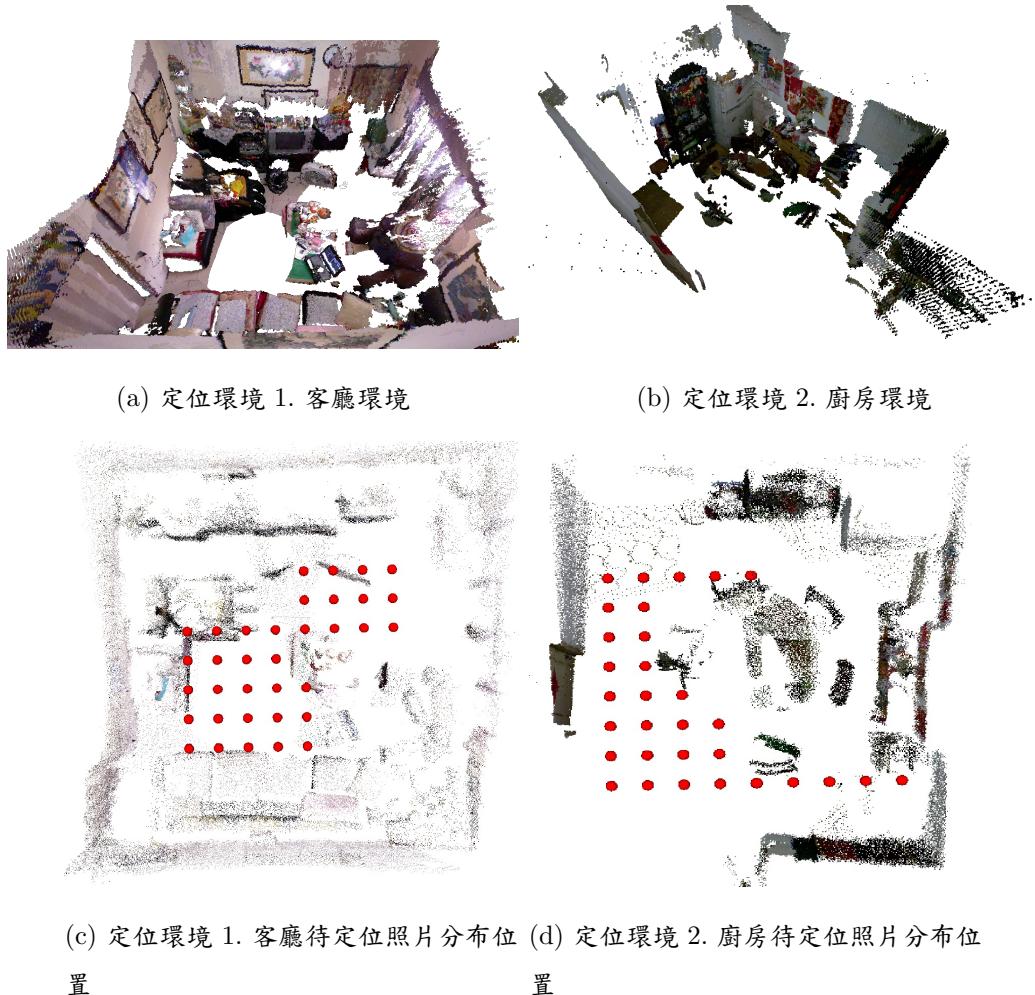
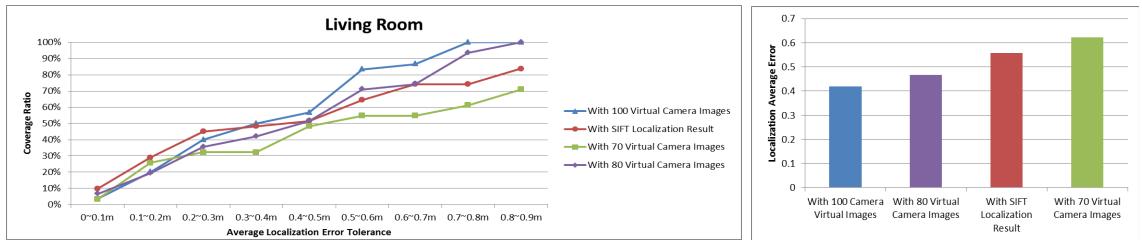
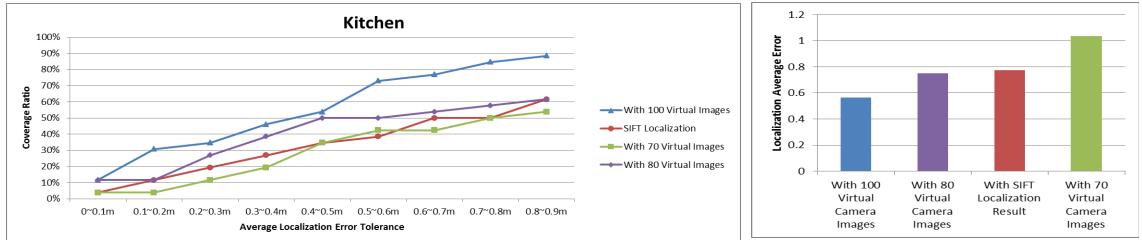


圖 3.5: 室內定位環境與待定位照片分布位置



(a) 定位環境 1. 客廳環境



(b) 定位環境 2. 廚房環境

圖 3.6: 室內環境定位成果

高，有 80% 左右的定位成功率，但是平面 2D 影像只有 60% 左右的成功率，增加了 20% 的覆蓋率。以平均誤差來說也有改善，在上個章節中我們發現平面 2D 影像定位誤差會有不穩定的情況發生，這種情況增加了整體的平均定位誤差，而在虛擬影像定位因為誤差會穩定在 1m 以內的範圍內，所以也會降低平均定位誤差。

以 3.6(b) 來看，待定位的照片集中分布於同一側，所以 2D 影像定位沒有足夠的環境資訊，在誤差範圍 1m 以內的覆蓋率也只有在 60% 上下，100 張虛擬照片的定位結果覆蓋率提升最為明顯，可以看出虛擬影像定位可以改善相片分布的位置以及優化相機取出的角度，也可以節省紀錄相機位置的時間。

Bibliography

- [1] C. Harris and J. Pike, “3d positional integration from image sequences,” *Image and Vision Computing*, vol. 6, no. 2, pp. 87 – 90, 1988.
- [2] J. Leonard and H. Durrant-Whyte, “Simultaneous map building and localization for an autonomous mobile robot,” pp. 1442 –1447 vol.3, nov 1991.
- [3] R. Chatila and J. Laumond, “Position referencing and consistent world modeling for mobile robots,” *Robotics and Automation. Proceedings. 1985 IEEE International Conference on*, vol. 2, pp. 138 – 145, mar 1985.
- [4] R. Smith, M. Self, and P. Cheeseman, “A stochastic map for uncertain spatial relationships,” *Proceedings of the 4th international symposium on Robotics Research*, pp. 467–474, 1988.
- [5] S. Betge-Brezetz, P. Hebert, R. Chatila, and M. Devy, “Uncertain map making in natural environments,” *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, vol. 2, pp. 1048 –1053, apr 1996.
- [6] A. Davison and N. Kita, “3d simultaneous localisation and map-building using active vision for a robot moving on undulating terrain,” vol. 1, pp. I–384 – I–391, 2001.
- [7] H. Durrant-Whyte, S. Majumder, S. Thrun, M. Battista, and S. Scheding, “A bayesian algorithm for simultaneous localisation and map building in springer tracts in advanced robotics,” vol. 6, pp. 49–60, 2003.
- [8] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [9] H. Bay, T. Tuytelaars, and L. Gool, “Surf: Speeded up robust features,” *Computer Vision – ECCV 2006*, vol. 3951, pp. 404–417, 2006.
- [10] G. Yu and J.-M. Morel, “A fully affine invariant image comparison method,” *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 1597 –1600, april 2009.

- [11] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, pp. 381–395, jun 1981.
- [12] H. Du, P. Henry, X. Ren, M. Cheng, D. B. Goldman, S. M. Seitz, and D. Fox, “Interactive 3D modeling of indoor environments with a consumer depth camera,” *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11*, pp. 75–84, 2011.