

國立交通大學  
電機與控制工程研究所

碩士論文

利用空間高斯混合模型進行影像地標判定  
及輔助定位



Image Landmark Registration and  
Localization

Using Spatial Gaussian Mixture Model

研究生：黃 恒 嘉

指導教授：胡 竹 生 博士

中華民國九十五年七月

# 利用空間高斯混合模型進行影像地標判定及輔助定位

研究生:黃恒嘉

指導教授:胡竹生

國立交通大學

電機與控制工程研究所碩士班

## 摘 要

在電腦視覺的領域中，要讓機器人能夠具有對於場景的認知，往往都需要在環境空間中，標注許多人工的特徵，或是紀錄場景中，特殊的形狀或是顏色，來幫助機器人確認自己的所在位置。然而在圖形識別的領域中，有許多功能成熟且強大的演算法，對於資料分析而言，並不需要對於輸入資料有很大的限制，也可以達到很高的辨識結果。本論文中採用高斯混合模型(Gaussian Mixture Model, GMM)來描述空間中的場景，將場景概念化，這樣任意的場景皆可以用高斯混合模型來表達，而無需加入大量的人工場景，接著將所獲取的每個希望校正點的資料採用由修改後的 Cyr and Kimia 的結合演算法 (combination algorithm) 來減低及統合資料，當場景接近原本的當初建立的場景，經由最大可能性(Maximum Likely)大體上會呈現單調的特色，將輔助機器人修正回原本的位置，以達到協助機器人定位的目標。

# **Image Landmark Registration and Localization**

## **Using Spatial Gaussian Mixture Model**

student: Heng-chia Huang      Advisor : Dr. Jwu-Shen Hu

Institute of Electrical and Control Engineering  
National Chiao Tung University

### **ABSTRACT**

To achieve the robot registration and localization using computer vision, artificial landmarks or specific shapes or specific colors in the image are usually used. However, there exist many robust algorithms in pattern recognition and image segmentation which do not require constraints on input data to achieve a good recognition performance. In other words, localization can be performed based on a general scene. In this thesis, we adopt Spatial Gaussian Mixture Model (GMM) in image segmentation to describe an image viewed by the robot in spatial domain without any artificial landmark. Secondly, we use a modified combination algorithm by Cyr and Kimia to combine similar data. It is found out that a monotonic relationship exists among the scene registered and its neighborhood in terms of distance. This phenomenon can be used to assist to localize the robot and this work demonstrate the feasibility by several experiments.

## 致謝

經過了披荊斬棘，焚膏繼晷，披星戴月終於皇天不負苦心人，終於在擠進研究所的大門後，又完成了人生另外一個重要的里程碑，理所當然，要感謝的人實在太多了，連天都被學長給謝走了，不過還是要八股的再一次感謝幫助過我的所有親朋好友。

給予我這篇論文最大指導方向的，當然屬於我的指導教授，胡竹生老師，他除了在研究方面給予一針見血的見解，更在邏輯思維方面，身教重於言教的以身作則，樹立了良好的典範；帥氣又古樸的龜毛宗敏學長，是我們影像組的棟樑，常常學長代師父職的給予我們最貼切的幫助與指導；幽默又搞笑的士奇跟穩扎穩打又喜歡抓頭的群祺，在影像組開出一條康莊大道；機器人與攝影高手立委；曾經史上最強的學士价呈；健身二人組的安喬兄與劉大人；英俊的貴族興哥；可愛但是臉很大的鏗元；害我接管理員的岑思；硬體高手晏榮；以及與我一同奮戰的夥伴，中流砥柱的螞蟻；超有信心與實力的朱木；帥氣耍色的耀賢；天真皮膚又好的鳥蕙；與影像可愛的夥伴的佩靜，要衷心的感謝一起走過多個報告的前夕，一起熬夜除蟲……，之間的點點滴滴可以出一部小說；還有學弟、同學，傻傻分不清楚的永融；影像組的新血弘齡；責任感太重的完美 alphas；運動好手楷祥；除此之外還要謝謝美麗又漂亮的淑玲，對於管理員的我在行政職務上面有莫大的幫助。

當然還要謝謝我所有的家人與死黨，在精神上給予我最大的協助，讓我可以更加專心的投入在研究領域上。在此至上我最大最大的感謝。

# 目錄

摘要.....	II
ABSTRACT.....	III
致謝.....	IV
目錄.....	V
圖目錄.....	VII
表目錄.....	VIII
<b>第一章 序論.....</b>	<b>1</b>
1.1 研究動機與相關背景.....	1
1.2 相關研究回顧.....	2
1.2.1 場景認知.....	2
1.2.2 機器人定位導航.....	3
1.2.3 電腦視覺.....	4
1.3 論文的主題.....	6
1.4 本篇論文的貢獻.....	7
1.5 論文的架構.....	7
<b>第二章 影像相似度判別演算法.....</b>	<b>9</b>
2.1 影像相似的量測準則.....	9
2.2 影像相似的方法概述.....	10
2.3 低通濾波器(LOWPASS FILTER).....	11
2.4 擴展直方統計圖(EXPAND HISTOGRAM).....	13
2.5 主軸分析(PRINCIPLE COMPONENT ANALYSIS).....	15
2.6 統計模型(STATISTICAL MODEL).....	16
2.7 利用影像相似度來進行影像認知與定位.....	20
<b>第三章 高斯混合模型建立資料庫.....</b>	<b>22</b>
3.1 高斯混合模型(GAUSSIAN MIXTURE MODEL).....	22
3.1.1 高斯混合模型簡介.....	22
3.1.2 模型的描述.....	23
3.1.3 模型的選擇.....	27
3.1.4 模型的參數初始化.....	28
3.2 最佳相似性估測演算法(EXPECTATION MAXIMIZATION).....	29
3.2.1 期望值(Expectation Step).....	30

3.2.2 最大值(Maximization Step) .....	31
3.3 高斯混合模型接近高群數的問題 .....	35
<b>第四章 資料庫集合與結合演算法 .....</b>	<b>39</b>
4.1 資料庫的建立 .....	39
4.1.1 Cyr and Kimia 提出的結合演算法 .....	39
4.1.2 修改過後的結合演算法 .....	41
4.2 辨識方法 .....	46
<b>第五章 實驗結果 .....</b>	<b>48</b>
5.1 系統架構 .....	48
5.2 實驗平台 .....	49
5.3 實驗結果 .....	51
5.3.1 比較各演算法之辨識結果 .....	51
5.3.2 高斯混合模型的容忍度測試 .....	55
5.3.3 影像相似與影像搜尋引擎 .....	58
5.3.4 環場影像的測試與分析 .....	59
5.3.5 利用高斯演算法大致上單調趨勢輔助定位 .....	61
<b>第六章 結論與未來發展方向 .....</b>	<b>66</b>
<b>參考文獻 .....</b>	<b>68</b>
<b>[23] .....</b>	<b>70</b>



# 圖目錄

圖 1	影像量側的準則	9
圖 2	低通濾波器可選擇的不同尺寸的遮罩	12
圖 3	利用擴展直方圖計算影像間的差異	14
圖 4	透過 PCA 運算所得到的模糊影像，中間為各自計算特徵向量，後側為合併計算	16
圖 5	不同群數去模擬資料分布的情形	17
圖 6	利用 KDE 進四真實資料的統計分佈	18
圖 7	頻寬係數對於 KDE 的影響	19
圖 8	擷取五維的特徵向量	23
圖 9	高斯混合模型之架構圖	24
圖 10	強度統計分布及高斯混合模型近似的分布	25
圖 11	不同共變矩陣的外觀之示意圖	25
圖 12	真實的將色塊影像透過不同的共變矩陣所得到的結果	26
圖 13	對角矩陣透過其分布將影像模糊化	27
圖 14	平均值分類法(K-MEANS CLUSTER)流程圖	29
圖 15	LIKELIHOOD FUCTION $E(\lambda)$ 最大化的示意圖	33
圖 16	高斯混合模型建立的架構	34
圖 17	一般的高斯分布與某一維度變異量為 0 的情況	36
圖 18	在計算機率值時，防止變異量過小之機制的流程圖	38
圖 19	在赤道面上取樣物體的 2 維影像	40
圖 20	利用攝影機選轉模仿多角度觀察物體	42
圖 21	利用物件辨識的結合演算所建立的減低資料庫的流程圖	46
圖 22	影像地標建立與未知影像比對	48
圖 23	完整的高斯混模型的建立與測試流程	49
圖 24	SONY EVI-D30 的攝影機與 PCI-EXPRESS CAMERA LINK 影像擷取卡	50
圖 25	建立影像地標的辦公室場景圖	50
圖 26	(A)環場攝影機；(B)環場攝影機所拍攝的影像	51
圖 27	置入卡通海賊王玩偶障礙物遮蔽影像	53
圖 28	(A)空間中 11 個場景(B)高斯混合模型之分布(C)其模糊化後之影像	54
圖 29	簡易搜尋引擎介面，由手繪的印象影像尋找相似的資料庫影像	58
圖 30	環場影像與展延影像示意圖	59
圖 31	某區段的環場影像透過延展的結果	59
圖 32	不同群數對於同一個場景的變化關係圖	62
圖 33	單調影像、局部最小值破壞單調、極值發生點錯誤	64
圖 34	當拍攝點向後移動時，像素的變化示意圖	64

# 表目錄

表 1	無參數核心密度估測 KDE 之核心 .....	18
表 2	各模型在同一位置不同角度的辨識率 .....	51
表 3	各演算法在不同地標的辨識率 .....	52
表 4	加入障礙物之後的辨識結果 .....	53
表 5	各位移量與各辨識方法搭配的結果 .....	55
表 6	加入障礙物遮蔽影像且偏離影像地標 .....	57
表 7	七個位置的展延影像與結合過的資料庫數量以及辨識率 .....	60
表 8	不同的群數對於場景的穩定區間 .....	61
表 9	單調關係存在的場景數量統計 .....	63
表 10	旋轉與左右移動測試單調的結果 .....	65



# 第一章 序論

## 1.1 研究動機與相關背景

在目前現有的機器人定位方式中，不論國內外，一般較為常見的方式，幾乎皆是標定特定的人工圖樣，或是特殊的顏色形狀，然而隨著機器人越來越普及，也越來越民生化，面對的不再只是單調的環境或是可以容易標示出特殊顏色或形狀的地標，取而代之的則是一般的家庭環境。當面對這樣的議題，必然存在著許多解決的方式，但是隨著科技日新月異，電腦速度越來越快，電腦視覺必定成為解決此問題的核心方法之一，因此我們反觀人類的行為，當一個近視的人，即使脫掉眼鏡，霧裡看花，也不至於在房間中迷路，因此當人類看到一個場景，而這個場景中局部細微的顏色變化、區塊的漸層，對於所在位置的影響，微乎其微，因此，回到本來的問題，對於任何的自然影像，如何簡易的表達出他們重要的區塊？如何概念化所獲的影像資料？

因此，重新審視這個問題，然而在機器人定位或者是電腦視覺 (computer vision) 這個領域，比較少人提到這一類的問題，於是向外尋找適合的解決方式，於是在圖形識別 (pattern recognition) 以及影像分割 (image segmentation) 等領域，卻有不少概念化，標示出影像中重要區塊的演算法，而這些演算法，提供了解決問題的遐想空間。試想當機器人可以不在受限於簡易或是特殊設計過的環境，皆能夠“認知”自己存在環境中的相對位置，未來對於無限可能的居家環境，或是甚至是醫院醫療、危險環境監控、無人區域的警衛、大型辦公室的助理...的可能性，無庸置疑的是一個很重要的議題，所有的行動必先決定自己的位置之後，才有可能接續下去完成；也就是說，我們希冀可以發展出一個較低限制的機器人視覺系統，來

模擬人類的行為，以概念化的場景為主軸，挑選適當的模型來闡述環境中的重要區塊。

而導航之初的定義：我在哪裡，我該如何去那裡[1]，無疑的是本論文中得著墨點也是最重要的議題，而在實驗的過程中，也意外的發現一些演算法中，大體上本來就存在的現象，但鮮少被人提及與應用，利用這個特色，希冀建立一個低限制的架構，來達到輔助機器人視覺定位導航的可行性。

## 1.2 相關研究回顧

透過上一節的描述，場景認知、機器人定位導航與電腦視覺，這三個領域跟我們有密切的關係，所以，研讀整理的論文，可分成三大類，如下：

### 1.2.1 場景認知

視覺認知，亦是機器人定位這領域的核心問題，對於人類來說，輕而易舉的事情，但是在電腦視覺領域，卻是一大難事，因此很多論文在這方面也提出許多解決方法；

以外在環境為考量。第一、空間中幾何位置的描述是透過完整的電腦繪圖模型，再利用機器人多視角之間的整合其關聯性，利用這些資訊做地標偵測或者與地圖模型相對應尋求吻合[2][3]，以達場景認知的效果；然而這類的認知，能否清楚正確的挑出地標，成為認知成敗與否的關鍵，所以，一個非人工刻意標示的地標物則比較難達到這樣的要求，第二、利用拓樸關係描述每個區域之間的關係[4][5]，取代了之前全域的地圖集以描述機器人周遭場景的影像，而不需去在意全域的幾何關聯性，其精確度稍微低一點，但是其運算量比較低，速度相對的也比較快。

而視覺基底的影像認知系統以概念來分類，可以分為感知以及非感知

兩大類。感知，是指可以辨識出影像中的內容物，藉由內容物的相對位置，來推算自己所在的區域；而非感知，相對而言，不對影像內容做分析，而是將整張影像的資料做分析，再透過資料分析比對，來達到影像認知的效果。

## 1.2.2 機器人定位導航

機器人導航，已經由來已久，相關的論文也不勝枚舉，對過去相關統合整理的論文也相當完善[2]，如同之前概述的論文提到，順著歷史潮流，回溯到西元 1979 年，在對戶外外在環境有極高限制之控制下，第一台自走車誕生[6]，同年，也有另一隻適用於室內環境的自走機器人問世[7]，而隨著時代的變遷，科技的演進，電腦視覺以及影像處理的技術，越來越趨於成熟，到了 1984 年，幾乎已奠定了影像處理在機器人定位導航上面，不可忽視的一環[8]，而到了 1996 年，美國首台無須動手控制的自走車自匹茲坦堡到聖地雅哥，橫渡整個美國[9]。

而視覺基底的機器人定位系統同上一段所介紹，同樣可以分為感知以及非感知兩大類，只是將影像認知實現在機器人定位導航系統中。感知，是指機器人在環境中可以相對的甚至是絕對的知道自己處於環境中的位置；而非感知，相對而言，就是在設計上，不特別考慮中間過程，有簡單的行徑指標，來達到自走車或是自走機器人的概念。

而對於機器人視覺的感知能力要能夠實現，無庸置疑的，就是需要建立環境的地圖模型，而依照建立地圖模型的方式，又可再分類成為，一、預先建立地圖拓撲模型(Map Based)：在欲導航環境中建立空間中的相對或絕對位置(CAD model)，爾後，再利用機器人視覺，對所拍攝到的場景，進行辨識審核，來決定目前所在位置，進而推論[10]或是事先規劃好行進路徑。要達到定位導航，可分為四大步驟，依序是擷取待測場景影像，偵測地標或是摘取影像特徵，接著是，將抽取特徵或地標，與資料庫的影像

做比對，最後，估算其目前所在的位置。而這方面的研究，也有許多論文在強調定位的重要性[11][12][13]，畢竟，定位為導航不能不解決的問題，要能導航，必先面臨定位問題，本篇論文也會以這方面為探討主題。二、則是在行進間，建立地圖模型，並利用該圖來做為導航的依據(Map Building)，當然，在單純視覺基底的智慧型演算，要做到邊移動邊建立環境拓撲關係，在運算上，勢必比較緩慢，但是其更加近似人類的行為，卻是其優勢[14][15]。

而沒有環境資訊的導航系統(Mapless Navigation)，要能夠導航，勢必讓機器人沿著特定物體或是特定建築物的邊緣移動，才有可能達到自走車或是智慧型機器人的可能性[16]。

### 1.2.3 電腦視覺



提到電腦視覺，影像認知等等問題，就要從影像定址(Image Registration)談起[17]，而其處理主要可分別為對抗三大類的問題，所做的改良；其一、在一連串序列的影像，如果沒有加入空間中幾何關係和出現的順序，對於影像定址比對，勢必多了許多困難處，因此，找出空間關係或是空間轉移函數，勢必成為解決該問題重要的里程碑。其二、當影像中存在著光影或是霧氣.....的影響，在影像比對時，將會面臨極大的挑戰，於是乎，自然而然的加入光影強度(intensity)的考量，然後轉換色彩空間到，可以清楚分離該特徵的空間域，方便對於光影變化有所處理。其三、則是當影像畫面中的物體有所改變，又為了方便影像追蹤於是各式各樣的特徵抽取，成了這類論文探討的主軸，其中包括了，背景濾除，影像追蹤，等等的議題。

影像定址比對，一般都是為了特殊的應用而產生其解決方案，所以要比較適當的分類這類的議題，必須針對其所需要解決的問題來分類，第

一、多感測元件的整合，需要影像定址技術，將各個不同感測元件在同一個拍攝角度所擷取的影像，做整合分析，以便標示出影像中有哪個地方可疑。譬如，醫學影像，可結合核磁共振、電腦斷層、X光影像.....的影像，來辨識出人體是否有病變，當然，類似的用法也可以用在，大樓探勘、地層分析、蔬果分類的超音波、電磁頻帶的偵測。第二、為了找到待測影像，與資料庫影像，局部或是全部圖形區塊相似，也是利用定址技術，找出關聯性。其方式大約有：模型基底近似、事先分析擷取合宜特徵、熟習感興趣物件的特色，或是階層式比對。譬如，自動停車系統、航空定位、飛彈導航，又或者是人形、圖案、數字等特徵比對，以及訊號分析處理的自動化，也常常使用此技術。第三、利用視覺定址技術，將不同視角的影像，做影像序列的分析。而電腦視覺，也常常是這類問題的常客，甚至物體追蹤，或是影像序列分析，些微差異影像的偵測都屬於這個範疇。第四、與之前的差異在於空間的不同，替換成時間的差異，主要是在偵測，一段時間前後的差異，以醫學影像為例，觀察身體異常腫瘤的成長，就是這一類的核心問題，同理可知，天然資源的監控，核災之後的植物觀察，也可以算是這類型的題目主軸。然而，以上就應用層面分成四類，但是，一旦遇到問題，並沒有刻意定義成某個類型，完全隨實際遇到的問題來分析，所以，面對以上兩個甚至三個種類的合併的問題，也不在少數[14]。

以本篇論文為例，以應用層面應該屬於第三類的電腦視覺，但是核心部分卻是利用影像相似度的偵測，比對連續影像序列的差異性。至於量測影像相似度，則會受到特徵選取不同，而跟著有不同的效果與方式，主要可以區分成，是否對於影像內容做分析與判讀：對於影像內容做特徵抽取，譬如說，判讀影像內容中的物件、特定形狀、特定顏色.....作為比對特徵，再利用這些影像內容的資訊，接著使用類神經網路，模糊控制[18]等理論進行定位，或者在利用，空間中幾何關係，來進行路徑規劃，進而達到導航的效果[19]；而這類型的定位技術，其優劣成敗關鍵，幾乎完全

取決於對於影像內容，能否完整抽取物件、形狀、顏色等特徵。而相對的，則是對於擷取到的影像，並不特別對影像內容做分析，而直接對於整張影像，做資料分析[13]，利用資料分析的結果，來判定目前的位置；而這類型的定位系統，通常都會是把原本的資料(raw data)經過各式各樣的運算，擷取出部份甚至全部當作影像的特徵，在比對彼此之間的差異性，來達到定位的要求，換句話說，也就是常用各式各樣的模型，來描述闡述原本的影像，而本篇論文，也將採用這樣的方式來描述所擷取的影像。

### 1.3 論文的主題

本論文主要分為兩個主題，影像地標的認知，與利用演算法中偏離影像地標越遠，其相似度越低的單調特色，來輔助機器人定位。

其一、在實驗環境中，我們可以先對於我們所感興趣的位置建立影像地標，然而，因為我們並不預期機器人一定會朝向特定方向，所以，我們對於同一個感興趣的空間位置，我們將在該位置建立多角度的影像集合，做為該點影像地標的資訊，利用多個角度的影像地標，來更加準確的輔助機器人，使機器人對於影像地標集合的認知更加正確；我們並藉由此實驗來分析不同的影像量測演算法，再進行盲目搜尋(blind search)有較好的結果，並挑選其中高斯渾合模型，來進行資料整合與資料統整的步驟，做一個更有系統的影像認知系統。

其二、當我們可以透過上述的方式對於影像地標的認知比對，使機器人約略知道自己處於環境中的特定區域範圍內，那在假使影像地標並沒有改變的情況下，是否演算法本身有一定程度的輔助機器人，可以使機器人更加接近影像地標建立的位置，而達到輔助定位，爾後將透過一維空間中的數個實驗，來描述演算法本身即存在一定的變異特色，可以輔助機器人

回到當初建立影像地標位置。

## 1.4 本篇論文的貢獻

本論文利用影像分割中，具參數模型(Parametric Model)的概念化區塊化空間中的自然場景，首先可以幫助機器人對於環境中的影像地標，有比較高的辨識率，可以讓機器人認知道自己落在環境中的某個區域中。

爾後，把機器人定位這個動作，仰賴概念化後的影像相似程度，把機器人修正回本來的位置，而不在採用環境中的點對點，或是物體對物體的映射對應的方式，希望可以使機器人定位能減少更多限制而更加的有彈性。而要能夠達到這個目的，必須在局部區域存在演算法的數值變化具有單調函數的特色。本文首先透過適當的選擇合宜的參數模型，並針對多群數回歸問題提出一個偏差較少的解決方式；並利用實驗方式驗證，概念化影像在一個自由度中幾乎存在單調函數的特色。

## 1.5 論文的架構

本論文將以影像地標的認知為基本問題，並希望透過演算法本身來達到輔助定位的效果，而組織架構如下：

第二章：先行給予影像判別的準則，再接著簡介一些常用的演算法來擷取影像特徵空間，並描述這些演算法的特色，與各演算法在特徵空間中比較變異量的方法。

第三章：以高斯演算法為主軸，介紹其獲得參數的演算法，並分析當高斯以較高群數在描繪影像資訊時，會遇到的問題，並提出解決方法。

第四章：當在建立影像地標，以提供之後所需的認知與定位資料庫時，我

們往往希望能夠減少資料庫的數量，所以透過修改過的結合演算法來達到減少資料量的目的。

第五章：主要有影像認知，與輔助定位在一維自由度中兩部分，將透過影像認知中所選取的演算法做為之後的輔助定位的演算法。

第六章：對本論文做一個結論，並提出未來可以繼續發展的方向。



## 第二章 影像相似度判別演算法

### 2.1 影像相似的量測準則

兩張影像彼此的相似度，如何量測？又或者是兩張影像的差異性有多大？如何公平且適度的定義這個量值，就牽涉到量測的準則。然而，影像相似的判別方式，完全依賴選取的影像特徵為何，而有不同的比對方式。但是，無論採用哪種比對方法，都必需遵照一個準則，那也就是，不論採用何種特徵空間中以及其相對應的比對方式，差異越大的影像，差異的程度或是差異的值，也要越大，如圖 1 所示，前面兩張差異小，影像差異值就低。越能夠滿足這個準則，其資料分析等過程中，誤判的可能性就越低。

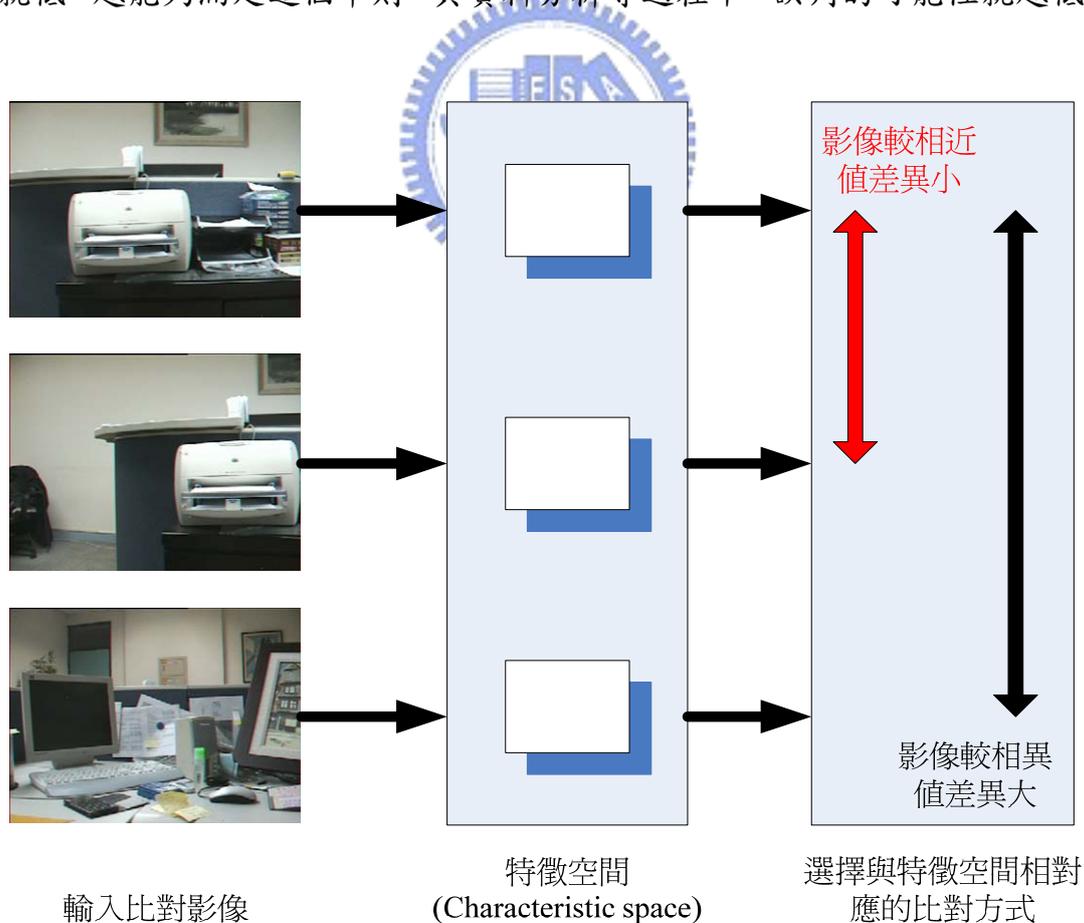


圖 1 影像量側的準則

## 2.2 影像相似的方法概述

然而，根據上一章所提到的，在選擇影像資料的特徵空間時候，有提過可以分為對影像內容有所認知的特徵空間，以及不對影像內容做分析，而對整張影像資料做統計的特徵空間；而第一種方法，就必須對影像做較多的前處理，以便有效的擷取影像內容來達到較優良的比對方式，爾後，當順利擷取出影像中所有重要特徵，將會需要臆度空間中的幾何關係，利用空間中的幾何關係，來判別影像的相似度，然而，如果採用這類的比對模式，將面臨兩個問題，第一，在極大部分的情況下，對影像進行前處理，必須對於外在環境很熟稔，對於可能的問題，利用影像處理技巧，來各個解決，譬如在光源不足的地方拍攝，容易產生高斯雜訊，於是在光源不足的地方，可以加入專門移除高斯雜訊的濾波器，來得到較清晰的影像。第二，當需要利用空間中的幾何關係時，也往往需要其他的感測器，或是比較特殊的感測元件，來對於外在環境的幾何關係，進行推論，方能進行影像比對。因此，基於以上的概念，如果對於感測元件沒有特殊要求，又希望能夠在較低限制的環境下，簡單的達到輔助定位的目的，一個盲目的搜尋方式(blind search)，理論上比較能夠滿足這樣的需求。

而所謂的盲目的搜尋，根據之前的描述，也就是對於影像，不做分析抽離影像內容物，而透過統計或是資料分析等方式，來選擇描述該影像適當的特徵空間，把影像資料，透過映射函數，統計模型.....方式，轉換到所選定的特徵空間。而這樣特徵空間的抽取，當需要面對前處理這個問題時後，對於外在環境認知要求比較低，通常都是比較一般的處理方式，如：通過低通濾波器，來移除些許隨機雜訊；或是轉換彩色空間，從一般的紅、綠、藍彩色空間(RGB Color space)轉換到色調、飽和度、亮度彩色空間(HIS Color space).....，而這類的前處理，對於環境的熟悉度不需要太高，不論任何環境都適用這類的處理方式，因此，對於低限制的輔助工具，以這類

的比對方式，為本論文所探討的主題。

盲目的搜尋比對方式，在尋找影像相似的過程中，可以視為兩大類，轉換至特徵空間的，則在特徵空間中，尋找足夠表達影像的特徵向量，利用彼此特徵向量的差異性，來量測影像的相似度；而另外一類，則是選取統計模型等模型型態作為特徵空間，來表達影像資訊。而比較模型的差異，則透過資料統計的比較方式，譬如：Kullback-Leibler distance，最大可能性(Maximum likelihood) .....的方式來判別影像彼此間的差異、或是相似度。

此後數各章節，將會介紹一些常見的特徵空間或是統計模型，並探討這些特徵空間，比較影像相似度的方式，而且討論其優缺點，並且對於該類特徵空間，探討未來可以衍生的附加優勢，來選取一個適度的特徵空間，作為本篇論文核心探討的特徵空間。



## 2.3 低通濾波器(Lowpass filter)

不論數位訊號處理，或是自二十世紀開始受人注意的影像處理[20]，低通濾波器(low pass filter)，一直以來都是最基本也最常見的濾波器，而此濾波器最初是用來減低資料量，與減低影像中的雜訊。然而，他可以公平的減少影像內每個區域的資料量，透過低通濾波器的轉換，我們可以將所得到影像，視為一個新的特徵空間，而這個特徵空間，和原本影像的特徵空間基底相同，所以是映射在同一個基底的轉換函式。

而要實現低通濾波器這個轉換函式，可以分兩個角度來探討，時域及頻域：

◎時域：即空間濾波的一類，一個線性系統的轉移函數和沖激響應做螺旋積分。先選擇一個適當的空間遮罩，此遮罩內所有的係數為正數，常見的如圖 2 所示，接著，將此遮罩對影像做螺旋積分，所得到的影像，即

是通過低通濾波器的影像，簡而言之，在時域做低通濾波器，就是對影像像素平均，而平均後的每一個像素值，可以視為一種特徵空間的特徵向量，來表示此影像的資訊。

◎頻域：相對於時域技巧，根據數位訊號處理的基礎，在線性系統的環境下，即是對轉移函數和影像函數分別做二維傅立葉轉換(Fourier Transform)，並將此轉換後的兩個函式相乘，其結果再透過反傅立葉轉換，即為平滑影像。

當然，不論透過何種方式，皆可以轉換投影影像資料到所設計的特徵空間，而在獲得表示影像的特徵向量之後，要比對兩張影像的相似度，最直接的方式，就是透過影像相減，而獲得的差異量，即可以表示兩張影像的差異。接著，當低通濾波器的頻寬越來越寬，也就表示可以通過保留的資訊越多，當頻寬大到全部的影像資料可以保留的時候，即是做螺旋積分的遮罩為 1，即不失真的完全保留整張影像資訊(包括雜訊)，此時，所需紀錄的地標影像參數(影像本身)會很大，但是對於資料卻是最完整的保留，而這本身是一個取捨的問題。

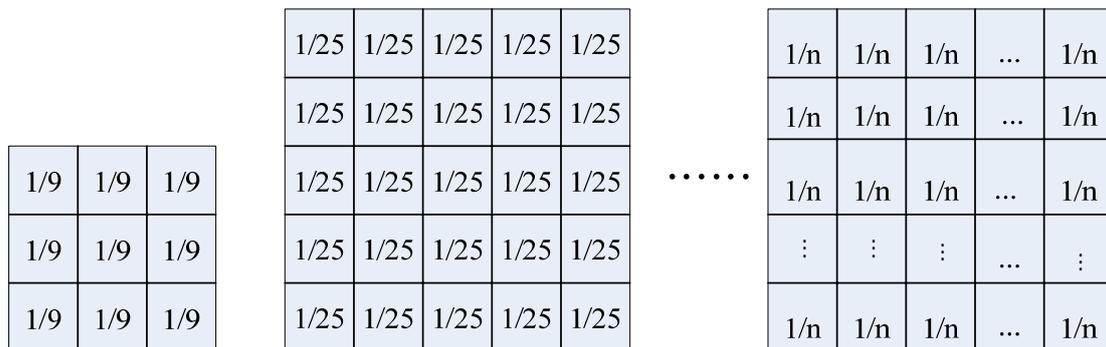


圖 2 低通濾波器可選擇的不同尺寸的遮罩

## 2.4 擴展直方統計圖(Expand Histogram)

影像直方圖，一直都是表達影像資訊分布的主要方法之一，而利用直方圖的值域大小做為影像資訊分布的特徵向量，亦可表達影像中的重要訊息，其表示方式如下所示：

$$(b_{\#}^1, b_{\#}^2, \dots, b_{\#}^i, b_{\#}^{i+1}, \dots, b_{\#}^{2i}, b_{\#}^{2i+1}, \dots, b_{\#}^{3i}) \quad (2.1)$$

其中  $b_{\#}^i$  表示第  $i$  個顏色範圍的所有各數，而以彩色影像而言，就分別有 RGB 三元色。但是，試想一下影像內容中，要是有一個球，自左而右，依序出現在一連串的影像序列中，若是不考慮攝影機白平衡的漂動、測光受到影響、光線受到移動物體的反射、折射、攝影機像素漂動等等問題，在一個極度理想的狀態下，可以輕易的認知到，在這一連串的影像序列中，影像直方圖是一模一樣的，為了解決此類的問題，在[21]中提到，加入直方圖的每個統計範圍的中心座標，將表達的特徵向量修改如下：

$$(b_x^1, b_y^1, b_x^2, b_y^2, \dots, b_x^{3i}, b_y^{3i}) \quad (2.2)$$

其中  $(b_x^i, b_y^i)$  是經過傳統的  $b_{\#}^i$  改良，表示第  $i$  個顏色範圍的所有各數位置的平均，稱為擴展直方圖(Histogram Expand)，便可以輕鬆解決前述的問題，因此，此方法，在連續影像的相似度判別上，也可見到這類方法的蹤跡，因此，如(2.2)所示，即可以求得該特徵空間的特徵向量，用此特徵向量描述該影像資訊。

當取得影像的特徵向量之後，就可以藉由高維度空間中，兩向量的相似度，作為影像的相似度；而兩相量的相似度判別有：最小平方法(least square)、歐基理德距離(two norm)、向量元素的差值合 (one norm)、最大值差異(infinite norm)，向量長度和向量角度.....方法不勝枚舉，而以歐基里德以距離為例，如下圖 3：影像差異越大，特徵向量距離越遠。

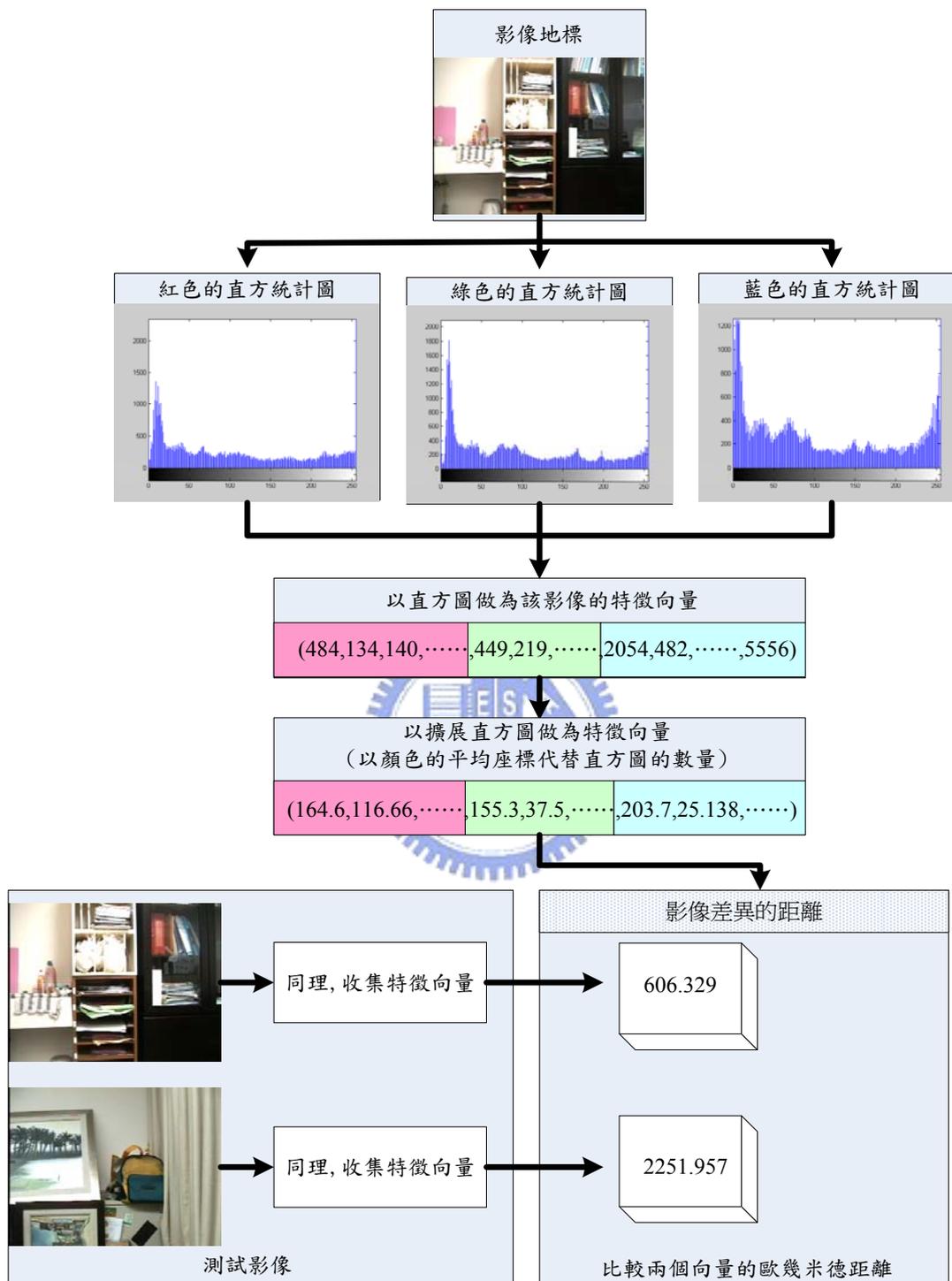


圖 3 利用擴展直方圖計算影像間的差異

## 2.5 主軸分析(Principle Component Analysis)

若將影像視為一份高維度的資料，那對於高維度的資料直接做處理，除了運算比較繁瑣以外，還有可能對雜訊的部份做分析，反而得到負面的影響，因此，降低資料維度，是解決此問題的最直覺的方式，而投影到的低維度空間必須滿足：1.資料各個群數的平均值要夠遠；2.資料各個群數的變異量要夠小。

主要分量分析(Principle Component Analysis)就是在尋求一個最佳的低維度空間，讓降低維度的資料，能夠保留最多重要的訊息，他的概念就是先將影像資料，透過線性代數的概念，投影到特徵空間中，接著利用特徵向量，與特徵值所表示該特徵向量上的分量，來選取適當的特徵向量，做為低維度的投影空間。

有一組  $N$  張影像資料，而每一張影像有  $d$  個像素值，可以將每張影像拉成一維的列向量，並將  $N$  張影像資料排列成：

$$A = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_d^1 \\ x_1^2 & x_2^2 & \cdots & x_d^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^N & x_2^N & \cdots & x_d^N \end{bmatrix} \quad (2.3)$$

其中  $x_j^i$  表示第  $i$  張影像的第  $j$  的像素值。由於影像的張數，常常不等於每張影像的像素值的個數，所以矩陣  $A$  也往往不是方陣，所以，要抽取矩陣  $A$  的特徵向量，就利用奇異值分解法(singular value decomposition)：

$$A_{N \times d} = U_{N \times N} D_{N \times d} V_{d \times d}^T \quad (2.4)$$

簡單來說就是求  $A^T A$  的特徵值與特徵向量，然後將特徵向量依照所對應的特徵值由大到小排列，然後選取，要保留的維度，將其特徵向量依序排列，即為轉換函數，由(2.4)式可整理成：

$$Z_{N \times q} = A_{N \times d} V_{d \times q} = U_{N \times N} D_{N \times q} \quad (2.5)$$

而其中  $q$  是降低維度後剩下的維度大小，所以， $q \leq d$ ，而  $Z$  即為降低維度後的資料。

而此後的每張測試影像都先轉換成  $a_{1 \times d}$  的向量，再透過映射矩陣  $V_{d \times q}$ ，轉換成特徵空間中的一個向量，再利用前一節介紹的比對向量差異的種種方式來比較影像向量的差異，而圖 4 則是為單獨與合併運算 PCA 後的模糊影響。

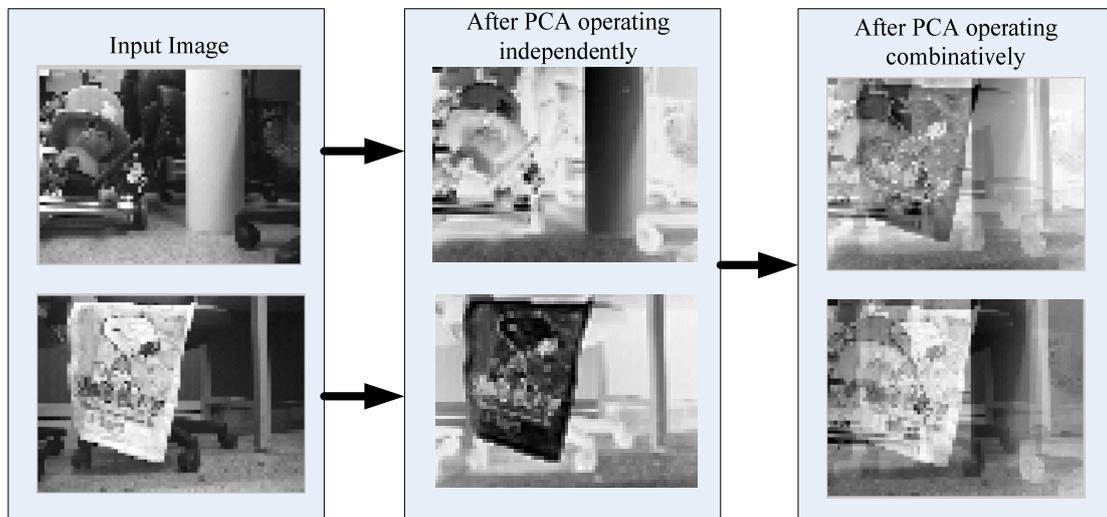


圖 4 透過 PCA 運算所得到的模糊影像，中間為各自計算特徵向量，後側為合併計算

## 2.6 統計模型(Statistical Model)

在圖形識別、影像認知等領域中，對於影像中的顏色分部及其所在位置，以統計的數學分布來代表該影像，稱為統計模型，在這類模型中，我們又可以將其分為有參數模型與無參數模型兩類：

### (1)有參數模型(Parametric Model)

根據所收到的影像資訊，以及其估測的參數，使用「特定」的統計分佈來近似其真正的分布，主要有：

- 單一高斯模型(Single Gaussian Model)

這是最基本的統計模型，但是其自由度太低，可調變參數太少，不足以表達影像資料的分布，所以有必要將它增加到多個高斯分佈。

●高斯混合模型(Mixture Gaussian Model-GMM)

組合多個模型，來滿足空間中資料分布的趨勢，如此一來可以大大的提升自由度，而隨著越多群的高斯模型，也可以越來越滿足影像資料的分布，將足夠來表達空間中的影像資訊，而且可以只需紀錄幾個參數，便可以輕鬆的紀錄該點的位置資訊，而他的效果，如下圖 5 所示：

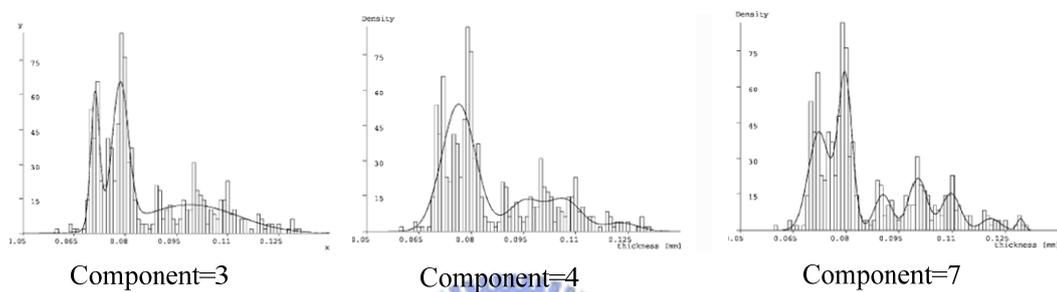


圖 5 不同群數去模擬資料分布的情形

(2)無參數模型(Non-Parametric Model)

不需要任何的理論假設，直接去預測它的密度函數去做近似真正分佈的動作，就是一種無參數的資料分布估測的方法。它們利用無參數核心密度估測(nonparametric kernel density estimation)的方法，簡稱 KDE，他們對於影像中每個像素的位置都給予一個核心函數。而此種 KDE 的方法，其簡單來說，就是針對每一個位置，利用同一張影像中某像素位置的顏色資訊，以及一個具有頻寬係數  $\sigma$  的核心函數  $K_\sigma$  (例：Gaussian、Uniform、Triangle...等)，來算出此像素現在影像資訊的機率值，其數學方程式為：

$$P_r(x_t) = \frac{1}{N} \sum_{i=1}^N K_\sigma(x_t - x_i) \quad (2.6)$$

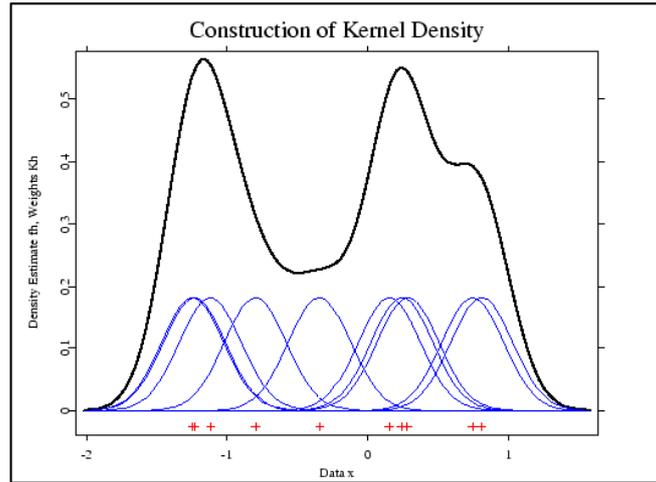


圖 6 利用 KDE 進四真實資料的統計分佈

圖 6 是在每一個資料點(紅色+)為中心，給予一個高斯的核心函數，並做加和及取權重，所得到的機率分佈函式。核心函式(Kernel Function)的種類有很多，下面表 1 中列出常見的種類及其數學表示式。

表 1 無參數核心密度估測 KDE 之核心

Kernel Name	$K(u)$
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$
Uniform	$\frac{1}{2} I( u  \leq 1)$
Triangle	$(1- u ) I( u  \leq 1)$
Epanechnikov	$\frac{3}{4} (1-u^2) I( u  \leq 1)$
Quartic	$\frac{15}{16} (1-u^2)^2 I( u  \leq 1)$
Triweight	$\frac{35}{32} (1-u^2)^3 I( u  \leq 1)$
Co sinus	$\frac{\pi}{4} \cos(\frac{\pi}{2}u) I( u  \leq 1)$

值得注意的是，在核心函數  $K_{\sigma}$  裡的頻寬係數  $\sigma$ ，若其越小，則會造成估測的機率曲線過於鋸齒化，加入了很多高頻的雜訊；反之，若其越大，

則會造成估測的機率曲線太過平滑，而忽略了資料分佈中較精細的部分，所以如何選取一個適當的頻寬係數  $\sigma$ ，使估測的機率曲線最接近實際的情況，是 KDE 中重要的一項工作，依序如下圖 7 所示：

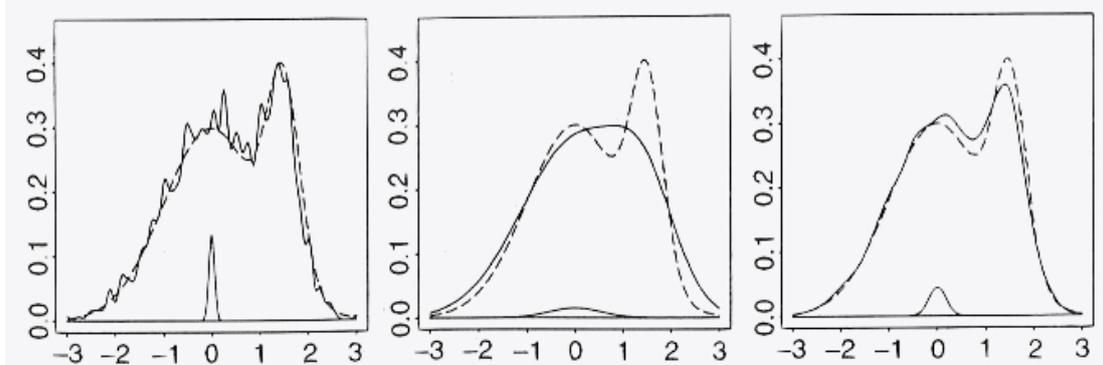


圖 7 頻寬係數對於 KDE 的影響

簡而言之，參數模型是利用多群高斯核心去模擬資料的分布，而無參數模型則是在影像的每個像素點上面放上一個核心函數，不斷的累加這些核心函數，使其分布類似影像資料的分布。當影像資料透過各種方式模型化後，比對兩個模型間的差異性或是比對未知位置的影像資料與模型的符合度的機制，在統計學中不勝枚舉，而我們就針對上述的兩個情形，各舉一個我們所採用的機制來說明。

◆當我們透過之後章節會介紹的 EM 演算法獲得統計分布的影像資訊參數後，如何判別兩個影像間的差異性，對我們來說是個刻不容緩的問題，在[22]中變介紹了最常見的 Kullback-Leibler distance，此量測方式，一開始的定義為

$$D(p_1 \parallel p_0) \approx \sum_{i=1}^L \{p_1(x_i) \cdot \log(\frac{p_1(x_i)}{p_0(x_i)})\} \quad (2.7) \text{其}$$

中， $p_i(x_i)$  表示機率模型中某點  $x_i$  的機率值，而影像資訊量大小共有  $L$  個像素點，而累加這些機率值的交互作用，做為模型 1 與模型 2 的差異；然而，可以清楚的看到這樣計算的結果，模型 1 與模型 2 的差異將不會等於模型 2 與模型 1 的差異，這樣的結果比較不如我們所預期，所以，經過不斷的

改進，最後 KL distance 所採取的公式如下：

$$D(p_1 \parallel p_0) \approx \sum_{i=1}^L \left( p_1(x_i) \cdot \log \left( \frac{p_1(x_i)}{m(x_i)} \right) + p_0(x_i) \cdot \log \left( \frac{p_0(x_i)}{m(x_i)} \right) \right) \quad (2.8)$$

透過上式的計算，便可以修正之前較不合理的部分，也可以透過此公式，將計算出兩個模型間的差異。而其中  $m(x_i) = \frac{p_1(x_i) + p_0(x_i)}{2}$  為兩個模型同一個位置的機率值的平均。

◆另外一個問題就是當影像地標已經建立好統計模型之後，當測試影像需要去測試與統計模型的相似度，如果採用再建立測試影像的統計模型再去利用 kl 距離比對，往往需要大量的運算時間，所以，我們採用將測試影像的資訊套入影像地標的統計模型中，如果符合模型分布，將會得到一個較高的機率值，如果與模型分布有些微的偏差，則會得到較低的機率值，並把這些機率值透過  $\log$  函數調整，在累加起來，避免值域太小，無法表達，而這樣的機制稱為最大期望值(Maximum likelihood)，而一般我們通常偏好正值，所以會再取絕對值，公式如下：

$$ML = \left\| \sum_{i=1}^L \log(p(x_i)) \right\| \quad (2.9)$$

## 2.7 利用影像相似度來進行影像認知與定位

第一、透過之前所介紹的影像相似的準則與量測方式，我們便可以利用這些方法，再利用盲目比對(blind search)，對於影像地標，進行認知的測試；因此，在室內環境中，當我們對於所感興趣的地點，擷取影像，利用此影像建立影像地標來取代特定的人工地標，然後利用上述的各種方法，將影像地標中的特徵空間挑選出來，接下來當再回到該空間附近，可以再次透過擷取目前位置的影像資訊，透過轉換矩陣或其他轉換方式，將目前的影像資訊轉換至特徵空間中所表示的特徵向量，比較各向量間的相似度，或是考慮模型間彼此相似的程度(如：KL distance)，亦或是考慮新

進資料與模型的適當度(如：ML value)，來判別未知位置的影像與各個影像地標的相似度，依此相似度來歸類未知影像應該屬於某個影像地標所涵蓋的區域內。

第二、當一個未知的影像透過之前的比對，可以將其歸屬於某個影像地標附近的區域，但是，我們希冀可以更加的精準的回到當初建立影像地標的位置。因此，我們將觀察當空間中偏離影像地標拍攝點越遠，其特徵空間中，待測影像與地標影像的差異幾乎越大，這樣單調(monotonic)的大體趨勢，將被我們使用來輔佐於影像定位，我們接下來會在之後的實驗，分析討論這單調的趨勢，我們將透過實驗嘗試解釋與分析單調趨勢的特色。



## 第三章 高斯混合模型建立資料庫

### 3.1 高斯混合模型(Gaussian Mixture Model)

高斯混合模型(GMM)，通常被使用與語音辨識系統與圖形識別之中，亦是影像分割(image segmentation)中主要方法之一，經過高斯混合模型處理的影像，會有兩個特色：1.影像中，太過細微的影像資訊會被移除；2.影像中，重要的輪廓，會更加的清楚。而這兩個效果，可以將場景概念化；所謂的概念化即是，對於影像內容中大區塊的漸層效果，以及太過細節變化的小區塊，忽略不紀錄，然而，反觀當人類忽略這些變化，並不會因此在空間中迷路，因此，利用高斯混合模型，來當影像資料的統計模型，可以預期利用該資訊來輔助定位，有著一定效果，此外，高斯混合模型所附帶的效果，會降低所需要紀錄的量值非常少，可以減少儲存空間。

#### 3.1.1 高斯混合模型簡介

高斯混合模型具備參數型機率密度函數之特性，但是並沒有被限制於機率分佈的種類，換句話說，它具備參數定義機率密度函數且無論任何機率密度函數都可適用；另一方面，高斯混合模型也擁有非參數型機率密度函數之特點，可以自由符合資料型態之任意機率密度[23]。

然而，當擷取到一張影像資訊時，首先面臨是該影像的訊息如何表示，先不論其他彩色空間，單以 RGB 彩色空間而言，若以對角協方差矩陣的高斯混合模型去描述三維的影像資訊，這樣等同於是把影像直方圖透過高斯混合模型去表達，當面臨同一個色塊，卻出現在不同位置這問題時，這樣的作法，不足以表達其影像的差異程度，所以色塊的位置資訊也不容小覷，因此，本篇論文採用每個像素的 RGBXY 的五維空間中向量來

描述單一影像的資訊，如圖 8。

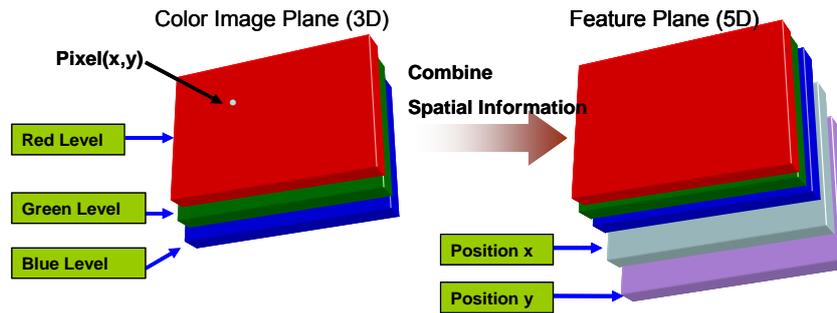


圖 8 擷取五維的特徵向量

再以數學的觀點來看，對任一具有多類別的樣本(Pattern)而言，高斯混合模型具有極佳的近似能力，與傳統的單一高斯分佈(Single Gaussian Mixture)及向量量化(Vector Quantization)兩種模型比較，單一高斯分佈模型，僅能用一個平均值向量來代表一堆樣本在向量空間的中心位置，用共變異矩陣來近似這些樣本在向量空間中所分佈的形狀，其效果當然不好。而向量量化的模型，是用幾個重要的位置來代表整個向量空間，但模型本身並沒有把這些樣本在空間中的分佈大小、形狀描述出來，因此此種方法也不理想。而高斯混合模型使用多個高斯來代表特徵向量的分佈，以數學的觀點來看，它不但精準地紀錄樣本的各种類別、在向量空間中的位置，也能描述出這些類別在空間中的大小及形狀，因此，高斯混合模型適合描述特徵向量在顏色空間的分佈。

### 3.1.2 模型的描述

高斯混合模型，顧名思義，則是由一組權重大小不同的高斯曲線，來描繪、近似資料分布，而高斯的公式如下[24]：

$$g_i(x) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)} \quad (3.1)$$

其中， $g_i(x)$  是資料點  $x$  輸入至第  $i$  群所得到的機率值， $D$  為所輸入  $x$  的資料點的維度， $\mu_i$  和  $\Sigma_i$  是高斯的平均值和共變異矩陣或稱為共變異矩陣或是協

方差矩陣(Covariance Matrix)；其中， $x$  和  $\mu_i$  的維度為  $D$ ，而  $\Sigma_i$  的維度為  $D \times D$ 。

而一個高斯混合模型還需要另外一個權重的參數，因此具有三個參數，分別是混合加權值(mixture weights)、平均值向量(mean vector)以及共變異矩陣(covariance matrix)，將這些參數集合起來並賦予新的符號，如下所示：

$$\lambda = \{w_i, \mu_i, \Sigma_i\}, i = 1, 2, \dots, M \quad (3.2)$$

其中  $w_i$  表示混合加權值， $\mu_i$  表示平均向量， $\Sigma_i$  表示共變異矩陣，而  $M$  則是高斯分佈的個數，對每一個影像像素而言，都可以用  $\lambda$  來表示像素的模型。若我們的資料  $X_N = \{X_1, X_2, \dots, X_n\}$  在  $D$  維空間中分佈，其高斯混合模型的相似度表示如下：

$$p(x_N | \lambda) = \sum_{i=1}^M w_i g_i(x_N) \quad (3.3)$$

而  $g_i(x_N)$  如上式(3.2)所示，且混合加權值也必須  $w_i$  滿足  $\sum_{i=1}^M w_i = 1$  的條件，而其架構圖如下圖 9 所示：

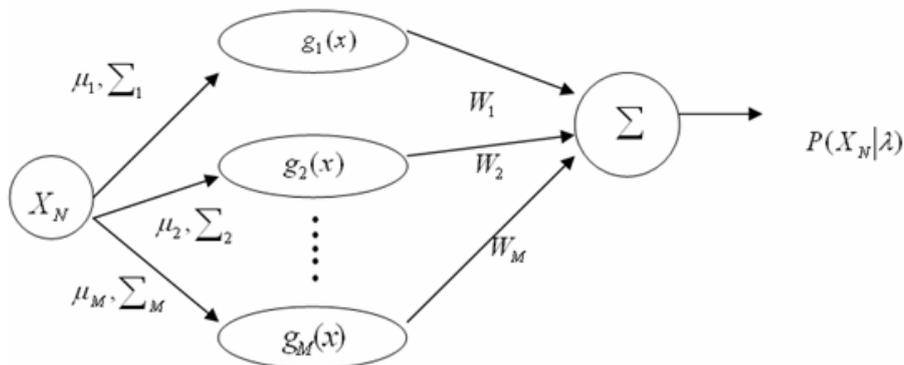


圖 9 高斯混合模型之架構圖

此外，從圖 10 又可以觀察出，左圖是某筆資料的分布情況，可以看出來大約使用三群高斯可以模擬他的分布情況，而右圖則是近似後的曲線

圖。

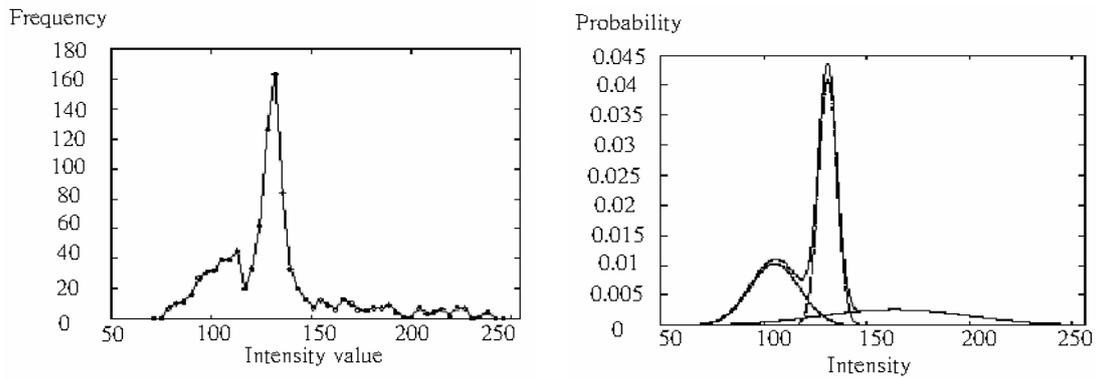


圖 10 強度統計分布及高斯混合模型近似的分布

然而高斯混合模型，除了群數可以選擇以外，對於共變異矩陣 $\Sigma$ ，我們亦有選擇的空間，主要分成三種，實矩陣(Full Matrix)、對角矩陣(Diagonal Matrix)、單位矩陣(Identity Matrix)，外觀如下圖 11，分別對其行為討論如下：

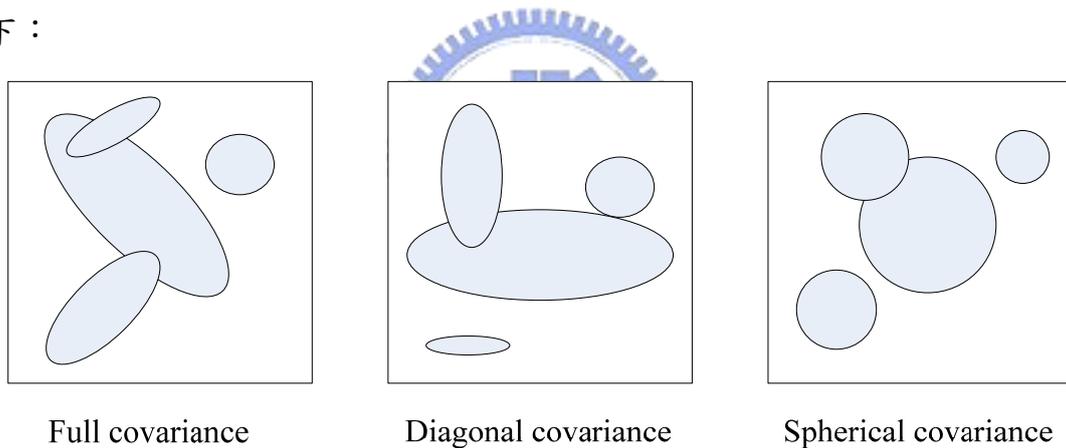


圖 11 不同共變矩陣的外觀之示意圖

◎實矩陣(Full Matrix)：所產生出來的共變異實矩陣(Full covariance)，因為是變異矩陣，所以必為對稱矩陣，在空間中，會呈現成斜橢圓球，除了主軸的變異量分布外，主軸亦可以傾斜，有極高的自由度；它也是這三類共變異矩陣中最可以充分描述資料分布的詳細情況，尤其是高維度空間中，資料分布比較廣，比較散亂，變動亦比較劇烈，一個自由度較大的共變異矩陣，自然而然的可以呈現的比較好。然而，由於自由度大，所以相對的會面臨到大量的運算，極花時間。

◎對角矩陣(Diagonal Matrix)：所產生的共變異對角矩陣(Diagonal covariance)，因為本身就只剩下對角元素，所以本來就是對稱矩陣，而在空間中呈現正橢圓球，僅剩下主軸上有變異量，主軸本身不會傾斜，各個元素彼此沒有關聯性；他運算量介於實矩陣與單位矩陣，他保留了每個元素自己的伸展空間，但是卻彼此不會影響對方，以運算量的角度而言，當計算到共變異矩陣的逆矩陣 $\Sigma^{-1}$ 時候，所需要的運算量，是遠比實矩陣來的簡易很多，速度也提升很多。

◎單位矩陣(Identity Matrix)：所對應的共變異單位矩陣，亦可稱為共變異球矩陣(Spherical covariance)，跟對角矩陣道理相通，必為對稱矩陣；且觀其名而知其意，在空間中分布，的共變異球矩陣，分布為球狀，不只僅剩下主軸上面允許有變異，且各主軸間的變異量亦限制相同；他的運算量是最簡易的，但是和對角矩陣而言相比較，每次計算高斯值只減少  $D-1$  個除法，增快不多，而要能夠表達出資料分布，在相同條件下，要達到相同的描繪能力，需要更高的群數，甚至，有時候再高的群數都很難去適度的描述資料分布。

而其不同共變異矩陣所得到的參數分布如下圖 12

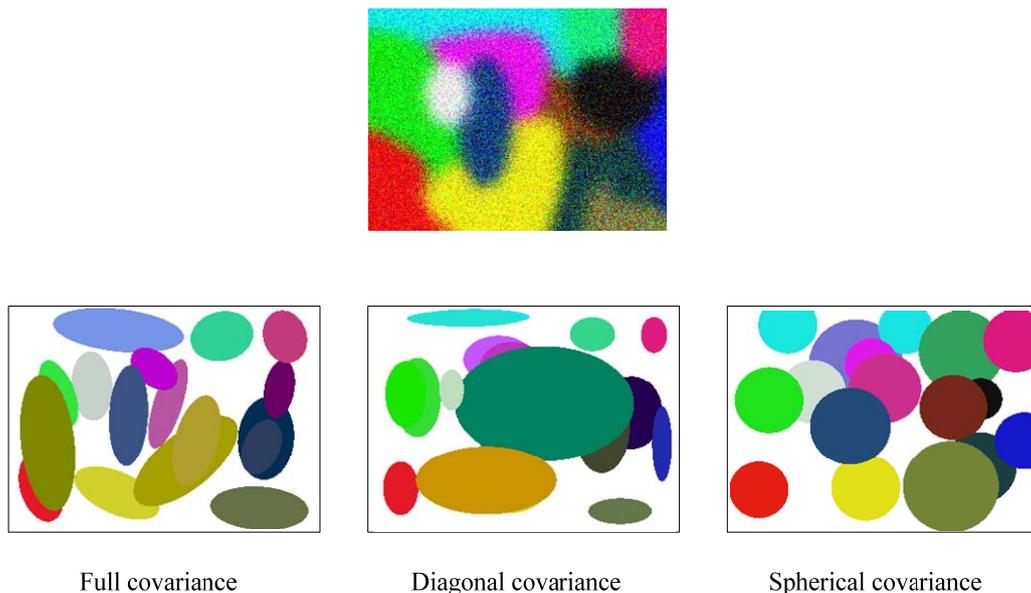


圖 12 真實的將色塊影像透過不同的共變矩陣所得到的結果

### 3.1.3 模型的選擇

如果希望辨識率或是之後的輔助定位能夠更加精確，理論上應該是要選取實對稱矩陣，但是基於下面兩個考量，所以並不選擇實矩陣：

第一、實對稱矩陣除了轉置矩陣與原本矩陣相同，並沒有其他的特色，而且由於它可以任意變形，所以可以去滿足影像地標的任何一個區塊，有時候如果群數不夠多，反而費力去描述細節的變化，反而忽略主要色塊。

第二、實對稱矩陣在計算反矩陣時，其計算量遠大於只有對角元素有值的矩陣，運算量太大。

第三、當實矩陣在計算反矩陣時，如果遇到絕對值(determine)接近或是等於 0 則整個運算就會發散，而對角矩陣如果遇到類似的問題，我們將在 3.3 節討論，而實矩陣如果要仿效對角矩陣，則必須還要將整個資訊投影到共變異矩陣的特徵空間，其繁瑣程度，不言而喻，所以採用對角矩陣。

而不選取單位矩陣做為共變異矩陣，則是因為選擇五維空間的影像資訊做為考量，所以共變異矩陣在影響這個兩個維度的變化的數量級幾乎可以是差不多，但是其像素值則會有很大的差異，而如果強迫其變異量必須一致，則模型的準確度則大打折扣，但是在簡單比對的條件下，可以考慮單位矩陣，而如果需要更精細的比對，或是資料分布描繪需要很清楚的情況下，將採用對角矩陣。而下圖 13 為一場景透過對角矩陣模糊化。

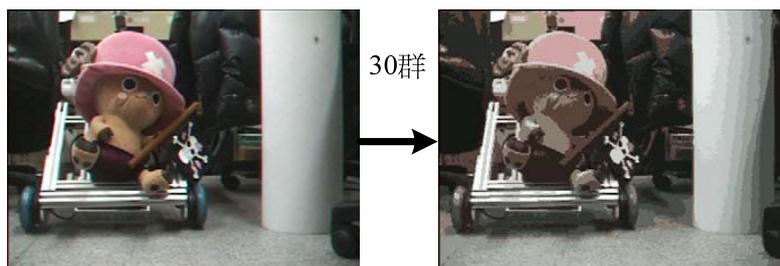


圖 13 對角矩陣透過其分布將影像模糊化

### 3.1.4 模型的參數初始化

我們如果希望快速地、精確地找出高斯混合模型的最佳參數，使得系統有最佳的表現，則在尋找最佳參數之前我們必須對參數做初使化的動作；在[25]中也提到，要加速提高 EM 演算法的效率，適當的初始化是非常重要的。向量量化(VQ)是一項運用非常廣泛的技術，它能將一堆特徵向量的資料，濃縮成幾個具代表性的類別(class)或群集(cluster)，所以這裡我們先採用 VQ 的技術，將我們得到的影像像素值，做初步的分群，得到高斯混合模型參數的初始化值(包括群的個數、群的中心)，以利於後面做參數的最佳化。VQ 的方法有很多種，我們採用 K 平均值分類法(K-means Cluster)，其流程如圖 14 所示，詳細的步驟說明如下：

#### 0、收集資料：

經過一段時間的收集，獲得  $N$  個欲做訓練的特徵向量。

#### 1、初始化：

假設希望的高斯群數是  $K$ ，並隨機地取  $K$  個向量當成每群的中心點。

#### 2、以新的群中心來分群：

其他( $N-K$ )個向量對這  $K$  個群中心做距離測量，以距離做為分群的依據，每個向量被分類到距離最短的中心。

#### 3、更新群中心：

接著對每一群算出新的向量平均值，以此做為新的群中心。

#### 4、判斷分群是否收斂：

將新的群中心與舊的群中心作比較，如果不再有變動，表示已收斂，則做步驟 5；反之，則重複步驟 2、3。

#### 5、得到初始化參數：

將最後分群的個數、群的中心、群的變異數以及每一群的資料個數當



作高斯混合模型的初始參數( $M$ 、 $\mu$ 、 $\Sigma$ 、 $w$ )。

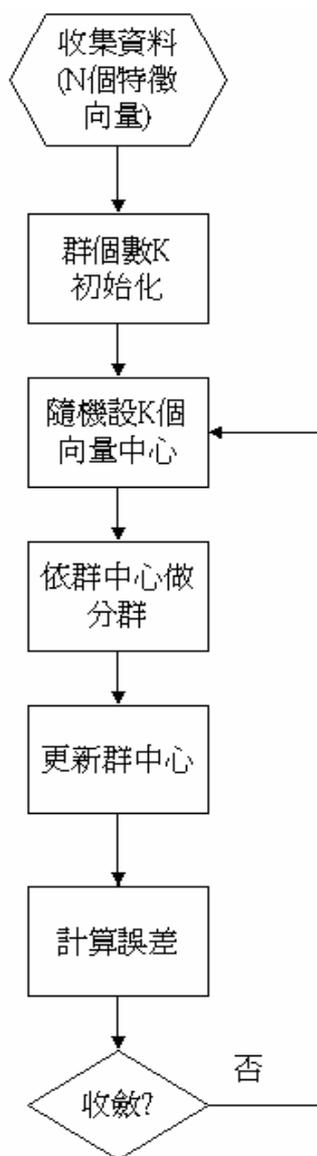


圖 14 平均值分類法(K-means Cluster)流程圖

### 3.2 最佳相似性估測演算法(Expectation Maximization)

我們在做資料模型訓練時，最終的目的是估測最佳的高斯混合模型參數 $\lambda$ ，所謂的『最佳』指的是，影像資訊的五維空間的真正的分佈，與模

型參數  $\lambda$  估測出來的分佈有最大的相似度，估測最佳參數的方法有很多，但最受歡迎、最適合的方法是『最佳相似性估測法』(Maximum Likelihood Estimation, MLE)。

在 3.1.2 節高斯密度函數的假設下，當  $x = x_i$  時，其機率密度為  $P(x_i|\lambda)$ ，如果  $x_i, i=1\sim n$  之間是互相獨立的事件，則發生  $X = \{x_1, x_2, \dots, x_n\}$  的機率密度之相似函數(likelihood function)可以表示成：

$$P(X|\lambda) = \prod_{i=1}^n P(x_i|\lambda) \quad (3.4)$$

由於  $X$  是確定的，因此 MLE 主要就是找出使得高斯混合模型的相似函數值為最大時的參數  $\lambda'$ ，也就是  $\lambda' = \arg \max_{\lambda} P(X|\lambda)$ ，但是(2.4)式對  $\lambda$  而言是一個非線性的方程式，無法直接最大化相似函數，所以我們採用期望值最大演算法(Expectation Maximization Algorithm)，利用疊代的方式找出 MLE 的估測參數  $\lambda'$ 。

EM 演算法的基本做法是先由之前 K 平均值分類法找出的初始化參數  $\lambda$ ，再利用 EM 估計出新的參數  $\bar{\lambda}$ ，使得滿足  $P(X|\bar{\lambda}) \geq P(X|\lambda)$ ，令  $\lambda = \bar{\lambda}$  重新疊代估計新的  $\bar{\lambda}$ ，直到  $P(X|\lambda)$  收斂或是達到某個門檻值才停止。EM 演算法主要分成 2 個部分，與 likelihood 函數有關的 Expectation Step (E-Step)，以及更新參數方程式的 Maximization Step (M-Step)。

### 3.2.1 期望值(Expectation Step)

目的是測試我們所求的 likelihood 函數值，是否達到我們的要求，若符合要求，EM 演算法就停止，反之就繼續執行 EM 演算法。這裡為了數學推導的方便，假設我們的模型是由三個高斯分佈函數所構成，則其密度函數可表示成：

$$P(x) = w_1 g(x; \mu_1, \Sigma_1) + w_2 g(x; \mu_2, \Sigma_2) + w_3 g(x; \mu_3, \Sigma_3) \quad (3.5)$$

其中共變異矩陣部分  $\Sigma_j$ ，為了推導方便而且如果以運算量效能比來考量，選擇對角矩陣比較有利，因此推導就由對角矩陣來考量，所以只剩對角有值， $P(x)$  的參數  $\lambda = [w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \Sigma_1, \Sigma_2, \Sigma_3]$ ，參數個數為  $(1+1+1+d+d+d+d+d+d)=3+6d$  個，依前述 MLE 原則，求出 likelihood 的最大值：

$$\begin{aligned} E(\lambda) &= \ln \left( \prod_{i=1}^n P(x_i) \right) = \sum_{i=1}^n \ln P(x_i) \\ &= \sum_{i=1}^n \ln [w_1 g(x_i; \mu_1, \Sigma_1) + w_2 g(x_i; \mu_2, \Sigma_2) + w_3 g(x_i; \mu_3, \Sigma_3)] \end{aligned} \quad (3.6)$$

為簡化討論，再引進另一個數學符號稱事後機率(post probability):

$$\begin{aligned} \beta_j(x) &= p(j|x) = \frac{p(j \cap x)}{p(x)} = \frac{p(j)p(x|j)}{p(x)} \\ &= \frac{p(j)p(x|j)}{p(1)p(x|1) + p(2)p(x|2) + p(3)p(x|3)} \\ &= \frac{w_j g(x; \mu_j, \Sigma_j)}{w_1 g(x; \mu_1, \Sigma_1) + w_2 g(x; \mu_2, \Sigma_2) + w_3 g(x; \mu_3, \Sigma_3)} \end{aligned} \quad (3.7)$$

### 3.2.2 最大值(Maximization Step)

主要目的是為了要找到使 likelihood 函數最大化的參數，因此我們分別對  $w_i$ 、 $\mu_i$ 、 $\Sigma_i$  做偏微分，再做後續的運算，於是我們便可以得到所求的參數，接著返回 E-Step 繼續做。

假設初始參數是  $\lambda_{old}$ ，我們希望找出新的  $\lambda$  值，滿足  $E(\lambda) > E(\lambda_{old})$ ，因為根據  $\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b)$ ， $E(\lambda) - E(\lambda_{old})$  可以延伸成下式：

$$\begin{aligned}
E(\lambda) - E(\lambda_{old}) &= \sum_{i=1}^n \ln \left[ \frac{w_1 g(x_i; \mu_1, \Sigma_1) + w_2 g(x_i; \mu_2, \Sigma_2) + w_3 g(x_i; \mu_3, \Sigma_3)}{w_{1,old} g(x_i; \mu_{1,old}, \Sigma_{1,old}) + w_{2,old} g(x_i; \mu_{2,old}, \Sigma_{2,old}) + w_{3,old} g(x_i; \mu_{3,old}, \Sigma_{3,old})} \right] \\
&= \sum_{i=1}^n \ln \left[ \frac{w_1 g(x_i; \mu_1, \Sigma_1) \beta_1(x_i)}{D(\lambda_{old}) \beta_1(x_i)} + \frac{w_2 g(x_i; \mu_2, \Sigma_2) \beta_2(x_i)}{D(\lambda_{old}) \beta_2(x_i)} + \frac{w_3 g(x_i; \mu_3, \Sigma_3) \beta_3(x_i)}{D(\lambda_{old}) \beta_3(x_i)} \right] \\
&\geq \sum_{i=1}^n \left[ \beta_1(x_i) \ln \frac{w_1 g(x_i; \mu_1, \Sigma_1)}{D(\lambda_{old}) \beta_1(x_i)} + \beta_2(x_i) \ln \frac{w_2 g(x_i; \mu_2, \Sigma_2)}{D(\lambda_{old}) \beta_2(x_i)} + \beta_3(x_i) \ln \frac{w_3 g(x_i; \mu_3, \Sigma_3)}{D(\lambda_{old}) \beta_3(x_i)} \right] \\
&= Q(\lambda) \tag{3.8}
\end{aligned}$$

上式中，因為  $\ln(x)$  是一個凸函數 (Convex Function)，滿足下列不等式：

$$\ln[\alpha x_1 + (1 - \alpha)x_2] \geq \alpha \ln(x_1) + (1 - \alpha) \ln(x_2) \tag{3.9}$$

推廣上式到「傑森不等式」(Jensen Inequality)：

$$\ln \left( \sum_{i=1}^n \alpha_i x_i \right) \geq \sum_{i=1}^n \alpha_i \ln(x_i), \sum_{i=1}^n \alpha_i = 1 \tag{3.10}$$

因為  $\sum_{j=1}^3 \beta_j(x_i) = 1$ ，所以可以將傑森不等式套用在 3.8 式，最後得到下式：

$$E(\lambda) \geq E(\lambda_{old}) + Q(\lambda) \tag{3.11}$$

只要  $Q(\lambda) > 0$ ，必滿足  $E(\lambda) > E(\lambda_{old})$ ，但我們通常希望  $E(\lambda)$  越大越好，最直接的方式就是找出使得  $Q(\lambda)$  最大的  $\lambda$  值，那  $E(\lambda)$  也會跟著變大，見下圖 15。

$Q(\lambda)$  是  $\lambda$  的函數，將一些與  $\lambda$  不相關的部分併入常數項，並重新整理  $Q(\lambda)$  成下式：

$$\begin{aligned}
Q(\lambda) &= \sum_{i=1}^n \sum_{j=1}^3 \beta_j(x_i) [\ln w_j + \ln g(x_i; \mu_j, \Sigma_j)] + c1 \\
&= \sum_{i=1}^n \sum_{j=1}^3 \beta_j(x_i) \left\{ \ln w_j + \ln \left[ \frac{1}{(2\pi)^{d/2} [\det \Sigma_j]^{1/2}} \exp \left( -\frac{(x_i - \mu_j) \Sigma_j^{-1} (x_i - \mu_j)^T}{2} \right) \right] \right\} + c1 \\
\text{對 } \mu_j \text{ 偏微分, } \partial_{\mu_j} Q = 0 &\Rightarrow \mu_j = \frac{\sum_{i=1}^n \beta_j(x_i) x_i}{\sum_{i=1}^n \beta_j(x_i)} \tag{3.12}
\end{aligned}$$

$$\text{對 } \Sigma_j \text{ 偏微分, } \partial_{\Sigma_j} Q = 0 \Rightarrow \Sigma_j = \frac{\sum_{i=1}^n \beta_j(x_i)(x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^n \beta_j(x_i)} \quad (3.13)$$

欲得到最佳之  $w_j$  值，須將  $w_j$  的總和為 1 的條件加入，引進 Lagrange Multiplier，並定義新的目標函數(object function)為：

$$E_{new}(\lambda) = E(\lambda) + \alpha(w_1 + w_2 + w_3 - 1) \quad (3.14)$$

將  $E_{new}$  對 3 個 weighting 做偏微分，可得到下面 3 個方程式：

$$\frac{\partial E_{new}}{\partial w_j} = -\frac{1}{w_j} \sum_{i=1}^n \beta_j(x_i) + \alpha = 0, j = 1, 2, 3 \quad (3.15)$$

最後將 3.15 的 3 個式子相加，可得到：

$$\begin{aligned} (w_1 + w_2 + w_3)\alpha &= -\sum_{i=1}^n [\beta_1(x_i) + \beta_2(x_i) + \beta_3(x_i)] \\ \Rightarrow \alpha &= -\sum_{i=1}^n 1 = -n \\ \Rightarrow w_j &= \frac{1}{n} \sum_{i=1}^n \beta_j(x_i), j = 1, 2, 3 \end{aligned} \quad (3.16)$$

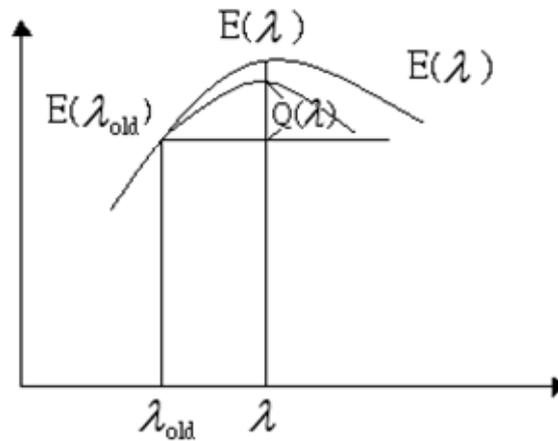


圖 15 Likelihood function  $E(\lambda)$  最大化的示意圖

綜合前面各步驟的說明，GMM 建立的整個流程如圖 16 所示，先將 N 個準備拿來訓練模型的資料點，經過 K-means Clustering 後得到初始的參數，再由 EM 演算法得到的三個方程式，

$$\mu_j = \frac{\sum_{i=1}^n \beta_j(x_i) x_i}{\sum_{i=1}^n \beta_j(x_i)} \quad \Sigma_j = \frac{\sum_{i=1}^n \beta_j(x_i) (x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^n \beta_j(x_i)} \quad w_j = \frac{1}{n} \sum_{i=1}^n \beta_j(x_i)$$

(3.17)

進行參數的更新，並計算新的相似函數的值，如此不斷的疊代，不斷地更新模型的參數，直到相似函數的值已經沒什麼變動，或是疊代的次數超過某個門檻值，才停止疊代。而上述(3.17)式中， $\Sigma_j$ 是通式的寫法，也就是實矩陣的表達模式，但是如果是對角矩陣或單位矩陣，其運算可以簡化成為下式：

$$\sigma_{kj}^2 = \frac{\sum_{i=1}^n \beta_j(x_i) (x_{ki} - \mu_{kj})^2}{\sum_{i=1}^n \beta_j(x_i)}, k=1,2,\dots,D \quad (\text{diagonal covariance})$$

$$\sigma_k^2 = \frac{\sum_{i=1}^n \beta_j(x_i) \|x_i - \mu_j\|^2}{\sum_{i=1}^n \beta_j(x_i)} \quad (\text{spherical covariance})$$

(3.18)

而平均和權重的計算方式，不會受到所選取的共變異矩陣不同而有所不同，所以公式不變。

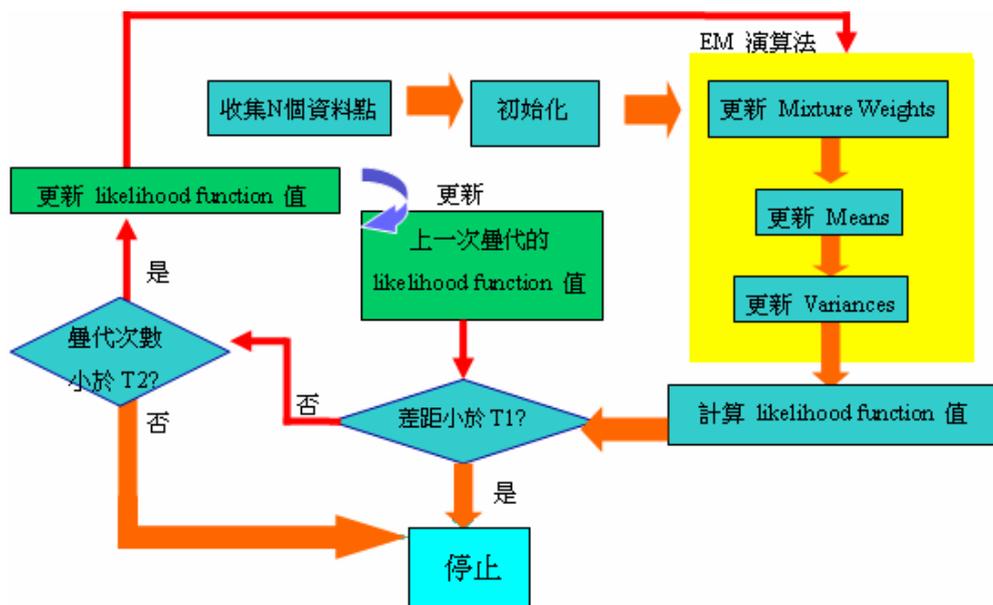


圖 16 高斯混合模型建立的架構

### 3.3 高斯混合模型接近高群數的問題

一般的資料分布，幾乎很難真的是跟高斯分布一模一樣，即使是混合高斯模型，亦難百分百去滿足任何的資料模型分布。然而，先不考慮資料分布中有雜訊以及過度緊貼資料(over-fitting)分布；理論上，我們可以很直覺的意識到越多群數，可以越清楚的描述資料分布的情形，但是，影像資料中，出現色塊的情形，並不算是罕見，也就是當五維資訊中，隨著 K means 的方法所得到的初始值不同，當高斯混合模型中的某一群高斯模型恰巧完全落在該色塊區域上，如此一來，該五維空間中，色彩的部分，變異量極小，甚至是 0 變異量，有些時候，即使顏色是一漸層分布，但是以 RGB 的分布來看，也是可能有一個彩色維度的顏色值域都維持不變，亦有可能受到編譯器(compiler)的影響，當某些值，過小的時候，編譯器已經無法分辨他的值，使得輸出結果為 0；但是不論是哪一種編譯器，都不可能去處理無限小的情形，所以變異量為 0 似乎是不可避免的問題，當遇到這樣的問題，當某次 EM 疊帶的結果(3.18)式中的  $\sigma_{kj}$  為 0 時，下次進入到 E-step 的時候會需要計算高斯的值，需要帶入(3.1)式，其中  $\Sigma^{-1}$  項會因為其 determine 為 0 而使得高斯值發散，進而破壞了整個 EM 的演算法。

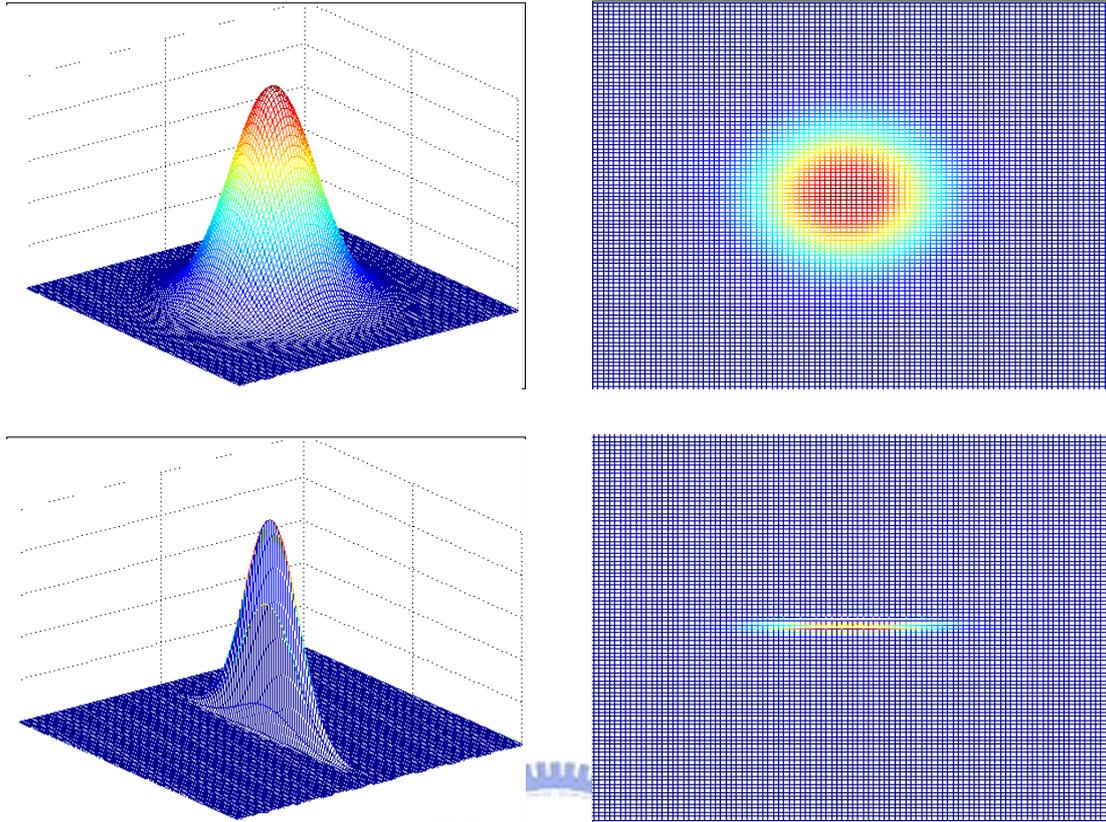


圖 17 一般的高斯分布與某一維度變異量為 0 的情況

而這問題的解決方式，透過多方搜尋的結果，我們有發現部分的解決方法是設立門檻值(threshold)，當判別變異量  $\sigma_{kj}$  小於一個門檻值時，則強制定義該變異量為門檻值，但是一旦設立門檻值，可能面臨兩個問題：第一、門檻值設立太高，無法收斂；第二、即使收斂了，因為每次遇到變異量太小就會強制將變異量改成門檻值，其模型精確度大幅下降，這不是我們所樂見的，所以，回到最基本的定義來探討變異量很小甚至為零的情況：當變異量很小而趨近於 0，表示該維度的分量沒有什麼變化，所以以高斯分布來看，在該維度的貢獻非零即一；我們再假設各維度間是線性獨立的關係(以對角矩陣為例)，可以看出，以圖 17 為例：

再搭配條件機率：

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\therefore P(A \cap B) = P(A|B)P(B)$$

$$P(A \cap B \cap C \cdots \cap N) = P(A)P(B|A)P(C|B \cap A) \cdots P(N|N-1 \cap \cdots \cap C \cap B \cap A) \quad (3.19)$$

且因為是對角矩陣，為獨立矩陣， $P(A|B) = P(A)$  且  $P(B|A) = P(B)$  為線性獨立，所以上式可以整理成：

$$P(A \cap B \cap C \cdots \cap N) = P(A)P(B)P(C) \cdots P(N) \quad (3.20)$$

所以透過條件機率的觀念，只要線性獨立，可以透過單獨計算每個元素(維度)的貢獻，來求取最後的機率值。

因此，為了要得到可以個別運算各為度的機率值，我們勢必對高斯的公式(3.1)稍做整理：

$$g_j(x) = \frac{1}{(2\pi)^{\frac{D}{2}} \left[ \begin{matrix} \sigma_{1j}^2 & 0 & 0 & 0 \\ 0 & \sigma_{2j}^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_{3j}^2 \end{matrix} \right]^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_j)^T \begin{bmatrix} \sigma_{1j}^2 & 0 & 0 & 0 \\ 0 & \sigma_{2j}^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_{3j}^2 \end{bmatrix}^{-1} (x-\mu_j)}$$

$$\Rightarrow g_j = \frac{1}{(2\pi)^{\frac{D}{2}} (\sigma_{1j}^2 \sigma_{2j}^2 \cdots \sigma_{Nj}^2)^{\frac{1}{2}}} e^{-\frac{1}{2} \left[ \frac{(x_1-\mu_{1j})^2}{\sigma_{1j}^2} + \frac{(x_2-\mu_{2j})^2}{\sigma_{2j}^2} + \cdots + \frac{(x_N-\mu_{Nj})^2}{\sigma_{Nj}^2} \right] (x_1-\mu_{1j}, x_2-\mu_{2j}, \dots, x_N-\mu_{Nj})^T}$$

$$\Rightarrow g_j = \frac{1}{(2\pi)^{\frac{D}{2}} (\sigma_{1j}^2 \sigma_{2j}^2 \cdots \sigma_{Nj}^2)^{\frac{1}{2}}} e^{-\frac{1}{2} \left[ \frac{(x_1-\mu_{1j})^2}{\sigma_{1j}^2} + \frac{(x_2-\mu_{2j})^2}{\sigma_{2j}^2} + \cdots + \frac{(x_N-\mu_{Nj})^2}{\sigma_{Nj}^2} \right]}$$

$$\Rightarrow g_j = \frac{1}{(2\pi)^{\frac{1}{2}} \sigma_{1j}} e^{-\frac{1}{2} \frac{(x_1-\mu_{1j})^2}{\sigma_{1j}^2}} \frac{1}{(2\pi)^{\frac{1}{2}} \sigma_{2j}} e^{-\frac{1}{2} \frac{(x_2-\mu_{2j})^2}{\sigma_{2j}^2}} \cdots \frac{1}{(2\pi)^{\frac{1}{2}} \sigma_{Nj}} e^{-\frac{1}{2} \frac{(x_N-\mu_{Nj})^2}{\sigma_{Nj}^2}} \quad (3.21)$$

透過上式的整理，便可以將各維度對於整個機率值的貢獻拆開來，再搭配條件機率的觀念，分別計算每個元素的值再累乘即答案，那當遇到變異量的值為 0 時，就可以獨立處理；而變異量為 0 的高斯函數，從定義著手，其機率值為 1，因此一旦在收斂的過程中，任一變異量  $\sigma_{kj}$  過小(可能導致編

譯器無法辨識)，我們就認為該維度其實已經為 0 變異量，這時候就需要拆開各維度的貢獻，變異量大的維度維持採用高斯的公式計算，而變異量小的維度，就跟著判斷，該影像資訊中被縮減維度的值，與該維度中心的差異是否落在殘存的變異量九倍之內(variance)，若是三倍偏移量之內，則認定該維度貢獻維 1，以外則為 0，如此就可以順利求出整個高斯分布的值，那整個 EM 的演算法就可以繼續疊代，只到我們希望停止的門檻值，而如此所得到個 EM 的混合高斯分布，其效果將比一般的設立變異量的門檻值的方法，更加能夠描述影像資訊的分布。而其計算最大期望值的部份流程圖如下圖 18：

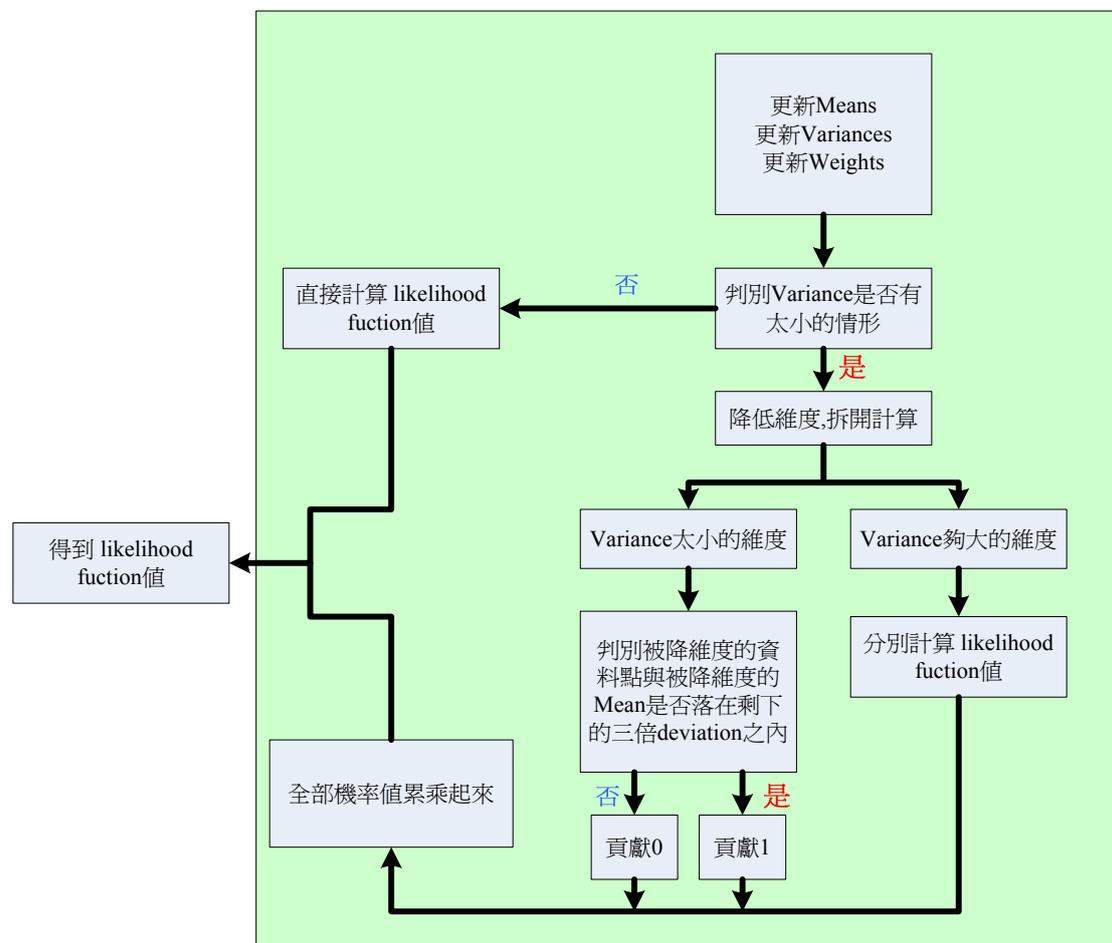


圖 18 在計算機率值時，防止變異量過小之機制的流程圖

## 第四章 資料庫集合與結合演算法

### 4.1 資料庫的建立

當我們希冀結合場景中多角度之影像，來提高辨識率，所以我們勢必在同一個場景拍攝多個角度的影像資訊，當將來再回到該場景附近時，不會因為過度缺乏資訊而造成誤判，但是拍攝過多的影像資訊，反而會造成誤判，或是運算量大增，於是乎，如何選擇適當且適量的不同角度的影像，是建立影像地標資料庫的主要目標。而我們透過物件辨識的概念，來縮減我們的資料量。

#### 4.1.1 Cyr and Kimia 提出的結合演算法

接下來我們就來說明 Cyr and Kimia [26]提出的外觀結合演算法，首先我們說明一些代表符號，以便後面的說明。 $V_m^n$ 指的是第  $n$  個物體的第  $m$  個面， $A_p^n$ 則是第  $n$  個物體的第  $p$  個外觀，它就是一個範圍內面的集合，可用下式來表示它

$$A_p^n = \{V_{m-k^-}^n, \dots, V_{m-1}^n, V_m^n, V_{m+1}^n, \dots, V_{m+k^+}^n\} \quad (4-1)$$

其中  $V_m^n$  是此外觀的特徵面 (characteristic view)， $(k^-, k^+)$  是此外觀的邊界。

Cyr and Kimia 提出的外觀結合演算法是一種基於相似度的區域成長的方法，在這裡它所應用的對象，是對在觀測物體的外側球面上的赤道線上每隔 5 度，按照由小到大排列的視角順序看到的 72 個面 (如下圖 19)

做外觀結合的動作，藉此來得到其適當的特徵面描繪(aspect-graph)

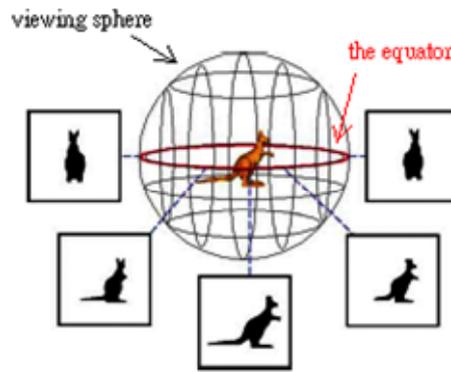


圖 19 在赤道面上取樣物體的 2 維影像

這個方法主要要符合兩個條件，分別敘述如下。

條件 1：對於每個特徵面  $V_m^n$  而言，存在有一個整數  $\delta_0 > 0$ ，使得下式成立。

$$d(V_m^n, V_{m \pm i}^n) < d(V_m^n, V_{m \pm j}^n) \text{ if } i < j \leq \delta_0 \quad (4-2)$$

這式其實就是代表特徵面  $V_m^n$  與其他附近的面所算出的距離會形成一個區域最小值且維持一個單調 (monotonicity) 的特性，而符合此條件的面，就有可能合併成一個外觀。而在此處他又做了一個假設，就是他假設我們所使用來代表物體的 72 個面是取樣足夠的，且沒被取樣到的點，也都一定會符合條件 1。但如果這個假設不成立時，辨識時就會發生錯誤。

條件 2：對於每一個外觀  $A_m^n$ ，其特徵面  $V_m^n$  必須滿足下式

$$\max_{V_i^n \in A_m^n} d(V_i^n, V_m^n) < \min_{V_j^n \notin A_m^n} d(V_j^n, V_m^n) \quad (4-3)$$

也就是說此外觀的特徵面  $V_m^n$  到在此外觀裡的其他面  $V_i^n$  的距離要小於此外觀的特徵面  $V_m^n$  到不在此外觀裡的面  $V_j^n$  的距離。

至於在一個外觀裡要如何選取其代表此外觀的特徵面，則是選取到此外觀裡的每個面的距離最小且符合構成外觀的兩個條件的那個面當做是特徵面。而此特徵面就是將來用來辨識的比較對象，也就是我們要建立的

資料庫。

以上就是此法的一些條件的介紹，但我們可以發現其對於有對稱性質的物體，其就會因條件 2 而產生過多的外觀，所以我們多加了一個門檻值  $threshold\_1$  去限制在條件 2 中，需要被比較大小的不屬於此外觀的面的個數。此法還有另一個問題，就是它在符合條件的情況下所做出的外觀，它們之間的距離（由兩個特徵面的距離來定義）可能會太大，因而導致分的外觀過少，而容易在辨識時產生誤判，所以我們又增加一個門檻值  $threshold\_2$  去限制當兩個外觀之間的距離大於此設定的門檻值時，不管其有無符合兩個條件，我們都不將其合併成一個外觀。

接著，我們將列出如何實際使用外觀結合演算法來做出足夠且適當的觀特徵面描繪(aspect-graph)，以利用其特徵面建立我們要的物體的資料庫。

1. 將由物體照順序（由小到大排列的視角順序）取樣的  $M$  個面，每個面都當做是一個外觀，且每一個外觀的特徵面都是它自己本身。
2. 利用計算相似度的方法，計算每一個相鄰的外觀之間的距離。
3. 選擇在所有相鄰的外觀之間的距離中，擁有最小距離的那一對外觀。假如它們包含的面組成的區域符合上述的判斷條件，則它們可以結合變成一個新的外觀，且此新外觀的特徵面為到此外觀裡的每個面的距離最小的那個面；相反的，若不符合上述的判斷條件，則不合併此二者。
4. 利用計算相似度的方法，再計算形成新的外觀後，每一個相鄰的外觀之間的距離。
5. 重覆步驟 3、4 直到所有相鄰的外觀都不能被合併，我們就得到所要的特徵面描繪(aspect-graph)，也就是資料庫。

### 4.1.2 修改過後的結合演算法

上述的方式是在描述一個物體的資料量縮減方式，但是，我們若想像

一個畫面，即是一個物體，那再這個畫面附近變動的時候，勢必和繞著一個物體選轉拍攝，有異曲同工的效果，如圖 20 所示：

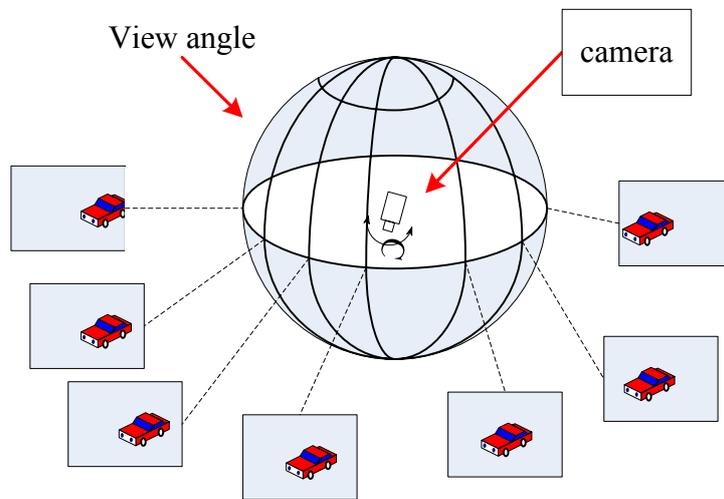


圖 20 利用攝影機選轉模仿多角度觀察物體

所以將上述演算法稍做改良，勢必對我們縮減資料量，有很大的助益。上述的概念是將一連串序列的影像，透過尋求最佳的特徵面，來描述該物體的外觀，然而仍存在幾個缺點

1. 每個物體在建立資料庫時，都必須使用  $M$  張照取樣順序排列的面，即照拍攝角度由小到大排列的視角順序的面，才能代入演算法去建資料庫。
2. 由於此法假設取樣的物體的面是足夠多去表示此物體的，但實際上我們知道一個物體可能投影出的面是無限多的，所以，此假設是會有問題的，因此必須有一個可以更新資料庫的機制，使的資料庫變的更完善。不幸的是，此法若要更新資料庫，則必須把物體全部  $M$  張面加上新進用來更新資料庫的面，重新按照拍攝角度由小到大排列的視角順序排好，全部再重新代入演算法去建資料庫，即無法簡單地一張一張的更新，而是不管有幾張要更新，都要加上之前的全部重算。
3. 若物體具有一些對稱性，雖然我們已對此法多加了一個門檻值來使

此法不致因物體的對稱性的關係而多出一些不必要的外觀，但這樣做還是不能完全避免不必要的外觀，例如像完全對稱的物體就還是會產生不必要的外觀。

因此有[27]提出的改善機制，他保留了 Cyr and Kimia 提出的外觀結合演算法分開建立資料庫的好處。然而，新的外觀結合演算法也是一種基於相似度的方法，所以利用其相似度相近者合併的原理，來輔助我們減低資料量，而且不用考慮到影像資訊所獲得的順序，對於將來資料庫升級(update)來說，有很大的助益，以下，就將其減低物體特徵面描述的機制，改成我們減低場景影像資料的機制：

首先，先定義一些用於下列說明的符號， $V_{new}$ 指的是新進來要建立資料庫的某場景的某個影像， $C_m$ 是此物體資料庫中某場景的第 $m$ 個角度之特徵影像， $C_{m^{min+1}}$ 則是此新進來的影像與所有此場景資料庫中距離最近的那個特徵影像的相鄰左右兩個特徵影像的角度， $m^{min}$ 代表此新進來的影像與所有此場景的特徵影像中距離最近的那個特徵影像。

接著列出此新的外觀結合演算法(用於場景資料庫建立)運作的步驟如下：

1. 當有一個新進影像要建立某場景資料庫時，先判斷此時此場景中的資料庫裡已存在多少個特徵影像。
2. 依據外觀的數目，來做其建立此場景的資料庫的動作。

(a) 特徵影像的數目=0：

此新進來的影像直接形成一個外觀，且此特徵影像就是此新進來的影像。

(b) 特徵影像的數目=1 或 2：

若下式成立，則不增加新的特徵影像，且將此新進來的面直接併入此擁有最小距離的影像，且此外觀的特徵面還是保持原來的。

$$\min_{all C_m} d(V_{new}, C_m) < threshold\_1 \quad (4-4)$$

若不符合上式，即此新進來的影像與所有此場景中的特徵面距離最近的那一個距離比門檻值  $threshold\_1$  大，而此時，此新進來的影像就會形成一個新的特徵影像，且此新特徵影像的就是此新進來的影像。

(c) 特徵影像的數目  $\geq 3$ ：

設定兩個門檻值，來縮減資料量：

$$\min_{all C_m} d(V_{new}, C_m) > threshold\_2, \quad threshold\_2 > threshold\_1 \quad (4-5)$$

$$\min_{all C_m} d(V_{new}, C_m) < threshold\_2 \text{ and } d(V_{new}, C_{m^{\min \pm 1}}) > threshold\_2 \quad (4-6)$$

若 (4-5) 或 (4-6) 其中一式成立且 (4-4) 式不成立，則此新進來的影像就會形成一個新的特徵影像，且此新特徵影像就是此新進來的影像。至於此新特徵影像的位置，則依據 (4-7) 式來判定，若成立，則此新外觀加在  $m^{\min}$  和  $m^{\min-1}$  兩個影像之間，反之，加在  $m^{\min}$  和  $m^{\min+1}$  兩個影像之間。

$$d(V_{new}, C_{m^{\min+1}}) > d(V_{new}, C_{m^{\min-1}}) \quad (4-7)$$

若此時 (4-5) 或 (4-6) 兩式皆不成立或是 (4-4) 成立，則不增加新的特徵影像，且將此新進來的影像直接併入此擁有最小距離的特徵影像，且此特徵影像還是保持不變。

步驟 2 的 (c) 的含義就是說，此新進來的影像與所有此場景的特徵影像中距離最近的那一個距離，

1. 若比  $threshold\_1$  小，則合併此新進來的影像。
2. 若比  $threshold\_1$  大、比  $threshold\_2$  小，則依據跟  $m^{\min}$  左右兩個特徵影像的關係來決定，也就是說跟  $m^{\min}$  左右兩個特徵影像的距離都大於  $threshold\_2$ ，則要用此新進來的影像新增一個特徵影像，反之，

則合併此新進來的影像。

3. 若比  $threshold\_2$  大，則要用此新進來的影像新增一個特徵影像。

所以，由上述步驟 2 的 (c) 的說明可知，這兩個門檻值會影響整個新的外觀結合演算法最後得到的結果，因此若希望適當的資料庫，則此處兩個門檻值的設定就變的很重要。值得注意的是， $threshold\_1$  的值不能設定的與  $threshold\_2$  的值比例差太大，否則會導致此場景的資料庫裡的特徵影像的數目不斷的增加，而使得用來表達場景的特徵影像過度表達。而若  $threshold\_2$  設定的值越小，場景就會被表達的越細微，即特徵影像的數目會越多，當然此時用來表達場景也可能會過度表達。

如此可以透過上述的方式，將同一場景內不同角度所攝影到的點，透過相似度的判定，而本論文採用的相似度是透過 Kullback-Leibler 距離來量測，利用 KL 距離來跟兩個門檻值做比較；當兩個模型透過 EM 演算法建立出來之後，要判別兩個模型是否都留下來，就再利用此機制，可以輕鬆的幫助我們選擇適當的保留特徵影像來描述該場景的位置，而不必人工去挑選資料庫，可以提升後段的辨識率，減低運算時間，更減少資料量，一石三鳥，而唯一稍微需要嘗試的則是兩個門檻值的調變，將會影像資料庫的大小與辨識的結果。而修改過後的結合演算法，其流程圖如下圖 21 所示：

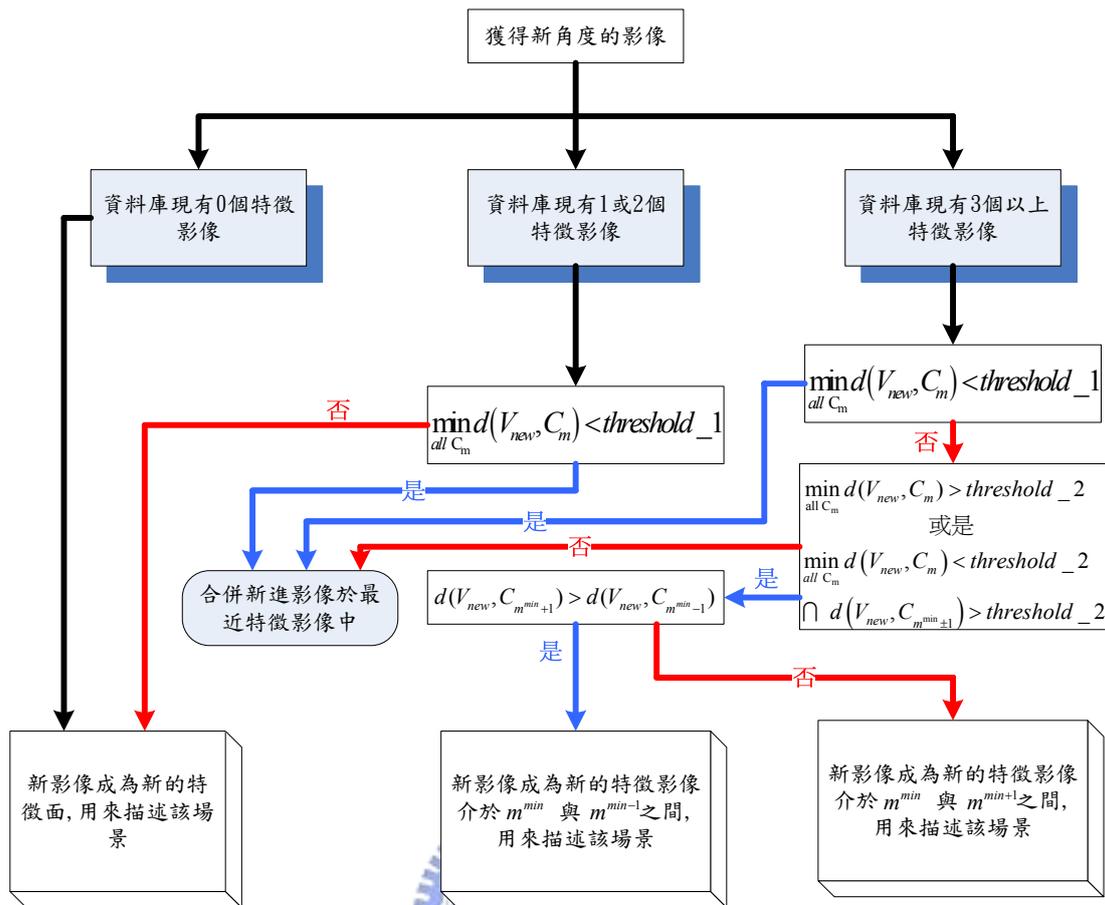


圖 21 利用物件辨識的結合演算所建立的減低資料庫的流程圖

## 4.2 辨識方法

當建立影像資料庫之後，我們採用三個方式來比較影像地標認知的情況：

第一、當在未知區域拍攝一張影像，透過該影像與影像地標集合進行比對；而這裡稍微注意的是，如果採用高斯混合模型，則採用比對的方法為計算最大期望值，而不是計算 KL 距離；而當該未知影像與某影像地標最為接近，則判別該影像是否屬於該區域，並統計其辨識率，並透過分析挑選後，選擇高斯混合模型為爾後實驗的主軸，並進行之後的兩個辨識方法，希冀可以再提高辨識結果。

第二、當未知區域的影像與影像地標集合比對之後，我們並不百分百相信第一相近的影像地標，其第二、第三相近的影像地標，可能只是角度

不同，但是坐落在同一個影像地標集合中的機率也頗高，所以我們整合前三個最相似的影像座標，當某兩個結果為同一影像地標點，則認定該點為此區域，如果前三個最近距離的影像地標分屬不同位置，則僅以最近的表達該未知區域的位置，希望可以提升辨識結。

第三、當只看一眼無法對所在位置有所認知的情況下，反觀人類的行為，將會左顧右盼，透過其他角度的影像來對自己所在位置進行判別，所以我們如果增加未知區域但是不同角度的影像，進行比對，勢必可以在提升辨識率：所以我們每隔一個角度擷取一張影像，選擇不同角度的三張影像來判別，如果這時候這三張影像中有兩張屬於同一影像集合中，則判定此未知影像為該區域，否則，我們再採用第二個方法的概念，如果最相近的三張影像不為同一點的影像集合，則比較此三個角度，第二相近的影像地標，觀察是否有相同的結果，如果仍沒有，再同理比對第三相近的影像地標，如此可以更加提高區域認知的正確性。



# 第五章 實驗結果

## 5.1 系統架構

整個實驗分成兩大類，影像地標的認知與單調特色協助機器人定位的分析：首先，透過比較第二、第三章所介紹的演算法，比較其準確性，之後我們將挑選高斯混合演算法做為主軸，將會透過結合演算法整理出適合的資料庫，而測試的部分，我們不僅僅移動機器人的位置，並甚至透過繪圖軟體嘗試將影像地標遮蔽，高斯混合模型的辨識率也有不錯的表現，並討論其辨識率結果。而初步的架構圖如下圖 22：

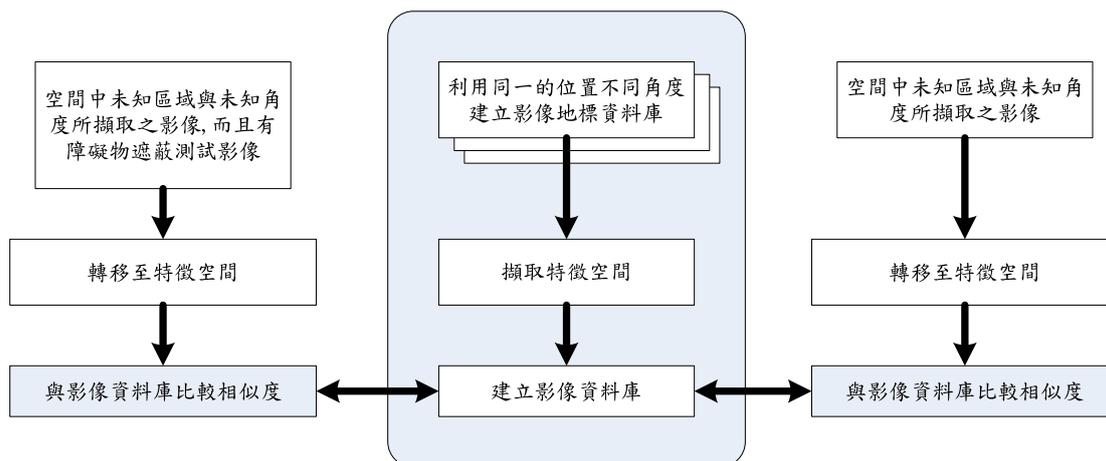


圖 22 影像地標建立與未知影像比對

而其細節的部分，將如圖 23 所示，每一個建立影像地標的位置，拍攝數個角度的資訊，每張影像透過 EM 演算法，計算出其模型參數，並將此參數透過結合演算法，決定是否保留此角度的影像模型。而每個影像地標的位置，各自依此法建立其影像地標資料庫；接著我們以四種情況測試高斯混合模型的容忍度，甚至插入障礙物改變影像地標，來檢驗其容忍度。

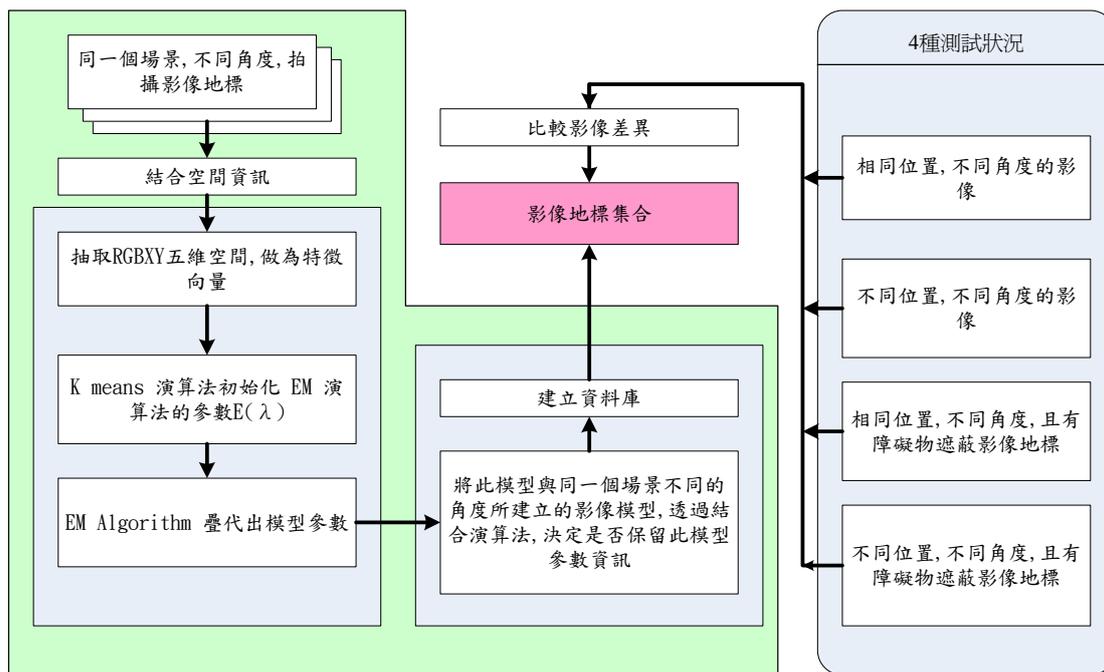


圖 23 完整的高斯混模型的建立與測試流程

第二個部分，影像地標認知之後，我們再假設影像定標不被改變的情況下，觀察分析影像資料在影像地標附近一維空間移動時，觀察其影像差異的變化，並分析此變化的結果，希冀可達輔助定位的特色。

## 5.2 實驗平台

我們實驗的平台是在 CPU 2.8 GHz 的電腦上，Visual C++ 6.0 的環境下來完成整個演算法，並透過 Sony EVI-D30 的攝影機 PCI-Express Camera Link 影像擷取卡，如圖 24 所示，去獲得影像大小是 320 x 240，格式是 bmp 檔，顏色的解析度是 RGB 各 8bits 的影像。



圖 24 Sony EVI-D30 的攝影機與 PCI-Express Camera Link 影像擷取卡

而我們拍攝的環境，如圖 25 所示，為一般的辦公室環境，環境中並沒有特別的人工標記，在第一類辨識率的實驗中，我們將在該辦公室任意選取 11 點做為影像地標。

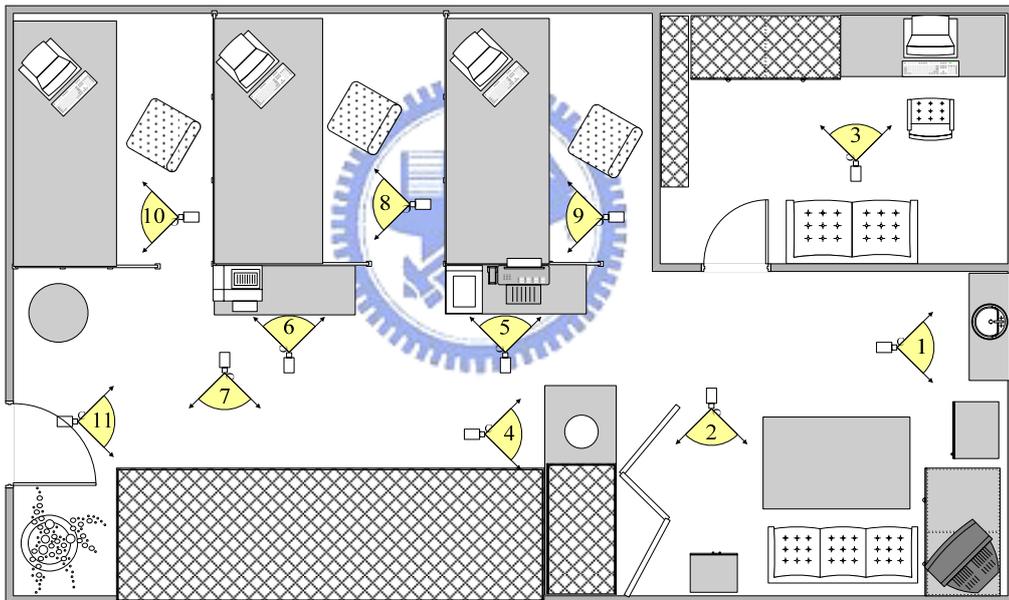
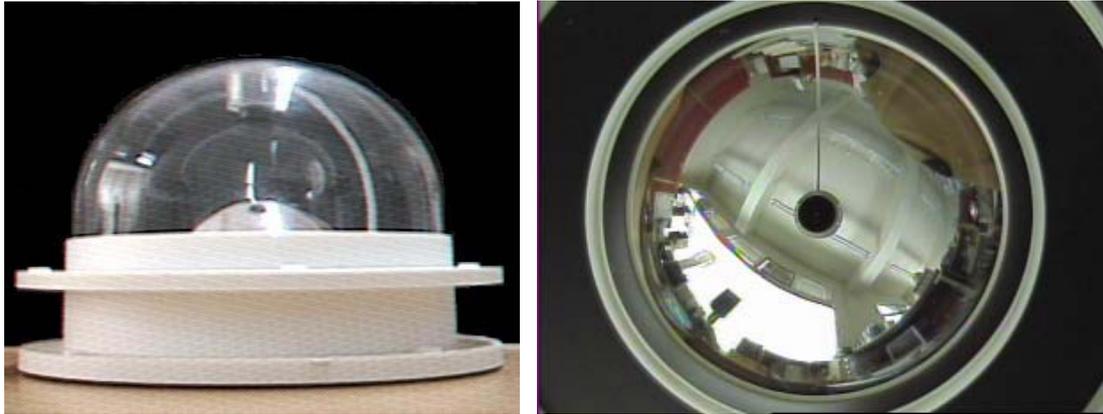


圖 25 建立影像地標的辦公室場景圖

此外，我們也嘗試為了獲取大範圍的影像資訊，我們將採用了環場攝影機(Omni-directional camera)，此後會在後面的小節討論其優缺點，而其外觀如下圖(a)，而其所拍攝的影像，將會透過置於攝影機前的雙曲面鏡，而攝影機與該曲面鏡的焦點，坐落於雙曲面的兩個焦點上，因此攝影機可以擷取到周圍 360 度的影像，而其拍攝到的影像如下圖(b):



(a)

(b)

圖 26 (a)環場攝影機；(b)環場攝影機所拍攝的影像

### 5.3 實驗結果

依照第一節所介紹的流程，我們將透過分析影像地標認知的辨識率，來選擇精度較高的演算法，再分析此方法在影像地標附近的變化情形。

#### 5.3.1 比較各演算法之辨識結果

在第一個實驗，我們比較四個演算法的結果：我們在環境中拍攝 11 個場景，每個場景拍攝 13 個角度，每個角度間隔 5 度，將此影像群做為影像地標的資料庫；而測試部份，在同一個位置拍攝 62 個不同角度的影像，每個角度相差 1 度，所以測試影像與影像地標，本身就有些微的差異，然後將統計分析這些辨識率：

表 2 各模型在同一位置不同角度的辨識率

演算法	辨識率
低通濾波器	98.8%
擴展直方圖	100%
主軸分析	97.3%
高斯混合模型	100%

此表所採用的參數分別是：低通濾波器是將原本的影像資料 $320 \times 240$ 個像素點，變成 $40 \times 30$ 個像素點，而主軸分析採用前面 150 個特徵向量，而擴展直方圖所採用的解析度為每個 8bits 的像素值(即 0~255 的整數)，為直方圖的基本單位；而高斯混合模型採用 12 群；其中可以透過觀察，低通濾波器和主軸分析這兩項的辨識率比較低一點，所以我們嘗試將低通濾波器的模糊程度稍微提高，將影像資料分別變成： $80 \times 60$ 、 $160 \times 120$ 甚至是 $320 \times 240$ ，其辨識率分別為 99.4%、99.7%以及 99.4%，所以如果希望能夠再提高此方法的辨識率，可以將影像的解析度維持在 $80 \times 60$ 個像素點以上，但是其效果仍不夠好，同理，如果將主軸分析採用的特徵向量改成取前 200 個特徵向量，其辨識率將變成 97.31%，僅僅提萬分之一，所以前 150 個特徵向量在此實驗應該足以表示此影像地標，至於擴展直方圖與高斯混合，其效果不錯，所以以下的參數將延續上個實驗。

接著我們要嘗試將位置前後左右偏移分別為 5cm、10cm、15cm、20cm、50cm，而角度仍然採用-30 到 30 度，每隔一度擷取一張影像，與 11 的影像地標進行比對，其結果如下表所示：

表 3 各演算法在不同地標的辨識率

演算法 \ 偏移距離	5cm	10cm	15cm	20cm	50cm
低通濾波器	99.7%	99.25%	98.89%	96.5%	74.07%
擴展直方圖	100%	100%	99.85%	98.65%	79.13%
主軸分析	95.3%	93.29%	89.26%	84.2%	61.3%
高斯混合模型	100%	100%	99.68%	98.77%	83.16%

其中，低通濾波器的部分透過上一個實驗，發現，至少保留 $80 \times 60$ 像素以上的資訊其辨識率較高，所以這個實驗，就採用 $80 \times 60$ 像素做為特徵空間中所描繪的分布；而擴展直方圖、主軸分析、高斯混合模型則同上採用同樣的參數設定分別為：8bits 的像素值為直方圖基本單位，前 150 個特徵

向量，高斯混合群數為 12 群。而其效果，不言而喻的可以發現擴展直方圖與高斯混合模型有比較優良的結果，但是觀察其它演算法發現，影像內容物對於擴展直方圖有很直接的影響，所以如果影像地標中稍微有障礙物進入阻隔，其受影響的程度會遠比高斯混合模型來的大，我們將利用影像處理軟體，將障礙物玩偶，如圖 27 所示；放置在影像地標前面，來測試其結果，而所選用的場景為建立影像地標的原點，每隔 1 度拍攝一張影像，從-30 到 30 度，共 61 張影像，：

表 4 加入障礙物之後的辨識結果

障礙物 \ 演算法	擴展直方圖	高斯混合模型
遮蔽 5%	90.46%	100%
遮蔽 10%	87.05%	100%

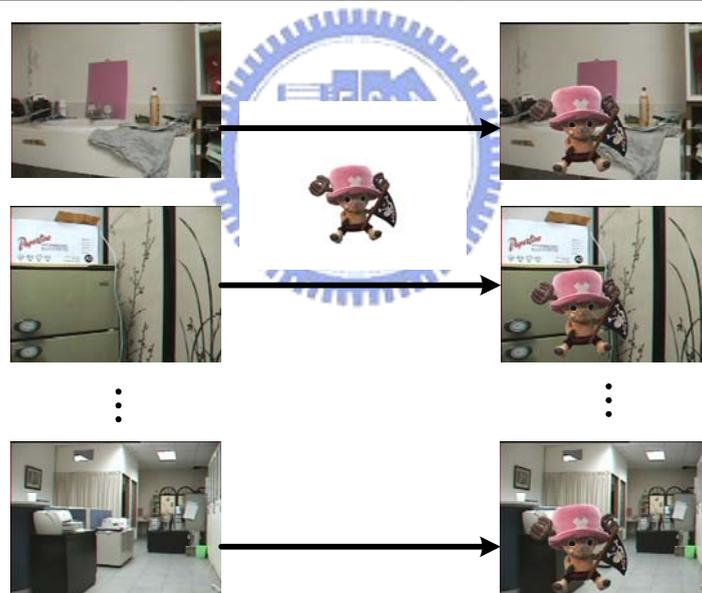


圖 27 置入卡通海賊王玩偶障礙物遮蔽影像

因此，當偏離影像地標近，而且影像地標不被遮蔽改變的情況下，擴展直方圖，的確是比較佔優勢，但是擴展直方圖，對於雜訊的變異有比較大的反應，對於有物體被改變時，也是有很大的反應，容忍度較低，而高斯模糊則能夠辨識出自己該歸屬的影像地標資料庫，而較少誤判。

所以總結以上數個實驗的結果，可以分成數個考量面來分析：

### 第一、速度：

建立資料庫：以高斯混合模型與主軸演算法在建立資料庫時比較耗費時間，而低通濾波器為最快。

影像地標認知測試：以低通濾波器與高斯演算法的辨識比較耗時，而擴展直方圖為最快的比對。

### 第二、儲存空間：

以低通濾波器的結果為最耗費儲存空間，需要紀錄接近整張影像資料大小的數量級，而其餘三者只需要紀錄特徵向量，或是紀錄模型參數即可。

### 第三、精確度與容忍度：

透過之前種種的實驗，高斯的確在這方面有比較優良的結果，而且高斯模糊化影像並不是將所有的影像資訊等份的濾除以降低資料量，而是濾除比較不重要的局部區域變化，所以得到減低資料量的影像反而更加銳利，如圖 28 所示，對於未來可以發展的空間，則有比較大的衍生空間，來從事其他利用或更加準確定位。

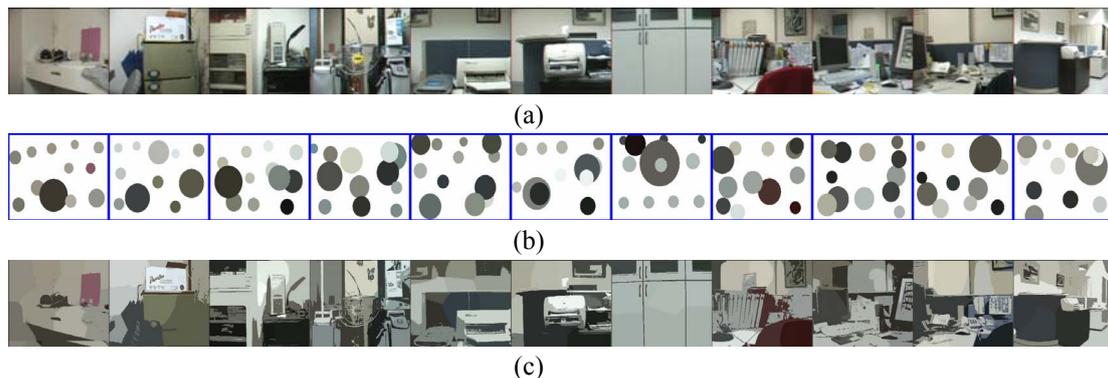


圖 28 (a)空間中 11 個場景(b)高斯混合模型之分布(c)其模糊化後之影像

### 5.3.2 高斯混合模型的容忍度測試

透過上一節連續的實驗結果，我們將選取高斯混合模型為以下實驗的主軸，首先，我們將透過 4.3 節所介紹的方式，來嘗試提高辨識率，我們將重複上一個小節的實驗，來驗證辨識率的提升。如下表所示，R1 為基本的辨識方法，4.3 節所介紹的第一個方法，只判斷一個面且只判斷第一個最相似的可能，即上一節採用的方式；而 R2 與 R3 分別為 4.3 節所介紹的方法，將透過辨識率，分析其效果：

表 5 各位移量與各辨識方法搭配的結果

偏移距離與方向		辨識方法與辨識率		
距離	方法	R1	R2	R3
0 cm		100%	100%	100%
5 cm	向前偏移	100%	100%	100%
	向後偏移	100%	99.85%	100%
	向左偏移	100%	100%	100%
	向右偏移	100%	100%	100%
10cm	向前偏移	100%	100%	100%
	向後偏移	100%	99.85%	100%
	向左偏移	100%	99.11%	100%
	向右偏移	100%	100%	100%
15cm	向前偏移	100%	100%	100%
	向後偏移	100%	99.40%	100%
	向左偏移	99.55%	97.77%	100%
	向右偏移	99.18%	97.87%	99.67%
20cm	向前偏移	100%	100%	100%
	向後偏移	100%	99.26%	100%
	向左偏移	97.47%	95.53%	98.21%
	向右偏移	97.62%	95.68%	98.36%
50cm	向前偏移	84.95%	84.95%	85.54%
	向後偏移	89.72%	85.40%	93.44%
	向左偏移	75.26%	74.81%	77.35%
	向右偏移	82.71%	79.73%	85.25%

我們將之前四個偏移方向，分開條列出來，可以更容易看到其中的變化，其中 R1 平均，即是上一小節的結果；而 R2 其辨識率，反而下降，此方法的概念是希望透過第二、第三相近的影像地標，在第一相似的影像地標誤判時，可以輔助修正，但是第二、第三誤判的機率反而來的高，所以此方法辨識率會下降；第三個方法 R3，其辨識率則可以使辨識率上升，畢竟當本身影像地標判斷錯誤時，可以透過其他角度的影像幫助認知，其辨識率的提高是可以預期的。

透過上述三個方法比較，越多不同角度的影像資訊被比較，其結果越可以被信賴，就如同人類的行為一般，如果在陌生的環境，無法單從眼前的影像認知自己所處的位置，勢必會東張西望，利用這些影像資訊來幫助影像地標的認知可以更加準確。

同上面的概念，我們將分析一下高斯混合模型的容忍度，當障礙物存在，使得影像被遮蔽分別為遮蔽掉 5%、10%、15%、20%、50%，觀察此情況下，是否能夠分析辨識影像地標，我們將依循 5.1 小節中所介紹的流程圖（圖 23），同樣的，我們將四個方位且依序偏離 5 公分、10 公分、15 公分、20 公分、50 公分的結果，分開條列出來，其結果如下表所示，其結果在測試位置與影像地標相隔 20 公分，障礙物遮蔽整張影像百分之二十的情況下，都還有超過九成以上的辨識率，如果採用第三個方式來增強影像地標認知的效果，其效果在此差異性中更可以將整個辨識率拉高，使最低辨識率也達到百分之九十七點五，可見在 20cm 與 20% 的影像差異中，也不容易與其他位置的影像地標有所混淆；而且即使是 50cm 的差異與百分之五十的影像被遮蔽，在如此巨大的差異下，其辨識率也大於五成，因此高斯混合模型，在這領域有很大的優勢。而此實驗的影像地標資料庫跟之前一樣沒有受到遮蔽，仍為-30 度到 30 度，每隔 5 度拍攝一張，共 13 張，而共有 11 個點。

表 6 加入障礙物遮蔽影像且偏離影像地標

偏移距離		不同遮蔽情況下的辨識率(%)														
		5%			10%			15%			20%			50%		
距離	方法	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3
	0 cm		100	100	100	100	100	100	100	100	100	100	100	100	80	76.9
5 cm	前	100	100	100	100	100	100	100	100	100	100	100	100	79.7	77.5	80.9
	後	100	100	100	100	100	100	100	100	100	100	100	100	79.4	76.3	80.9
	左	100	100	100	100	100	100	100	100	100	100	100	100	79.4	77.9	80.6
	右	100	100	100	100	100	100	100	100	100	100	100	100	79.1	75.4	81.8
10cm	前	100	100	100	100	100	100	100	100	100	100	100	100	79.6	78.4	80.2
	後	99.7	97.9	100	100	99.7	100	100	99.7	100	99.7	99.6	100	78.5	73.8	80.2
	左	100	99.3	100	100	99.6	100	100	99.6	100	100	99.3	100	79.6	77.2	80.5
	右	100	100	100	100	100	100	100	99.7	100	100	99.9	100	77	74.7	81.7
15cm	前	100	100	100	100	100	100	100	100	100	100	100	100	78.4	76.9	79.7
	後	100	99.6	100	100	99.3	100	99.7	98.8	100	99.4	98.7	100	76.3	72.6	78.1
	左	99.7	97.9	100	99.7	97.9	100	99.7	97.9	100	99.4	97.5	100	77.9	73.6	79.4
	右	99.2	98.2	99.7	99.2	97.7	99.7	99.2	97.9	99.7	99.2	97.7	99.7	72.6	70.5	75.7
20cm	前	100	100	100	100	100	100	100	100	100	100	100	100	76.5	75.1	78.2
	後	99.9	98.7	100	99.4	98.2	100	99.1	97.9	100	98.8	97.6	100	73.3	71.1	74.8
	左	97.6	96	98.7	97.9	97	99.3	97.5	96.1	97.9	97.5	96.3	97.9	74.8	69.9	77.6
	右	97.5	95.5	98.4	97.9	97	99.3	97.6	95.1	98.4	97.5	95.1	98.4	71.4	69.4	74.4
50cm	前	84.5	83.8	85.4	84.5	84.4	84.9	84.2	82.9	84.5	83.2	81.5	83.9	50.8	50.3	52.3
	後	88.1	83.9	92.5	84.8	82.1	89.6	83	80.9	87.2	79.6	78.5	83.3	52.5	50.8	55.3
	左	75	74.1	77.5	74.8	74.2	76.8	73.3	72.9	75.4	72.3	71.1	73.5	50.2	50.1	53.1
	右	81.1	79.4	84.1	79.9	78.4	82.7	78.1	77	81.7	75.3	76.3	77.8	53.2	50.2	53.1

### 5.3.3 影像相似與影像搜尋引擎

透過之前的影像相似，高斯混合模型，具備了概念化場景，與頗高容忍度的搜尋比對，而且，當障礙物遮蔽影像資料庫的資訊時，其正確可以用作比對的像素資訊就變的更少，因此，我們換個角度去想，不用百分之百正確的影像，也可以找出真正的影像，那如果是一張模糊，只是趨勢大約相似的影像，同理，也可以在之前的影像資料庫中找出相似的對應影像。而這樣的概念，也是現在影像比對這個領域的一個研究主軸之一，我們也建立一套簡單的搜尋介面，只要利簡單的繪圖軟體，概略畫出影像內容，其介面如下：



圖 29 簡易搜尋引擎介面，由手繪的印象影像尋找相似的資料庫影像其比對的結果會不斷的替換掉最像的幾張影像，直到全部比對完畢，就可以得到還可以接受的答案，對於影像相似，也是一種應用。

### 5.3.4 環場影像的測試與分析

以上的實驗，通通以 Sony 公司的 EVI-D30 的 PTZ(Pan Tile Zoom)攝影機，其視角有限而且因為建立資料庫的張數太少，透過結合演算法，往往不容易看出結合演算法的功效，所以希望能夠有效率的拍攝到大範圍且多張影像，因此，一個大視角的影像攝影裝置，是我們感興趣的，於是，我們採用了宜昇科技公司的環場攝影機，其拍攝範圍除了中央的攝影機架設的位置，涵蓋了以攝影機為中心，整個 360 度的環境影像，我們將拍攝到扭曲的影像展延成為一個圓柱狀影像[28][29]，其延展的概念大約如下圖所示：

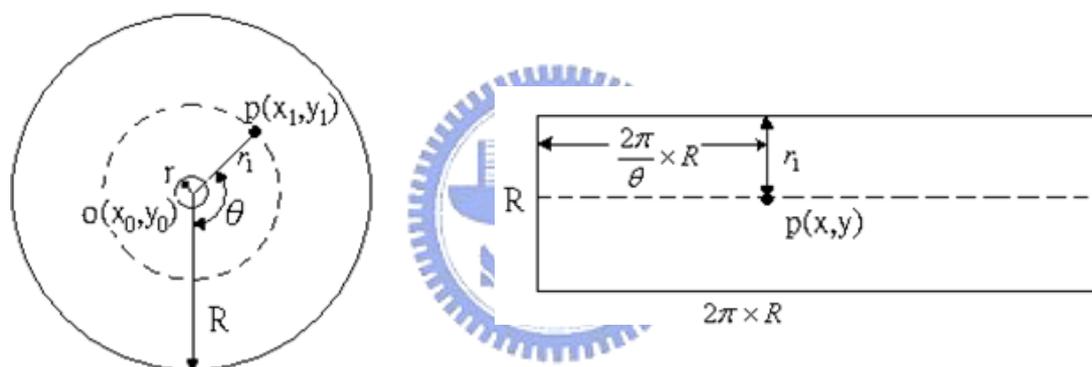


圖 30 環場影像與展延影像示意圖

其中環場攝影機中任意像素值  $p(x_1, y_1)$ ， $r$  與  $R$  理論上為同心圓，分別表示可見區域的最小及最大極限， $O(x_0, y_0)$  為同心圓之圓心，而圖 31 則是任意一扇型區域透過延展將會呈現在延展影像的某區段：



圖 31 某區段的環場影像透過延展的結果

而此延展影像的長寬比等於環場影像的半徑周長比，大約為  $2\pi$ ，則如此轉

換失真會較少。

透過此攝影器材可以拍攝到環場影像，因為拍攝環場影像做為影像地標時候需要大範圍內無人物走動，所以拍攝上的限制比較大，所以僅以七個點來進行以下的實驗：

首先拍攝環場影像，並延展該影像，再進行切割，每隔五度切一張影像，共有 72 張影像，接下來，透過結合演算法計算出縮減影像資料庫，其結果如下表所示，可以發現，在資料統整上面有不錯的成效。

表 7 七個位置的展延影像與結合過的資料庫數量以及辨識率

七個場景的展延影像	結合後資料庫數	辨識率
	54	73%
	49	78%
	47	76%
	49	88%
	26	75%
	45	80%
	51	80%

然而辨識率卻是下滑，主要原因有幾個：第一、環場攝影機是透過劇烈扭曲周圍場景投影到攝影機之成像平面，之後再透過內插方式，將影像還原至原本的影像，其失真大幅上升而解析度大幅下降，第二、僅僅以分割後的每一個區間的影像來判斷，因為過度失真情況下，不足以描述該位置資訊，而導致辨識率下滑；此外拍攝時，使用者必須低於攝影機，在拍攝時候難度大大提升，所以還是以一般 PTZ 攝影機為比較方便的選擇。

### 5.3.5 利用高斯演算法大致上單調趨勢輔助定位

透過之前的實驗，已經可以將環境中的場景透過影像近似的演算法，歸類至所歸屬的影像地標，藉而判斷出所屬的空間區域，然而，如果希冀不再外加其他方式，而單純的以現有的資訊是否可以程度上的輔助機器人定位，因此我們將透過以下的結果來探索這個問題。

在此實驗之前，我們先透過分析群數對於這個議題的影響，於是，採用 11 個場景，沒有偏轉而取前後每隔 5 公分、10 公分、15 公分、20 公分、50 公分拍攝一張的影像，觀察群數對於趨勢分布的影響，如下表所示：

場景	開始穩定的群數	開始擾動的群數
第 1 個場景	35	
第 2 個場景	30	75
第 3 個場景	35	90
第 4 個場景	30	
第 5 個場景	45	90
第 6 個場景	40	75
第 7 個場景	25	70
第 8 個場景	35	
第 9 個場景	40	95
第 10 個場景	40	85
第 11 個場景	25	

表 8 不同的群數對於場景的穩定區間

上面實驗是選擇 5 到 95 群，每 5 群建立一個模型，觀察其在一維空間中的曲線變化，所呈現的將如下圖所示。下圖為環境中的第二個場景，由左而右，由上而下，依序是 5、10 群……，中央點為影像地標建立的位置，而判斷的方式是透過將前後兩個區段透過一維線性回歸，用直線方程式去逼近，然後利用此兩條直線方程式的參數做為向量，比較兩群數向量關係，可以填出上表的關係。在低群數的時候，如果兩個群數變異量太大則認為此群數還不夠穩定，而如果群數過多，則會發現之後會有擾動，如下圖，以第二個場景為例，低於 30 群或是超過 75 群之後，每 5 群之間的差異，其線性回歸參數差異大，所以，不同群數所產生的結果都不太一樣，不像中間那段穩定。

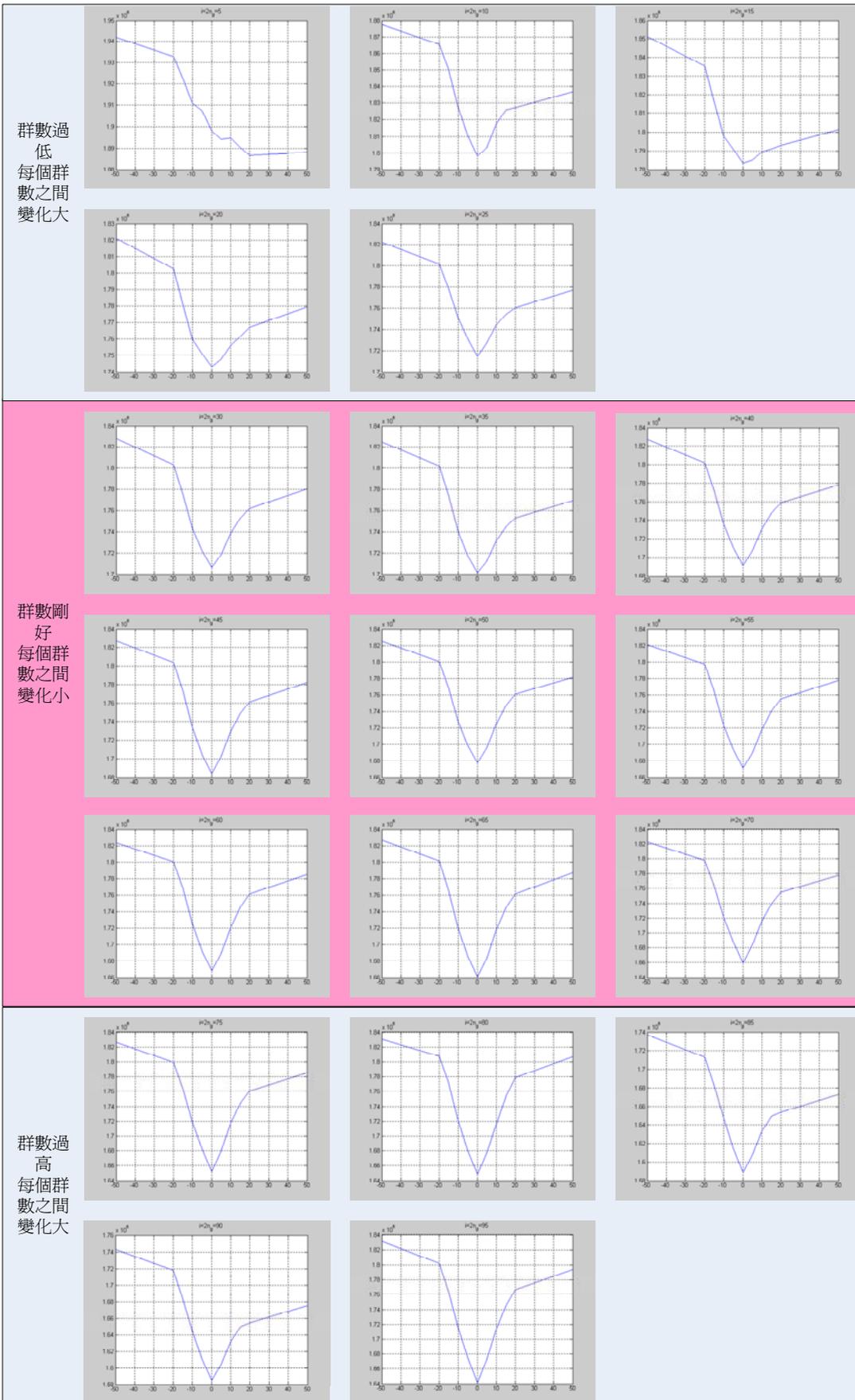


圖 32 不同群數對於同一個場景的變化關係圖

透過上面的觀察，三十、四十幾群是比較適合來描述其模型分布的群數，因此爾後的實驗，大約都把群數設定在這個範圍。

我們先將實驗鎖定在一個自由度的空間來探討，分別是旋轉，前後移動，與左右移動，三種變化，在沒有誤判及影像地標不被改變的前提假設下，觀察在影像地標附近的影像與影像地標的差異變化，可以發現部份場景，會存在著單調的特色，而極大部份的場景，大體上也都存在著單調的趨勢，我們希望可以藉由此性質來輔助機器人可以更加貼近影像地標。

首先，先觀察一維空間中前後移動的情形，為了鎖定一維空間，所以我們從新拍攝了一組影像地標共 20 個場景，前後移動，前後共 30 公分，每隔 3 公分，拍攝一張影像，且經過多次，觀察其變化如下表所示：

表 9 單調關係存在的場景數量統計

完全呈現單調的場景		8 個場景
多次實驗中約超過 50% 呈現單調		4 個場景
多次實驗中約超過 30% 呈現單調		3 個場景
無單調 場景	僅有一個點破壞單調特色	2 個場景
	超過兩個點以上破壞單調特色	1 個場景
	最低點不在影像地標本身	2 個場景

以上實驗群數為 45 群，每次測試時重新透過 EM 演算法建立該位置的影像地標，其中有八個場景，不論資料庫如何變化，其單調的特色一直能夠維持，其部份結果如圖 33 所示，而另有七個場景，在某些參數模型會存在單調的特色，而另外有五個場景則不論資料庫如何變化，都會有一兩個點存在著局部最小值(local minimum)，而其錯誤的點，並不是隨機的出現，而是固定某些位置一定會產生局部最小值，可見與模型參數無關。

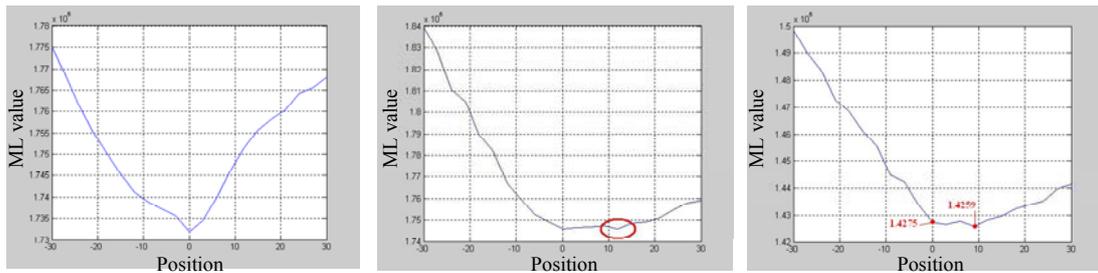


圖 33 單調影像、局部最小值破壞單調、極值發生點錯誤

首先，假設模型參數可以百分百表達影像地標，那當鄰近位置的影像，勢必在五維空間中有所偏移，其示意圖如下所示：

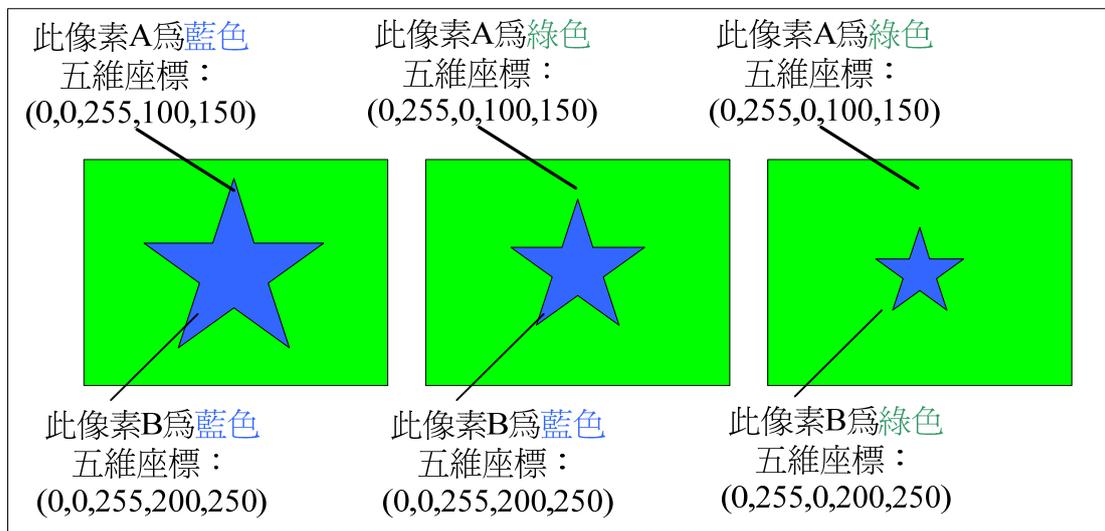


圖 34 當拍攝點向後移動時，像素的變化示意圖

如果以第一張建立影像模型，而以後面兩張做為測試影像；我們挑選兩個點來觀察其變化，A 點在第一個偏移之後，顏色劇烈改變，所以在五維空間中的該點會大幅度的變動，在套模型參數，此點的符合度會大幅下降，而 B 點在第一個偏移後，其顏色仍然維持不變，也就是雖然偏移，但是該點再五維空間是不會被移動，所以此點與模型的符合度就會維持很高，直到再度偏移，此點五維空間中的分布才會被改變，而符合度會更低；因此，只要色塊影像類似此分布，單調的情況是可以預期的，但是實驗結果並沒有百分百的單調，所以可能下有下面幾個：

第一、模型沒有真的百分百描述出該影像，或是該影像地標在五維空間中

的分布並不是能夠以高斯來描述，而偏移數個位置之後的影像，反而更能符合高斯的趨勢，導致在測試相似度的時後，反而有比影像地標更接近模型參數的測試影像，這是影像本身分布的問題，比較無法控制。

第二、當移動的時候，假設有左右有區域是類似或是重複的，當左側的影像越來越消失時，右側的影像反而補到原本左側的影像；或是類似此情形，都會在過程中有局部最小值產生，然而一般自然場景中，非人工刻意安排，要有這樣的影像比較不容易，所以巧合所造成的局部最小也不會很誇張。

第三、當影像中有較大面積的物體因為移動而完全消失在測試影像中時，用該影像去套模型參數時，其變異也會比較沒有規律，所以有些非單調是發生在距離影像地標太遠的地方。

同樣的，我們觀察另外兩個情況，左右移動，與旋轉角度情況如下表所示：

表 10 旋轉與左右移動測試單調的結果

旋轉(11 個場景)			左右移動(14 個場景)		
完全呈現單調		4 個場景	完全呈現單調		7 個場景
部份單調		5 個場景	部份單調		3 個場景
無單調	局部最小存在	1 個場景	無單調	局部最小存在	3 個場景
	最低點錯誤	1 個場景		最低點錯誤	1 個場景

其中，旋轉取-15 到 15 度每隔一度拍攝一張影像，共 25 個測試點，而左右移動的則是每隔 3 公分拍攝一張，從偏離左側 30 公分到偏離右側 30 公分，共有 21 個點。

## 第六章 結論與未來發展方向

本論文引用了場景概念化的機制，將一個場景模糊，以建立影像地標來描述該位置大約可見的景色，並利用影像相似演算法機制，判別一個未知的場景，將其歸類到附近的影像地標，藉此判別未知位置坐落在空間中的某個區域，並希望可以藉由演算法本身存在的關係，使機器人往影像地標處移動。

而整個流程是透過各種不同的演算法去評估較佳的盲目搜尋比對系統，且把顏色以外的空間座標一起併入考量，使得區隔性更大，以降低誤判；而且當面臨大資料量的時候，也提出了一套結合演算法來降低資料量；並透過之前的數據顯示，只要挑選一個適宜的演算法，即使是處於一個資料庫沒有的位置，面對一個資料庫沒有的角度，且有障礙物遮蔽影像地標的訊息，在一定範圍內，能有很高的辨識率，所以在影像地標附近的影像，可以輕鬆的辨認出自己所屬的區域。

而在輔助定位的部份，目前還僅僅限制在一個自由度的變化可預期(前後偏移，左右偏移，左右旋轉)，所以如果可以限制機器人的活動在一維空間(如：有軌道)，或是結合兩個或三個一維空間，再搭配其他感測元件輔助定位，還是可以有一定成效。

本篇論文採用了高斯混合模型來描述影像地標，然而，高斯混合模型在影像處理的議題中，也被使用在影像分割(image segmentation)，可見他分離影像內容物的能力也很不錯，由我們之前所實驗的結果圖，如圖 28 所示，其將影像邊緣銳化的能力也很不錯，如果加入邊緣偵測演算法，將原本的盲目搜尋模式，近而對影像內容物的分析抽離比對，不但可以偵測物件的位置，甚至可以利用這些位置來進行路徑規劃，如此需要考量的則是速度，值得慶幸的是隨著科

技發達，運算速度越來越快的晶片也推陳出新，因此，對於此演算法的加速極其發展的可能性，有莫大的幫助。



## 參考文獻

- [1] J. Leonard, H. Durrant-Whyte, “Mobile robot localization by tracking geometric beacons.” *IEEE Trans. On Robotics and Automation* Vol.7, pp. 89-97, June 1991.
- [2] G.N. Desouza, A.C. Kak, , “Vision for mobile robot navigation: a survey,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 237-267, 2002
- [3] A. Kosaka and AC Kak, “Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties,” *Computer Vision, Graphic, and Image Processing -- Image Understanding*, Vol. 56, No. 3, November, pp.271-329, 1992.
- [4] I. Ulrich and I. Nourbakhsh, “Appearance based place recognition. for topological localization,” in *IEEE Conf. on Robotics and Automation*, November 2000, pp. 1023–1029.
- [5] L. Pierre, T. Adriana, G. Etienne, T. Nicola and S. Roland, “Environmental modeling with fingerprint sequences for topological global localization,” in *IEEE International Conf. on Intelligent Robots and Systems*, October 2003, pp. 3781–3786.
- [6] G. Giralt, R. Sobek, and R. Chatila, “A Multi-Level Planning and Navigation System for a Mobile Robot; A First Approach to Hilare,” *Proc. Sixth Int’l Joint Conf. Artificial Intelligence*, vol. 1, pp. 335-337, 1979.
- [7] S. Tsugawa, T. Yatabe, T. Hirose, and S. Matsumoto, “An Automobile with Artificial Intelligence.” *Proc. Sixth Int’l Joint Conf. Artificial Intelligence*, pp. 893-895, 1979.

- [8] N.J. Nilsson, "Shakey the Robot," Technical Report 323, SRI Int'l, Apr. 1984.
- [9] D.A. Pomerleau and T. Jochem, "Rapidly Adapting Machine Vision for Automated Vehicle Steering," IEEE Expert, Intelligent Systems and Their Applications, vol. 11, no. 2, Apr. 1996.
- [10] A.A. Argyros, K.E. Bekris, S.C. Orphanoudakis, L.E. Kavraki "Robot Homing by Exploiting Panoramic Vision." Journal of Autonomous Robots, July 2005, in press
- [11] M. Meng and A.C. Kak, "Mobile Robot Navigation Using Neural Networks and Nonmetrical Environment Models." IEEE Control Systems, pp. 30-39, Oct. 1993.
- [12] C. Drocourt, L. Delahoche, C. Pegard, A. Clerentin, "Mobile Robot Localization Based on an Omnidirectional Stereoscopic Vision Perception System." IEEE Robotics and Automation, May 1999.
- [13] B.J.A. Krose, N. Vlassis, R. Bunschoten, Y. Motomura "A probabilistic model for appearance-based robot localization." Image and Vision Computing 19 pp381-391, 2001
- [14] H.P. Moravec, "The Stanford Cart and the CMU Rover," Proc. IEEE, vol. 71, no. 7, pp. 872-884, July 1983.
- [15] J. Borenstein and Y. Koren, "The Vector Field Histogram—Fast Obstacle Avoidance for Mobile Robots," IEEE Trans. Robotics and Automation, vol. 7, no. 3, pp. 278-288, June 1991.
- [16] J. Santos-Victor, G. Sandini, F. Curotto, and S. Garibaldi, "Divergent Stereo for Robot Navigation: Learning from Bees," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, 1993.
- [17] L.G. Brown "A Survey of Image Registration Techniques." ACM

Computing Surveys, Vol.24, NO.4, December 1992.

[18] H. Tang, T. Fang, P.F. Shi “Image Feature Ratio Model: Similarity Measures with Both Semantic and Visual Features in Interactive Image Retrieval” MIPPR 2005, Image Analysis Techniques, Oct 31-Nov 2 2005.

[19] A.A. Argyros, K.E. Bekris, S.C. Orphanoudakis, L.E. Kavradi “Robot Homing by Exploiting Panoramic Vision” Autonomous Robots, p 7-25,v 19, July 2005.

[20] R.C. Gonzalez and R.E. Woods “Digital Image Processing”, second edition, Feb,1993.

[21] L. Jan, Latecki “Image Similarity Measure for video Analysis” Proceedings of SPIE - The International Society for Optical Engineering, p219-227, V 5018, 2003.

[22] H. Greenspan, J. Coldberger. “A continuous probabilistic framework for image matching.” Computer Vision and Image Understanding, p 384-406,v 84, December 2001



[23] <http://islab.ee.fju.edu.tw/member/92/steven/ch6.htm>

[24] S.C. Jeng, J.S. Hu “A GMM-based Method For Dynamic Background Image Model Construction with Shadow Removal.” Electrical and Control Engineering, National Chiao Tung University, July 2005, Taiwan, Diploma thesis.

[25] T. Deselaers, D. Keysers, and H. Ney “Clustering visually similar images to improve image search engines” Lehrstuhl für Informatik VI, Computer Science Department University of Technology, D-52056 Aachen, Germany ,Diploma thesis.

[26] C. M. Cyr and B. Kimia, “A Similarity-Based Aspect-Graph Approach to 3D Object Recognition,” in *International Journal of Computer Vision*,

57(1):5–22, 2004

[27] C.C. Lin、J.S. Hu、Dr. Chi-Cheng Jou “Shape Memorization and Recognition of 3-D Objects Using A Similarity-Based Aspect-Graph Approach.” Electrical and Control Engineering, National Chiao Tung University, July 2005, Taiwan, Diploma thesis.

[28] T.E. Boulton, X. Gao, R. Micheals, M. Eckmann, “Omni-directional visual surveillance”, Image and Vision Computing, p515-534, v22, July 2004

[29] *LIU Hong, PI Wenkai, ZHA Hongbin* “Motion Detection for Multiple Moving Tracking by Using an Omnidirectional Camera.”

