

國立交通大學

資訊學院資訊科技(IT)產業研發碩士專班

碩士論文

室內畫展導覽系統及參觀路徑追蹤

A Painting Exhibition Guide System
and Tour Tracking

研究 生：貝佳瑄

指導 教授：陳稔 教授

中華民國九十九年六月

室內畫展導覽系統及參觀路徑追蹤

A Painting Exhibition Guide System
and Tour Tracking

研究 生：貝佳瑄

Student : Chia-Hsuan Pei

指 導 教 授：陳 稔

Advisor : Zen Chen

國 立 交 通 大 學

資訊學院資訊科技(IT)產業研發碩士專班

碩 士 論 文

A Thesis
Submitted to College of Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
1896

Industrial Technology R & D Master Program on
Computer Science and Engineering
June 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年六月

室內畫展導覽系統及參觀路徑追蹤

學生：貝佳瑄

指導教授：陳稔

資訊學院資訊科技(IT)產業研發碩士專班

摘要

在本論文中，利用智慧型手機的照相功能，提出了一套基於影像檢索以及相機定位功能的室內畫作導覽系統及參觀路徑追蹤。本系統主要可分為兩部份，第一部份為針對單張獨立影像，當使用者對著感興趣的畫作拍照，系統會將所輸入的影像進行畫作特徵點擷取和特徵點比對，得到初步查詢結果後，再加以驗證查詢結果之正確性，並利用單張影像進行相機校正，估算法出相機位置，依據使用者拍攝視角變化，回傳不同的畫作相關資訊解說，以增進參觀民眾對於展出作品的了解；第二部份主要是建立出展場中畫作配置情形，追蹤使用者相機位置與多幅畫作之相對關係，透過歐式座標空間轉換，將原本各自獨立的畫作串聯，呈現一完整的展場畫作空間配置關係，此功能不僅能當作室內導航系統，提示使用者目前在展場中的位置，提供個人化的參觀路徑記錄回顧及建議，對於館內工作人員統計熱門作品或是動線規劃等問題也將有很大的助益，如此的概念提供了有別於傳統導覽方式的全新體驗。

A Painting Exhibition Guide System and Tour Tracking

student : Chia-Hsuan Pei

Advisors : Zen Chen

Industrial Technology R & D Master Program of
Computer Science College
National Chiao Tung University

ABSTRACT

In this thesis, the proposed system based on image retrieval and camera calibration is used to provide the indoor painting exhibition guide and tour tracking. The system consists of two parts. The first part aims at user's searching for images of paintings. Users can get the information about paintings of interest by simply inputting a picture of them. The system provides the inquiry service about the input query image. Once the query image identity is verified and the camera pose is estimated, the system returns the pertinent information of the query painting image. The second part is about constructing the location of the exhibited painting through the pose estimation and the recoding of the tour path. This service not only can be an indoor positioning system but also provides useful suggestions for viewing the missing painting exhibition. The proposed system provides a brand new experience in visiting galleries and museums.

誌 謝

能夠完成本篇論文，首先要感謝陳稔老師這段時間來的諄諄教誨，從一開始對於電腦視覺這領域完全陌生的我，可以在課堂中吸收到許多攝影機相關的論文研究與實作方法，在日常生活中也能給予我最大的包容。還要感謝賈叢林老師以及周宏隆學長在口試時提供了許多寶貴的意見，讓論文結構更為完整。

其次，要感謝實驗室裡成員們，學長文昭、彥旭，啟銘以及學弟嘉峻的陪伴，特別是彥旭學長，亦師亦友的特質，不論是在學業、生活、或是玩樂上，總是不能缺少他，讓我可以更有毅力與決心可以走到今天。

最後，要感謝我的主管Kevin，讓我有機會可以到這優秀的環境深造，感謝我的父母在這段期間給予我的肯定，以及我的男友Lopin，總是往返台北新竹兩地接送我，讓我免於舟車勞頓之苦，體諒我這段時間沒能好好陪他，用盡最大的包容心陪我走過許多風風雨雨，也要感謝這兩年內關心我、幫助我的人，能夠完成論文，這份喜悅的成果是屬於你們的。



貝佳瑄 謹誌於
中華民國 99 年 6 月

目 錄

中文摘要.....	i
英文摘要.....	ii
誌 謝.....	iii
目 錄.....	iv
表 目 錄.....	vi
圖 目 錄.....	vii
第一章 緒論	1
1.1 研究動機與目標	1
1.2 相關研究文獻	2
1.3 研究內容概述	3
1.4 論文架構	3
第二章 室內畫展導覽之辨識基礎	4
2.1 畫展特徵點擷取	4
2.2 畫展特徵點描述子	5
2.2.1 ZMs 之強度以及相位介紹	5
2.2.2 ZMs 之相似度計算	6
2.3 畫展畫作特徵點描述子k-度空間樹狀(k-d tree)索引之建立與投票機制	7
第三章 系統架構及畫展資料庫內容索引建立	9
3.1 系統架構介紹	9
3.2 資料庫製作	11
3.2.1 畫作實際尺寸丈量與畫作 3D 點制定	11
3.2.2 畫展畫作 ZMs 描述子與特徵點位置之資料庫建立	12
第四章 拍攝畫作之影像查詢	14
4.1 拍攝畫作之kd-tree查詢結果投票名次	14
4.2 畫作查詢結果之驗證	14
4.2.1 資料庫影像與查詢影像之Homography估測	14
4.2.2 拍攝畫作之特徵點篩選與驗證正確率之計算	16
4.3 利用 Homography matrix H_{image} 之查詢結果驗證	18
第五章 畫作與相機之相對位置估測	20
5.1 畫作與相機之空間座標相對位置的計算	20

5.1.1 內部參數之計算.....	20
5.1.2 外部參數之計算.....	22
5.2 畫展畫作空間配置圖之建立.....	24
5.3 使用者參觀路徑記錄.....	27
第六章 實驗結果	28
6.1 畫展影像辨識實驗結果.....	28
6.2 畫展影像相機定位實驗結果.....	31
6.2.1 單張資料庫影像之Homography驗證及相機軌跡定位.....	31
6.2.2 展場畫作空間配置圖及使用者參觀路徑估測.....	35
第七章 結論與未來發展	38
參考文獻	39

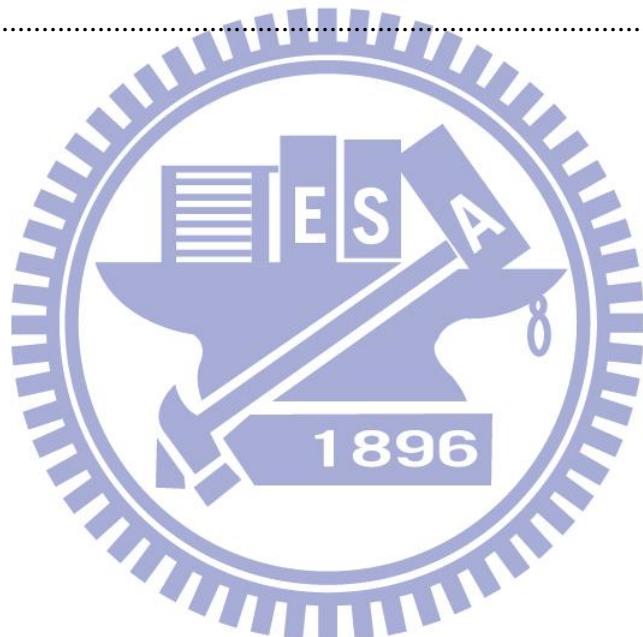


表 目 錄

表 1 將Homogarphy驗證結果轉為灰階圖之誤差	31
表 2 在圖 20 (a)、(b)、(c) 中之projected points誤差(pixels)之最小前 10 名 ...	33



圖 目 錄

圖 1 由左至右為ARToolKit與ARTag.....	1
圖 2 MSER示意圖	4
圖 3 畫作辨識流程圖	9
圖 4 相機定位流程圖	10
圖 5 畫作配置及參觀路徑追蹤流程圖	10
圖 6 實體畫作座標示意圖	11
圖 7 資料庫影像之MSER特徵點區域示意圖	13
圖 8 Homography檢驗流程示意	19
圖 9 Homography驗證結果示意圖	19
圖 10 相機定位流程示意圖	23
圖 11 串連畫作空間配置關係示意圖	24
圖 12 畫作空間配置建立方法示意圖	25
圖 13 三個座標系觀察共同一點 P 示意圖	26
圖 14 使用者介面之畫面	28
圖 15 輸入影像開始查詢	29
圖 16 查詢辨識結果	29
圖 17 Homography驗證查詢結果之一	30
圖 18 Homography驗證查詢結果之二	31
圖 19 資料庫影像之高更的《帶光環的自畫像》	32
圖 20 (a)、(b)、(c)為估測查詢影像之Homography	32
圖 21 相機軌跡估測結果	35
圖 22 畫作空間配置圖	36
圖 23 (a)、(b)、(c)為估測畫作空間配置圖	36
圖 24 (a)、(b) 分別為使用者所輸入之查詢影像	37
圖 25 使用者參觀路徑估測	37

第一章 緒論

1.1 研究動機與目標

在日常生活中，人們去參觀美術、文物、人文歷史展，常因為作品旁邊沒有適當的詳細說明或隨行駐館解說人員，使得觀賞的民眾失去許多可以深入瞭解作品的機會，大幅降低了參觀的趣味性及教育意義。即使在展場中常可看到大量印刷文宣品，許多參觀民眾隨手拿了文宣，卻往往沒有發揮紙本解說的最大效益，也忘了這些紙本可以回收再利用，這些舉動往往在無形中造成了資源的浪費。

一般傳統虛擬導覽對於使用者來說，屬於較為被動的方式，都是由使用者依照指示按下滑鼠後才會有下一步的動作。然而隨著科技進步以及取像設備普及化，標榜著虛實合一的擴增實境(Augmented Reality，簡稱 AR)已成為近年來許多展場的主要應用技術。AR 是將真實環境與虛擬實境結合，提供額外輔助資訊，增加使用者對於觀察對象的瞭解，該領域的核心技術挑戰點為：系統如何得知使用者的觀察對象(辨識)和使用者的觀察角度(定位)，來給予相對應的額外資訊。目前 AR 技術又可分為有無使用特定 marker 作為區分，分別為制式標記式(fiducial marker)技術以及無標記式(markerless)技術。

一般常見的 fiducial marker 式 AR 技術(ARToolKit[1]或 ARTag[2])，為使用一正方形黑框，內含各式各樣之預先制定好的 marker 圖案(如圖 1)，其中正方形黑框用來搜尋在影像中 marker 的位置，而 marker 中的圖案用來估測使用者的視角以利後續做攝影機定位。



圖 1 由左至右為 ARToolKit 與 ARTag

然而 fiducial marker 式技術，必須在真實世界中設置額外的 marker 在欲觀察對象的旁邊，並在使用時，攝影機除了須拍攝到觀察對象之外，也須同時將 marker 入鏡，系統才能給針對目標物給予額外的資訊說明。因此，fiducial marker 的存在除了給予使用者不自然的感受，在操作上也較為不便。而 markerless[3-8]技術，則是強調直接利用現有環境的物件(牆壁、桌面)或欲觀察的對象(畫作、照片)來

做辨識與定位，不需置放額外的 marker，讓 AR 技術能有更廣的應用。Markerless AR 主要是利用特徵點擷取與對應技術，來比對攝影機拍攝的畫面是否有資料庫已知的物件，以提供額外的資訊給使用者了解。

本論文提出一套有別於傳統導覽的方式，以無標記式室內導覽技術概念，使用者只需透過取像設備的鏡頭對著展場裡的景物拍照，在畫面上除了呈現實際影像之外，系統還可以自動標示出展出作品說明及介紹，使得展場資訊更能夠無遺漏地傳達給參觀民眾。此外，除了提供展出作品查詢功能外，也利用相機定位原理，記錄使用者已看過的展覽品，呈現出參觀民眾曾走過的路線，也可以更進一步引導使用者前往還未參觀的區域，能夠更加提升民眾參觀作品的完整性。如此的無接縫服務功能，將帶給使用者更豐富、更多元的寓教於樂成效。藉由此論文的構思以及技術，便能同時兼顧教育性、娛樂性，也能達到環保的訴求。

1.2 相關研究文献

就目前的趨勢而言，一般常見的導覽系統須具備辨識以及空間定位功能。定位方式主要分為兩種，分別為絕對定位與相對定位，其中絕對定位又可細分為點定位和方向定位，常見點定位裝置如:GPS、Zigbee、Infrared ray 等，這些裝置因訊號遭建築物阻擋、易受無線電干擾、或被室內展示燈光擾亂等缺點而不適用於室內導覽技術，而方向性定位的儀器如：gyroscope、電子羅盤等，則是因為造價太過昂貴或易受磁性物品干擾，也無法將如此的室內導覽設備推廣到一般使用者身上。[9-10]所使用的定位設備為 RFID 以及 Wi-Fi，除上述缺點外，也無法提供使用者與畫作間的相對位置，因此不考慮採用這些設備。

現今無標記式導覽系統[11-13]在特徵點擷取與比對的技術上，多半是使用 Scale Invariant Feature Transform (SIFT)[14-15]，SIFT主要是結合了 scale invariant region detector 和 gradient-distribution based descriptor 的觀念。然而 SIFT 只能針對 similarity transformation 做處理，且無法對抗在 wide-baseline 下因視角變化而造成的幾何變形。法國 INRIA (Institut national de recherche en informatique et

automatique)研究機構過去曾經評估常見的affine-invariant region detectors (如 Harris-Affine、Hessian-Affine[16] 、edge-based region detector (EBR)[17] 、image-based region detector (IBR)、MSER[18])之效能與精確度，最終結論是以 MSER及Hessian-Affine為最好的 2 個detectors。至於Zernike moments(ZMs) 目前已經被廣泛地應用在物體辨識 [19-22]的主題上，其不怕位移、放大縮小及旋轉等影像變化等多項優點，都非常適合用來做為特徵描述元的方法。

1.3 研究內容概述

本論文以辨識及相機定位為兩大主軸，在畫作辨識方面使用有別於常見的 SIFT 或 Adaboost 等方法，而是基於 Maximally stable extremal regions(MSER)和 Invariant Zernike Moment 的架構上針對物體特徵點進行處理，定位方面則是克服一般點定位缺乏與標的物相對位置，利用影像透過相機校正的方法估測出使用者拍攝當時相機與畫作的關係，進而建立出使用者的參觀路線。

1.4 論文架構

本論文架構一共分為七個章節，第一章為緒論，第二章為本論文使用之辨識方法概念介紹，含括特徵點擷取、特徵點描述以及索引建立，第三章為描述資料庫製作，第四章為查詢影像辨識結果之驗證，第五章為相機定位，說明相機內部參數的估測方法，第六章是實驗結果呈現，第七章為結論。

第二章 室內畫展導覽之辨識基礎

2.1 畫展特徵點擷取

本論文利用 MSER 作為畫作特徵點擷取方式，其優點誠如 1.2 節所提及，可對抗例如位移、放大縮小、旋轉、視角改變等影像變化。它藉由影像的灰階值，並使用以下兩種方法萃取出特徵點：1. 在影像中某一灰階值以下找出最大相連 (maximally connected) 區域，2. 將灰階值增減一個範圍後，其區域的面積變化率相對最小。MSER 將影像視灰階值視為一地形剖面圖，如圖 2 所示， $R, R_{+\Delta}, R_{-\Delta}$ 及所謂的 extremal regions，橫軸為影像座標，縱軸為灰階值，縱軸上的 $I(R)$ 表示 region R 裡所包含的最大灰階值。

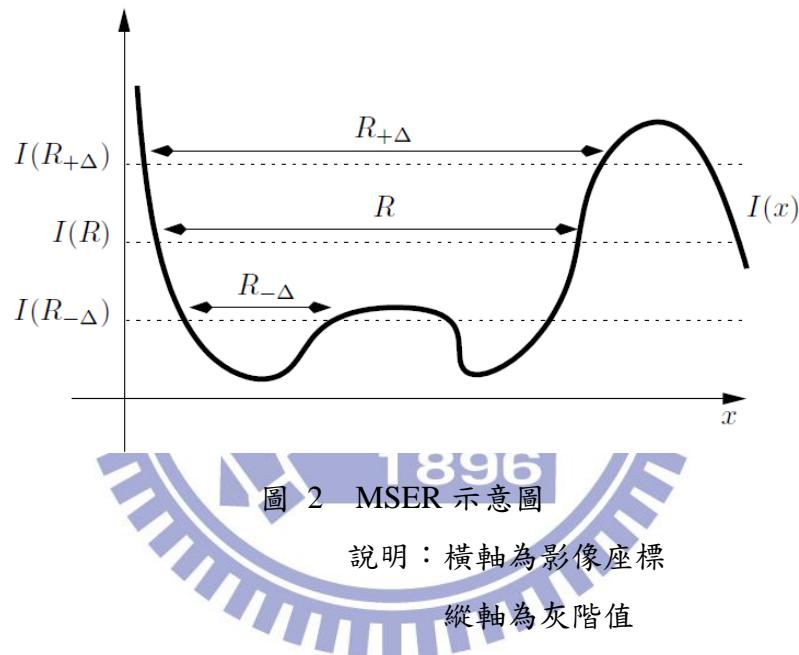


圖 2、MSER 示意圖，橫軸為影像座標，縱軸為灰階值。

Extremal regions 的成長是將灰階值以遞增方式並找尋 4-connected component 為原則，灰階值較大的區域會包含灰階值較小的，呈現階層式的關係。對於每個 extremal region 都可計算出一相對應的面積變異值(variation)，仍舊以圖 2 為說明範例，假設 $\Delta > 0$ ，將 $R_{+\Delta}$ 定義為一 extremal region，此 region 不僅包含 R ，且其灰階值超過 $I(R)$ 至少 Δ 之最小面積。假設定義 Q_1 為一個 extremal region，則定義如下：

$$R_{+\Delta} = \arg \min \left\{ |Q_1| : Q_1 \supset R, I(Q_1) \geq I(R) + \Delta \right\} \quad (2.1)$$

至於 $R_{-\Delta}$ 一樣被定義為一 extremal region，此 region 不僅被 R 所包含，且其灰階值低於 $I(R)$ 不大於 Δ 之最大面積。假設定義 Q_2 為一個 extremal region，則定義如下：

$$R_{-\Delta} = \arg \min \{ |Q_2| : Q_2 \supset R, I(Q_2) \geq I(R) - \Delta \} \quad (2.2)$$

藉由 (2.1)、(2.2) 來計算前面所提到的面積變異值 ρ ，計算方式如下：

$$\rho(R, \Delta) = (|R_{+\Delta}| - |R_{-\Delta}|) / |R|$$

當一個 extramal region R 的 ρ 同時小於 $R_{+\Delta}, R_{-\Delta}$ ，就定義 R 為最穩定的 extremal region。

2.2 畫展特徵點描述子

在第一階段擷取出畫作特徵點後，還需要將特徵點區域描述成特徵向量，以利後續的畫作辨識以及相似程度比對。本論文利用 Zernike Moments (ZMs)[2-8] 的方法進行特徵向量描述。由於 ZMs 擁有對抗位移、尺寸縮放、以及方向的不變性，因此常被應用在圖形辨識以及影像分析上。

2.2.1 ZMs 之強度以及相位介紹

本論文用到 ZMs 之基礎概念可分為兩個主軸，分別為強度 (magnitude) 以及相位 (phase)，而強度與相位的預估值來自與 basis function 的比較結果。將 basis function 視為一事先定義好的 42 維特徵向量 (每一維表示一種特徵)，這 42 綴向量互相正交且獨立，對於每一目標 pattern，利用若干個橢圓區域去包覆其擷取出來的特徵點，並且評估該特徵是否與 basis function 相似。在與這 42 綴的 basis function 比較後，則可求出這目標 pattern 的 42 綴強度以及相位向量，若與 basis function 越相近，則值越大。

Zernike Moments 是利用 Image function $f(\rho, \theta)$ 與 basis function $V_{nm}(\rho, \theta)$ 做積分得到，且將 basis function 定義在一 unit circle 中，order 為

n 、repetition 為 m ，且須符合 $n - |m|$ 為偶數以及 $|m| < n$ 的條件。

$$V_{nm}(\rho, \theta) = R_{nm}(\rho, \theta)e^{jm\theta}, \rho \leq 1$$

在此

$$R_{nm}(\rho) = \sum_{s=0}^{\lfloor(n-|m|)/2\rfloor} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}-s)!} \rho^{n-2s}$$

Zernike basis function $\{V_{nm}(\rho, \theta)\}$ 間為互相正交，式子如下：

$$\int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) V_{pq}(\rho, \theta) \rho d\rho d\theta = \frac{\pi}{n+1} \delta_{np} \delta_{mq}$$

$$\delta_{ab} = \begin{cases} 1 & a = b \\ 0 & \text{otherwise} \end{cases}$$

計算二維數位影像 Zernike Moments 的數學式如下：

$$Z_{nm} = \frac{n+1}{\pi} \sum_{(\rho, \theta) \in \text{unit disk}} f(\rho, \theta) V_{nm}^*(\rho, \theta) = |Z_{nm}| e^{i\phi_{nm}}$$

定義影像的輸入影像的特徵向量為 P

$$P = [|Z_{11}| e^{i\phi_{11}}, |Z_{31}| e^{i\phi_{31}}, \dots, |Z_{n_{\max} m_{\max}}| e^{i\phi_{n_{\max} m_{\max}}}]^T$$

Z_{nm} 為強度， ϕ_{nm} 為相位。本論文將用兩個特徵向量來表示一個 MSER region，分

別為

$$mag = [|Z_{11}|, |Z_{31}|, \dots, |Z_{n_{\max} m_{\max}}|]^T$$

$$phase = [\phi_{11}, \phi_{31}, \dots, \phi_{n_{\max} m_{\max}}]^T$$

2.2.2 ZMs 之相似度計算

假設存在兩組 ZMs 特徵向量 P^1 : mag^1 、 $phase^1$ 與 P^2 : mag^2 、 $phase^2$ ，

其中 $mag^1 = [|Z_{11}^1|, |Z_{31}^1|, \dots, |Z_{n_{\max} m_{\max}}^1|]^T$ ， $phase^1 = [\phi_{11}^1, \phi_{31}^1, \dots, \phi_{n_{\max} m_{\max}}^1]^T$ ；

$mag^2 = [|Z_{11}^2|, |Z_{31}^2|, \dots, |Z_{n_{\max} m_{\max}}^2|]^T$ ， $phase^2 = [\phi_{11}^2, \phi_{31}^2, \dots, \phi_{n_{\max} m_{\max}}^2]^T$

相似度計算數學式子如下：

- (1) 強度的相似度定義為兩向量的夾角餘弦值，變動範圍為 $[0,1]$ ，夾角餘弦值越大，代表相似程度越高。

$$S_{mag}(mag^1, mag^2) = \frac{mag^1 \cdot mag^2}{\|mag^1\| \|mag^2\|}$$

- (2) 相位的每個維度為一個角度，由於相位有循環性(circularity)，這意味 θ 與 $\theta + \gamma * 2\pi$ 是一樣的表示，其中 γ 為任意整數，而本論文定義相位的相似度為

$$S_{phase}(phase^1, phase^2) = 1 - \sum_m \sum_n w_{nm} \frac{\min\{|\Phi_{nm} - (m\hat{\alpha}) \bmod(2\pi)|, 2\pi - |\Phi_{nm} - (m\hat{\alpha}) \bmod(2\pi)|\}}{\pi}$$

在此，

$$w_{nm} = \frac{|Z_{nm}^1| + |Z_{nm}^2|}{\sum_{n,m} |Z_{nm}^1| + |Z_{nm}^2|}$$

$$\Phi_{nm} = (\Phi_{nm}^1 - \Phi_{nm}^2) \bmod 2\pi$$

是實際相位差值，旋轉角度 $\hat{\alpha}$ 是藉由遞迴性(iteratively)地計算 $\hat{\alpha}_m = (\Phi_{nm} - \hat{\alpha}_{m-1}) \bmod 2\pi$ 所求得，而初始 $\hat{\alpha}_0 = 0$ ， $S_{phase}(phase^1, phase^2)$ 的範圍為 $[0,1]$ ，值越大代表越相似。

$$S_{phase} = (\rho_{Phase^1} \quad \rho_{Phase^2}) = 1 - \sum_m \sum_n w_{nm} \frac{\min\{|\varphi_{nm}|, 2\pi - |\varphi_{nm}|\}}{\pi}$$

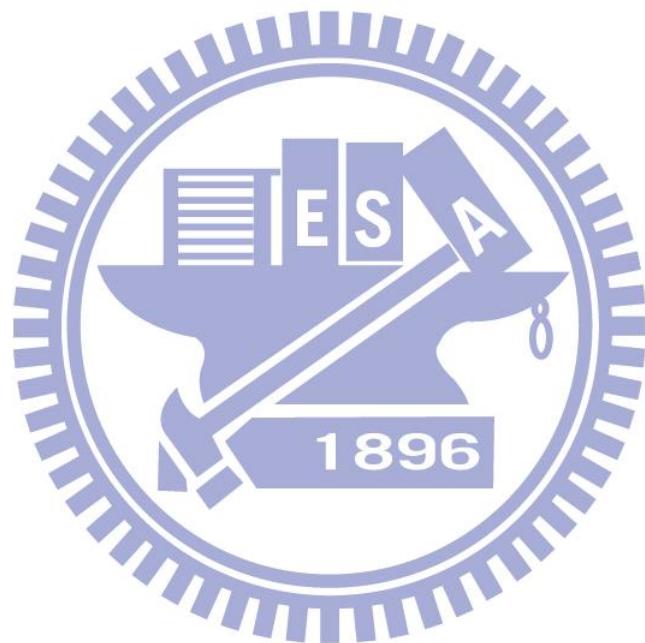
2.3 畫展畫作特徵點描述子 k-度空間樹狀(k-d tree)索引之建立與投票

機制

藉由 2.2 節 descriptor 相似度比較的計算方法，已可達成 2 張影像間的點對應求取，但對導覽系統來說，需要去搜尋出查詢影像的點對應到影像資料庫中哪一張影像，倘若用暴力式搜尋與資料庫內的每一張影像做點對應比對，顯然非常耗時。, 本論文利用建立索引(indexing)機制，事先為資料庫中的影像建立索引，以相對簡易的方式先搜尋較可能的對應影像後，再作細部的特徵點對應，由於畫作特徵點的資料量龐大，必須使用適合快速搜尋最近點的資料結構來儲存這些資訊，以加快特徵點比對時的速度，而 kd-tree 演算法正好提供了一個 k 維空間快

速搜尋最相近點的資料結構。

Kd-tree 為一均衡二元數，假設已制定 A, B, C, \dots 等 n 種特徵向量，則訂定樹的第一層 root 為特徵向量 A 之強度向量的平均值，第二層兩個 roots 為特徵向量 B 之強度向量的平均值，依序類推下去，當一查詢特徵向量進入 kd-tree 檢索時，會先搜尋出最接近的強度向量，接著再比對位向向量，若兩種向量的相似度皆大於某一門檻值，則將樹中該特徵向量投下一票，再遞迴性地依序完成其他特徵向量比對及投票動作。由於畫作通常是由一群特徵點所構成，所有資料庫影像的每個特徵向量都有可能獲得票數，而得票數最高的影像代表與查詢影像有最多相似的特徵點，則以該資料庫影像做為辨識結果。



第三章 系統架構及畫展資料庫內容索引建立

本章節承接第二章辨識基礎理論，將介紹系統執行流程以及利用實例畫作建立資料庫的方法，包含畫作尺寸丈量與在歐式空間的位置制定，畫作特徵點描述子索引建立為畫作辨識的依據，以及畫作特徵點描述子之 Homography 矩形調整之計算，作為後續相機軌跡估算之基礎。

3.1 系統架構介紹

為了使本論文目標說明以更清楚的方式呈現，以下將使用者在展場中實際使用系統之情境區分為三大功能，分別為畫作辨識、相機定位以及參觀路徑追蹤。當使用者使手持具有拍照功能的智慧型手機對著展場中感興趣的畫作拍照，傳送至系統後即自動偵測查詢影像的特徵點，並與資料庫中影像搜尋比對，得到初步辨識結果後，透過 Homography 機制來驗證查詢結果，以確保所欲回傳資料之正確性。由於 Homography 需要正確的特徵點對應關係，在此加入了特徵點篩選步驟，可使該系統更為強健。流程如圖 3 為示：

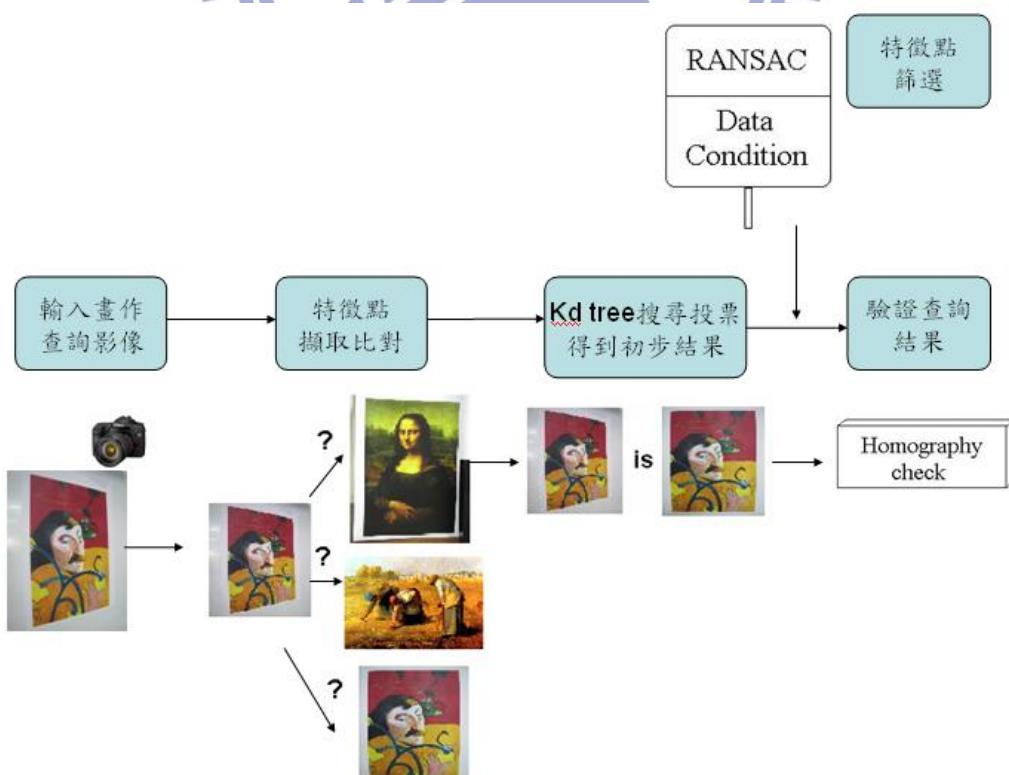


圖 3 畫作辨識流程圖

為了提供互動式的畫作相關資訊解說，系統利用相機定位原理，估算出使用者拍攝的視角，當拍照的位置改變，即回傳不同面貌的作品解說，使參觀民眾對於畫作具有更深入的了解。此外本系統另外規劃展場畫作的空間配置關係，不僅可以作為展場室內導航功能，協助參觀民眾前往感興趣的展示區，並且在使用者即將結束當日的參觀行程時，系統還提供了參觀路徑追蹤功能，回顧當日所欣賞過的作品，並且提醒使用者還有哪些遺漏的推薦作品，作為下次參觀路徑的建議，提供更為客製化的服務。示意流程分別為圖 4、5 為示：

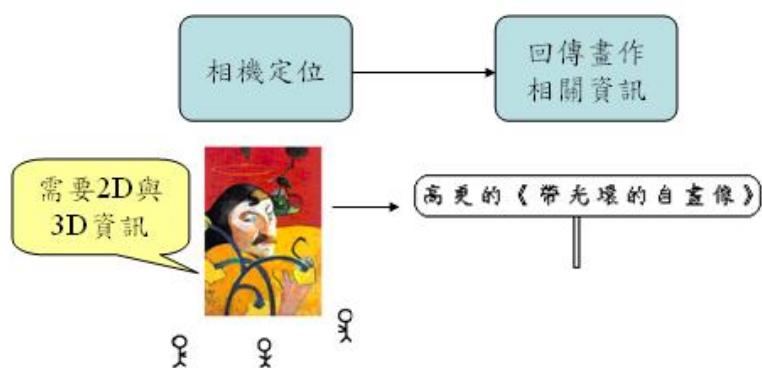


圖 4 相機定位流程圖

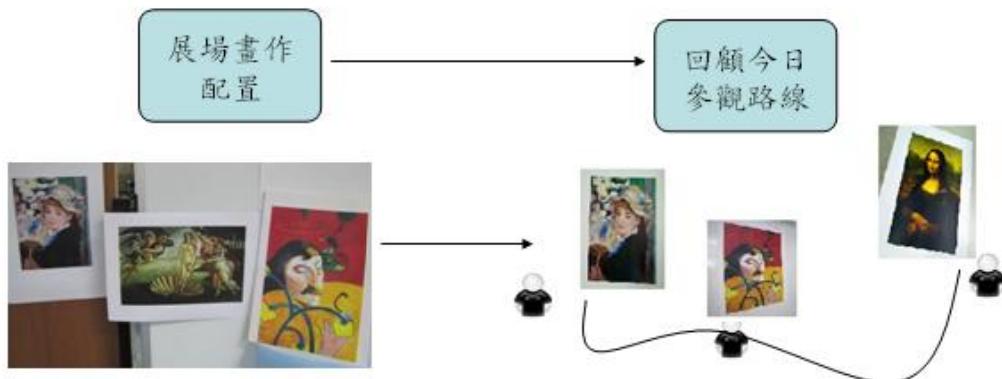


圖 5 畫作配置及參觀路徑追蹤流程圖

3.2 資料庫製作

為了達成上述系統功能目標，必須事先規劃好資料庫，完善的資料庫製作才能提供使用者更佳的服務功能，其製作內容包括實際畫作資料設定、資料庫影像取得、資料庫影像特徵點擷取，執行方法將在以下章節分別介紹。

3.2.1 畫作實際尺寸丈量與畫作 3D 點制定

由於世界座標系統沒有統一的規格，必須透過假設來制定物體在世界座標下的 3D 點座標，首先需測量畫作的實際長寬，若畫作尺寸為 $x * y (\text{mm}^2)$ ，令畫作上的每一點都躺在 $z = 0$ 的平面上，則訂定 3D 點座標的方法如圖 6 所示：

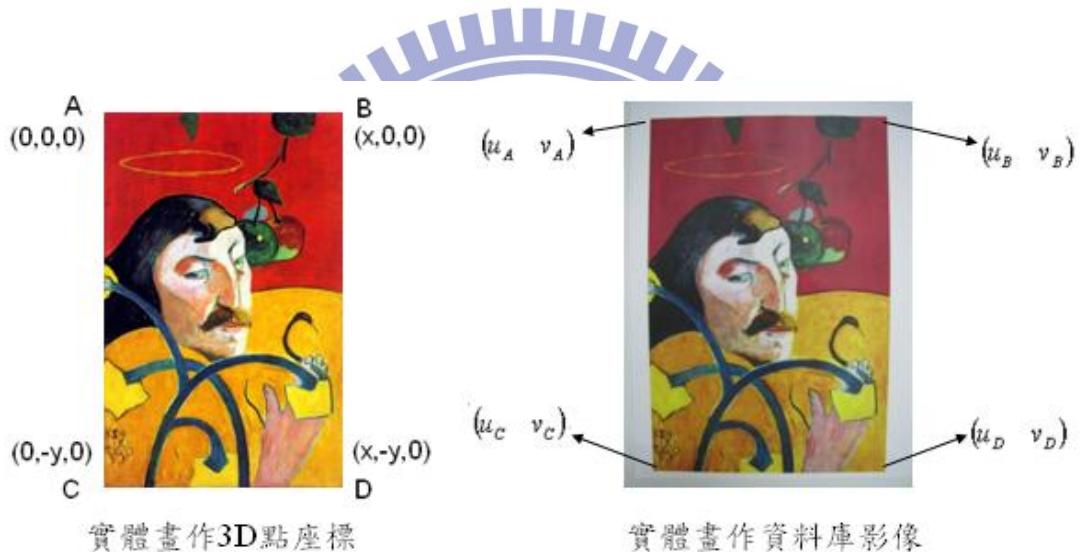


圖 6 實體畫作座標示意圖

由圖 6，另左圖點 A 為世界座標原點 $(0,0,0)$ ， \overrightarrow{AB} 與 x 軸平行， \overrightarrow{CA} 與 y 軸平行，利用[24]之假設，令物體上的點都躺在 $z=0$ 的平面上，考慮到畫作實際長寬比例，可定 A,B,C,D 4 個 corners 的 3D 點座標分別為 $(0,0,0), (x,0,0), (0,-y,0), (x,-y,0)$ 。對著畫作拍攝一張影像，畫作面積需盡量填滿整個畫面，影像需清晰避免手震造成模糊，每張影像中只含有一張畫作，在影像中可得知畫作 4 個 corners 的 2D 點座標，基於 4 個 2D 點、對應的 3D 點之關係

以及所有 3D 點都躺在 $z=0$ 平面上的假設，可列出數學式子如下：

$$s \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}_{3*1} = K \begin{bmatrix} r1 & r2 & r3 & t \end{bmatrix}_{3*4} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}_{4*1} = K \begin{bmatrix} r1 & r2 & r3 & t \end{bmatrix}_{3*4} \begin{bmatrix} X_i \\ Y_i \\ 0 \\ 1 \end{bmatrix}_{4*1} \cong K \begin{bmatrix} r1 & r2 & t \end{bmatrix}_{3*3} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix}_{3*1},$$

$i = 1 \sim 4 \wedge \wedge (3.1)$

其中 $\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}$ 為成像平面的 4 個 corners 的 2D 點座標， s 為深度資訊， K 是相機內部參數， $[r1 \ r2 \ r3 \ t]$ 為旋轉以及位移矩陣， $\begin{bmatrix} X_i \\ Y_i \\ 0 \\ 1 \end{bmatrix}$ 為物體的 4 個 corners 3D 點座標，由公式 (3.1)，令 $H_{DB} = K \begin{bmatrix} r1 & r2 & t \end{bmatrix}$

故 $s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \cong H_{DB} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \wedge \wedge (3.2)$

因為實際畫作與拍攝影像皆為平面，利用該特性可以將此對應關係視為平面與平面的對應，得到資料庫影像與實際畫作(2D ground truth)間的 homography matrix H_{DB} 後，在 3.2.2 節裡將利用 H_{DB} 找出畫作上所有 2D 特徵點所對應到的 2D ground truth 座標。

3.2.2 畫展畫作 ZMs 描述子與特徵點位置之資料庫建立

對於資料庫中拍攝的每一張影像，利用 2.1 節 MSER 特徵點擷取與 2.2 節 Zernike Moment 特徵點描述子，可求得資料庫裡每一張影像的特徵向量集合，如圖 7 所示，紅色橢圓表示特徵區域。每個橢圓所包含的特徵點區域，可計算出一個相對應的 Zernike Moment 的特徵，如 2.2 節所提到的 magnitude 向量與 phase 向量兩部份以及該特徵點的 2D 點座標。



圖 7 資料庫影像之 MSER 特徵點區域示意圖

說明：紅色橢圓表示所偵測出的特徵區域

此為高更的《帶光環的自畫像》

當搜尋完畢資料庫畫作的特徵點後，必須先為這些特徵點建立其相對應的 2D ground truth 資訊，以便後續查詢影像在找完特徵點時，可以很快的計算出查詢影像與實際畫作的 Homography 關係。要建立資料庫畫作每一個特徵點的 2D ground truth 資訊，可以先用手點方式點選畫作的四個邊框之 corners 的 2D 特徵點位置(pixels)，而這四個 corners 分別又對應到 3.2.1 節描述的位置 $(0,0,0), (x,0,0), (0,-y,0), (x,-y,0)$ ，由公式(3.2)得知資料庫畫作與實際畫作之間的 Homography H_{DB} 矩陣關係式，當已知資料庫畫作特徵點的 2D 座標時，可以用公式(3.2)反算出該 2D ground truth 座標，求出 2D ground truth 座標後便可利用此資訊作為後續相機定位之用。

$$H_{DB}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \Lambda \Lambda \quad (3.3)$$

第四章 拍攝畫作之影像查詢

上一章節所介紹的是系統流程架構以及畫作資料庫前置作業等相關設定，而第四章所要討論的是關於查詢影像，包含當使用者拍攝一張影像傳送到系統後，系統該如何回傳查詢結果以給予使用者回應，並且驗證系統回傳結果的正確性。

4.1 拍攝畫作之 kd-tree 查詢結果投票名次

由 2.1 至 2.3 節所提方法，當使用者輸入影像後，系統對於查詢影像會先偵測 MSER 橢圓特徵點區域並計算 ZMs 的特徵描述向量，利用強度與相位兩種向量，搭配 kd-tree 之索引結構，將資料庫中與查詢影像最相似的特徵向量搜尋出來，並統計各張資料庫影像擁有與查詢影像相似的特徵點數，將這些點數視為投票票數，最後取出最高票的資料庫畫作影像，當作初步辨識結果。

4.2 畫作查詢結果之驗證

4.2.1 資料庫影像與查詢影像之 Homography 估測

由 4.1 節的初步結果，可得知查詢影像所對應到的資料庫影像，利用這兩張影像間相互對應的 2D 特徵點，即可計算出彼此之間 Homography matrix H_{image} 的關係式：

$$s \begin{bmatrix} u_{DB} \\ v_{DB} \\ 1 \end{bmatrix}_{3*1} = H_{image} \begin{bmatrix} u_{query} \\ v_{query} \\ 1 \end{bmatrix}_{3*1} \Lambda \Lambda \quad (4.1)$$

其中 s 表示一 scalar， $\begin{bmatrix} u_{DB} \\ v_{DB} \\ 1 \end{bmatrix}$ 是資料庫影像中的某一特徵點像素座標，該點對應到

查詢影像上像素座標為 $\begin{bmatrix} u_{query} \\ v_{query} \\ 1 \end{bmatrix}$ 。將公式 (4.1) 展開可整理成式子如下：

$$\text{Let } H_{image} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}_{3 \times 3}$$

將公式(4.1)重新整理為

$$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} u_{query} \\ v_{query} \\ 1 \end{bmatrix} = \begin{bmatrix} su_{DB} \\ sv_{DB} \\ s \end{bmatrix}$$

$$\Rightarrow h_{11}u_{query} + h_{12}v_{query} + h_{13} = su_{DB} \quad (1)$$

$$h_{21}u_{query} + h_{22}v_{query} + h_{23} = sv_{DB} \quad (2)$$

$$h_{31}u_{query} + h_{32}v_{query} + h_{33} = s \quad (3)$$

將式(3)代入式(1)(2)可再整理成：



$$\begin{bmatrix} u_{query} & v_{query} & 1 & 0 & 0 & -u_{DB}u_{query} & -u_{DB}u_{query} & -u_{DB} \\ 0 & 0 & 0 & u_{query} & v_{query} & 1 & -v_{DB}u_{query} & -v_{DB}u_{query} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 0 \quad (4.2)$$

由公式(4.2)，假設 $h_{33}=1$ ，剩餘 8 個未知數，要求解一個 $3 \times 3 H_{image}$ 至少需要 4 組對應點，而這 4 組對應點，必須要沒有包含對應錯誤的情形(outliers)，且座標越精確，算出來的 H_{image} 才會越正確。考慮到 MSER 所找到的特徵點座標，座標難免會有偏差，且 ZMs 有時也會發生對應點對錯的情形(如圖案有重覆或很相似的時候)，為了讓本論文計算 H_{image} 能更為可靠，4.2.2 節將介紹使用 RANSAC 和

data condition 二方法來讓計算結果更為強健(robust)。

4.2.2 拍攝畫作之特徵點篩選與驗證正確率之計算

使用者所拍攝的查詢影像，除了過度模糊或主體佔鏡頭畫面面積太小之外，該畫作所擷取出的特徵點總數通常會大於 4 點，而這些特徵點當中又可能包含了對應錯誤的情形，因此需要加入篩選特徵點的機制以求出最佳 Homography matrix，以下將介紹三個步驟，分別為 RANSAC 、搜尋反投影點之最小 median error 以及 data condition ，以求出最佳的 Homography matrix H 。

4.2.2.1 利用 RANSAC 與 Minimum Median Error 檢驗找出最佳 H

假設資料庫影像與查詢影像之間一共有 m 組對應點，在計算 H_{image} 時如前面所述，只有當至少 4 組點都不包含錯誤的對應點時，結果才會好。因此利用 RANSAC 的精神，從 m 組對應點隨機選 4 組來算，重覆此動作 n 次，然後選取這 n 次中最好的結果來做為 H_{image} 的解。為了評估每一次所估測結果的好壞，先從 m 組對應點隨機選 4 個點來計算 Homography matrix H ，接著利用該 H ，將查詢影像其餘 $m - 4$ 點反投到資料庫影像上而得到： $D_{estimated,j}, j = 5 \sim m$ ，這些反投影點與資料庫影像的 2D 對應點： $D_{ground_truth,j}, j = 5 \sim m$ 必定存在些許誤差，為了避免極端狀況，取出這些誤差的中間值(median error)，記錄為 $medErr_i, i = 1 \sim n$ ，做為評估這組解好壞的依據。若希望從 n 個 Homography matrices $H_i, i = 1 \sim n$ 中挑選出最佳估測結果，可將 $medErr_i$ 排序並且取出 $\min(medErr_i)$ ，再將該 $\min(medErr_i)$ 所對應到的 H_i 定為最佳的 Homography matrix H_{best} ，以下為演算法流程：

Input : m corresponding feature points,

n iteration rounds

Output : H_{best}

if $m \geq 4$

for $i = 1 \sim n$

$set_i = get_candidate_featurePt(m)$

$H_i = getHomography(set_i(1:4))$

for $j = set_i(5:m)$

$D_{estimated}(j) = H_i * set_i(j)$

$Err_j = D_{ground_truth}(j) - D_{estimated}(j)$

end

$medErr_i = median(Err)$

end

$index = \min(medErr)$

$H_{best} = H_{index}$

$get_candidate_featurePt(featurePtnumber\ x)$

$result = randomPermute(1 \sim x)$

return result

1896

4.2.2.2 使用 Data Condition 減少數值誤差

假設有一張 $M \times N$ pixels 的影像，影像中任一點 p 的座標為

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \text{ for } u = 1 \sim M, v = 1 \sim N, \text{ 由於一般習慣將影像原點設定在影像的左上角，則}$$

影像 4 個corners到原點的距離不一，距離越遠的數值變化越大，以上情形將會造成計算時的誤差。為了避免這種數值誤差(numeric error)，利用Hartley所提出之方法[25]，在利用對應的 2D 特徵點座標計算Homography matrix時，可先將影像原點移至影像中心點，使 4 個corners到影像原點的距離皆相等，再將 2D 特徵點座標的長、寬(u, v)可變動的範圍都壓縮至 $[-1 \ 1]$ 之間以避免數值計算誤差，使

估算出的Homography matrix能夠更精準。

4.3 利用 Homography matrix H_{image} 之查詢結果驗證

由 4.1 節的投票結果，可找出得票數最高的候選影像，但仍無法確定該候選影像是否就是使用者所輸入之查詢影像，因此在得到辨識階段初步結果後，利用 rectification[26]的觀念，將查詢影像搭配 4.2 之 H_{image} ，使查詢影像轉正後與資料庫回傳影像疊合，若兩張影像間的差值越小則越相似，回傳影像之可信度也越高。方法描述如下：

令資料庫回傳影像為 I_{DB} ，查詢影像為 I_{query} ，先產生一張空白影像 I_{DB}' ，其尺寸

大小與 I_{DB} 相同。 I_{DB}' 中的每一影像座標 $\begin{bmatrix} u_{DB}' \\ v_{DB}' \\ 1 \end{bmatrix}$ ，可利用公式(4.1)找出在 I_{query} 中

所對應到的影像座標 $\begin{bmatrix} u_{query} \\ v_{query} \\ 1 \end{bmatrix}$ ，則可將每一點 $\begin{bmatrix} u_{DB}' \\ v_{DB}' \\ 1 \end{bmatrix}$ 的 RGB 資訊用 $\begin{bmatrix} u_{query} \\ v_{query} \\ 1 \end{bmatrix}$ 來取

代，若遇到轉換後的座標資訊不是整數，可用 2 維內插法求出該對應點之資訊。

如此便可製造出一張與 I_{DB} 拍攝視角相同且內容是取自於 I_{query} 的影像 I_{DB}' 。將

I_{DB} 與 I_{DB}' 相減，若 RGB 值相減的平均誤差越小，也就代表 I_{DB}' 與 I_{DB} 越相近。

圖 8 為檢驗流程示意圖，左子圖為查詢影像對對應到的前三名，右子圖為查詢影像轉正後的結果；圖 9 為 Homography 驗證結果示意圖。



圖 8 Homography 檢驗流程示意

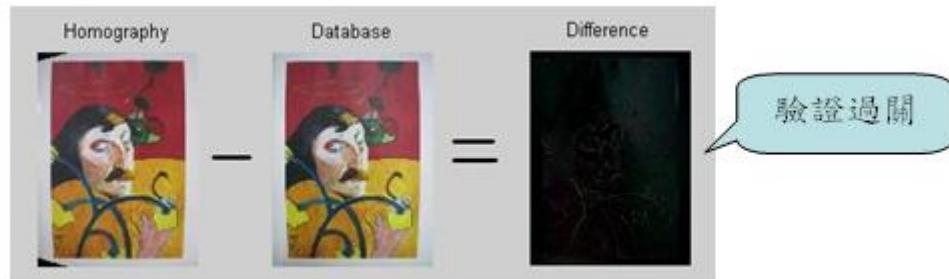


圖 9 Homography 驗證結果示意圖

其演算法如下：

```

 $H_{image} = getHomo(I_{DB}, I_{query});$ 
 $(size\_y \ size\_x) = getSize(I_{DB});$ 
 $I_{DB}' = createEmptyPic(size\_y \ size\_x);$ 
 $\text{for } i = 1 : size\_y$ 
     $\text{for } j = 1 : size\_x$ 
        
$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = H_{image}^{-1} * \begin{bmatrix} j \\ i \\ 1 \end{bmatrix};$$

         $I_{DB}'(i, j) = get\_RGB(I_{query}(u, v));$ 
     $\text{end}$ 
 $\text{end}$ 
 $difference = I_{DB} - I_{DB}';$ 

```

第五章 畫作與相機之相對位置估測

第二至四章節部分所討論的是有關畫作辨識、驗證辨識結果正確性以及 Homography matrix 的估測方法，而本章節所要陳述的主題為相機定位，為了使系統與使用者之間更具有互動性，必須估測出畫作與使用者相機位置間的相對關係。以下分為三個子章節，分別為相機定位估測、畫作空間配置圖製作以及使用者參觀路徑重建。

5.1 畫作與相機之空間座標相對位置的計算

5.1.1 內部參數之計算

由公式(3.1)(3.2)可得知，Homography matrix 是由內部參數 K 乘以 projection matrix $[R \ t]$ 的第一、二、四 columns 所得來，為了與 4.2 節所討論的 H_{image} (指兩張影像間的 Homography matrix) 作區別，本章節定義 Homography matrix 為 H_R ，所代表意義為實際畫作與單張 2D 影像間的 Homography matrix。

[24]所提到分解相機內部參數的方法，主要是利用 2 至 3 張影像，也就是使用 2 到 3 個 Homography matrices 來分解出 K ，然而在現實中無法規定使用者所拍攝的影像張數，因此將分解 K 的方法做適當彈性調整，只需使用一張影像即可

推算出 K 。令相機內部參數 $K = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$ ， α 、 β 為 u 、 v 兩軸的 aspect ratio，

γ 為兩軸的歪斜程度，若影像座標的兩個座標軸呈 90° ，則 $\gamma=0$ ， $(u_0 \ v_0)$ 為影像中心點，由公式(4.2)可得知一張影像可提供兩個等式，因此假設 $\gamma=0$ 以及 (u_0, v_0) 為影像之正中央位置，利用上述資訊來求解剩下兩個未知數 α 、 β 。令

$$B = K^{-T} K^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{\gamma}{\alpha^2 \beta} & \frac{(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta} \\ -\frac{\gamma}{\alpha^2 \beta} & \frac{\gamma}{\alpha^2 \beta^2} + \frac{1}{\beta^2} & -\frac{\gamma(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} - \frac{v_0}{\beta^2} \\ \frac{(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta} & -\frac{\gamma(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} - \frac{v_0}{\beta^2} & \frac{(v_0 \gamma - u_0 \beta)^2}{\alpha^2 \beta^2} + \frac{v_0^2}{\beta^2} + 1 \end{bmatrix}$$

將 $\gamma = 0$ 代入可整理成

$$B = \begin{bmatrix} \frac{1}{\alpha^2} & 0 & -\frac{u_0}{\alpha^2} \\ 0 & \frac{1}{\beta^2} & -\frac{v_0}{\beta^2} \\ -\frac{u_0}{\alpha^2} & -\frac{v_0}{\beta^2} & \left(\frac{u_0^2}{\alpha^2} + \frac{v_0^2}{\beta^2} + 1 \right) \end{bmatrix}$$

其中只有 α 、 β 為未知， u_0 、 v_0 為已知，因此只需解出 B_{11} 、 B_{22} 即可求出 K 。

藉由[24]所使用的旋轉矩陣中任兩軸向量互為正交，且兩軸向量長度相等的特性，可推導出以下式子：

令旋轉矩陣 $= R = [r1 \ r2 \ r3]_{3*3}$ ，位移矩陣 $= t_{3*1}$ ，因此 $H_R = K \begin{bmatrix} R(r1 \ r2) & t \end{bmatrix}_{3*3}$ ，其

中定義 $H_R = [h_1 \ h_2 \ h_3]$ 。利用上述旋轉矩陣之特性得知

$$h_1^T K^{-T} K^{-1} h_2 = 0 \quad \Lambda \Lambda (5.1)$$

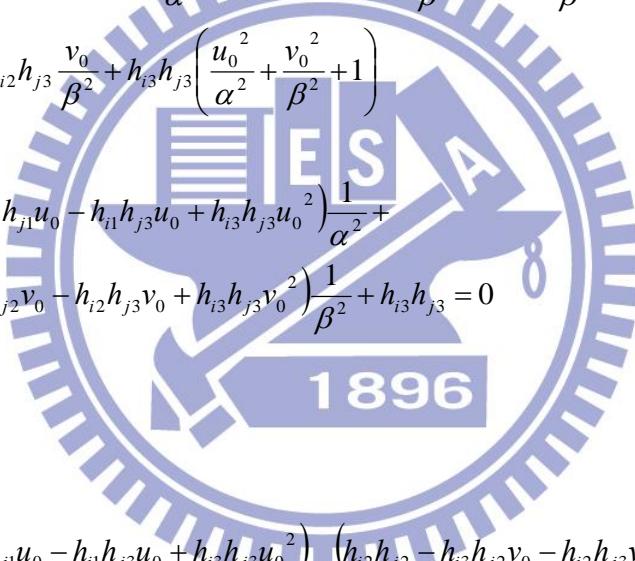
$$h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2 \quad \Lambda \Lambda (5.2)$$

並定義 H_R 第 i 個 column 為 $h_i^T = [h_{i1} \ h_{i2} \ h_{i3}]^T$ ，第 j 個 column 為

$h_j^T = [h_{j1} \ h_{j2} \ h_{j3}]^T$ ， $i \neq j$ ，由公式(5.1)(5.2)可列出以下式子：

$$\begin{aligned}
h_i^T B h_j &= \begin{bmatrix} h_{i1} & h_{i2} & h_{i3} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} \begin{bmatrix} h_{j1} \\ h_{j2} \\ h_{j3} \end{bmatrix} = 0 \\
&\Rightarrow \\
&\left(h_{i1}h_{j1}B_{11} + h_{i2}h_{j1}B_{12} + h_{i3}h_{j1}B_{13} \right) + \left(h_{i1}h_{j2}B_{12} + h_{i2}h_{j2}B_{22} + h_{i3}h_{j2}B_{23} \right) + \\
&\left(h_{i1}h_{j3}B_{13} + h_{i2}h_{j3}B_{23} + h_{i3}h_{j3}B_{33} \right) = 0
\end{aligned}$$

將 $B_{11} = \frac{1}{\alpha^2}$, $B_{22} = \frac{1}{\beta^2}$, $\gamma = 0$ 代入再將上式整理成：



$$\begin{aligned}
&h_{i1}h_{j1}\frac{1}{\alpha^2} + h_{i2}h_{j1}0 - h_{i3}h_{j1}\frac{u_0}{\alpha^2} + h_{i1}h_{j2}0 + h_{i2}h_{j2}\frac{1}{\beta^2} - h_{i3}h_{j2}\frac{v_0}{\beta^2} - \\
&h_{i1}h_{j3}\frac{u_0}{\alpha^2} - h_{i2}h_{j3}\frac{v_0}{\beta^2} + h_{i3}h_{j3}\left(\frac{u_0^2}{\alpha^2} + \frac{v_0^2}{\beta^2} + 1\right) \\
&= (h_{i1}h_{j1} - h_{i3}h_{j1}u_0 - h_{i1}h_{j3}u_0 + h_{i3}h_{j3}u_0^2)\frac{1}{\alpha^2} + \\
&(h_{i2}h_{j2} - h_{i3}h_{j2}v_0 - h_{i2}h_{j3}v_0 + h_{i3}h_{j3}v_0^2)\frac{1}{\beta^2} + h_{i3}h_{j3} = 0 \\
&\therefore \\
&\left[\begin{array}{c} (h_{i1}h_{j1} - h_{i3}h_{j1}u_0 - h_{i1}h_{j3}u_0 + h_{i3}h_{j3}u_0^2) \\ (h_{i2}h_{j2} - h_{i3}h_{j2}v_0 - h_{i2}h_{j3}v_0 + h_{i3}h_{j3}v_0^2) \\ h_{i3}h_{j3} \end{array} \right] \begin{bmatrix} \frac{1}{\alpha^2} \\ \frac{1}{\beta^2} \\ 1 \end{bmatrix} = 0 \quad \Lambda \Lambda (5.3)
\end{aligned}$$

觀察 (5.3) 式可發現該式為 $Ax = 0$ 的型式，其中 A 為已知， x 為未知，若要求解 x ，可對 A 做 singular value decomposition 來得到 x 之解，也就是求得 α 、 β ，還原出 K 。

5.1.2 外部參數之計算

當求出內部參數 K 後，藉由之前定旋轉矩陣 $R = [r1 \ r2 \ r3]$ ，且已知

homography matrix 式子為 $H_R = K \begin{bmatrix} R(r_1 & r_2) & t \\ 0 & 0 & 0 \end{bmatrix}$ ，當 H_R 已知，並且承 5.1.1 節計算出內部參數 K ，利用以上資訊便可以計算出將旋轉矩陣以及位移矩陣，利用內部參數反矩陣分別乘上 homography matrix 的第一、二、三個 column，可求出旋轉矩陣的第一、二 columns，以及位移矩陣，而剩下未知的旋轉矩陣第三軸，可藉由前面所提兩軸正交之特性，利用旋轉矩陣已知的兩軸進行向量外積計算，即可得到第三軸資訊，其公式如下：

$$\begin{aligned} r_1 &= \lambda K^{-1} h_1 & r_2 &= \lambda K^{-1} h_2 & r_3 &= r_1 \times r_2 \\ t &= \lambda K^{-1} h_3 \end{aligned}$$

在此 λ 為一 scalar 且定義為 $\lambda = \frac{1}{\|K^{-1}h_1\|} = \frac{1}{\|K^{-1}h_2\|}$ ，藉由以上步驟即可順利求出相機外部參數，當計算出外部參數後，也就代表系統得知使用者相機對於實體畫作的相對位置，利用這樣的資訊才能藉由相機視角的變化做為擴增實境的基礎，以提供互動式的畫作解說。圖 10 為相機定位流程示意圖，左子圖為模擬使用者站在不同角度拍照以及所輸入的查詢影像，右子圖為系統所估測出的相機位置。

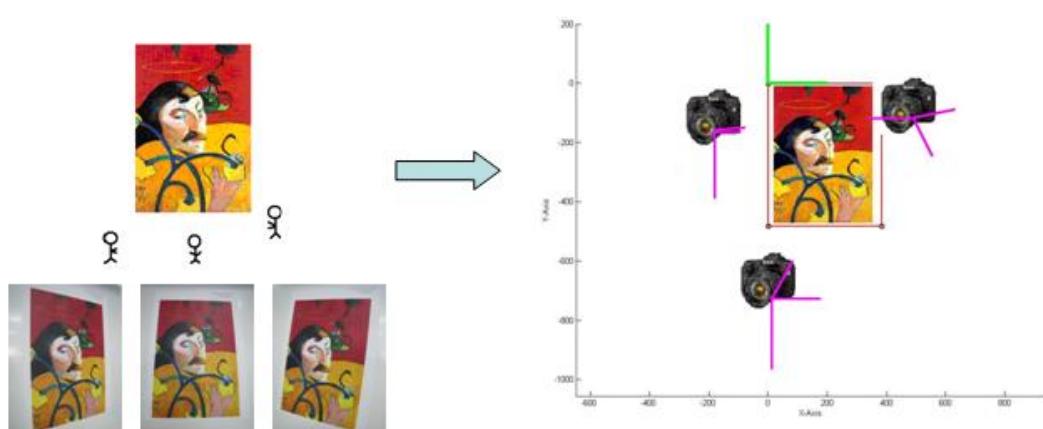


圖 10 相機定位流程示意圖

5.2 畫展畫作空間配置圖之建立

在 5.1 節中，系統是將畫展中的所有畫作視為獨立個體，每張畫作都擁有各自的座標系統，然而為了提供使用者室內參觀方向導覽以及參觀路徑追蹤的服務，必須事先知道在展場中所有畫作的空間配置圖，也就是需要串聯起原本各自獨立的畫作，定義出一個共同的世界座標原點後，再找出彼此之間的空間相對關係，圖 11 為串連畫作空間配置關係示意圖。



圖 11 串連畫作空間配置關係示意圖

為了使原本各自獨立的畫作與其他畫作產生關聯，需要以拍照方式並至少以兩張畫作同時入鏡為原則，才能找出畫作間的相對關係，如圖 12 所示，(a)至(c)分別為單張畫作影像，同時拍攝兩張畫作，串聯起展場畫作空間配置關係。



圖 12 畫作空間配置建立方法示意圖

(a)獨立影像

(b) 串聯畫作

(c)所有畫作空間配置

假設在展場中存在兩張畫作 A 、 B ，各自有擁有獨立的座標系，且原點分別為 O_A 、 O_B ，當同時對著 A 、 B 兩張畫作拍照，取得兩張畫作同時入鏡之影像後，藉由[24]利用兩張影像的方式求出一組共同的 K ，並分解出 $\begin{bmatrix} {}^C R & t_A \end{bmatrix}$ 、

$\begin{bmatrix} {}^C R & t_B \end{bmatrix}$ ，分別為畫作 A 與相機 C 的相對位置以及畫作 B 與相機 C 的相對位

置，並將相機 C 之原點定義為 O_C ，其中 $t_A = {}^C(\overrightarrow{O_C O_A})$ ， $t_B = {}^C(\overrightarrow{O_C O_B})$ ，由於要在畫面中同時呈現 A 、 B 兩張畫作配置圖，若以 A 畫作為基準，則需要將 B 畫作之

$\begin{bmatrix} {}^C R & t_B \end{bmatrix}$ 轉換到以 A 為參考原點的座標系。假設在空間裡存在一點 P ，由 A 畫

作之座標系看 P 定義為 ${}^A P$ ， B 畫作之座標系看 P 定義為 ${}^B P$ ，相機 C 之座標系看

P 定義為 ${}^C P$ ，圖 13 為上述關係示意圖。

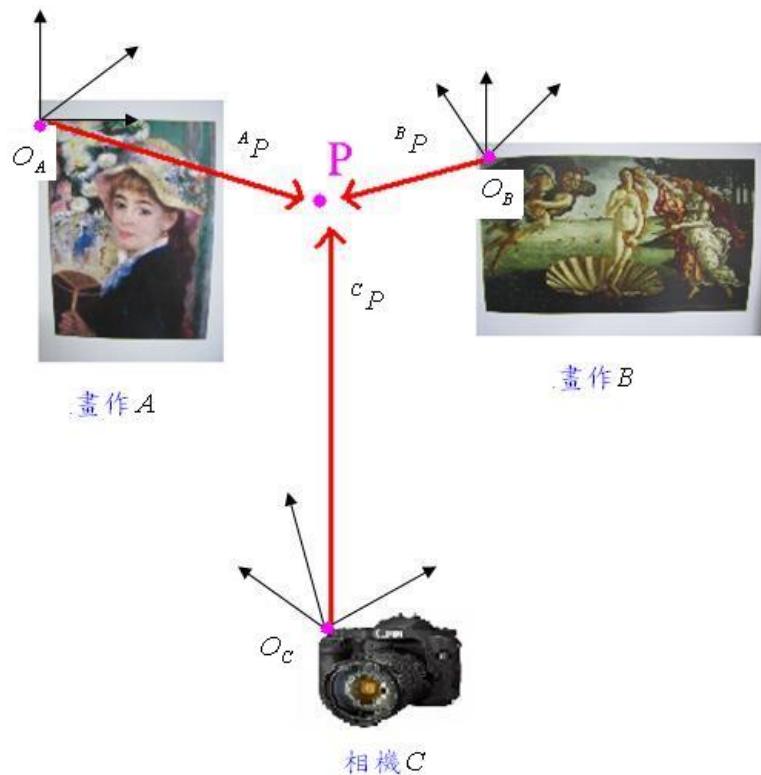


圖 13 三個座標系觀察共同一點 P 示意圖

藉由 $[{}^C R \ t_A]$ 及 $[{}^B R \ t_B]$ 可列出關係式如下：

$${}^C P = [{}^C R \ t_A] {}^A P, \quad {}^C P = {}^C R {}^A P + t_A \quad \text{AA} \quad (5.4)$$

$${}^C P = [{}^B R \ t_B] {}^B P, \quad {}^C P = {}^C R {}^B P + t_B \quad \text{AA} \quad (5.5)$$

由公式(5.4)可整理成

$$\begin{aligned} {}^C P - t_A &= {}^C R {}^A P \\ \Rightarrow {}^A R^T ({}^C P - t_A) &= {}^A P \quad \text{AA} \quad (5.6) \end{aligned}$$

利用公式(5.5)的 ${}^C P$ 代入公式(5.6)

$$\begin{aligned} & {}^A R^T (({}^C R {}^B P + t_B) - t_A) = {}^A P \\ \Rightarrow & {}^A R^T {}^C R {}^B P + {}^A R^T t_B - {}^A R^T t_A = {}^A P \\ \Rightarrow & {}^A R^T {}^C R {}^B P + {}^A R^T (t_B - t_A) = {}^A P \quad \text{AA} \quad (5.7) \end{aligned}$$

因為 ${}^C_R, {}^C_R, t_B, t_A$ 皆為已知，當已知 ${}^A P$ 與 ${}^C P$ 的轉換關係式，也已知 ${}^B P$ 與 ${}^C P$ 的相對關係，由公式(5.7)便可推導出 ${}^B P$ 轉換到以 A 為參考原點之座標，亦可得知 ${}^A P$ 轉換到以 B 為參考原點之座標。承上述方法，在找出畫作 A 、 B 的相對關係後，即可建置出展場畫作的空間配置圖。

5.3 使用者參觀路徑記錄

本節承接 5.1 和 5.2 的結果，求出相機內外部參數以及展場畫作的配置圖後，便可將使用者在參觀畫展時的拍照軌跡建立出來，藉由相機軌跡與展場畫作配置之間的相對位置，不僅能夠提供參觀行進路線導覽，也能針對使用者進行個人化參觀行程建議，提醒使用者是否還有未欣賞之畫作。當使用者對著某一張畫作拍照，系統所計算出的旋轉及位移矩陣 $[R \ t]$ ，這裡的 $[R \ t]$ 表示畫作轉換到相機的相對關係，但為了呈現相機軌跡時是以畫作為基準點，因此要將 $[R \ t]$ 的關係式修改為相機轉換到畫作。假設畫作上存在一點 $A = \begin{bmatrix} X_A \\ Y_A \\ Z_A \end{bmatrix}$ ，畫作到相機的轉換關係為 $[R \ t]$ ， A 點在相機座標裡表示為 $\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix}$ ，則關係式如下：

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = R \begin{bmatrix} X_A \\ Y_A \\ Z_A \end{bmatrix} + t$$

$$\Rightarrow R^{-1} \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} - R^{-1}t = \begin{bmatrix} X_A \\ Y_A \\ Z_A \end{bmatrix} \quad \Lambda \Lambda (5.8)$$

從公式(5.8)便可得知相機轉換到畫作座標的關係式，當知道相機位置如何對應到展場畫作後，即可建立出使用者之參觀路徑相機軌跡。

第六章 實驗結果

本實驗系統所使用之軟硬體配備為華碩桌上型主機，其規格為 Pentium(R) Dual-Core CPU E6500 2.93 GHz，2G RAM，600G WDC 硬碟，作業系統為 Windows XP，搭配相機 Canon SX1 IS，系統程式是使用 Matlab R2008a 所撰寫。

6.1 畫展影像辨識實驗結果

本實驗藉由一操作簡易的使用者介面，如圖 14 所示，事先建立好資料庫影像的資訊，當使用者輸入查詢影像如圖 15，並按下”search”按鍵，系統便可自動搜尋出與查詢影像最相似的資料庫圖片，並按照得票數的高低以遞減方式排列下來如圖 16。在得到最相似的資料庫影像後，還需檢驗其正確性，利用 4.3 節所提方法，將轉正後的查詢影像與資料庫影像相減，並觀察相減後的結果(圖 17、18)，表格 6.1 為將相減後的結果圖檔轉為灰階，並計算該灰階圖之平均值及變異數。



圖 14 使用者介面之畫面



圖 15 輸入影像開始查詢

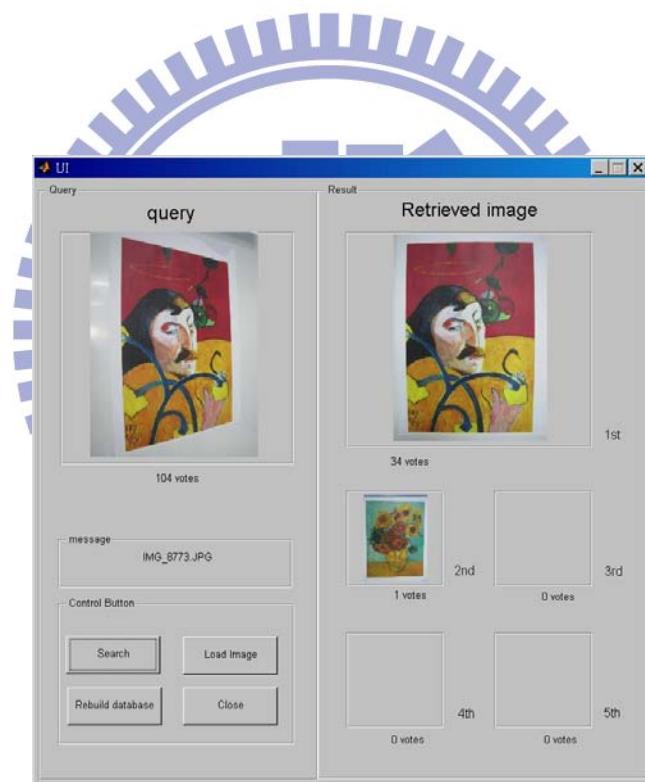


圖 16 查詢辨識結果

說明：右上角為得票數最高之資料庫影像

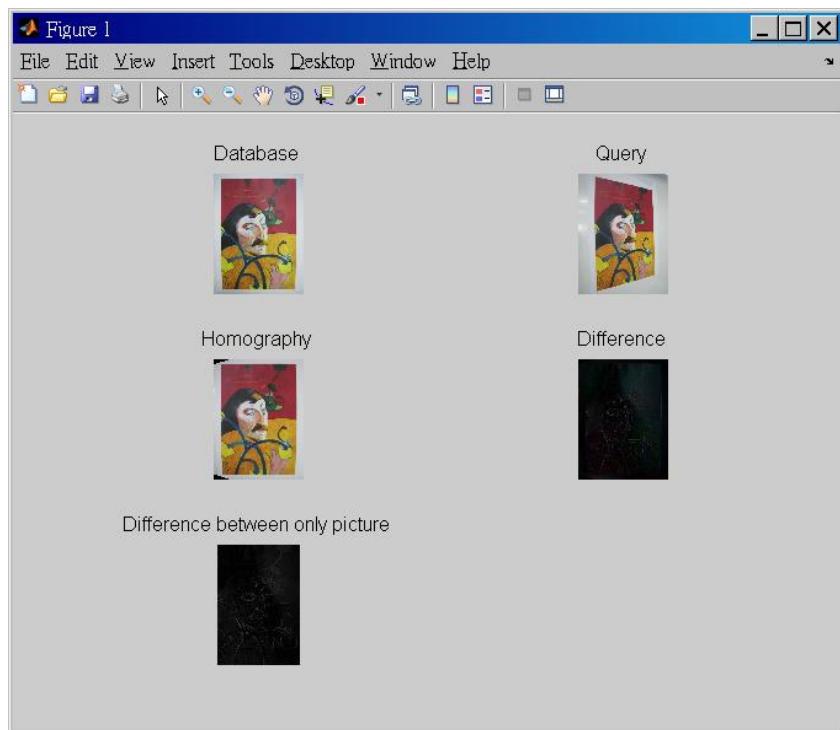


圖 17 Homography 驗證查詢結果之一

說明：(a) 辨識結果，(b) 查詢影像

(c) 將查詢影像轉正結果

(d) (a)-(c)後的影像

(e) 針對(a)、(c)中僅畫作部分相減結果

a^{φ}	b^{φ}
c^{φ}	d^{φ}
e^{φ}	φ

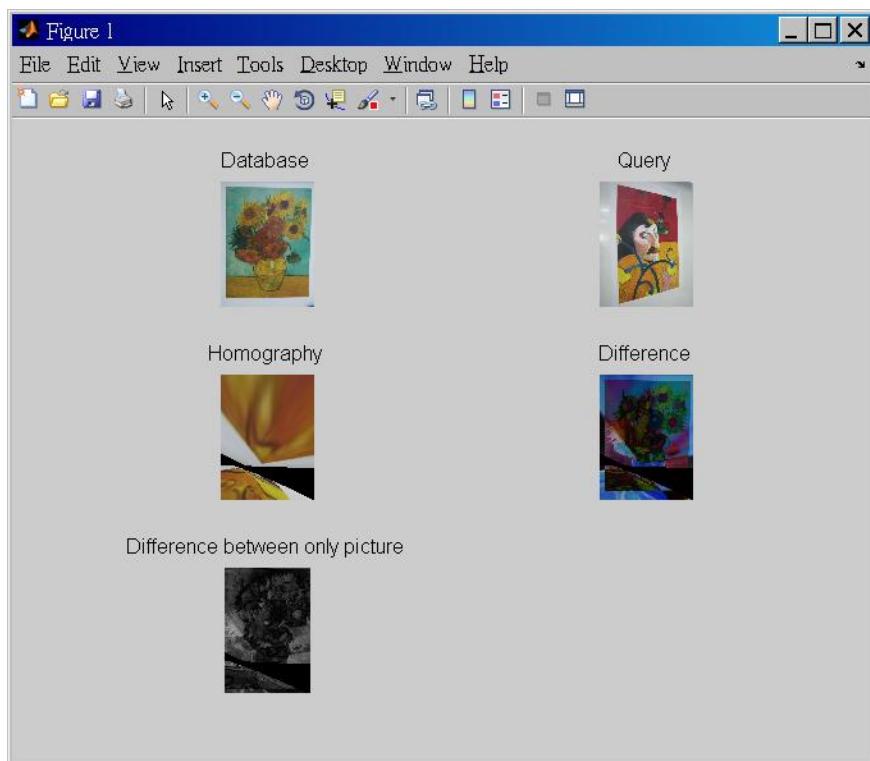


圖 18 Homography 驗證查詢結果之二

說明：(a) 辨識結果，(b) 查詢影像

a	b
c	d
e	

(c) 將查詢影像轉正結果，(d) (a)-(c)後的影像

(e) 針對(a)、(c)中僅畫作部分相減結果

表 1 將 Homogaphy 驗證結果轉為灰階圖之誤差

	轉為灰階圖之平均值	轉為灰階圖之變異數
圖 17 (e)	0	0
圖 18 (e)	44.8740	175860

6.2 畫展影像相機定位實驗結果

6.2.1 單張資料庫影像之 Homography 驗證及相機軌跡定位

相機定位結果的好壞取決於 Homography 是否估算精確，本實驗利用 Homography 將畫作的 4 個 corners 端點 2D ground truth 座標反投影至查詢影像上，並以綠色圓圈標示，若是 Homography 正確無誤，則 projected corners 也應該落在正確的影像座標處，圖 19 是資料庫影像，圖 20 為模擬使用者在畫展拍攝畫作的情況，站在三種不同的角度對畫作取像，利用 5.1.1 節之 H_R 估測出以綠色圓圈畫出 4 個 corners 之反投影座標並用紅色線段連線，表格 6.2 為列出估測之反投影座標誤差之最小前 10 名，圖 21(a)、(b)、(c)分別為估測查詢影像之相機軌跡之 front-view 、bottom-view 、top-view 。



圖 19 資料庫影像之高更的《帶光環的自畫像》



圖 20 (a)、(b)、(c)為估測查詢影像之 Homography

說明：綠色圓圈為估測的 projected corners

a²	b²	c²
----------------------	----------------------	----------------------

表 2 在圖 20(a)、(b)、(c)中之 projected points 誤差(pixels)之最小前 10 名

	圖 20(a)	圖 20(b)	圖 20(c)
1	0.0651	0.0546	0.0265
2	0.0858	0.0859	0.0285
3	0.1008	0.0907	0.0405
4	0.1173	0.1046	0.0533
5	0.1317	0.1133	0.0608
6	0.1366	0.1215	0.0793
7	0.1384	0.1245	0.0836
8	0.1556	0.1262	0.0840
9	0.1827	0.1268	0.1160
10	0.1914	0.1317	0.1176

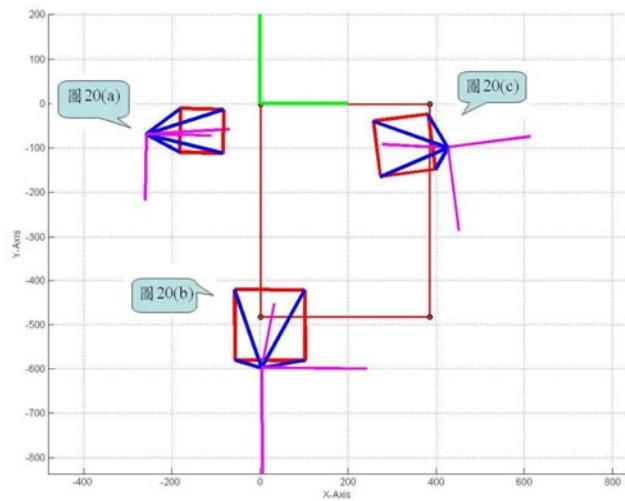


圖 21(a)相機軌跡估測 front-view

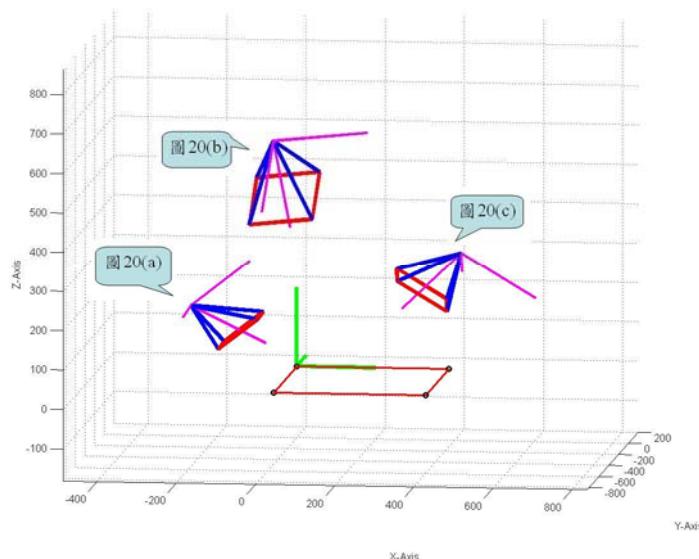


圖 21(b)相機軌跡估測 bottom-view

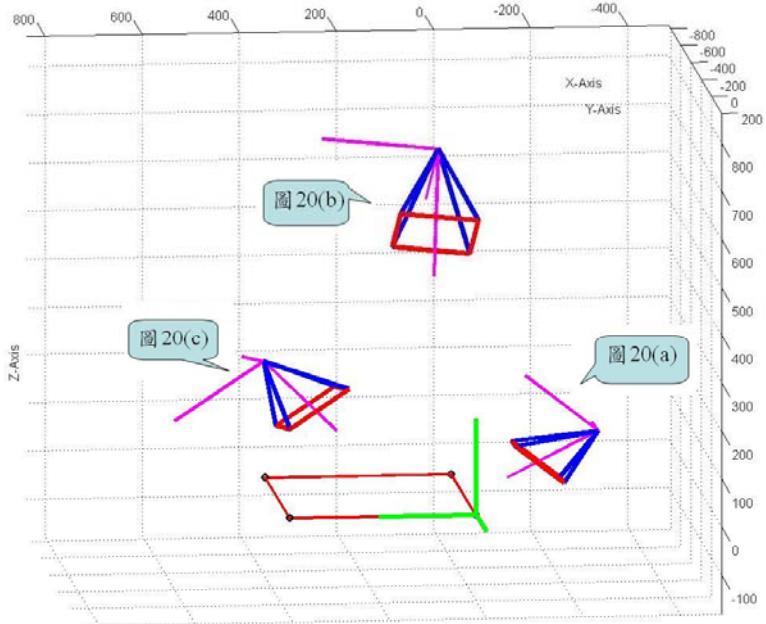


圖 21(c)相機軌跡估測 top-view

圖 21(a)、(b)、(c) 相機軌跡估測結果

說明：紅色細線邊框為畫作，定義畫作左上角頂點為世界座

原點，並以綠色三軸表示，相機座標以粉紅色三軸表示，

Z 軸朝內，分別對應至圖 20(a)、(b)、(c) 之相機位置

a
b
c

6.2.2 展場畫作空間配置圖及使用者參觀路徑估測

圖 22 為畫作空間配置圖，定義左邊畫作一雷諾瓦的《拿扇的少女》之左上角頂點為參考基準，圖 23 (a) 、(b) 、(c) 分別為 front view 、 bottom view 以及 top view 之估測畫作配置圖。



圖 22 畫作空間配置圖

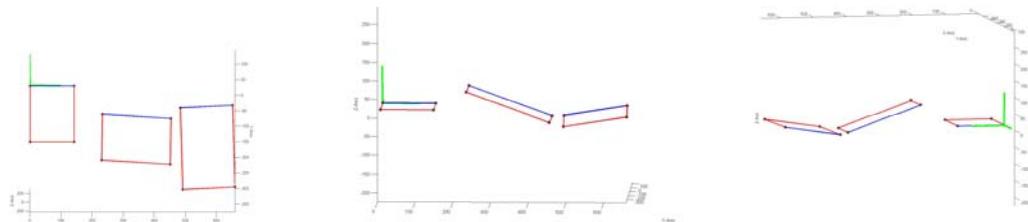


圖 23 (a)、(b)、(c)為估測畫作空間配置圖



假設使用者所輸入圖 24 畫作空間中之兩張之查詢影像分別為圖 24(a)、(b)，則
圖 25 為使用者在圖 25 畫作空間中之參觀路徑估測。

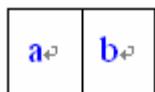


圖 26 (a)、(b) 分別為使用者所輸入之查詢影像

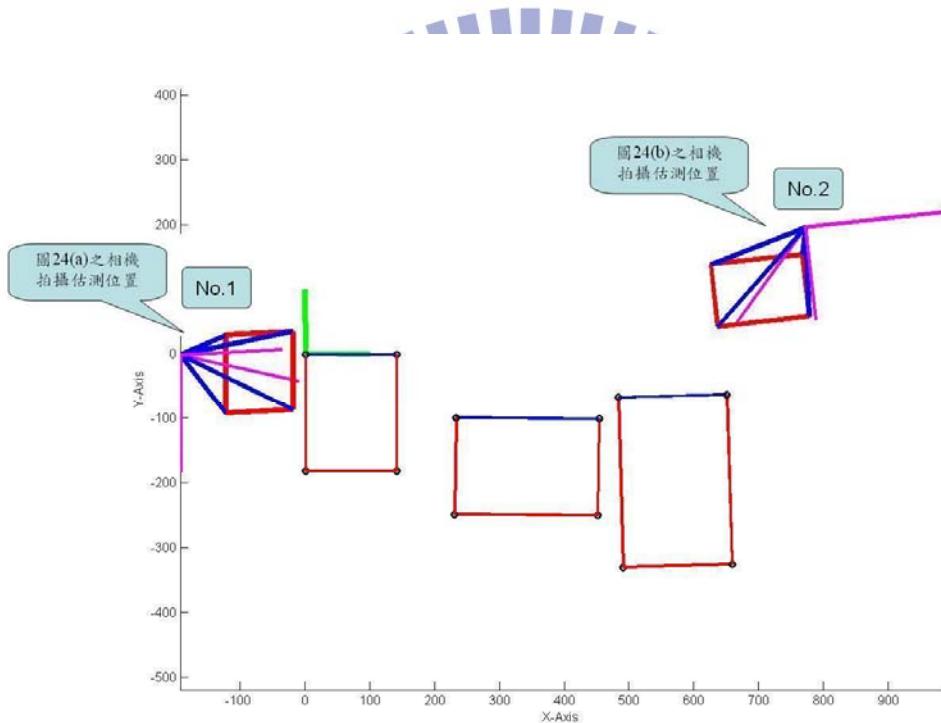


圖 27 使用者參觀路徑估測

說明：制定左一畫作之左上角頂點為世界座標之參考點，
如綠色三軸座標表示，左邊三角錐為圖 24(a)之估測
相機位置，粉紅色三軸為相機系統座標，並且 z 軸朝
內，右邊三角錐為圖 24(b)之估測相機位置

第七章 結論與未來發展

本論文提出一套室內畫作導覽系統的概念，分別以辨識以及定位為兩大目標，利用辨識效能較為優秀的 MSER 以及 ZM 作為第一階段的辨識結果，並搭配相機校正之原理，利用單張影像來完成相機定位與軌跡估測之成果。但本論文實驗目前只針對理論之可行性，未來還需將這概念拓展到手持裝置上，並充實畫展說明內容，包括動畫甚至加入多國語音導覽。未來希望能夠融入 tracking 之技術，將單張影像查詢定位功能延伸至影片，使系統功能更具完整性，也能更符合使用者實際應用需求。



參考文獻

- [1] H. Kato and M. Billinghurst, “Marker tracker and HMD calibration for a video-based augmented reality conferencing system,” in Proceedings of the 2nd International Workshop on Augmented Reality, San Francisco, USA, 1999.
- [2] M. Fiala, “ARTag, a fiducial marker system using digital techniques,” in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, 20-25 June 2005, pp. 590-596.
- [3] G. Simon, A.W. Fitzgibbon, and A. Zisserman, “Markerless tracking using planar structures in the Scene,” in Proc. IEEE/ACM Int'l Symp. Augmented Reality (ISAR '00), pp. 120-128, 2000.
- [4] V. Ferrari, T. Tuytelaars, and L.J.V. Gool, “Markerless Augmented Reality with a Real-Time Affine Region Tracker,” in Proc. IEEE/ACM Int'l Symp. Augmented Reality (ISAR '01), pp. 87-96, 2001.
- [5] I. Skrypnyk and D.G. Lowe, “Scene Modelling, Recognition and Tracking with Invariant Image Features,” Proc. IEEE/ACM Int'l Symp. Mixed and Augmented Reality (ISMAR '04), pp. 110-119, 2004.
- [6] T. Lee and T. Höllerer, “Multithreaded hybrid feature tracking for markerless augmented reality,” IEEE Trans. Vis. Comput. Graph., vol. 15, no. 3, pp. 355–68, May–Jun. 2009.
- [7] R. O. Castle and D. W. Murray, “Object Recognition and Localization While Tracking and Mapping,” In Proceedings of the 8th IEEE/ACM International Symposium on Mixed and Augmented Reality
- [8] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg. Pose tracking from natural features on mobile phones. In Proc 7th IEEE and ACM International Symposium
- [9] Koutsou, A.D.; Seco, F., Jimenez, A.R., Roa, J.O., Ealo, J.L., Prieto, C., Guevara,

J..” Preliminary Localization Results With An RFID Based Indoor Guiding System ,” Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium, pp.1-6

[10] Lassabe, F., Canalda, P., Chatonnay, P., Spies, F., Charlet, D.” Refining WiFi Indoor Positioning Renders Pertinent Deploying Location-Based Multimedia Guide ,” Advanced Information Networking and Applications, 2006. AINA 2006. 20th International Conference on Volume 2, pp.126-132

[11] K.-C. Chen and W.-H. Tsai,” Vision-Based Autonomous Vehicle Guidance for Indoor Security Patrolling by a SIFT-Based Vehicle Localization Technique,” Vehicular Technology, IEEE Transactions on Volume: PP , Issue: 99 , pp. 1-1

[12] N. Zhang , M. Li and B. Hong,” Simultaneous Localization and Mapping Using Invariant Natural Features,” Robotics and Biomimetics, 2006. ROBIO '06. IEEE International Conference , pp . 1682 – 1687

[13] Y. Ogawa , N. Shimada and Y . Shirai,” Environmental mapping for mobile robot by tracking SIFT feature Points using trinocular vision,” SICE, 2007 Annual Conference, pp.1996 - 2001

[14] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int'l J. Computer Vision, vol. 60, no.2, pp. 91–110, 2004

[15] K. Mikolajczyk and C. Schmid,“A Performance Evaluation of Local Descriptors,” IEEE Trans.Pattern Analysis and Machine Intelligence, 2005, 27(10), p1615-1630

[16] K. Mikolajczyk and C. Schmid, “Scale & affine invariant interest point detectors,” *Int'l J. Computer Vision* , vol. 60, no. 1, pp. 63–86, 2004.

[17] T. Tuytelaars and L. Van Gool, “Matching widely separated views based on affine invariant regions,” *Int'l J. Computer Vision*, vol. 59, no. 1, pp. 61-85, 2004.

[18] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide-baseline stereo from maximally stable extremal regions,” Image and Vision Computing , vol. 22,

pp.761–767, 2004.

- [19] A. Khotanzad and Y. H. Hong, “Invariant image recognition by Zernike moments,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 5, pp.489–497, 1990.
- [20] C. S. Won, D. K. Park, and S.-J. Park, “Efficient use of MPEG-7 edge histogram descriptor,” ETRI Journal, vol.24, no.1, pp. 23-30, 2002.
- [21] S. Li, M.-C. Lee, and C.-M. Pun, “Complex Zernike moments features for shape-based image retrieval,” IEEE Tran. Systems, Man, and Cybernetics—Part A: Systems and Humans, vol. 39, no. 1, pp.227 - 237, 2009.
- [22] S.K. Sun, “Image Representation, Matching, and Recognition Using Invariant Zernike Moment Descriptors,” 博士論文，國立交通大學，資訊科學與工程研究所，新竹市，臺灣, 2009.
- [24] Z. Zhang, “A Flexible New Technique for Camera Calibration,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334, 2000.
- [25] Richard I. Hartley . "In Defense of the Eight-Point Algorithm". IEEE Transaction on Pattern Recognition and Machine Intelligence
- [26] A. Fusiello, E. Trucco, and A. Verri, “A compact algorithm for rectification of stereo pairs,” Machine Vision and Applications, pp.16-22,2000.