

# 國立交通大學

資訊科學與工程研究所

## 碩士論文

以內容為基礎的建築物影像檢索  
Content-Based Building Image Retrieval

研究生：黃啟銘

指導教授：陳 稔 教授

中華民國九十九年六月

以內容為基礎的建築物影像檢索  
Content-Based Building Image Retrieval

研究生：黃啟銘

Student : Chi-Ming Huang

指導教授：陳 稔

Advisor : Zen Chen



A Thesis  
Submitted to Institute of Computer Science and Engineering  
College of Computer Science  
National Chiao Tung University  
in Partial Fulfillment of the Requirements  
for the Degree of  
Master  
in  
Computer Science

June 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年六月

# 以內容為基礎的建築物影像檢索

學生：黃啟銘

指導教授：陳 稔

國立交通大學

資訊科學與工程研究所



本論文目的在使用影像區域特徵來建立一個建築物影像檢索系統。此檢索系統分成資料庫與查詢兩個部份，資料庫部份按照處理順序又可分為三個步驟，第一步驟使用可抗視角變化的 Maximally Stable Extremal Region 做特徵區域擷取；第二步驟使用旋轉不變的 phased-based Zernike Moment 做特徵區域描述；第三步驟使用 kd-tree 建立特徵向量的索引。建立資料庫時，使用同一棟建築物相鄰的影像特徵互相比對，去除不穩定出現的特徵區域，並使用 Density-Based Spatial Clustering of Applications with Noise 分群法，以減少資料庫中存在的儲存重覆特徵問題。查詢部分採用 kd-tree 找最近點與鄰近點的便利性，以直觀的投票機制找出資料庫中與查詢影像最相似的建築物。

# Content-Based Building Image Retrieval

Student : Chi-Ming Huang

Advisor : Zen Chen

Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

## ABSTRACT

The goal of this thesis research is to construct a building image indexing and retrieval system. This system consists of two parts: the database organization (indexing) and the query part (retrieval). The database part is further composed of three modules. In the first module, view-invariant feature detection, Maximally Stable Extremal Region (MSER), is used to extract the regions of interest. In the second module, the phased-based Zernike Moment is used to describe the regions. In the third module, a kd-tree structure is used to establish the index of Zernike Moment feature vectors. When constructing the database, in order to eliminate the unstable regions, a trick of comparison of the features extracted from the neighboring views of the same building is used. To reduce the problem of redundancy, the clustering algorithm, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), is used. In the query part, the kd-tree provides a convenient way to find the nearest neighbor. And then an intuitive voting mechanism is used to find the building from the database which is most similar to the query image.

## 誌 謝

本篇論文的完成，首先要感謝我的指導教授陳稔老師，在我的碩士研究生涯中提供了知識上以及生活上的許多幫助與教誨，老師的研究態度與精神是我學習與敬佩的榜樣。感謝口試委員周宏隆先生與賈叢林教授，在論文撰寫方面給予了很多指教，讓論文能順利完成。

接著要感謝博士班的樹國學長、彥旭學長與文昭學長，有了樹國學長對於 Zernike Moment 的深厚研究基礎，使得這篇論文在特徵描述子方面並未遭遇很大的困難，樹國學長也常在我研究焦頭爛額之際，提供很有用的建議；彥旭學長不僅教導我如何思考一個問題，在待人處事上也給與了我很多的幫忙；文昭學長則對於我的視野與做事態度，有很多正面的影響。

感謝實驗室的成員，文軒學長、岳樺學長、佳瑄和嘉峻學弟，一同修課、討論、生活，從你們身上給了我很多勇氣讓我能夠堅持下去。感謝我的家人，你們是我精神上最重要的支柱，謝謝你們在經濟上對我的幫助與諸般容忍。最後，感謝一同來新竹念書的瑞桓、室友致翔、彥后，還有許多一路上曾經陪伴我、幫助我的人，有和你們的分享，研究生活才能夠多姿多采，僅將這篇論文獻給你們。

黃啟銘 謹誌於  
中華民國 99 年 7 月

# 目 錄

摘 要.....	i
ABSTRACT.....	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 問題陳述.....	1
1.3 相關研究.....	2
1.4 論文架構.....	3
第二章 單張影像特徵偵測與描述.....	5
2.1 MSER 特徵區域擷取.....	5
2.1.1 MSER 概述.....	6
2.1.2 過濾不適用於辨識的 MSER 橢圓區域.....	8
2.2 Zernike Moment 特徵區域描述.....	11
2.2.1 Zernike Moment 概述.....	12
2.2.2 Zernike Moment 相似度的衡量方法.....	13
第三章 建築物影像與特徵資料庫建立.....	14
3.1 建築物影像拍攝.....	14
3.1.1 建築物資料庫影像拍攝考慮因素.....	16
3.1.2 建築物查詢影像拍攝考慮因素.....	17
3.2 建築物資料庫影像特徵偵測與描述.....	18
3.2.1 使用相鄰影像去除不穩定的建築物影像特徵區域.....	18
3.2.2 建築物多面貌的表示法.....	22
3.3 使用 DBSCAN 做建築物資料庫影像的特徵區域分群.....	22
3.3.1 DBSCAN 概述.....	23
3.3.2 運用 DBSCAN 對影像特徵區域的分群.....	24
3.4 建築物影像特徵的 kd-tree 資料庫建立.....	26
第四章 建築物影像檢索機制.....	29
4.1 kd-tree 最近點搜尋與範圍搜尋.....	29
4.1.1 利用 kd-tree 找強度最接近的特徵向量.....	29
4.1.2 利用 kd-tree 範圍搜尋找強度接近的特徵向量.....	31
4.2 建築物檢索結果的產生.....	32
第五章 辨識實驗結果與評估.....	33
5.1 交大建築資料庫的辨識結果.....	33

5.1.1 交大建築資料庫介紹.....	33
5.1.2 交大建築物影像不同拍攝因素下的辨識結果.....	36
5.1.3 時間和空間複雜度評估.....	36
5.2 ZuBud 建築資料庫的辨識結果 .....	39
5.2.1 ZuBud 建築資料庫介紹 .....	39
5.2.2 ZuBud 建築資料庫的辨識結果 .....	39
第六章 結論與未來發展.....	41
參考文獻.....	42



## 表 目 錄

表 1、使用相鄰影像去除不穩定的建築物影像特徵區域演算法.....	20
表 2、Density-Based Spatial Clustering of Applications with Noise 演算法 .....	24
表 3、特徵向量分群結果與影像關係的資料結構.....	26
表 4、kd-tree 資料庫建立演算法 .....	28
表 5、交大建築資料庫影像與建築物查詢影像的比較.....	33
表 6、交大建築資料庫涵蓋的建築物列表.....	33
表 7、交大資料庫影像範例.....	34
表 8、交通大學查詢影像分類範例表.....	35
表 9、交通大學查詢影像範例表.....	35
表 10、交大建築物查詢影像辨識結果.....	36
表 11、不同解析度的測試影像做特徵偵測與描述所需時間.....	36
表 12、各個會影響資料庫中特徵向量個數的步驟對時間的影響.....	37
表 13、bounding box 的範圍大小對於查詢時間的影響 .....	38
表 14、ZuBud 建築物影像範例 .....	39
表 15、ZuBud 辨識結果 .....	40



## 圖目錄

圖 1、方法流程圖.....	4
圖 2、建築物影像含有許多較一致性且邊界明確的區域.....	5
圖 3、MSER 示意圖.....	6
圖 4、4x4 的 8 灰階圖 extremal region 示意圖.....	7
圖 5、按照產生順序與包含關係建成的 Extremal Region 階層結構.....	7
圖 6、初始偵測出來的 MSER 橢圓區域.....	8
圖 7、去除太大與太小區域後的 MSER 橢圓區域.....	9
圖 8、 $\Delta$ 提高、去除面積變異大的區域之後的 MSER 橢圓區域.....	9
圖 9、消除重覆區域之後的結果.....	10
圖 10、再去除過長的 MSER 區域的結果.....	10
圖 11、兩種型式的 MSER 的特徵區域.....	11
圖 12、橢圓標準化成圓示意圖.....	11
圖 13、嚴重遮擋的建築物影像範例.....	14
圖 14、建築物近照範例.....	15
圖 15、建築物遠照範例.....	15
圖 16、建築物表面凹凸不平造成遮擋範例.....	16
圖 17、建築物一個面資料庫儲存影像範例.....	17
圖 18、MSER 特徵區域示意圖，交通大學竹軒.....	18
圖 19、MSER 特徵區域示意圖，交通大學工程三館.....	18
圖 20、有較小視角變化時的 MSER 特徵區域萃取結果.....	19
圖 21、使用相鄰影像去除不穩定的建築物影像特徵區域結果.....	21
圖 22、交通大學工程五館適合使用者拍攝的三個面.....	22
圖 23、DBSCAN 演算法分群示意圖.....	23
圖 24、DBSCAN 演算法特徵區域分群結果.....	25
圖 25、建立 kd-tree 的流程示意圖.....	27
圖 26、kd-tree 近似最近點示意圖.....	29
圖 27、當 approximate nearest distance 比 $\overline{mag^q}$ 到分割平面的距離大的示意圖.....	30
圖 28、當 approximate nearest distance 比 $\overline{mag^q}$ 到分割平面的距離小的示意圖.....	30
圖 29、排除掉不可能含有更近點的空間。.....	31

# 第一章 緒論

## 1.1 研究動機與目標

現在是數位相機普及的時代，人人隨手可得的手機、PDA 等行動裝置，也大量配備相機，如果可以直接使用影像來幫助使用者獲得環境的資訊，將會讓生活更加便利。

本論文希望能由相機拍攝到的建築物影像，以基於影像特徵擷取與比對方法，辨識所拍攝建築影像為影像資料庫中哪一棟建築，進而能提供相關的資訊給使用者，如目標建築的文字、影音介紹等等。

本研究之影像檢索系統的定位功能與一般常用來定位的裝置 GPS(Global Positioning System)或羅盤的不同點有下：在高樓之間、橋梁下方，GPS 容易訊號不良，羅盤會因為鋼筋建築物影響而有角度偏差的現象，影像檢索則無此限制；另外，GPS 與羅盤只能在目標物旁當場查詢出所在位置，而影像檢索可以在使用者不在目標物身旁仍然可以作檢索，相較之下，以影像來做檢索與定位有其重要性。

另外，因為基於特徵點擷取與對應的方法不僅能夠提供圖片檢索的功能，還能夠藉由對應點的資訊，做空間的定位，藉由估算出使用者所看見的影像與資料庫中的影像之間相機位置的轉換，可以更進一步應用在 Augment Reality 下，例如：古蹟導覽、商店 3D 的廣告。一個古蹟導覽的可能場景如下：當使用者拿著手機所附的相機朝著一棟古蹟拍攝，經過影像搜尋，可以判斷這棟建築影像是對應到資料庫中哪一棟建築的影像，螢幕更可進一步顯示這棟建築的文字介紹，以及古代人們生活在此處的三維模型，利用 Augment Reality 技術將影像顯示在古蹟的門口，而且可以隨著使用者的手機照到這棟建築物的不同方向，變換三維模型的角度，使導覽更加活靈活現，充滿趣味。

## 1.2 問題陳述

為了做建築物影像的檢索，必須事先拍攝建築物影像資料庫，影像資料庫中含有各棟建築物不同角度的影像，當使用者拍攝一查詢影像，與資料庫中影像做比對，檢索系統可回報查詢影像與資料庫中哪一棟建築物的影像最相近。

在做建築影像比對時，需要克服的挑戰有(1)視角改變，如相機拍攝時鏡頭的縮放、位移、旋轉；(2)光線變化，如陰晴、相機光圈、快門大小；(3)遮擋物，路樹、路燈、旗子的遮擋。

觀察一棟建築物的外型，較顯著的特徵有：形狀、顏色、樓層數，以及紋理等等，其中形狀若以輪廓描述，會隨著相機視角不同而改變，難以有一致性描述；顏色在陽光強烈照射的情況下飽和度會過低，造成顏色不穩，樓層數則在不同建築的樣式之下，使用影像判斷，無法有一致性的規則。本研究使用建築的紋理資訊來做影像識別，為了克服遮擋的問題，選擇以基於區域(local-based)的影像特徵資訊來描述建築物的影像。

### 1.3 相關研究

以內容為主的影像檢索，是由使用者提供一張欲搜尋的類似或相關的圖片，然後根據影像的顏色、紋理(texture)、形狀(shape)或其它特性來做影像的檢索。而除了使用 spatial domain 的特性外，還有將影像資訊轉換到 frequency domain 尋找特徵係數的做法，例如 wavelet-based 演算法[3]

有關於影像檢索的論文，可以分成兩大類，一類是影像的分類(classification)，一類是物體的辨識(identification)。

影像的分類方面，Feng Jing 等人提出一套基於以區域為主的影像檢索架構[6]：使用 JSEG 做影像的切割，以 color moment 來做區域的描述，並使用區域權重的方式，動態的調整、學習人對影像上的哪個區域是有興趣的，就給定比較高的權重，並以 Generalized Lloyd Algorithm 將區域分群，建立 codebook，以 codeword 與此 codeword 的權重來表示這張影像，並使用 Earth Mover's Distance(EMD)來比對兩張影像的相似度。這類型影像檢索的目標是要將影像分類，"A retrieved image is considered a match if it belongs to the same category as the query image"，針對已經分類完成的 database，如：COREL，測試他們分類的能力，這種資料庫中會含有人為分門別類好的影像，例如：氣球、蝴蝶、狗、花、獅子、山、火車、瀑布等等種類。

Jutta Willamowski 等人則使用了 affine invariant 描述子來描述影像[7]，先對描述子做向量量化，接著使用 Naïve Bayes 與 SVM(support vector machine)等等分類器做分類，目的是要分別出臉、樹木、車子、電話、腳踏車、書等等的圖片類別。

物體的辨識(identification)方面，Wen Wu 等人針對街上的廣告招牌做辨識[8]，查詢影像是人工在影像上標示出的廣告牌，對資料庫中的影像或是影片做搜尋，方法是對欲搜尋的影像做 salient 區域偵測，再對 salient region 找 canny edge、SIFT 特徵點，再比較影像分割的結果，從資料庫中影像的街道上的景物找出相同廣告牌的圖案與位置。

建築物的辨識方面有兩種方式，一個是使用全域性(global)的特徵如顏色資訊來做辨識，例如：Wei Zhang 等人對建築物做辨識[10]，分成兩個階段，第一階段利用建築物結構上紋理較具方向性的特性，找出建築物影像上屬於建築物的像素，利用其顏色資訊，轉成 YCbCr 空間，接著以 1D 的飽和度做索引，第二階段加上以 SIFT 為特徵來做影像比對，然而實際拍攝影像時，會面臨陽光太強或是陰影造成顏色飽和度過低的情況。N.J.C. Groeneweg 等人，使用直接以整張影像的 rgb color histogram 來做特徵[13]，辨識 ZuBud 資料庫有 94% 的辨識率，而這種方式在另一個影像資料庫 Roeterseiland database 只能達到 24% 的辨識率，因為 ZuBud 資料庫中建築物拍攝的視角、與天氣變化都不是很明顯，資料庫影像與查詢影像顏色資訊太相近等因素造成的。

另外一個方式是使用區域性(local)的特徵，例如：Obdrzalek 等人對建築物做辨識[9]，從影像上擷取出 local affine 特徵，接著將特徵做光學與幾何的標準化，之後以 compact DCT-based representation 來表示找到的特徵，使用 DCT(discrete cosine transform)的係數當作特徵區域描述子來做比對，因為使用 DCT 來做描述子本身無法抗視角、旋轉等變化，缺點是對於一個特徵區域，需要儲存多個 local affine frame 來做比對，越複雜的區域就要

使用越多的local affine frame來表示。Gerald Fritz使用MAP(Maximum A Posteriori)萃取出較具區辨力的SIFT特徵i-SIFT來對建築影像做辨識[14]。

當使用特徵點偵測與對應技術來做影像檢索時，在相似的影像當中，我們感興趣的區塊，會隨著相機在不同的時間、位置、角度拍攝，而改變其位置、大小、光線、形狀等等性質，造成比對的困難，而 Scale Invariant Feature Transform (SIFT)[4, 5]提供了人們一個好的特徵比對資訊，因此成為了目前經常被使用來尋找強健的特徵點方法。它結合了 scale invariant region detector 和一個以 gradient distribution 為基礎的 descriptor。然而，SIFT 只能處理 similarity transformation，無法應付在 wide-baseline 下因視角變化下所造成的幾何變形，故為了強化特徵比對，發展能抗幾何變形的 CBIR 檢索技術，本研究採用 affine-invariant region detector, affine-invariant region detector 所萃取出來的特徵區域，來做影像的比對。

本研究挑選特徵偵測子時選擇 MSER，因為建築物上含有較多較一致性的 (Homogeneous)且含有明顯邊界的區域，相較於其他的 affine invariant 特徵偵測子如 Hessian-Affine、Harris Affine 有較好的表現[3]。

選擇特徵描述子方面本研究選擇使用 phase-based 的 Zernike Moment, Chen Z.與 Sun S.K.在[2]有針對目前最常見的特徵描述子 SIFT、PCA-SIFT、GLOH、steerable 與 phase-based Zernike Moment 在影像模糊、亮度改變、雜訊、JPEG 壓縮、旋轉、位移、視角改變等因素下的特徵區辨能力做比較，而 Zernike Moment 在這些因素之下的特徵點比對準確度，明顯優於其他的特徵描述子。

特徵索引結構方面，kd-tree[12]常用在影像檢索的最相近特徵搜尋上，而其所提供的資料結構除了可尋找最近點以外，也可以快速找尋一定範圍內的資料點。

## 1.4 論文架構

圖 1.1 為本研究方法流程圖，論文第二章涵蓋本研究對一張建築影像做特徵擷取與描述的方法，對應到流程圖的特徵擷取模組；第三章敘述如何建立建築物資料庫的完整方法，涵蓋如何拍攝資料庫影像、偵測並描述影像特徵、建立可供快速查詢的資料結構等，對應到流程圖的特徵擷取模組與資料庫產生模組；第四章描述本研究檢索一張查詢影像的方法，對應到流程圖的影像檢索模組；第五章展示實驗結果，測試資料分成兩部分：本研究自行建立的國立交通大學建築影像資料庫以及 Swiss Federal Institute of Technology 所建立的 ZuBud(Zurich Buildings Database)兩個建築影像資料庫。

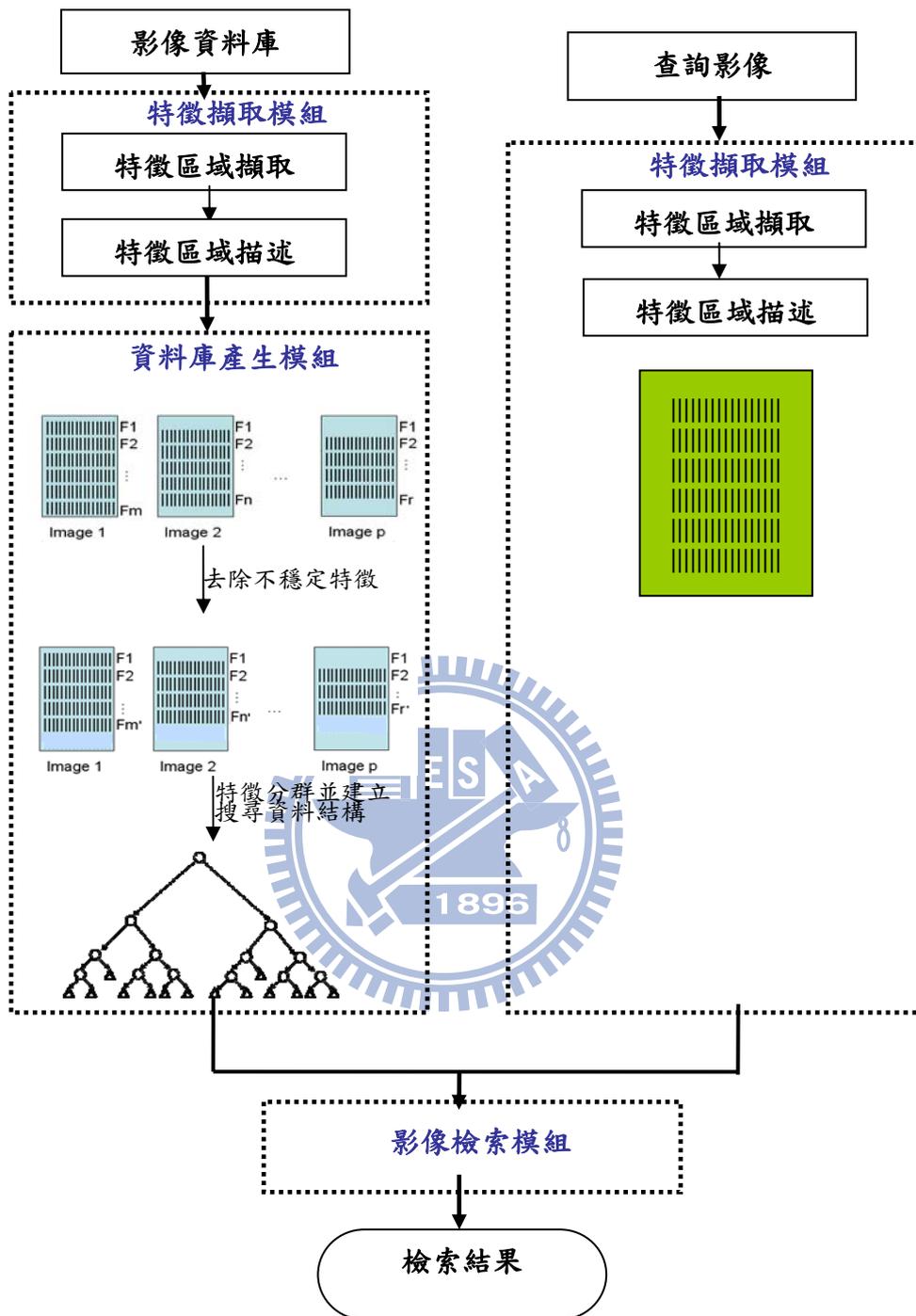


圖 1、方法流程圖

## 第二章 單張影像特徵偵測與描述

本章節涵蓋如何對影像做 MSER(Maximally Stable Extremal Region)特徵區域擷取與對特徵區域使用 Zernike Moment 做特徵區域描述，以得到特徵向量的方法。

### 2.1 MSER 特徵區域擷取

MSER[17]為一種特徵區域擷取方法，優點是在抵抗視角變化的能力特別優異，並且在含有許多較一致性(homogenous)且有明確邊界區域的影像表現較穩定[1]。因為建築物影像有許多邊界明確的結構，如窗戶、色塊等等，使用 MSER 能夠較穩定的擷取出這些區域，如圖 2。擷取的原則有二：在一個強度值以下為最大相連(maximally connected)的區域、強度值增減一個範圍內區域的面積變化率相對最小。



(a) 建築物影像含有許多較一致性且邊界明確的區域，交通大學田家炳光電大樓；(b) 偵測 MSER 橢圓區域的結果。

### 2.1.1 MSER 概述

MSER 處理的是灰階影像，如果是彩色 RGB 影像，可先利用公式 2.1 將彩色影像轉成灰階影像。

$$I = \frac{R+G+B}{3} \quad (\text{公式 2.1})$$

MSER 將影像的灰階值(Intensity)視作類似地形圖的曲面，如圖 3，為一個一維影像的示意圖，橫軸為影像座標，縱軸為灰階值。 $R$ 、 $R_{+\Delta}$ 、 $R_{-\Delta}$  為 extremal region，是符合一些特定條件的相連像素集合，關於 extremal region 的詳細定義請見後文，定義  $I(R)$  為  $R$  這個區域包含的像素灰階值最大值，當作  $R$  這個 extremal region 的灰階值。

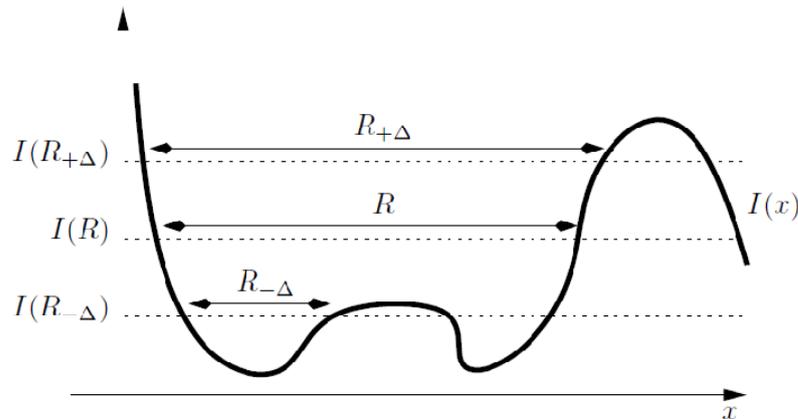


圖 3、MSER 示意圖，橫軸為影像座標，縱軸為灰階值。

extremal region 指的是灰階值小於或等於某一個 intensity(0~255 任一個灰階值)的連通成份(connected component)，在本研究中，偵測 MSER 時二維的影像連通成份定義為 4-connected。extremal region 間會有互相包含的關係，如灰階值小於或等於 255 的 extremal region 就是整張影像，包含了所有可能的 extremal region。

如圖 4，展示了找出所有 extremal region 的過程，灰階值由小到大漸增，一次找一組小於等於此灰階值的最大連通成份，圖 4(a)、(b)、(c) 找到的分別是灰階值為 0、1、2 的 extremal region  $ER_0$ 、 $ER_1$ 、 $ER_2$ ，當灰階值由小變大，偵測出來的一系列 extremal region 互相之間有包含關係，而按照偵測出來的順序，可建立出階層關係，如圖 5，越上層的越晚產生，包含越多像素，可看出  $ER_0$  包含的像素亦被  $ER_1$  所包含， $ER_1$  包含的像素亦被  $ER_2$  包含，產生順序相鄰的兩 extremal region，在階層架構下互為親子關係(parent and child)，定義互為親子關係的兩區域，上層 extremal region 直接包含(immediately containing)下層的 extremal region；而下層 extremal region 為被上層 extremal region 直接包含(immediately contained)。

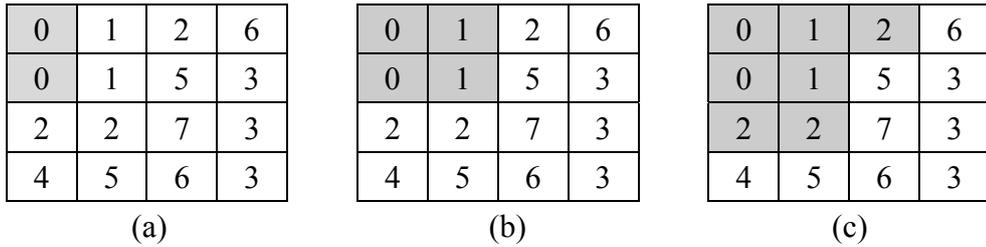


圖 4、4x4 的 8 灰階圖 extremal region 示意圖(a)最大灰階等於 0 的 extremal region : ER0 (b) 最大灰階等於 1 的 extremal region : ER1 (c)最大灰階等於 2 的 extremal region : ER2

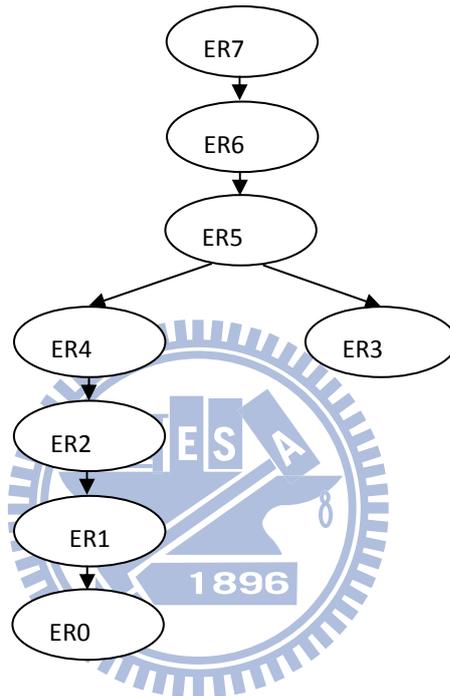


圖 5、按照產生順序與包含關係建成的 Extremal Region 階層結構。

每個 extremal region 都會有一個面積變異值(variation)，為了方便解釋面積變異值，回到一維的例子，如：圖 3，假設 $\Delta > 0$ ， $R_{+\Delta}$ 定義為包含 R 且灰階值超過 $I(R)$ 至少 $\Delta$ 的最小面積之 extremal region

$$R_{+\Delta} = \operatorname{argmin}\{|Q|: Q \supset R, I(Q) \geq I(R) + \Delta\}, Q \text{ 為 extremal region} \quad (\text{公式 2.2})$$

同樣的， $R_{-\Delta}$ 定義為被 R 所包含且灰階值低於 $I(R)$ 不少於 $\Delta$ 的最大面積之 extremal region。

$$R_{-\Delta} = \operatorname{argmax}\{|Q|: Q \supset R, I(Q) \leq I(R) - \Delta\}, Q \text{ 為 extremal region} \quad (\text{公式 2.3})$$

面積變異值 $\rho$ 的計算的方式如下：

$$\rho(R; \Delta) = \frac{|R+\Delta| - |R-\Delta|}{|R|} \quad (\text{公式 2.4})$$

當一個 extramal region 的面積變異值同時小於直接包含他與他直接包含的 extramal region 時，就定義這個 extramal region 為最穩定的 extramal region(maximally stable extramal region)。

由於找出來的MSER為不規則形狀，為了比對不同的MSER區域，本研究以一個橢圓來套合此不規則區域，橢圓的方程式可由計算MSER包含的所有影像點座標之first與second moment來推得[17]。

以上機制找出來的區域為中間深外面淺的橢圓區域，為了也將中間淺外面深的區域考慮進來，本研究將原始影像的灰階值反白之後得到的反白影像，再套用一次上述找MSER的步驟，可以得到更多的特徵區域。

### 2.1.2 過濾不適用於辨識的 MSER 橢圓區域

2.1.1 偵測出初步的 MSER 橢圓區域之後，結果如圖 6，其中包含了太多不適用於辨識的區域，需要進一步過濾：過大、過小、過長、面積變異值過高、接近重覆的橢圓區域，以下以交通大學圖書館為例，展示如何消除這些不適用的區域。

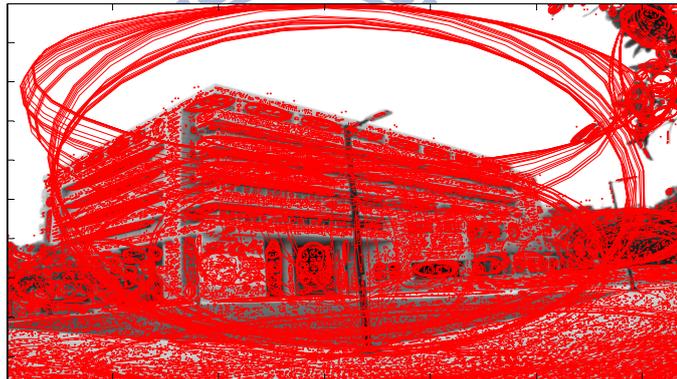


圖 6、初始偵測出來的 MSER 橢圓區域，交通大學圖書館。

可以調整的參數有以下五個：size threshold、 $\Delta$ 、variation threshold、duplicate threshold、aspect ratio threshold。

首先以 size threshold 過濾掉太大與太小的區域，因為太大的區域會包含到不共面的建築物表面，太小的區域資訊少，內容也較單調，本研究只保留未套合橢圓前面積介於影像面積的 $\frac{1}{144}$ 到影像面積的 $\frac{1}{10000}$ 之間的 MSER，過濾結果如圖 7。

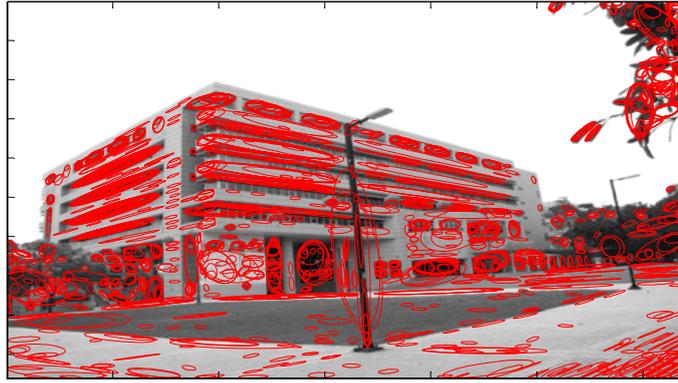


圖 7、去除太大與太小區域後的 MSER 橢圓區域，交通大學圖書館。

$\Delta$  的值越大，邊緣對比度低的區域就會較不容易出現，圖 7 的  $\Delta=1$ ，而圖 8 的  $\Delta=3$ ，一般來說，本研究將  $\Delta$  值設在 3~5，一方面可以保留較具有特色的區域，一方面不致於使特徵區域數量太少。

圖 8 同時也使用 variation threshold 將面積變異值高於此臨界值的區域去除，因為符合 MSER 的條件的區域之面積變異的值不見得低。



圖 8、 $\Delta$  提高、去除面積變異大的區域之後的 MSER 橢圓區域。

為了消除 MSER 橢圓的重覆現象，當找到兩個恰互相包含的未套合橢圓的 MSER 區域之面積太相近時，也就是  $\frac{area(R_{smaller})}{area(R_{bigger})} < duplicate\ threshold$ ，只留下較大的 MSER

區域，消除重覆之後的結果如圖 9。



圖 9、消除重覆區域之後的結果。

因為之後做特徵描述時，需將各橢圓區域標準化成為圓，而過長的區域包含的訊息比較不對稱，需要剔除，使用 aspect ratio threshold 來去除去  $\frac{\text{橢圓短軸長}}{\text{橢圓長軸長}}$  值太小的 MSER 橢圓區域，結果如圖 10。

到此為止，得到的橢圓區域即為本研究所使用的特徵區域。



圖 10、再去除過長的 MSER 區域的結果。

2.1.1 節最後所提到 MSER 特徵區域之兩種型式，一個為直接對原始影像偵測特徵區域，如圖 11(a)，另一個為將原始影像反白之後，再偵測特徵區域，如圖 11(b)，本研究所使用的特徵區域為此兩者之聯集，如圖 11(c)所示。

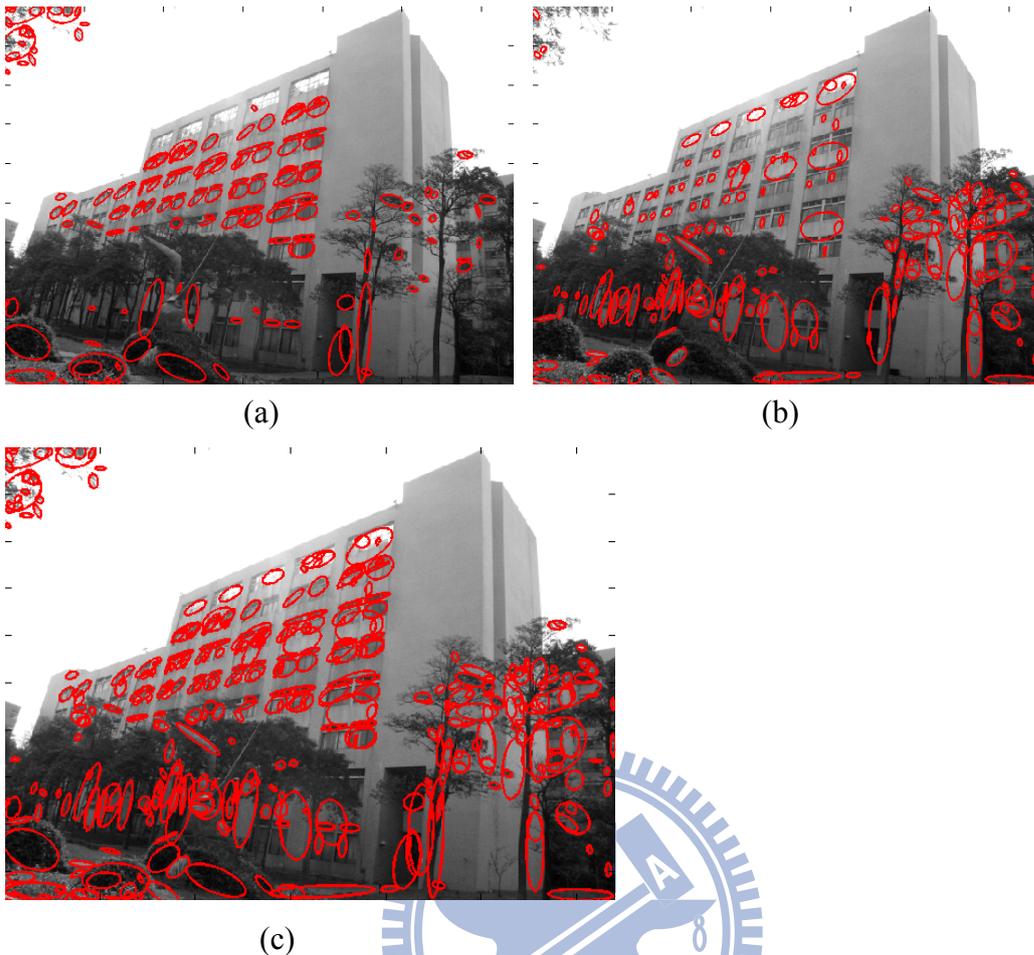


圖 11、(a)直接偵測 MSER；(b)反白之後偵測 MSER；(c)兩種型式的 MSER 的特徵區域聯集

## 2.2 Zernike Moment 特徵區域描述

取出特徵區域之後，還需以特徵描述子來將其轉換成特徵向量，以便於辨識或比較相似度。本研究選擇有旋轉不變特性的 phase-based Zernike Moment 來描述，Zernike Moment 可以將一個特徵區域轉成 magnitude 以及 phase 兩個向量。

因為各個特徵區域的橢圓大小、長短、寬扁、方向不一，需要先將各種橢圓標準化成一樣大小的圓，本研究將包含橢圓的小影像標準化成為  $41 \times 41$  正方形影像，接著才對標準化過後的圓形區域使用 Zernike Moment[2]來做描述，如圖 12。

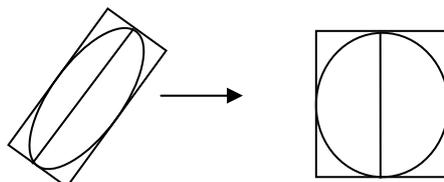


圖 12、橢圓標準化成圓示意圖

## 2.2.1 Zernike Moment 概述

Zernike Moment是由影像函數 $f(\rho, \theta)$ 與Zernike basis function  $V_{nm}(\rho, \theta)$ 做積分運算所得來的，其中Zernike basis function定義在一個單位圓，其次序(order)為 $n$ ，重覆(repetition)為 $m$ ， $m$ 與 $n$ 需要滿足下列兩條件：(1)  $n-|m|$ 為偶數 (2)  $|m| \leq n$ .

$$V_{nm}(\rho, \theta) = R_{nm}(\rho, \theta)e^{jm\theta}, \quad \rho \leq 1, \quad (\text{公式 2.5})$$

而 $\{R_{nm}(\rho)\}$ 定義為

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s} \quad (\text{公式 2.6})$$

Zernike basis function  $\{V_{nm}(\rho, \theta)\}$ 間為互相垂直(orthogonal)關係，如下：

$$\int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) V_{pq}(\rho, \theta) \rho d\rho d\theta = \frac{\pi}{n+1} \delta_{np} \delta_{mq}, \quad \delta_{ab} = \begin{cases} 1 & a=b \\ 0 & \text{otherwise} \end{cases} \quad (\text{公式 2.7})$$

二維的數位影像之 Zernike Moment 計算方式如下：

$$Z_{nm} = \frac{n+1}{\pi} \sum_{(\rho, \theta) \in \text{unit disk}} f(\rho, \theta) V_{nm}^*(\rho, \theta) = |Z_{nm}| e^{i\phi_{nm}} \quad (\text{公式 2.8})$$

定義輸入影像的特徵向量為 $\vec{P}$

$$\vec{P} = [|Z_{11}| e^{i\phi_{11}}, |Z_{31}| e^{i\phi_{31}}, \dots, |Z_{n_{\max} m_{\max}}| e^{i\phi_{n_{\max} m_{\max}}}]^T \quad (\text{公式 2.9})$$

$|Z_{nm}|$ 為 Zernike Moment 的強度(magnitude)， $\phi_{nm}$ 為 Zernike Moment 的相位(phase)。

詳細求得 Zernike Moment 的步驟如下：

- (1) 將每一個標準化成為圓後的目標影像區域所有像素灰階值，先減去所有此區域像素的平均灰階值，再除以灰階值的標準差，可得到一張亮度標準化之後的目標影像區域。
- (2) 將標準化的目標影像區域與 Zernike Moment 的 basis 做積分，可以求得 Zernike Moment 係數，以  $n=1,2,\dots,12$  為例，Zernike Moment 特徵向量的維度為 42。

本研究將一個 MSER 區域以兩個特徵向量表示：

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

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

### 2.2.2 Zernike Moment 相似度的衡量方法

已知兩組 Zernike Moment 的特徵向量為  $P^p : \overline{mag}^p, \overline{phase}^p$  ;  $P^q : \overline{mag}^q, \overline{phase}^q$  ,

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

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

相似度計算如下：

(1) 強度的相似度定義為兩向量的夾角餘弦值，範圍為[0,1]，值越大代表越相似。

$$S_{mag}(\overline{mag}^p, \overline{mag}^q) = \frac{\overline{mag}^p \cdot \overline{mag}^q}{\|\overline{mag}^p\| \|\overline{mag}^q\|} \quad (\text{公式 2.10})$$

(2) 相位的每個維度皆為一個角度，由於相位有循環性，也就是 $\theta$ 與 $\theta + \gamma * 2\pi$ 是一樣的，其中 $\gamma$ 為任意整數，因此，相似度需特別定義，本研究定義相位的相似度為

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

$$\text{其中, } w_{nm} = \frac{|Z_{nm}^p| + |Z_{nm}^q|}{\sum_{n,m} |Z_{nm}^p| + |Z_{nm}^q|} \quad (\text{公式 2.11})$$

$\Phi_{nm} = (\phi_{nm}^p - \phi_{nm}^q) \bmod(2\pi)$  是真實的相位差值，旋轉角度 $\hat{\alpha}$ 是以反覆性(iterative)

的計算 $\hat{\alpha}_{m'} = (\Phi_{nm'} - \hat{\alpha}_{m'-1}) \bmod(2\pi)$ 得來，其中初始的 $\hat{\alpha}_0 = 0$ ， $m' = 1, 2, \dots, m_{\max}$ ， $\hat{\alpha} = \hat{\alpha}_{m_{\max}}$ ，

$m\hat{\alpha}$ 為估計出來的位向差值， $\Phi_{nm}$ 是實際上的位向差值，若兩個 pattern 是一樣的，只存

在一個旋轉關係的話， $|\Phi_{nm} - (m\hat{\alpha}) \bmod(2\pi)| = 0$ ，若兩個 pattern 不同的話，

$|\Phi_{nm} - (m\hat{\alpha}) \bmod(2\pi)| \neq 0$ 。  $S_{phase}(\overline{phase}^p, \overline{phase}^q)$  的範圍為[0,1]，值越大代表越相似。

### 第三章 建築物影像與特徵資料庫建立

本章節涵蓋資料庫建立的所有流程，包含建築物影像拍攝、特徵點偵測與描述，以及使用 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)分群法[15]減少資料庫中存在的特徵重覆的問題，最後會說明如何將特徵向量儲存在可供快速查詢的 kd-tree 資料結構。

#### 3.1 建築物影像拍攝

本研究針對使用一般相機或手機所附相機的使用者，由行人能夠拍攝建築物之角度，來拍攝資料庫與查詢影像，不包括空照圖等特殊角度拍攝的影像；遮擋過於嚴重的建築影像也不在本研究的範圍內，如圖 13。

由於不同建築物長寬、高度、形狀各具，本研究欲以少數幾張影像來代表一棟建築物，除了需使用可抗視角變化的描述方式，在拍攝建築物影像的時候，還必須要拍攝對建築物具有代表性的影像才有意義，因此需要加上一些合理的拍攝條件。

同一棟建築物，不同遠近、角度之下可拍攝的照片有無窮多種可能。近照的時候會拍攝到比較多建築物上的細節，如：外牆磁磚上花紋的細節、外牆上的污漬、透明窗戶中的物品、玻璃的倒影、建築物外貼的海報上的圖案、路邊的旗子上的圖案等等，每張建築物的近照影像因為涵蓋建築物範圍很小，不易決定應使用幾張影像、要涵蓋多細的細節，因此較不具代表性，如圖 14。相對的，遠照的時候，會將許多不相干的物體、建築物照進影像中，難以分清楚誰是主體，建築物的影像包含的像素也會很少，紋理資訊顯的不足，因此也較不具代表性，如圖 15。

另外，建築物表面並非平面，具有凹凸起伏，不同角度拍攝同一個面，會有遮擋的現象，造成影像內容的不同，因此資料庫中需以多個角度記錄建築物同一面，才有代表性，如圖 16。



圖 13、嚴重遮擋的建築物影像範例，交通大學人社一館。



圖 14、建築物近照範例，交通大學工程五館近照。



圖 15、建築物遠照範例，交通大學工程五館遠照。



(a)



(b)

圖 16、(a) 建築物表面凹凸不平造成遮擋範例，交通大學科學二館角度一；(b) 建築物表面凹凸不平造成遮擋範例，交通大學科學二館角度二。

### 3.1.1 建築物資料庫影像拍攝考慮因素

為了保證資料庫中的建築物影像之代表性，本研究考慮以下拍攝因素：

#### A. 主體性：

- (a) 建築物資料庫中須包含各個沒有嚴重遮擋且使用者可能拍攝的面——使用者預設是行人，因此可拍攝的面，依各棟建築物情況有所不同，獨棟的建築物四面都適合拍攝；有些建築則包含不適合拍攝的因素，如與其他建築並排的建築，只有一、兩面適合拍攝。
- (b) 大小——建築物要盡可能占影像的 2/3 寬以上，並且將第一層到最後一層樓都拍進來。
- (c) 單棟——只拍進一棟建築物。
- (d) 位置——建築物的位置要在影像中間。

#### B. 照明度：

以白天的影像為主，包含陰天與晴天的影像，但不包含雨天。

#### C. 清晰度：

避免因手震而使影像模糊。

#### D. 曝光：

建築物主體曝光要適中，避免感光不足(under exposure)或是過度曝光(over exposure)的情況。

#### E. 拍攝角度：

由於有遮擋(Occlusion)的狀況，會造成影像內容的不同。資料庫中為了要儲存接近使用者拍攝的影像內容，建築物同一個面需拍攝多張影像來代表，本研究預設對建築物的一個面，由左到右取得共五張不同位置拍攝的照片，如圖 17。



圖 17、建築物一個面資料庫儲存影像範例，交通大學工程三館其中一個面，資料庫中儲存了五張不同角度的影像。

### 3.1.2 建築物查詢影像拍攝考慮因素

查詢影像除了需符合資料庫影像拍攝考慮因素的主體性、照明度、清晰度、等性質外，為了測試辨識的可靠度，必須與資料庫影像存在以下差異點：視角改變(縮放、位移、旋轉)、光線(陰晴或快門大小)、遮擋(樹、路燈、旗子)。

#### A. 視角改變：

對使用者而言，不同時間拍攝的影像必定會有視角的改變，由於視角改變的量難以定義，本研究採取使用大量不同時間隨機拍攝的查詢影像，測試視角改變對辨識的影響，拍攝的角度與使用者站立的位置皆為隨機，每個資料庫中建築物的面皆各拍攝 5 張查詢影像。

#### B. 光線

為了測試光線對影像辨識的影響，將在晴天與陰天各同時拍攝三組角度相近的

查詢影像，對於資料庫建築物的每一個面，選擇快門值偏小(亮度偏亮)、快門值偏大(亮度偏暗)、快門值適中的影像各拍攝五張查詢影像。

### C. 遮擋

拍攝此部分的查詢影像時，刻意在拍攝建築物的時候將遮擋物體，如：路燈、旗杆、樹葉等拍進畫面中，並且略為遮擋住建築物，對於資料庫建築物的每一個面，各拍攝五張含有遮擋的查詢影像。

## 3.2 建築物資料庫影像特徵偵測與描述

對於資料庫中拍攝的每一張影像，利用 MSER 特徵區域偵測與 Zernike Moment 特徵區域描述，可以求得每一張影像的 MSER 特徵區域與其對應的特徵向量集合，如圖 18、圖 19 所示，以紅色橢圓標出特徵區域。一個橢圓特徵區域，可計算出一個 Zernike Moment 的特徵，包含了 magnitude 向量與 phase 向量兩部份。



圖 18、MSER 特徵區域示意圖，交通大學竹軒。



圖 19、MSER 特徵區域示意圖，交通大學工程三館。

### 3.2.1 使用相鄰影像去除不穩定的建築物影像特徵區域

觀察特徵區域偵測結果，可以發現直接由 MSER 特徵區域偵測方法偵測出來的特徵區域，包含了很多非建築物上的區域，如圖 18 的樹葉、腳踏車、汽車與圖 19 的樹葉與地板上的區域，這些多餘的區域不屬於建築物上，而且如樹為非剛體，會隨風搖晃，每次拍攝都不太相同，因此，需要將這些特徵區域排除在資料庫之外。

建築物遠看時，由於表面的窗戶、外牆等等區域接近平面，較不容易隨著較小的視角變化而有太大的光影改變，因此，在較小視角變化下，建築物本身的橢圓特徵區域偵測出來穩定性較高；相對而言，樹木、腳踏車、汽車等等，形狀較凹凸有致、錯綜複雜，容易隨著小視角的改變，有較大的光影變化，因此，當有視角變化時，在樹木、腳踏車、汽車等等較無法偵測出相同的橢圓特徵區域，如圖 20。



圖 20、有較小視角變化時的 MSER 特徵區域萃取結果，交通大學工程三館。

利用上述特性，配合 Zernike Moment 的區辨力，可以發展出一個全自動去除不穩定特徵區域的簡單方法。本研究對於建築物的一個面拍攝五張照片，以圖 20 為例，五張照片鄰近之兩張間角度變化較小，如果一個特徵區域夠穩定出現，也會同時出現在角度變化較小的兩張鄰近影像。此演算法以建築物同一面的五張照片為一個處理單位。

表 1、使用相鄰影像去除不穩定的建築物影像特徵區域演算法

Algorithm 3.1：使用相鄰影像去除不穩定的建築物影像特徵區域演算法
<p><b>Input：</b></p> <p>由左到右拍攝的五張建築物影像 <math>I_1, I_2, \dots, I_5</math> 之特徵向量集合</p> $M = \{\overline{\text{mag}}_{i,j} \mid 1 < i < 5, 1 < j < N_{\text{original}_i}\},$ $P = \{\overline{\text{phase}}_{i,j} \mid 1 < i < 5, 1 < j < N_{\text{original}_i}\}$ <p>(<math>\text{mag}_{i,j}</math> 與 <math>\text{phase}_{i,j}</math> 為第 <math>i</math> 張影像的第 <math>j</math> 個特徵區域的 magnitude 與 phase 向量，<math>N_{\text{original}_i}</math> 為第 <math>i</math> 張影像初始偵測 MSER 特徵區域的數量)</p>
<p><b>Output：</b></p> <p><math>I_1, I_2, \dots, I_5</math> 之較穩定的特徵向量集合</p> $\bar{M} = \{\overline{\text{mag}}_{i,j'} \mid 1 < i < 5, 1 < j' < N_{\text{stable}_i}\},$ $\bar{P} = \{\overline{\text{phase}}_{i,j'} \mid 1 < i < 5, 1 < j' < N_{\text{stable}_i}\}$ <p>(<math>N_{\text{stable}_i}</math> 為第 <math>i</math> 張影像較穩定的特徵區域數量)</p>
<p><b>Algorithm：</b></p> $\bar{M} = \{\}; \bar{P} = \{\};$ <p>for image_index = 1 to 5</p> <p>  if image_index == 1</p> <p>    reference_set = <math>I_2</math> 之 mag 與 phase 向量集合;</p> <p>  else if image_index == 5</p> <p>    reference_set = <math>I_4</math> 之 mag 與 phase 向量集合;</p> <p>  else</p> <p>    reference_set = <math>I_{\text{image\_index}+1}</math> 與 <math>I_{\text{image\_index}+1}</math> 之 mag 與 phase 向量集合;</p> <p>  end</p> <p>  for j = 1 to <math>N_{\text{original}}_{\text{image\_index}}</math></p> <p>    if similarity(<math>\text{mag}_{\text{image\_index},j}</math>, reference_set 其中一個 magnitude) &gt; 0.7</p> <p>      if similarity(<math>\text{phase}_{\text{image\_index},j}</math>, reference_set 其中一個 phase) &gt; 0.7</p> <p>        將 <math>\text{mag}_{\text{image\_index},j}</math> 加入 <math>\bar{M}</math>、<math>\text{phase}_{\text{image\_index},j}</math> 加入 <math>\bar{P}</math>;</p> <p>      end</p> <p>    end</p> <p>  end</p> <p>end</p>

簡單來說，決定一個特徵區域是否保留的條件，就是檢查它的鄰近影像有無長相接近的 Zernike Moment 特徵向量，此演算法就是查詢鄰近的影像特徵相似度的過程，因為 Zernike Moment 的區辨力夠強，樹葉等等不穩定的區域並不容易在鄰近角度的影像中同時出現，此演算法的結果範例如圖 21。



圖 21、使用相鄰影像去除不穩定的建築物影像特徵區域結果，交通大學工程三館。

### 3.2.2 建築物多面貌的表示法

一棟建築存在多個適合拍攝的面，由於每一個面的長相都不相同，因此資料庫中必須儲存各個不同面之影像，而每面因為建築物本身構造為非平面，會造成遮擋現象，需要拍攝五個不同角度的影像，例如：圖 22 顯示的是交通大學工程五館適合拍攝的三個面，本研究就以十五張影像來表示工程五館這棟建築物。

這麼一來，資料庫中已經概略儲存了使用者可能拍攝的各種角度，但使用者實際拍攝查詢影像時，並不會與資料庫中的其中一張影像一模一樣，即使是資料庫中拍攝角度與查詢影像最接近的一張，仍存在旋轉、位移、光線等等的不同。本研究利用 MSER 與 Zernike Moment 的視角不變特性，找出資料庫中內容與查詢影像最相似的影像，其所代表的建築物，即是建築物辨識的結果。



圖 22、交通大學工程五館適合使用者拍攝的三個面。

### 3.3 使用 DBSCAN 做建築物資料庫影像的特徵區域分群

建築物的一個明顯特色就是存在重覆性的紋路樣式，如：窗戶、磚塊等等，由於 MSER 特徵擷取的特性為抓取灰階穩定區域，容易偵測到窗戶等等內暗外亮或是內亮外暗的結構，因此，資料庫中儲存的特徵向量也會有重覆的現象。

傳統分群法，如：k-means，必須要事先決定分群數，衡量分到哪群的根據是”相對”距離遠近，而每一群的代表點通常是特徵向量的平均向量，無法保留特徵空間中孤立的特徵向量維持自己為孤立群，而是有可能強迫將不同長相的特徵分成一群，取平均值當作代表，會降低特徵原有的區辨能力。由於要保留特徵向量的區辨能力的同時，又要能

適度減低資料庫中重覆的特徵向量造成的空間浪費與搜尋的時間浪費，本研究使用了 DBSCAN( Density-Based Spatial Clustering of Applications with Noise )分群法[15]來對特徵向量做分群。

### 3.3.1 DBSCAN 概述

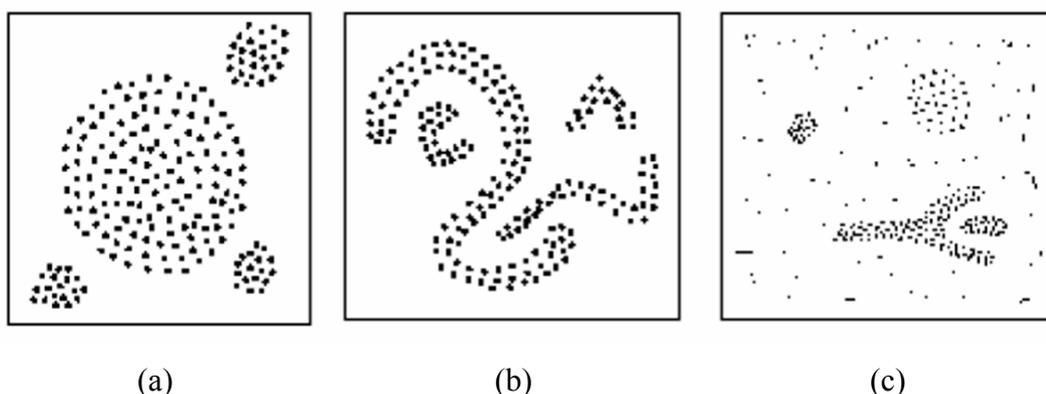


圖 23、 DBSCAN 演算法可分成(a) 4 群；(b) 4 群；(c) 4 群加上孤立點。

DBSCAN 為以基於分布的疏密為主(density-based)的一種分群法，具有三個特點：適用任意形狀、可以處理孤立點、只要一次掃描(one scan)。本研究選擇此方法是基於 DBSCAN 可以處理孤立點的特性。

DBSCAN 的輸入是  $k$  維空間的點座標集合，輸出是各點的分群結果，判斷是否為一群的依據是密度夠高連續成團的點集合為一群，如圖 23 (a)(b)(c)，在(c)的例子可以看出來，分布較分散的孤立點並不會被 DBSCAN 分到任何群中，而是視作孤立點。

實做 DBSCAN 的方法是利用兩個量測值來決定分群的結果，一個是 Eps，單位為距離；一個是 Minpt，單位為點數。想法是將某一點周遭 Eps 半徑內的點，視作此點的鄰近點，若某一群鄰近點含有 Minpt 個點以上，這些點就會被視作為分布的夠密，而被分為同一群，由此簡單想法，隨著串接就可以延伸出任何形狀的分群結果，簡單的 DBSCAN 實作演算法如下：

表 2、Density-Based Spatial Clustering of Applications with Noise 演算法

<p><b>Algorithm 3.2 :</b> Density-Based Spatial Clustering of Applications with Noise 演算法</p>
<p><b>Input :</b>  <math>Pt = \{ p_i   i = 0, 1, \dots, M \}</math> : k 維空間的 M 個點  distance_map : 記錄任兩點的距離，預設為歐氏距離  Eps : 單位為距離  Minpt : 單位為點</p>
<p><b>Output :</b>  M 個點皆被標記一個分群號碼，表示被分到哪一群，其中孤立點被視作 noise 標記成 default 號碼</p>
<p><b>Algorithm :</b>  For each point <math>p_i</math>, <math>i = 1: M</math> {      get the set of points, <math>N(p_i)</math> which are <math>p_i</math>'s neighbor within the distance <math>Eps</math>;      if <math>N(p_i)</math> contains less than MinPt points {          if <math>p_i</math> hasn't been labeled {              label <math>p_i</math> as a default number;          }      } else {          if <math>N(p_i)</math> contains more than or equal to 2 clusters' label {              set a new cluster label to all point in <math>N(p_i)</math> and merge the clusters that <math>N(p_i)</math> contains and let them be the same new label;          } else if <math>N(p_i)</math> contains only one label {              label all points in <math>N(p_i)</math> as the cluster label that <math>N(p_i)</math> contains;          } else {              set a new cluster label to all points in <math>N(p_i)</math>;          }      }  }</p>

### 3.3.2 運用 DBSCAN 對影像特徵區域的分群

本研究 3.2 節留下來較穩定的建築物影像特徵區域做 DBSCAN 的分群，使用的特徵向量為 Zernike Moment 的強度(magnitude)，相似度的衡量方式為兩向量夾角的餘弦值，見公式 2.10，相似度(similarity)的值介於 0 到 1 之間，值越大越相似。

Eps 的意義在本研究的詮釋下為相似度的 lower bound：與一個特徵向量相似度大於 Eps 的點，為此特徵向量的相似向量(鄰近點)，若相似向量(鄰近點)至少有 Minpt 個，就將這些特徵向量分成同一群。

由於要避免 DBSCAN 存在的一大缺點：可能串接不同長相的特徵向量，串接的太長，而使得同一群的兩個特徵向量可能會因此不相似，所以本研究採用部分分群的方法，將建築物同一面的每五張影像的特徵向量做一次 DBSCAN 分群，分群的結果範例請見圖 24。



圖 24、DBSCAN 演算法對交通大學工程三館特徵區域分群結果的其中一群。

本研究以同群特徵向量強度(magnitude)的平均值(群中心)來當作代表此群的特徵向量，並且計算被分成同一群的每一個特徵向量與群中心特徵向量的相似度，計算這些相似度的平均與標準差，判斷有無串接長相不同的兩種區域的現象，若相似度的平均太小，代表有 DBSCAN 串接過長的現象，此群中心不足以代表此群的特徵向量，無法成功分群，在此將此種群中的點皆歸為孤立點。

因為建築物也會存在穩定出現卻非重覆樣式(repeat pattern)的特徵區域，對於沒有被分到群的特徵向量，本研究不將其視為干擾點(noise)，由於這些特徵向量對辨識而言也有區辨力，所以將每個孤立點分別被視為一個一個不同的孤立群。

對建築物的資料庫每五張為一個處理單位做 DBSCAN 分群，每一群皆有一特徵向量來代表，接著將所有代表的特徵向量儲存在一起，就成為資料庫特徵向量集合。當一特徵向量代表一孤立群時，對應到某一張影像上的一個特徵區域；當一特徵向量代表一般群時，則可能對應到多張影像上的數個特徵區域，因此，本研究需以額外一個資料結構來儲存資料庫中每一個特徵向量與所出現在哪些圖上、出現的個數等等，如表 3，由於 DBSCAN 分群的處理單位是一次五張影像，所以一個代表的特徵向量最多出現在五張影像上。

表 3、特徵向量分群結果與影像關係的資料結構

特徵向量編號	共有多少名額	出現在幾張圖中	第一張影像編號	第一張影像幾個	第二張影像編號	第二張影像幾個	第三張影像編號	第三張影像幾個	第四張影像編號	第四張影像幾個	第五張影像編號	第五張影像幾個
1	10	3	1	6	2	3	4	1				
2												
...												

本研究對建築物各個面拍攝五張影像，每張影像都會記錄含有哪些特徵向量、有多少數量，以這些特徵向量與特徵向量的數量來代表此棟建築物。在最理想的情況下，當使用者對著一棟建築物拍攝一張查詢影像，拍攝內容理應會最接近資料庫中正確答案的建築物的其中一張影像，而偵測出來的影像特徵向量與數量也會接近這張影像，因此，只要能判斷查詢影像與資料庫中哪張影像最相似，就能判斷出查詢影像為哪棟建築物的影像。

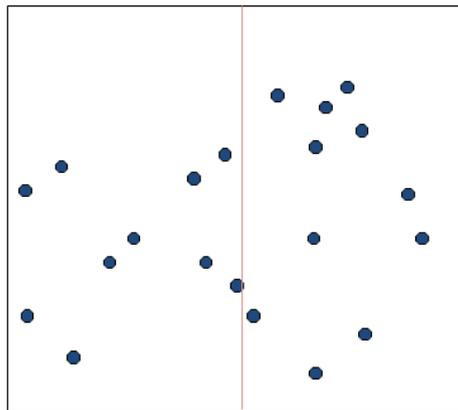
### 3.4 建築物影像特徵的 kd-tree 資料庫建立

3.2 節將較不穩定出現的特徵區域刪去之後，可得到 DBSCAN 分群的代表向量，所有代表向量可對應到資料庫中所有的特徵區域，將每一個代表向量，當作 k 維空間上的一個座標點。對於查詢影像上偵測到的每一個特徵區域，必須要比對出其與資料庫中哪一些特徵區域最相近，而這個比對方式若為一個一個窮比對需要花費大量時間，kd-tree[16]提供了一個 k 維空間快速搜尋最相近點的資料結構。

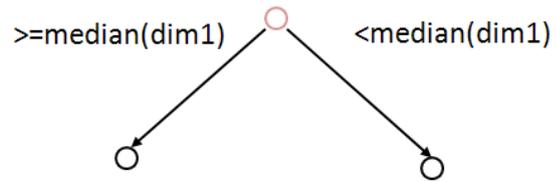
如圖 25 所示，kd-tree 是一種二元樹，每一層以一個維度(目標維度)來將資料點分成

兩部份，每次分割依照目標維度將子樹中資料點排序，取中位數的值當作分割的依據，目標維度的座標大於或等於中位數的資料點分到子樹的一側，目標維度的座標小於中位數的資料點分到另一側。當所有維度都已使用之後，再重覆從第一個維度開始切。

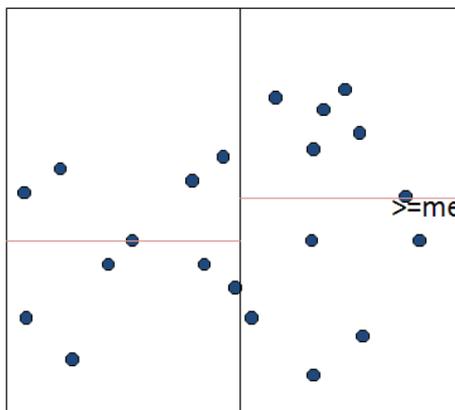
以三維空間為例，一開始按照所有的資料點的 x 座標作排序，取出 x 座標中位數當作基準，x 座標大於或等於中位數的分到一邊，小於中位數的分到另一邊，接著，對兩子樹以 y 座標當作目標維度分別用一樣的方法做分割，可將空間分成四個區域，接著使用 z 座標做分割，三個維度都用完之後，再從 x 座標為目標維度開始切割，以此類推。



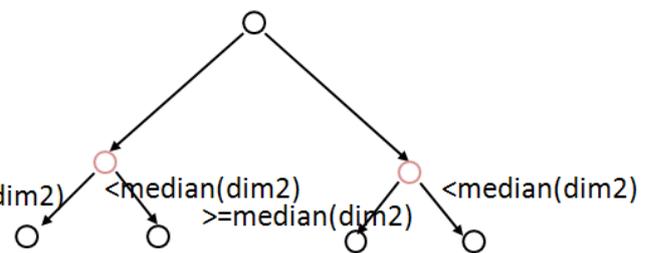
(a)



(b)



(c)



(d)

圖 25、建立 kd-tree 的流程示意圖 (a) kd-tree 的資料點第一維分割示意圖；(b) 對應於(a)建立出來的 kd-tree；(c) kd-tree 的資料點第二維分割示意圖；(d) 對應於(c)建立出來的 kd-tree。

表 4、kd-tree 資料庫建立演算法

<p><b>Algorithm 3.3 :</b> kd-tree 資料庫建立演算法</p>
<p><b>Input :</b> N 個 k 維的特徵向量(視為空間座標上的點座標)集合 <b>Dataset</b> , <b>dim</b> 為分割的維度</p>
<p><b>Output :</b> kd-tree , 每個葉節點儲存一個特徵向量</p>
<p><b>Algorithm :</b>  <pre> kd_tree_build(<b>Dataset</b> ,<b>dim</b>){   If <b>Dataset</b> contain only one point     make a node and set it as a leaf node containing the point, stop     branching   else     1. sort all points in <b>Dataset</b> according to feature dimension <b>dim</b>.     2. extract the median value of feature dimension <b>dim</b> in <b>Dataset</b>,     make a new node and save the median value information in it.     3. <b>Dataset_bigger</b> :The points in <b>Dataset</b> who's value of feature     dimension <b>dim</b> &gt;= median value     4. <b>Dataset_smaller</b> :The points in <b>Dataset</b> who's value of feature     dimension <b>dim</b> &lt; median value     5. set <b>Dataset_bigger</b> as the new node's right child     6. set <b>Dataset_smaller</b> as the new node's left child     7. call kd_tree_build (<b>Dataset_bigger</b> ,(dim+1)%k)     8.call kd_tree_build (<b>Dataset_smaller</b> ,(dim+1)%k) } </pre> </p>

建立 kd-tree 的演算法使用到 recursive 的概念，每個節點會儲存此節點的目標維度與分割時使用的中位數座標值，以供查詢時比對。

空間中的座標點，則儲存在葉節點上，當子空間只剩下一個座標點時，就會停止繼續分割。

本研究將 ZM 的強度特徵向量視做空間上的一個座標點，對資料庫中所有 ZM 的強度向量建立 kd-tree 資料結構，以供查詢影像的特徵向量可以很快的比對 kd-tree 的節點找到最相似的資料庫特徵。

## 第四章 建築物影像檢索機制

本研究對於影像檢索的概念為與查詢影像最相似的資料庫影像，其共同擁有相似特徵區域數量與其他資料庫影像相比會最多。

利用此一概念，先對查詢影像利用第二章的方法偵測 MSER 橢圓特徵區域，計算對應的 Zernike Moment 特徵向量，包含強度與相位兩向量，對於查詢影像的每一個特徵區域，利用 kd-tree 的索引結構，將資料庫中與其最相似的一個(一些)特徵區域找出來，最後統計資料庫中各張影像擁有與查詢影像相似的特徵區域數，取最高票的一張影像所屬的建築物，當作辨識的輸出。

### 4.1 kd-tree 最近點搜尋與範圍搜尋

#### 4.1.1 利用 kd-tree 找強度最接近的特徵向量

資料庫中所有的特徵區域都以 Zernike Moment 的強度(magnitude)與相位(phase)來表示，將資料庫中的特徵區域利用強度做分群，以各群的平均強度當作各群的代表，由於相位有循環性，無法以平均值當作代表，分群之後便以強度為主來做檢索。

將資料庫中分群之後的強度特徵向量利用 kd-tree 資料結構製作索引，查詢影像的一強度特徵向量為  $\overline{mag^q} = [ |Z_{11}^q|, |Z_{31}^q|, \dots, |Z_{n_{\max}m_{\max}}^q| ]^T$ ，利用 kd-tree 找出最相似強度特徵向量的方法如下(在此，將特徵向量視為點座標)：

- (1) 將  $\overline{mag^q}$  的每個維度與 kd-tree 的節點儲存的分割的中位數值  $m_h$  (h 為樹的高度)，做一一比較，快速找出  $\overline{mag^q}$  所在的葉節點，以  $\overline{mag^q}$  葉節點中含有的點座標  $\overline{mag^i}$ ，得到一近似的最近距離  $d_{\text{approximate}}$ ，距離計算方式為歐氏距離，圖 26 為一示意圖(實際上為一高維空間)，空心圓圈部分為  $\overline{mag^q}$ ，實心圓圈部分為資料庫中的特徵向量。

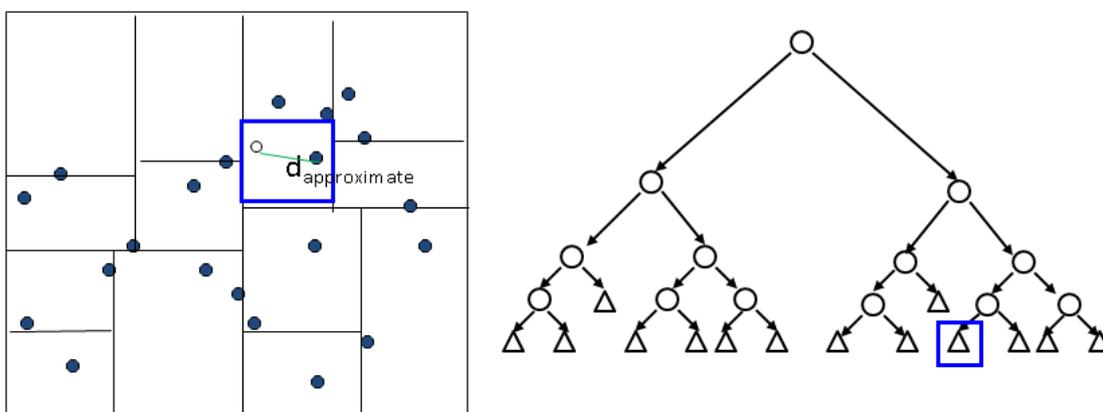


圖 26、kd-tree 近似最近點示意圖，空心圓圈部分為  $\overline{mag^q}$ ，實心圓圈部分為資料

庫中的特徵向量。

(2)重新走一次 kd-tree，從根節點開始，當遇到一個節點時，比較  $\overline{mag^q}$  和分割的中位數  $m_h$ (平面)的距離與  $d_{\text{approximate}}$  的大小：

- 若  $d_{\text{approximate}}$  較大，表示 kd-tree 的兩子樹皆有可能出現比  $d_{\text{approximate}}$  還近的点，兩邊都要繼續檢查，如圖 27
- 若  $d_{\text{approximate}}$  較小，表示分割平面的另一邊不會出現比  $d_{\text{approximate}}$  還近的点，只需檢查  $\overline{mag^q}$  所在的子樹即可，如圖 28。

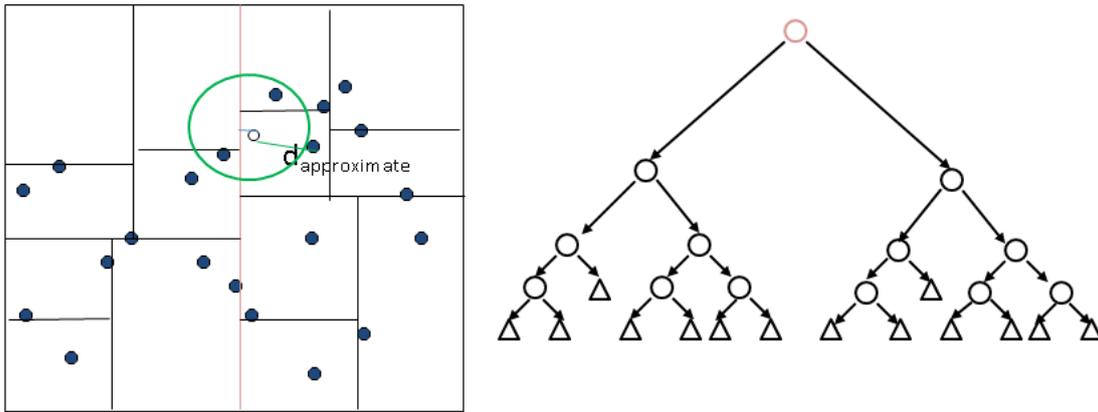


圖 27、當 approximate nearest distance 比  $\overline{mag^q}$  到分割平面的距離大的示意圖。

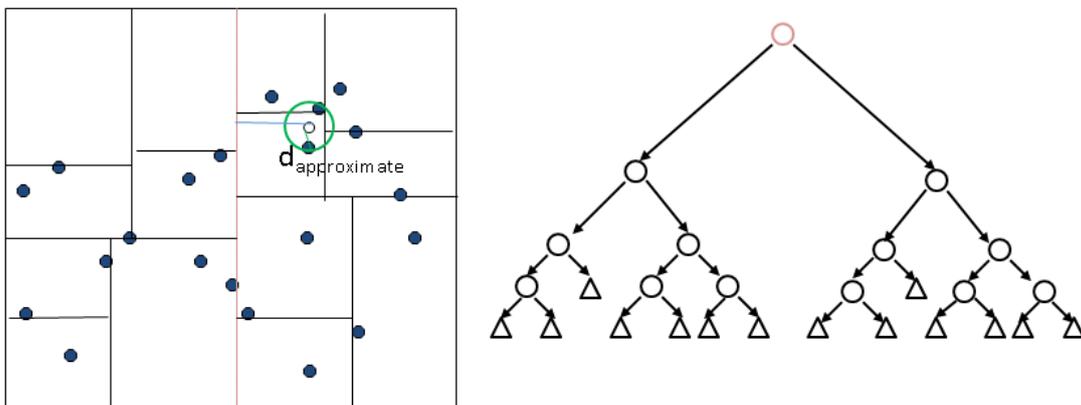


圖 28、當 approximate nearest distance 比  $\overline{mag^q}$  到分割平面的距離小的示意圖。

(3)利用(2)可以將不可能含有更近点的空間排除，如圖 29，當找到葉節點時，判斷  $\overline{mag^q}$  與葉節點包含的点之座標的距離是否小於  $d_{\text{approximate}}$ ，若此距離較小，則找到的為更近的点，更新  $d_{\text{approximate}}$  的值與記錄目前找到最近点的座標。

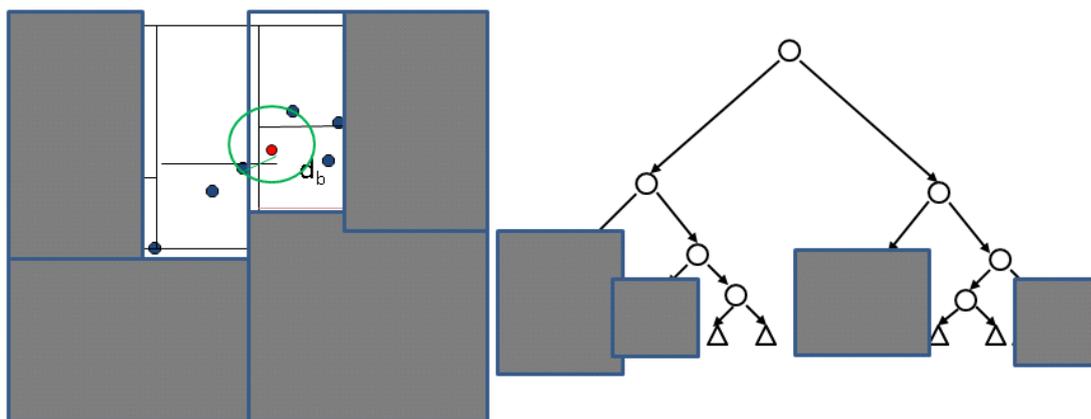


圖 29、排除掉不可能含有更近點的空間。

(4)檢查完所有可能存在較近點的分支之後，就可以找到最近點

#### 4.1.2 利用 kd-tree 範圍搜尋強度接近的特徵向量

利用 kd-tree 範圍搜尋可找到與查詢影像上的特徵區域相比，在資料庫中 Zernike Moment 強度特徵向量最接近的特徵區域。然而，因為資料庫含有許多類型的影像，資料庫中強度最接近的特徵向量，可能不是正確答案的建築物影像上的特徵區域，因此，需要放寬搜尋的標準，尋找以查詢影像強度特徵向量為中心，一定範圍內的資料庫特徵向量，方法如下：

(1)先利用 4.1.1 節所述方法，找到最近點與查詢點  $mag^q = [|Z_{11}^q|, |Z_{31}^q|, \dots, |Z_{n_{max}m_{max}}^q|]^T$  與最近點的距離  $d_{nearest}$ 。

(2)利用 kd-tree 的資料結構找出  $[|Z_{11}^q| \pm s \cdot d_{nearest}, |Z_{31}^q| \pm s \cdot d_{nearest}, \dots, |Z_{n_{max}m_{max}}^q| \pm s \cdot d_{nearest}]^T$ ，所包

含的點，稱此範圍為一個限定範圍的盒子(bounding box)，找尋方法如 4.1.1 的步驟(2)~(4)，經過每個 kd-tree 的內部節點時，判斷左子樹或右子樹是否可能含有範圍內的點，排除掉不含有範圍內點的路徑，最後便可以找出所有在範圍內的點，在此稱這些找到的點為接近點。

範圍訂定的條件參照最近點的距離的原因是為了保留彈性，當最近點的距離越近，查詢區域與資料庫的此區域相似度越高，越有可能是正確建築物上的區域，因此，要額外找的範圍就訂的小一些；反之，當最近點的距離越遠，查詢區域與資料庫的此區域相似度就越低，要額外找的範圍就訂的大一點。

限定範圍的盒子為一個多維空間的立方體，s 的值決定它的邊長，此盒子範圍內與查詢點最近距離為  $s \cdot d_{nearest}$ ，最遠距離為  $\sqrt{k} \cdot s \cdot d_{nearest}$ ，k 為維度。當  $s=0.5$  時，以 Zernike Moment 的 order  $n=1,2,\dots,12$  而言， $k=42$ ，最近距離為  $0.5 d_{nearest}$ ，最遠距離約為  $3.2 d_{nearest}$ 。

## 4.2 建築物檢索結果的產生

由 4.1 節所述的機制可以快速找到資料庫與  $\overline{mag^q}$  距離最近的點(最近點)或是附近的一群點(接近點)，因為是歐式距離相對較近，並不代表特徵向量實際上夠相似，因此，需要進一步判斷此最近點(接近點)的特徵向量與  $\overline{mag^q}$  的相似度，使用 2.2.2 節的強度相似度計算方法，當相似度大於一門檻值，就視兩特徵區域為相像，經驗值為 0.85~0.9。當滿足相像條件，就給此資料庫特徵向量所在的影像一票，若不滿足相似條件，代表此特徵區域並無出現在資料庫中，便不投票。

若在不用 DBSCAN 特徵分群的情況下，可以再進一步檢查相位的相似度，當找到強度特徵向量的最近點(接近點)時，檢查同一個特徵區域的相位向量，計算強度相似度與相位相似度，當兩者都大於門檻值時，才算滿足相像條件，相位的相似度門檻值經驗值為 0.75~0.85。

查詢影像中每一個特徵區域都按照上述機制，找到資料庫中最相似(接近)的特徵區域，並且投票，最後計算得票數，每張資料庫中的影像都會有一個得票數，得票數最大的影像代表與查詢影像含有最多相似區域，其所代表的建築物當作辨識的結果。



## 第五章 辨識實驗結果與評估

### 5.1 交大建築資料庫的辨識結果

#### 5.1.1 交大建築資料庫介紹

本研究以國立交通大學的建築物影像來實作檢索與辨識，影像共分成兩大部分，一部分為建築物資料庫影像，另一部分為建築物查詢影像，比較如表 5：

表 5、交大建築資料庫影像與建築物查詢影像的比較

	資料庫影像	查詢影像
拍攝相機	Canon SX11S 相機	Canon EOS 500D 相機
相片解析度	1600x1200	2352x1568
拍攝日期	2010 年 3 月到 4 月	2010 年 6 月 5 日到 6 月 8 日
影像張數	190 張	2280 張
涵蓋建築物	22 棟	22 棟

資料庫影像拍攝方式符合 3.1.1 節所述，每個面由建築物的左到右連續拍攝不同角度各 5 張照片，資料庫中一共 22 棟建築物，190 張照片，詳細列舉如表 6，拍攝條件為符合正確曝光、主體性、儘量避免遮擋、無手震等，資料庫影像範例請見表 7。

表 6、交大建築資料庫涵蓋的建築物列表

資料庫涵蓋的建築物(幾個面)		
工程三館	工程四館(3 面)	行政大樓(2 面)
中正堂	人社一館	電子資訊大樓
固態電子系統大樓	奈米電子研究大樓	女二舍(4 面)
圖書館(3 面)	工程五館(3 面)	十二舍(2 面)
十三舍	工程六館(2 面)	田家炳光電大樓(2 面)
交映樓(2 面)	研二舍	綜合一館(2 面)
竹軒	體育館	研一舍
科學二館(2 面)		

表 7、交大資料庫影像範例

	視角 1	視角 2	視角 3	視角 4	視角 5
工程 三館					
工程 四館 (面 1)					
工程 四館 (面 2)					
工程 四館 (面 3)					
行政 大樓 (面 1)					
行政 大樓 (面 2)					
中正堂					

查詢影像拍攝方式符合 3.1.2 節所述，共分成晴天與陰天兩種，每一種根據拍攝方式又可分為 6 個類別，所以總共 12 種查詢影像，舉例如表 8 所示。其中，每一種情況對資料庫所有的建築物之面隨機拍攝 5 個不同角度的影像，拍攝方式包含了旋轉(直拍、橫拍、斜拍)、大小縮放、位移等等，每種查詢影像各 190 張，如表 9。

表 8、交通大學查詢影像分類範例表

	晴天	陰天
類別 A (中等亮度)		
類別 B (偏亮)		
類別 C (偏暗)		
類別 D (中等亮度、 加上遮擋物)		
類別 E (偏亮、 加上遮擋物)		
類別 F (偏暗、 加上遮擋物)		

表 9、交通大學查詢影像範例表



### 5.1.2 交大建築物影像不同拍攝因素下的辨識結果

查詢影像一共分成 12 種不同的狀況，各 190 張，分別做查詢，以第一高票的影像所屬的建築物當作辨識的結果，得到的辨識率如表 10 所示，表 10 使用的辨識流程是資料庫有使用相鄰影像去除不穩定區域，且做 DBSCAN 分群之後，查詢特徵區域使用 kd-tree 找 bounding box( $s=0.5$ )之內的候選特徵區域，檢查 ZM 強度夠相近才投票的方法，其中，晴天的辨識率略低於陰天，而偏亮的影像辨識率略低於中等亮度與偏暗的影像。

比較類別 D、E、F 與類別 A、B、C 可知，加上部分遮擋物的影像對辨識率並無明顯的影響，可以顯示出本研究使用基於區域(region-based)的影像檢索機制的優點，在部份遮擋的情況下，仍然有足夠數量且具區辨力的特徵區域能夠讓正確的建築物影像得到最高票。

表 10、交大建築物查詢影像辨識結果

	晴天	陰天
類別 A(中等亮度)	178/190 = 93.6 %	186/190 = 97.8 %
類別 B(偏亮)	175/190 = 92.1 %	175/190 = 92.1 %
類別 C 偏暗	177/190 = 93.1 %	183/190 = 96.3 %
類別 D(中等亮度、 加上遮擋物)	178/190 = 93.6 %	183/190 = 96.3 %
類別 E(偏亮、 加上遮擋物)	175/190 = 92.1 %	179/190 = 94.2 %
類別 F(偏暗、 加上遮擋物)	176/190 = 92.6 %	184/190 = 96.8 %

### 5.1.3 時間和空間複雜度評估

本研究所使用的設備為 2.83GHz 的 Intel Q9550 CPU，3GB 的記憶體，使用程式語言為 Matlab，並且尚未對程式碼做向量最佳化的動作。

本研究影像辨識所需時間可以分成兩部份，一個是影像特徵偵測與描述，另一個是與資料庫做比對的時間。

影像特徵偵測與描述所需的時間與影像解析度有關之外，還與影像內容有關，影像越平滑，初始偵測的 extremal region 越少，整體偵測 MSER 所需時間就越少。一般而言，特徵偵測與描述，所需時間與影像解析度成正相關，如表 11，分別測試 92 張不同解析度的影像取其平均與標準差。

表 11、不同解析度的測試影像做特徵偵測與描述所需時間

解析度	2352x1568	1600x1200	640x480
時間(標準差)	13.8 sec (4.3 sec)	5.8 sec (1.58 sec)	1.8 sec (0.7 sec)

與資料庫比對所需時間與資料庫大小有關，如果是使用 kd-tree 找最近點，幾乎只與資料庫含有特徵向量多寡有關(因為 kd-tree 為一 balanced tree)，如果是使用 kd-tree 找一個範圍，則又與資料庫中特徵分布在查詢資料特徵附近的點數有關。一般而言，與資料庫做比對的時間與資料庫的特徵向量多寡成正相關，而資料庫特徵向量的多寡，又與資料庫影像多寡、有無使用 3.2.1 節所述使用相鄰影像去除不穩定的建築物影像特徵區域方法、有無做 DBSCAN 分群有關，比較如表 12，使用的測試資料為陰天、中等亮度、無遮擋的測試資料。

表 12、各個會影響資料庫中特徵向量個數的步驟對時間的影響

實驗編號	實驗 1	實驗 2	實驗 3
狀況	無去除不穩定特徵且無 DBSCAN，判斷找到的 bounding box 包含區域與查詢影像上區域相似性只看 ZM 強度	有去除不穩定特徵但無 DBSCAN，判斷找到的 bounding box 包含區域與查詢影像上區域相似性只看 ZM 強度	有去除不穩定特徵且有 DBSCAN，判斷找到的 bounding box 包含區域與查詢影像上區域相似性只看 ZM 強度
feature vector 數量	113194 個	68036 個	56089 個
bounding box 大小參數 s	0.5	0.5	0.5
辨識正確張數	180 張(共 190 張)	180 張(共 190 張)	186 張(共 190 張)
辨識正確率	94.7%	94.7%	97.8%
feature extraction and description 平均花費時間	17 sec	17.4 sec	17.6 sec
indexing 平均花費時間	39.7 sec	25.1 sec	22.4 sec
平均處理一張 query 影像的時間	56.7 sec	42.5 sec	40.0 sec

與資料庫做比對的時間也與 kd-tree 範圍搜尋時使用的範圍大小有關，4.1.2 節所述控制範圍大小的參數為 s，s 越小，所需要詳細比對相似度的資料庫特徵區域就越少，所需時間越少。然而，當 s 太小時 bounding box 中可能會不包含任何點，本研究測試當 s=0.2 時，範圍搜尋無法做辨識，因為此時最近點有太大機率不存在 bounding box 中，造成資料庫中影像得票數皆為 0 票。觀察可知，雖然縮小 s 可以讓辨識時間減短，辨識率卻會稍微下降，而縮小 s 也有所極限(s=0.2)，如表 13。

表 13、bounding box 的範圍大小對於查詢時間的影響

實驗編號	實驗 3	實驗 4	實驗 5
狀況	有去除不穩定特徵且有 DBSCAN，判斷找到的 bounding box 包含區域與查詢影像上區域相似性只看 ZM 強度	有去除不穩定特徵且有 DBSCAN，判斷找到的 bounding box 包含區域與查詢影像上區域相似性只看 ZM 強度	有去除不穩定特徵且有 DBSCAN，使用 <b>kd-tree 找最近點</b> ，判斷找到的最近點區域與查詢影像上區域相似性只看 ZM 強度
bounding box 大小參數 s	0.5	0.35	無使用 bounding box
feature vector 數量	56089 個	56089 個	56089 個
辨識正確張數	186 張(共 190 張)	182 張(共 190 張)	190 張(共 190 張)
辨識正確率	97.8%	95.8%	100%
feature extraction and description 平均花費時間	17.6 sec	18.2 sec	17sec
indexing 平均花費時間	22.4 sec	10.5 sec	8 sec
平均處理一張 query 影像的時間	40.0 sec	28.2 sec	25sec

空間複雜度的評估方面，由於 kd-tree 是 ballanced tree，儲存的容量或是節點數直接由資料庫儲存的向量多寡來決定，表 12 顯示本論文的去除不穩定區域與 DBSCAN 機制可以有效節省空間的使用，如果兩者皆不使用的話，特徵向量數為 113194 個(kd-tree 佔記憶體 22.0M)，見實驗 1；去除不穩定特徵區域後，特徵向量個數可以減少到 68036 個(kd-tree 所佔空間約 12.9M)，省下約 40%空間，見實驗 2；若去除不穩定特徵區域之後再加上 DBSCAN 分群，可以使得特徵向量數減少到 56089 個(kd-tree 所佔空間約 10.6M)，與初始個數相比減少了 50%左右，而在此實驗中，這些改進方法並不會對實驗辨識率有明顯不良影響。

## 5.2 ZuBud 建築資料庫的辨識結果

### 5.2.1 ZuBud 建築資料庫介紹

ZuBud(Zurich Buildings Database)是由 Swiss Federal Institute of Technology 所建立的建築物影像資料庫，以 Zurich 街上建築物的為對象，分成兩個部分，資料庫部分包含了 201 棟建築物，每棟建築物各五張任意視角的影像，共有 1005 張；查詢影像部分包含了 115 張建築物的影像。為了包含更多不同照明度的影像，所有影像皆在不同季節與天氣狀態之下由兩台不同相機所拍攝，如表 14。

表 14、ZuBud 建築物影像範例

查詢影像	資料庫影像				
					
					
					
					

### 5.2.2 ZuBud 建築資料庫的辨識結果

本研究測試 ZuBud 資料庫的辨識率時，發現測試兩區域相似性時，必須同時考慮 Zernike Moment 的強度與相位資訊才能獲得更好的辨識率，而使用 kd-tree 搜尋的時候找最近一點的辨識率略高於找一個範圍的辨識率，辨識結果如表 15。

表 15、 ZuBud 辨識結果

實驗編號	實驗 5	實驗 6	實驗 7
狀況	判斷找到的 bounding box 包含區域與查詢影像上區域相似性只看 ZM 強度	判斷找到的 bounding box 包含區域與查詢影像上區域相似性同時看 ZM 強度與相位	使用 kd-tree 找最近一點，之後同時判斷 ZM 強度與相位
bounding box 大小 參數 s	0.5	0.5	無
feature vector 數量	107744 個	107744 個	107744 個
辨識正確張數	82 張	108 張	109 張
辨識正確率	71.3%	93.9%	94.8%

觀察實驗結果可以發現，在 ZuBud 資料庫的辨識，因為資料庫較交通大學的資料庫大得多，含有更多形形色色的區域，當查詢影像的特徵區域使用 kd-tree 找到資料庫中一些強度接近的特徵點之後，只進一步細比較查詢影像與這些資料庫中特徵區域 ZM 的強度是否夠相似來決定投票與否，條件會不夠嚴，因為當兩區域 ZM 的強度相近，還可能存在如鏡射、反白等等不相似的情況，因此需要讓 Zernike Moment 發揮它更強大的辨識力，也就是：當 ZM 的強度與「相位」皆相近才能確定此兩區域夠相似，才能夠投票。

## 第六章 結論與未來發展

本研究結合 MSER 特徵區域與 Zernike Moment 描述子，並使用 kd-tree 的多維空間索引結構來做建築物的影像辨識。在交通大學建築資料庫可獲得 95% 左右的辨識率，在 ZuBud 資料庫最佳也可獲得 94.8% 的辨識率，辨識率皆相當不錯。

針對 MSER 的偵測在不同的光線條件下，會偵測到的不穩定區域，本研究利用其他張相同的建築影像偵測之區域，做互相消除的動作，有效增加的空間與時間的效率。本研究並利用 DBSCAN 分群法減少資料庫中的重覆區域，而為了保有特徵之間的區辨力，只將非常相似的區域分成同群，因此分群產生的效益並未非常巨大，未來若能找到能夠兼顧區辨力與有效減少資料庫中儲存的特徵向量的數量之分群法，在空間與時間上，一定會有更大的改進。

本研究發現面對較大的資料庫，如 ZuBud，需同時考慮強度與相位的相似性，確認查詢影像上的區域與資料庫的區域極相似，才能充分發揮 Zernike Moment 的區域辨識能力，獲得更好的辨識率。

本研究使用的程式語言為 Matlab，並且尚未對程式碼做最佳化的動作，因此主要著重在準確度上，而非辨識的速度，未來如果對程式碼做最佳化，並且使用編譯式的程式語言，或是將重覆的動作平行化，將可以獲得更佳效率。

然而，本研究所用辨識的方法，並未結合各個區域間的空間關係，為了獲得更高的區辨力，可再結合各區域間的空間資訊來做強化，如以資料庫影像與查詢影像的特徵區域點對應估計出 3D 的關係，來做更精確的比對。在建築物的 3D 的比對方面將會面臨到問題是特徵區域的重覆性，如何克服特徵點的重覆性做到有效的點對應，是建築物 3D 比對將面臨到的重要課題。

## 參考文獻

- [1] Wang J., G. Wiederhold, O. Firschein, and S. Wei, “Content-Based Image Indexing and Searching Using Daubechies’ Wavelets”, *Int’l J. Digital Libraries*, vol. 1, pp. 311-328, 1998.
- [2] Chen Z. Sun SK, “A Zernike Moment Phase-Based Descriptor for Local Image Representation and Matching”, *IEEE Transactions on Image Processing*, vol. 19, No. 1, pp. 205-219, 22 September 2009.
- [3] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky F., Kadir, T., and Van Gool, L, “A Comparison of Affine Region Detectors”, *International Journal of Computer Vision*, vol. 65, pp. 43-72, 2005.
- [4] D. G. Lowe, “Distinctive Image Features from Scale-invariant Keypoints”, *International Journal of Computer Vision*, vol. 60, no.2, pp. 91-110, 2004.
- [5] Mikolajczyk K., Schmid, C., “A Performance Evaluation of Local Descriptors”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, 2005.
- [6] Feng Jing, Mingjing Li, “An Efficient and Effective Region-Based Image Retrieval Framework”, *IEEE Transactions on Image Processing*, vol. 13, no. 5, pp. 699-709, 2004.
- [7] J. Willamowski, D. Arregui, G. Csurka, C.R. Dance, and L. Fan., “Categorizing Nine Visual Classes Using Local Appearance Descriptors”, *ICPR Workshop on Learning for Adaptable Visual Systems*, Cambridge, August 2004
- [8] W. Wen, J. Yang, “Object Fingerprints for Content Analysis with Applications to Street Landmark Localization”, *Proceedings of ACM International Conference on Multimedia*, pp.169-178, 2008.
- [9] Obdrzalek S, Matas J., “Image Retrieval Using Local Compact DCT-Based Representation”, *Pattern Recognition, 25th DAGM Symposium*, vol. 2781 of Lecture Notes in Computer Science. Magdeburg, Germany: Springer Verlag, pp.490–497, 2003
- [10] Wei Zhang, Jana Kosecka, “Hierarchical Building Recognition”, *Image and Vision Computing*, vol. 25, pp.704-716, 2007.
- [11] Shao, H., Svoboda, T., Van Gool, L., “ZuBuD—Zurich Buildings Database for Image Based Recognition”, Technical Report 260, Computer Vision Laboratory, Swiss Federal Institute of Technology, 2003
- [12] Jerome H. Friedman, Jon Louis Bentley, and Raphael Ari Finkel, “An Algorithm for Finding Best Matches in Logarithmic Expected Time”, *ACM Transactions on Mathematical Software*, vol. 3, no 3, pp. 209-266, 1977
- [13] N.J.C. Groeneweg, B. de Groot, A.H.R. Halma, B.R. Quiroga, M. Tromp, and F.C.A. Groen, “A Fast Offline Building Recognition Application on a Mobile Telephone”, *Advanced Concepts for Intelligent Vision Systems*, vol. 4179, pp. 1122-1132, 2006
- [14] Gerald Fritz, Christin Seifert, and Lucas Paletta, “A Mobile Vision System for Urban

- Detection with Informative Local Descriptors”, *In Proc. International Conference on Computer Vision Systems*, pp. 30-30, 2006.
- [15] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu., “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, *In Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD'96)*. AAAI Press, Menlo Park, CA, pp. 226-231, August 1996
- [16] Jerome H. Friedman, Jon Louis Bentley, Raphael Ari Finkel, “An Algorithm for Finding Best Matches in Logarithmic Expected Time”, *ACM Transactions on Mathematical Software*, vol. 3, no. 3, pp.209-226, September 1977
- [17] Andrea Vedaldi, “An Implementation of Multi-Dimensional Maximally Stable Extremal Regions”, February 7, 2007

