

國立政治大學資訊科學系
Department of Computer Science
National Chengchi University

碩士論文

Master's Thesis

影像內容檢索中以社群網絡演算法為基礎之
多張影像搜尋

Query by Multiple Images for
Content-Based Image Retrieval Based on
Social Network Algorithms

研究生：張瑋鈴

指導教授：沈錕坤

中華民國一〇一年七月

致謝

本論文得以完成，首先要由衷的感謝指導教授 沈錕坤 博士在研究所這段時間，給予我莫大的指導和鼓勵，由於老師給了我許多寶貴的建議，才讓我可以順利完成此篇碩士論文，同時也教導了我許多為人處事的道理，我想這些東西是我一輩子都受用無窮的。在研究的過程中，老師不厭其煩之指引我研究的方向，以及引導我思考許多本來沒有注意到的問題，使我在跟老師討論的過程中得到了許多靈感和啟發，讓我面對瓶頸時有了解決問題的能量，感謝老師對學生的耐心及包容，點燃了我對於做研究的興趣，讓我在這些年中獲益良多，老師對學問的嚴謹更是我學習的典範，謹此致上最誠摯的謝意。

同時也要感謝口試委員 柯佳伶 博士以及 蔡銘峰 博士對本論文的指導與建議，使得本論文得以更加完善。

感謝參與 Group Meeting 的同學們，因為有你們一路相伴和鼓勵，讓我在論文研究上不感覺孤單。

最後由衷的感謝我的家人，因為有家人的全力支持，讓我毫無後顧之憂，專心一致將論文完成，感謝曾經幫助過我的朋友們，我由衷地感謝你們。

影像內容檢索中以社群網絡演算法為基礎之多張影像搜尋

摘要

近年來，隨著數位科技快速的發展，影像資料量迅速的增加，因此影像檢索成為重要的多媒體技術之一。在傳統的影像內容檢索技術中，使用影像低階特徵值，例如顏色(Color)、紋理(Texture)、形狀(Shape)等來描述影像的內容並進行圖片相似度的比對。然而，傳統的影像內容檢索僅提供單張影像查詢，很少研究多張影像的查詢。因此，本研究提出一個可針對多張影像查詢的方法以提供多張影像查詢的影像內容檢索。本研究將影像內容檢索結合社群網絡演算法，使用MPEG-7中相關特徵描述子和SIFT做為主要特徵向量，擷取影像的低階影像特徵，透過特徵相似度計算建立影像之間的網絡，並利用社群網絡演算法找出與多張查詢影像相似的影像。實驗結果顯示所提出的方法可精確的擷取到相似的影像。

關鍵字：影像內容檢索、多張影像查詢、社群網絡。

Query by Multiple Images for Content-Based Image Retrieval Based on Social Network Algorithms

ABSTRACT

In recent years, with the faster and faster development of computer technology, the number of digital images is grown rapidly so that the Content-Based Image Retrieval has become one of important multimedia technologies. Much research has been done on Content-Based Image Retrieval. However, little research has been done on query by multiple images. This thesis investigates the mechanism for query by multiple images.

First, MPEG-7 image features and SIFT are extracted from images. Then, we calculate the similarity of images to construct the proximity graph which represents the similarity structure between images. Last, processing of query by multiple images is achieved based on the social network algorithms. Experimental results indicate the proposed method provides high accuracy and precision.

Keywords : Content-Based Image Retrieval 、 Multiple Images Search 、 Social Network

目錄

摘要	I
ABSTRACT	II
第一章 前言	1
1.1 研究背景與動機	1
1.2 研究方法	3
1.3 研究貢獻	4
1.4 論文架構	4
第二章 相關研究	5
2.1 MPEG-7	5
2.2 影像低階特徵值	6
2.3 影像內容檢索(Content-Based Image Retrieval)系統簡介	8
2.3.1 QBIC	8
2.3.2 VisualSEEK	9
2.3.3 VIR Image Engine	9
2.3.4 Blobworld	10
2.3.5 MARS (Multimedia Analysis and Retrieval System)	10
第三章 研究方法	14
3.1 方法架構	14
3.2 影像特徵擷取(Image Feature Extraction)	15
3.2.1 Color Layout Descriptor (CLD)	16
3.2.2 Scalable Color Descriptor (SCD)	16
3.2.3 Color Structure Descriptor (CSD)	17
3.2.4 Dominant Color Descriptor (DCD)	18
3.2.5 Homogeneous Texture Descriptor (HTD)	19
3.2.6 Edge Histogram Descriptor (EHD)	20
3.2.7 Region Shape Descriptor (RSD)	21
3.2.8 Scale-Invariant Feature Transform (SIFT)	21
3.3 相似度正規化	24
3.4 建立 Image Proximity Network	25
3.5 Community Search Algorithms	26
3.5.1 Maximizing the Minimum Degree	27
3.5.2 Communities with Size Restriction	28
第四章 實驗	35
4.1 實驗資料	35
4.2 實驗設計	36

4.3 實驗結果	38
4.4 實驗總結	44
第五章 結論與未來研究方向	45
5.1 結論	45
5.2 未來研究方向	45
參考文獻	47



圖目錄

圖 1.1	系統概念圖	2
圖 2.1	MPEG-7 標準定位圖	11
圖 2.2	MPEG-7 的應用	11
圖 2.3	DDL、DS 與 D 的關係圖	11
圖 2.4	MPEG-7 色彩描述	12
圖 2.5	QBIC 顏色查詢介面	12
圖 2.6	QBIC 形狀查詢介面	12
圖 2.7	Blobworld 系統查詢結果	13
圖 2.8	MARS 查詢介面	13
圖 3.1	系統流程圖	25
圖 3.2	影像特徵擷取流程架構	30
圖 3.3	CLD 擷取過程	30
圖 3.4	SCD 擷取過程	30
圖 3.5	CSD 擷取過程	31
圖 3.6	五種邊緣類型(Edge Type)	31
圖 3.7	影像分割圖示	31
圖 3.8	Difference of Gaussian (DOG)	32
圖 3.9	極值偵測示意圖	32
圖 3.10	SIFT 演算法特徵點描述圖	32
圖 3.11	Maximizing the Minimum Degree Algorithm	33
圖 3.12	Image Relationship Graph	33
圖 3.13	Image Proximity Network	34
圖 3.14	計算每一張圖片到查詢圖片的距離	34
圖 4.1	影像資料庫	36
圖 4.2	Query Image 為同類別時，建立相似度關係門檻值 m 分析實驗結果	41
圖 4.3	Query Image 為不同類別時，建立相似度關係門檻值 m 分析實驗結果	41
圖 4.4	Query Image 為同類別時，建立 k 個相似度關係分析實驗結果	41
圖 4.5	Query Image 為不同類別時，建立 k 個相似度關係分析實驗結果	42
圖 4.6	Query Image 為相同類別時，門檻值 m vs. k 個相似度分析實驗結果	42
圖 4.7	Query Image 為不同類別時，門檻值 m vs. k 個相似度分析實驗結果	42
圖 4.8	Query Image 為相同類別時，回傳圖片 n 張之準確率分析實驗結果	43
圖 4.9	Query Image 為不同類別時，回傳圖片 n 張之準確率分析實驗結果	43
圖 4.10	MPEG-7 查詢與 SIFT 查詢之準確率分析實驗	43

表目錄

表 3.1	Degree List	33
表 3.2	刪除 Minimum Degree 後的 Degree List.....	34
表 4.1	圖片類別簡介	35
表 4.2	各個參數說明	36



第一章

前言

1.1 研究背景與動機

隨著網路與數位科技快速的發展，加上多媒體資料取得容易，使得各種數位影像資料量迅速的增加，對於影像資料應用上也越來越廣泛，而使用者如何能迅速的從大量影像資料庫中找到想要的影像，已成了影像檢索的主要訴求。

數位影像並非像數位文件一樣有明確的字詞可以建立關鍵字索引，故在傳統的影像檢索技術皆是透過人工的方式，使用影像標題、檔名或關鍵字來描述影像的內容，將影像資料做註解、索引並分類管理，來進行文字搜尋。但面對日漸龐大的影像資料管理來說，此種以文字為基礎的搜尋方式需透過大量的人力給予圖片適當的語義，顯得十分費時，且每個人對於相同影像描述認知有所不同，往往影像檢索的結果有所差異，導致不能夠滿足使用者的需求。為了提高影像檢索的效能及準確性，因此許多以內容為基礎的影像檢索 (Content-Based Image Retrieval, 簡稱CBIR) [21][23][29]技術逐漸成為研究發展的重點。

所謂以內容為基礎的影像檢索，是擷取影像的低階特徵值來描述圖片的內容，以影像內容做為查詢依據。在大部分的研究中，影像的低階特徵值通常使用色彩、紋理、形狀三種類型的特徵來描述圖片的特性。傳統的以內容為基礎的影像檢索系統可以分成三個步驟：圖片特徵擷取、相似度比對和影像檢索。以Query By Image Example為例，使用者可能選擇任一張圖片當作查詢影像範例進行影像檢索，系統將擷取查詢影像低階特徵值來描述影像的特性，低階特徵[26]如顏色(Color) [6] [9]、紋理 (Texture) [16] [20]與形狀 (Shape) [1] [10] [12]，利用這些屬性和影像資料庫中所有圖片的低階特徵進行相似度比對，從影像資料庫中找到最相似的影像回傳給使用者。

而為了滿足人類觀感最接近的搜尋結果，有許多研究提出了關聯性回饋機制[28]，

如MARS系統[8]，此系統可讓使用者在系統回傳的相似影像中，選擇較符合使用者想要的影像並回饋(Feedback)給系統。系統在得到相關回饋圖片資訊後，便將查詢參數重新定義，反覆的進行調整，直到使用者得到滿意的結果為止。

然而當使用者想要針對多張影像做查詢，找出與多張查詢影像特徵相似的圖片時，多數CBIR的研究都是針對單張影像的查詢。少數的CBIR系統雖然提供多張影像的查詢，但其處理的方式是針對先個別做單張影像查詢，再把多個單張影像查詢後的結果做整合。如此透過分別查詢與結果整合的檢索技術，所檢索出的影像並沒有捕捉到使用者的需求。本論文因此研究多張影像查詢的處理機制，可以針對多張查詢影像做檢索，並且使得檢索後的影像皆滿足多張查詢影像的共同特徵。為了針對多張查詢影像做檢索，系統必須擷取每張影像的特徵值並計算影像與影像之間的相似度，來建立Image Proximity Network，透過Community Search的概念找出與多張查詢影像相似的影像。

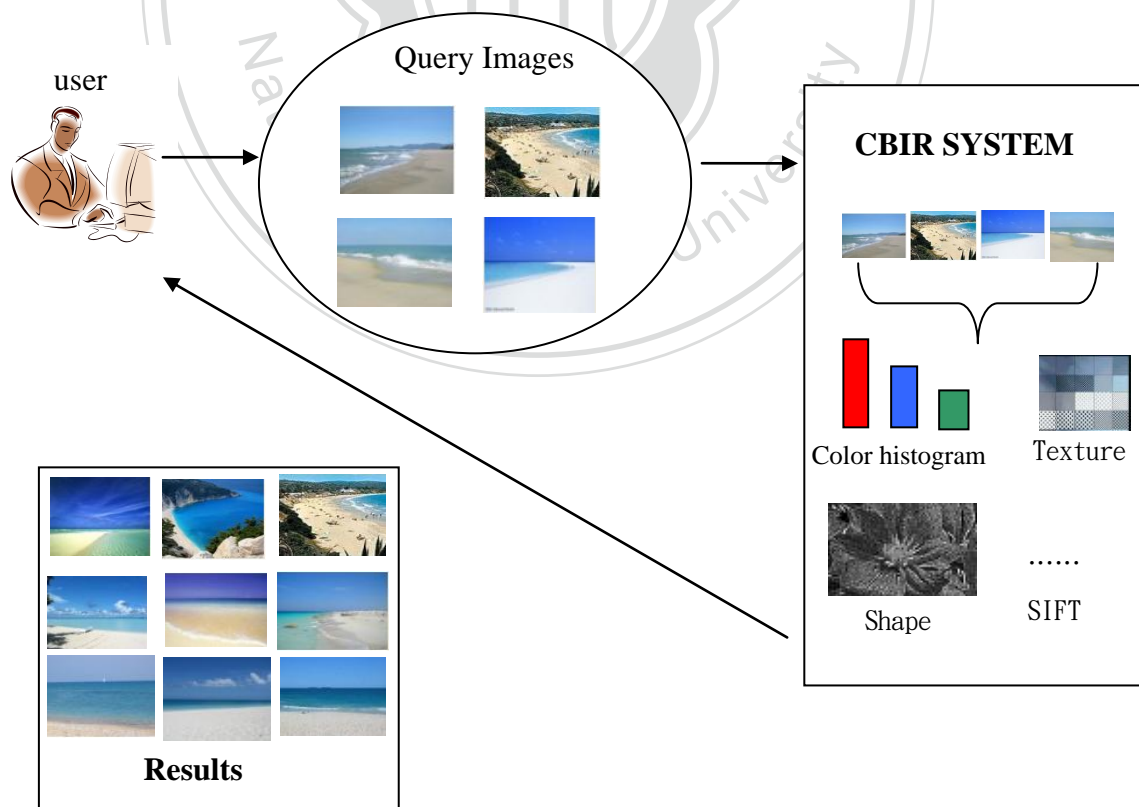


圖 1.1 系統概念圖。

1.2 研究方法

本論文所提出的方法可分成兩個階段來看，方法流程如圖3.1：

(1) 離線處理階段(Off-line Stage)

由於我們所開發的CBIR系統是利用Community Search的方式來做影像檢索，所以建立影像與影像之間的關係是很重要的。而我們希望藉由擷取影像特徵值計算出影像相似度並建立相似關係圖，來達到多張查詢影像檢索的結果。

- 特徵值擷取(Feature Extraction)：在本研究中，特徵值擷取部份採用MPEG-7官方網站[3]上所提供的特徵值擷取程式XM Software及Rob Hess之SIFT[7]演算法，採用了顏色結構 (Color Structure)、可調性色彩 (Scalable Color)、色彩分佈 (Color Layout)、主要色彩(Dominant Color)、齊次性紋理 (Homogeneous Texture)、邊界直方圖 (Edge Histogram)、區域形狀 (Region Shape)、尺度不變特徵轉換(Scale-invariant Feature Transform)等八種特徵值來對影像做描述。
- 相似度計算(Similarity Calculation)：利用這八種特徵值做影像相似度的計算與比對。將影像資料庫中的所有影像擷取特徵值，進行相似度計算，針對不同特徵值採用不同的相似度計算，關於相似度計算公式，將在3.2節中做更進一步的介紹。
- 相似關係建立：透過影像相似度來建立影像關係，形成一個影像的Proximity Network。每一張影像當成Proximity Network中的點，邊則是描述點與點之間的關係，所以影像與影像彼此的關係可以表示為邊，即可形成一個影像的Proximity Network，接下來以此影像資料庫的Proximity Network來做影像檢索，所以影像相似關係的建立將會對查詢結果產生很大的影響。故計算出相似度後，找出每張影像與其他所有影像相似度最接近的K張，並建立影像與影像之間的關係，形成影像的Proximity Network，此Proximity Network表示影像彼此之間

的關係。

(2) 線上處理階段(On-line Stage)

- 查詢處理階段(Query Processing Phase)：針對 Image Proximity Network，利用 Community Search[27]概念中的 Greedy 演算法運算後，產生包含查詢影像且緊密連結的 Subgraph。詳細的 Community Search 的 Greedy 演算法，我們將會在第三章中介紹。

1.3 研究貢獻

我們的研究主要貢獻在於，將既有的影像內容檢索技術融入了 Social Network 的概念，將影像建立 Proximity Network，並藉由 Community Search 技術，達到多張影像查詢的功能，找出多張查詢影像和其它影像特徵值之間隱含的關係。

因此本論文的目的在於透過多張查詢影像的共同低階特徵關係，檢索出更多相似影像，並應用 Community Search 的概念來實現影像檢索。實驗結果證明我們所提出的方法的確能達到多張查詢影像的功能，並且能快速的擷取到相似的影像。

1.4 論文架構

本論文其餘章節架構如下：第二章為相關文獻的探討，介紹影像特徵、影像內容檢索(Content-Based Image Retrieval)的概念；第三章詳細說明特徵擷取的方法與相似度的計算及如何利用 Community Search 的問題，來解決多張影像查詢後的結果更加符合使用者需求；第四章為相關實驗分析結果的討論；第五章將對本論文做結論並對未來的研究做一簡單的敘述。

第二章

相關研究

以內容為基礎的影像檢索影像(CBIR)主要分成三個步驟：影像低階特徵擷取 (Feature Extraction)、內容描述(Content Description)及搜尋與檢索(Search and Retrieval)。在研究方法中使用MPEG-7所定義的視覺特徵的描述子描述影像的特徵值及SIFT，來做為檢索的特徵。

2.1 MPEG-7

隨著多媒體技術的發展，面對大量的資源，準確且有效率的搜尋變得越來越重要了。因此為了滿足多媒體資料的搜尋、瀏覽、篩選等需求，於1996年10月MPEG (Moving Picture Experts Group)組織制定了MPEG-7 [15][18][21]標準。

目前MPEG組織制定的標準有：MPEG-1、MPEG-2、MPEG-4、MPEG-7與MPEG-21。MPEG-7與其他的MPEG主要差別在於，MPEG-1、MPEG-2、MPEG-4重點為研究視訊和音訊的壓縮編碼技術，而MPEG-7目的是定義了一個多媒體內容描述的標準，如圖2.1所示，主要的重點放在更有效地描述多媒體特徵，目標為提供基於內容之影像內容檢索，在應用範圍上只要是與影音資訊搜尋比對相關的應用皆可使用。

如同圖2.2中表示了MPEG-7其中的應用，藉由擷取特徵來建立MPEG-7描述，接著採用MPEG-7編碼器來編碼並儲存與傳送到媒體，透過搜尋引擎和條件過濾出使用者的需求，並將結果送至應用程式。

MPEG-7對多媒體內容描述標準化了描述器 (Descriptor, D)、描述方案 (Description Scheme, DS)、描述定義語言 (Description Definition Language, DDL)，以及對描述的方法和工具，DDL、DS與Descriptor的關係如圖2.3所示，以下簡述MPEG-7提供的各項工具：

- (1) Data：是MPEG-7所要描述的影像資料，不限定儲存方式或是某些編碼後的資料。
- (2) Feature：是從Data中擷取出來的特徵，與其它Data做差異化的屬性，是具有意義的。
- (3) Descriptor：Descriptor (D)是在MPEG-7中對Feature的定義，定義出如何表示Feature 的語法以及語意，例如色彩分佈、紋理、物件輪廓、物件移動軌跡等等。
- (4) Description Scheme(DS)：定義結構，可以定義出各個元件(Component)之間的語法和語意，即用來組合基礎特徵(Descriptor或Description Scheme)的樹狀資料結構。
- (5) Description：是Description Scheme的Instantiation，一組描述內容，儲存裝置及內容使用的結構及內容資料。
- (6) Description Definition Language：用來自行定義新的D或DS的MPEG-7語言，是根據XML Schema標準而來，主要目的是讓使用者能根據不同需求，擴充或修改已經有的DS。
- (7) Visual[17][22]：包含顏色、紋路、形狀、以及移動等相關低階特徵。
- (8) Reference Software：簡稱XM，為eXperimentation Model的縮寫。MPEG-7標準定義範圍內不包含特徵比對演算法，但XM提供了一套C++軟體，展示各D及DS的可行性與搜尋準確度。

2.2 影像低階特徵值

在人類的視覺裡可以很容易的判斷出一張影像的主題，而以電腦科學的角度來思考，要如何來判斷影像的內容及影像之間的是否相似，此時則必須依賴影像的低階特徵來描述圖片的內容。

在影像低階特徵部份採用了 MPEG-7 所定義的 Visual 的描述子[24][30][31]，Visual 描述子目前大致區分成：色彩、紋理、形狀、移動。我們採用以下七種特徵值，足以充分描述圖片的特性：色彩佈局 (Color Layout)、可調色彩 (Scalable Color)、色彩結構 (Color Structure)、主要色彩(Dominant Color)是屬於色彩類的特徵值；齊次性紋理 (Homogeneous Texture)、邊界直方圖(Edge Histogram)屬於紋理類的特徵值；而區域形狀 (Region Shape)則是屬於形狀類的特徵值；另外我們也採用了尺度不變特徵轉換 (Scale-Invariant Feature Transform ,SIFT)為描述影像的局部性特徵。以下針對這些影像低階特徵值做簡單的介紹：

- (1) 色彩佈局描述 (Color Layout Descriptor)：用來描述色彩在空間上的分佈狀態，所採用的顏色空間是 YCbCr，也就是 MPEG 與 JPEG 壓縮編碼中所用的色彩空間。
- (2) 可調色彩描述 (Scalable Color Descriptor)：用來記錄整張影像在 HSV 色彩空間的 Color Histogram，並統計顏色出現的次數，對於影像對影像比對 (Image-to-Image Matching) 的搜尋方式很有幫助。
- (3) 色彩結構描述 (Color Structure Descriptor)：描述影像的顏色內容 (Color Content) 和特徵，色彩空間的結構是以 4x4 像素 (Pixels) 為單位的方塊在影像內滑動，因此可以記錄相鄰畫素之間的顏色資訊及色彩直方圖，CSD 所使用的是 HMMD 的色彩空間。
- (4) 主要色彩描述 (Dominant Color Descriptor)：主要的目的是描述一張影像中的代表性顏色，只要用影像中某區塊中的顏色特徵就可以呈現整張影像的顏色訊息，這對基於顏色相似度的影像檢索是相當有用的。
- (5) 同質性紋理 (Homogeneous Texture)：主要的目的是可以描述出影像中紋理及材質的特性，此特徵值將紋理以量化的方式來表示。

- (6) 邊緣直方圖(Edge Histogram)：主要的目的是描述影像邊緣的空間分佈，可以擷取出空間邊界的局部性資訊，除了影像的色彩之外，人類視覺對於邊緣(Edge)感受也明顯，因此邊緣直方圖在不均勻的邊緣空間分佈上的影像比對上也是一種有用的影像特徵。
- (7) 區域形狀(Region Shape)：此特徵值主要的目的可以用來描述物件的形狀，能夠精簡及有效率的描述一個物件的多個區域，包含單一封閉式區域、複雜的形狀或是不相連的區域，所以在影像切割後仍然可以保留原始影像的特性，比對上也有不錯的效果。
- (8) 尺度不變特徵轉換(Scale-Invariant Feature Transform ,SIFT)：為 2004 年由 Lowe 所提出的特徵點演算法[14]，用來偵測與描述影像中的局部性特徵，在尺度空間偵測極值，提取出其位置、尺度、視角、光度、旋轉變化的不變性，並對於光線遮蔽、影像雜訊、物體移動的視角改變等的容忍度也相當高，達到良好的比對效果。

2.3 影像內容檢索(Content-Based Image Retrieval)系統簡介

由於傳統的影像搜尋方式需要耗費大量的人工來描述且比較無法滿足使用者在影像搜尋上的需求，因此後來提出許多關於影像檢索技術的研究，由原本的文字查詢發展至利用影像的色彩、紋理、形狀.....等低階特徵值做為查詢的依據，藉由特徵值的比對，來達到更好的檢索效果，在本節中將針對幾個著名的系統做個簡單的介紹。

2.3.1 QBIC

QBIC (Query By Image Content) [4][19]是由 IBM Almaden 研究中心所設計，是最早以內容為基礎的影像檢索系統，QBIC 系統提供顏色及形狀兩種查詢介面，使用者可以在色盤中直接挑選顏色或是在繪圖區畫下簡單的外型做為查詢的依據。

- (1) 顏色查詢(Color Search)：

主要以顏色的分佈比例來進行影像查詢，如圖2.5 所示，查詢介面分成左右兩部分，左邊為調色盤，使用者可以利用調色盤來挑選顏色，也可調整明亮度，或設定 RGB 三原色的比例來選擇顏色，顏色選定後，會在右上方顯示選擇的顏色，按下右上方三角型按鈕加入想搜尋顏色中，完成顏色選擇後，按下 Search 鍵即可得到搜尋的結果。

(2) 外型查詢(Layout Search)：

如圖2.6所示，查詢介面可分為左、中、右三個部份，左邊為調色盤，可透過調色盤挑選顏色，被選擇的顏色會顯示在中間上方之方塊；中間可以選擇形狀，在繪圖工具上可以使用矩形和圓形，並且可以調整圖層的順序；右邊為繪圖區，系統會依據繪圖區的形狀，從資料庫中找出與繪圖區相似的影像，也提供刪除及重新查詢的功能。

2.3.2 VisualSEEK

VisualSEEK是一個以內容為基礎的影像查詢系統，由哥倫比亞大學多媒體技術中心所設計[25]開發的，主要特色是結合影像特徵及影像的空間(Spatial)關係來做檢索，可透過空間資訊檢索出物件之間相對性的位置關係，以提升搜尋的能力，並提供了關鍵字全文檢索、不同的種類進行主題瀏覽和相關性回饋機制，可藉由使用者給予的回饋來調整 Vector Modify，以使檢索出的影像能更符合使用者的需求。

VisualSEEK系統中以WebSEEK的查詢介面為主要的特色，WebSEEK能夠針對WEB上的影像進行搜尋，在影像分析的過程中分為二個階段，第一個階段為擷取影像特徵並建立索引，第二個階段為分析影像相關的文字資訊，例如檔案名稱或URL，以做為影像的分類。

2.3.3 VIR Image Engine

VIR (Visual Information Retrieval) Image Engine[5]是由密西根大學、加州大學與麻省

理工學院媒體實驗室所開發之影像視覺資訊系統。VIR提供了影像分析、查詢比對、系統管理等三項主要的功能，主要的特色是除了可以查詢靜態影像（Image）外還可查詢動態視訊（Video）資料，並提供了色彩、紋理、形狀和範例圖做為查詢介面，也允許使用者對每一個內容特徵值給予不同的權值，以利用權重值選擇機制來滿足每個使用者的需求，進而找出最相似的影像。

2.3.4 Blobworld

Blobworld [2]是加州大學柏克萊分校所設計的，它提供了四種檢索方式包含顏色、外形、紋理和範例圖片，此系統主要以區域為基礎的影像做檢索，針對影像做區域的切割，使用者可以針對不同的區域給予不同的權重；首先讓使用者依照影像的主題來選擇範例圖片，接下來系統會把代表該影像的區域圖形的外形提供給使用者做選擇，並讓使用者設定影像特徵值的權重，最後系統根據使用者提供的資訊，進行檢索比對，進而檢索出相似的影像資料。

2.3.5 MARS (Multimedia Analysis and Retrieval System)

MARS (Multimedia Analysis and Retrieval System)[23]是由伊利諾大學 (University of Illinois at Urbana-Champaign) 發展的系統。系統擷取了顏色、紋理、形狀三種特徵值，主要特色結合了索引和檢索的技術與關聯性回饋的機制，透過使用者的回饋來分別調整不同影像特徵間的權重 (Feature Weighting) 及同一個影像特徵中各維度的權重，針對同一個特徵中各維度的權重而言，當在某個特徵維度所給予的值變異度越小，表示這個特徵維度相對的較重要，而對於不同影像特徵間的權重，如果只選擇某一個影像的特徵來計算相似度，所得到的結果與目前檢索出的結果重複越多，表示應提高此特徵的相對權重。

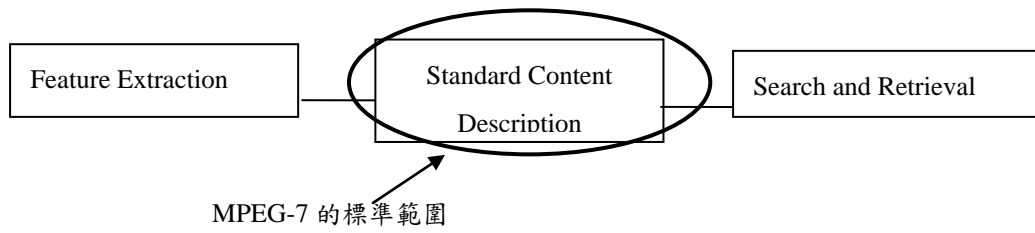


圖2.1 MPEG-7標準定位圖[18]。

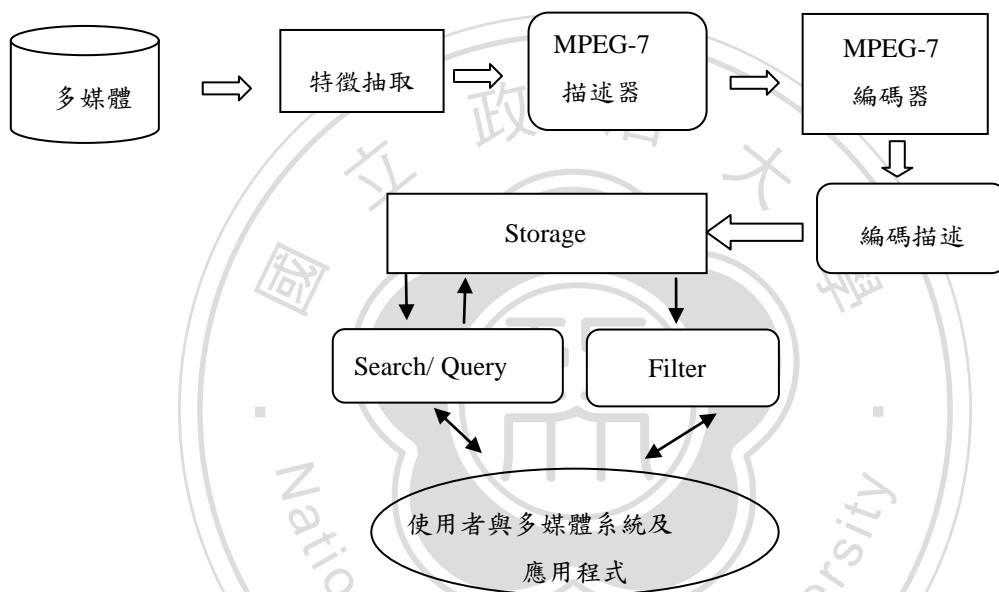


圖 2.2 MPEG-7的應用[18]。

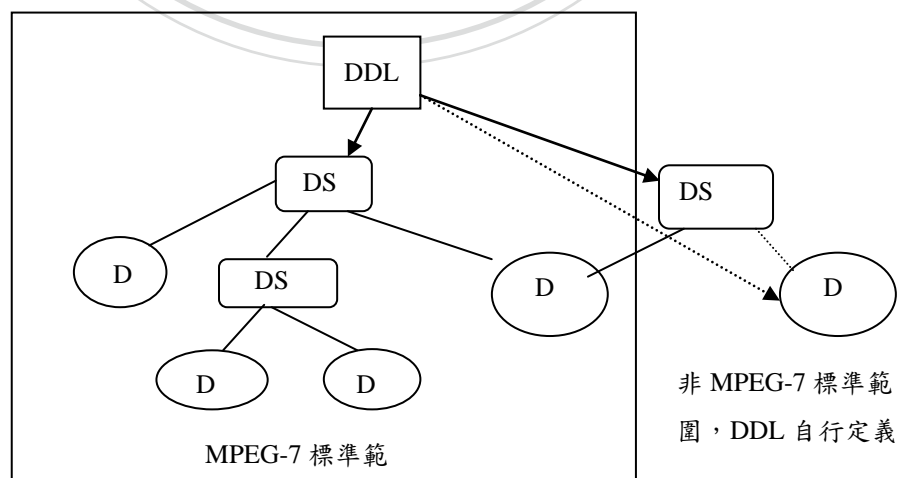


圖2.3 DDL、DS與D的關係圖[18]。

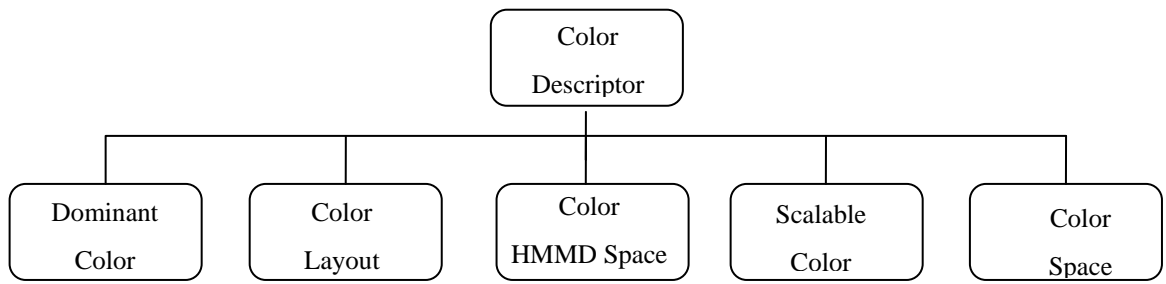


圖 2.4 MPEG-7 色彩描述[31]。

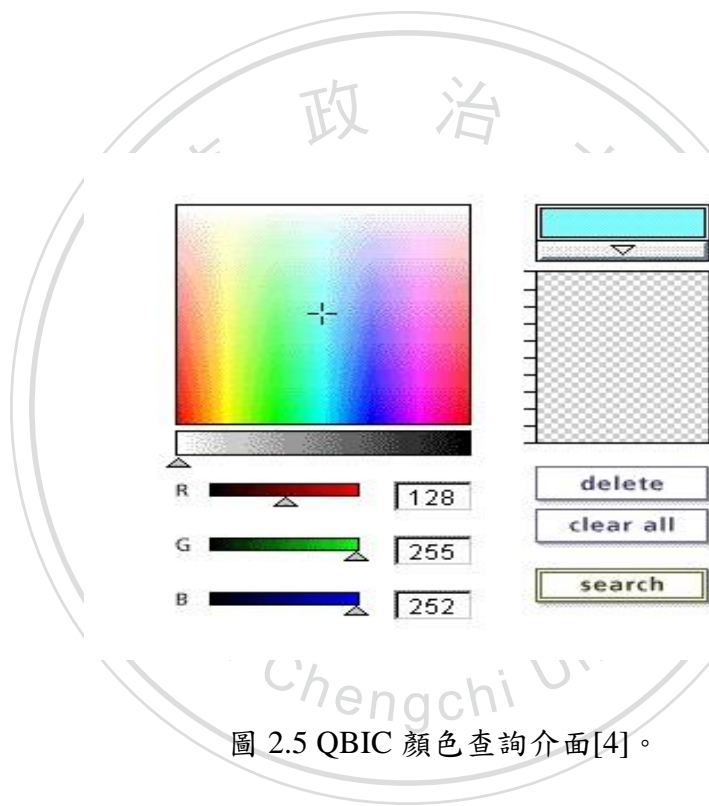


圖 2.5 QBIC 顏色查詢介面[4]。

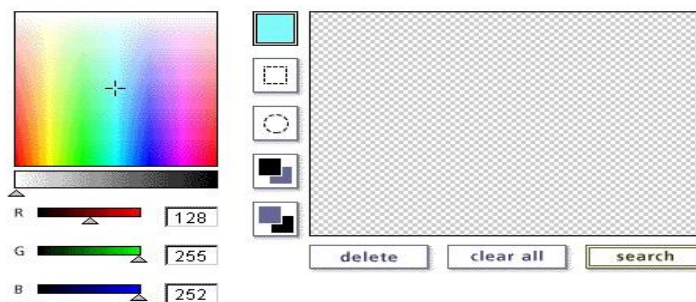


圖 2.6 QBIC 形狀查詢介面[4]。

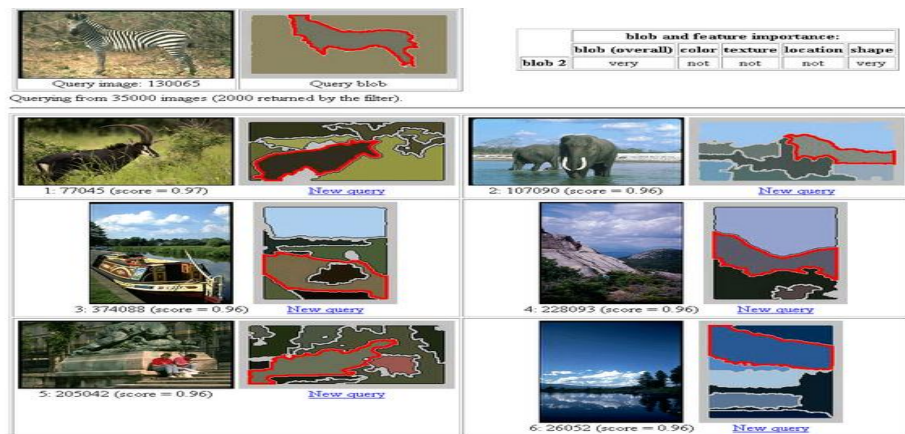


圖 2.7 Blobworld 系統查詢結果[2]。

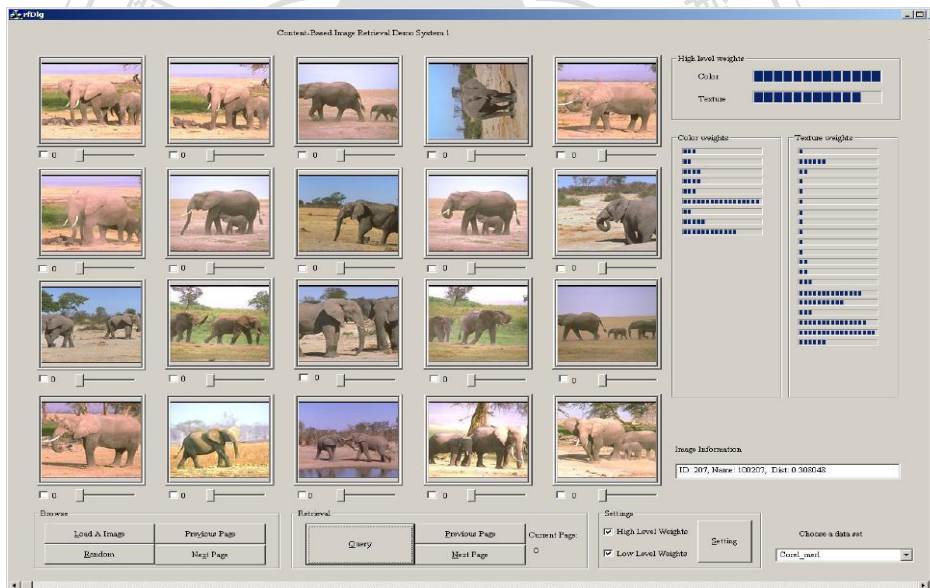


圖 2.8 MARS 查詢介面[23]。

第三章

研究方法

3.1 方法架構

傳統影像內容檢索(Content-Based Image Retrieval)系統中，當使用者選擇一張查詢圖片後，系統會擷取特徵值進行相似度的比對，並從影像資料庫中找出最相似的N張圖片回傳給使用者;假設我們希望能選擇多張影像做查詢，並回傳的結果影像與查詢影像有共同主題的影像，因此我們必須思考，系統要如何才能達到此目的。

在本系統中，我們先擷取出影像的低階影像特徵，除了色彩特徵、紋理特徵、形狀特徵外，並採取具有尺度與旋轉不變性的 SIFT 特徵，利用八種特徵值來對影像做描述並計算相似度的，再比對出每張影像其共同的特徵，來提升相似度的比對。

而研究中的 CBIR 系統與傳統 CBIR 不同之處在於計算特徵值相似度後，依據每張影像的相似度建立影像與影像之間的關係，形成一張 Image Proximity Network，並使用了 Community Search Problem 的概念，利用 Greedy 演算法來進行多張 Query Image 的影像內容檢索(Content-Based Image Retrieval)，找出與多張影像共同關係的影像，以滿足多張影像查詢的功能。

綜合上述方法，於本論文中提出了一個整合的架構，將特徵值擷取與影像的 Proximity Network 應用到原有的 CBIR 系統中來達到多張影像查詢的目的。本系統的流程如下圖3.1所示。

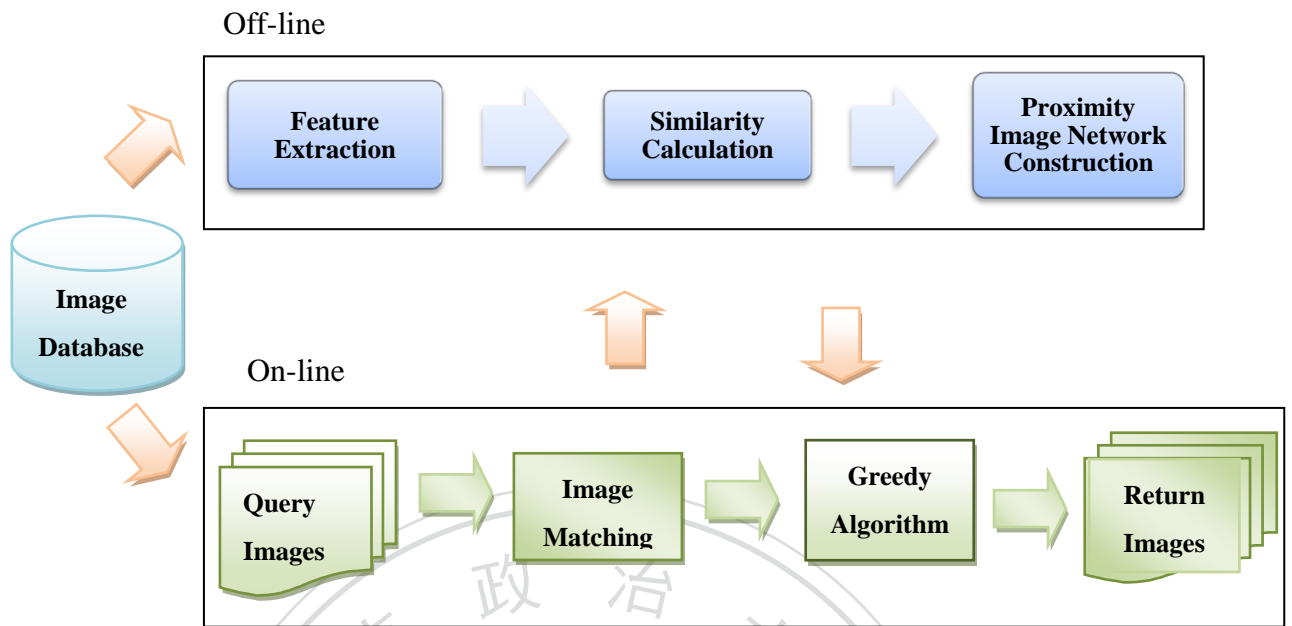


圖 3.1 系統流程圖。

在3.2 節中我們將介紹影像特徵擷取與相似度計算的方法，在3.3 節中詳述 Community Search Problem 的概念，最後在3.4 節裡將介紹如何將這兩項機制整合到影像檢索系統。

3.2 影像特徵擷取(Image Feature Extraction)

許多影像檢索利用影像的色彩、紋理、形狀……等低階特徵值做為查詢的基礎，藉由特徵值的比對，系統能夠找到更符合使用者需求的結果。

在本論文中，特徵值擷取方面，採用MPEG-7 官方網站上所提供的特徵值擷取程式 XM Software。我們擷取出顏色 (Color)、紋理(Texture)、形狀(Shape)等特徵，除此之外，另擷取尺度不變特徵轉換(Scale-Invariant Feature Transform ,SIFT)[14]之特徵，來建立影像圖片的各描述特徵向量，其處理步驟包含特徵的擷取、特徵向量值的相似度正規化，如圖 3.2所示，輸入影像擷取影像特徵，經過特徵的擷取、正規化，計算出每張影像在不同色彩特徵描述下各組的向量值，並寫入不同的特徵檔中，以下說明各特徵擷取演算法與相似度計算。

3.2.1 Color Layout Descriptor (CLD)

Color Layout Descriptor 記錄空間上的分佈資訊。在MPEG-7 XM中擷取Color Layout 分成4個部分，分別為Image Partitioning、Dominant Color Selection、DCT 轉換和 Zigzag-Scanned，其特徵擷取演算法如圖3.3所示：

- (1) 輸入的影像分割成8×8共64個區塊，每個區塊的色彩以YCbCr色彩空間表示，然後利用紋理映射將原始影像做轉換為8×8的方塊大小矩陣。
- (2) 再將8×8的矩陣套用離散餘弦轉換 (Discrete Cosine Transform) 來得到DCT係數矩陣的特徵值。
- (3) 對DCT矩陣進行作Zigzag scan 和Quantization。
- (4) 得到一組Y, Cb, Cr 各64 個係數的特徵向量{ DY, DCr, DCb }。在MPEG-7 中，預設的長度是Y 為6, Cb與Cr 則為3。

在特徵值相似度計算部份，假設二張影像 Img_1 和 Img_2 ，經過計算分別得到二組CLD 特徵向量 $F_1=\{DY_1, DCr_1, DCb_1\}$ 、 $F_2=\{DY_2, DCr_2, DCb_2\}$ ，其相似度計算公式，使用權重式歐基理德距離計算

$$Dis(F_1, F_2) = \sqrt{\sum_i w_{yi} (DY_i - DY'_i)^2} + \sqrt{\sum_i w_{cbi} (DCb_i - DCb'_i)^2} + \sqrt{\sum_i w_{cri} (DCr_i - DCr'_i)^2}$$

i ：為Zigzag-Scanning順序的系數， i 越大表示越低頻，則權重值也會隨著 i 值的增加而變小。

3.2.2 Scalable Color Descriptor (SCD)

Scalable Color Descriptor記錄Color Histogram，在HSV色彩空間上透過Haar轉換而產生色彩直方圖，根據不同的轉換係數，可以產生不同精確度的特徵值，其特徵擷取演算法如下：

- (1) 將影像轉換成HSV色彩空間。
- (2) 對HSV色彩空間為主的影像，線性量化進行直方圖統計。
- (3) 對統計後的直方圖使用Haar 轉換(Haar Transform)來編碼得到Haar係數。
- (4) 針對Haar係數進行均勻量化。
- (5) 最後輸出特徵為 128 維度的 SCD向量。

在特徵相似度計算部份，假設二張影像 Img_1 和 Img_2 ，經過計算分別得到長度為 N 的Haar係數陣列 $Haar_{F_1}(N)$ 與 $Haar_{F_2}(N)$ ，其兩組SCD特徵向量的相似度計算公式使用L1距離計算為 $Dis(F_1, F_2)$ ，距離越近則越相似，反之越不相似。

$$Dis(F_1, F_2) = \sum_{i=1}^N W_i |Haar_{F_1}(i) - Haar_{F_2}(i)|$$

3.2.3 Color Structure Descriptor (CSD)

Color Structure Descriptor記錄顏色內容在空間上的結構，是以 8×8 畫素為單位視窗，在影像內滑動並記錄該區域內的顏色資訊，因此，它可以記錄相鄰畫素之間的特性，

其特徵擷取演算法如下：

- (1) 將影像轉換成HMMD 色彩空間，並將所有數值量化。
- (2) 計算結構單元，結構單元的大小可以根據影像的大小做動態的改變。
- (3) 收集所出現的顏色值，並統計在直方圖 (Histogram) 中。
- (4) 最後將所有直方圖 (Histogram) 值量化，並輸出特徵向量維度為32的CSD向量。

假設二張影像 Img_1 和 Img_2 ，經過計算分別得到兩組維度為 N 的 CSD 特徵向量係數 $F_1(N)$ 與 $F_2(N)$ ，在特徵相似度計算部份使用 L1 距離計算距離公式。

$$Dis(F_1, F_2) = \sum_{i=1}^N W_i |F_1(i) - F_2(i)|$$

3.2.4 Dominant Color Descriptor (DCD)

主要色彩描述子 (Dominant Color Descriptor) 主要的目的是描述整張影像的色彩分佈，藉由影像局部性的顏色特徵即可呈現整張影像的顏色資訊。主要用在顏色的影像內容檢索，其特徵擷取演算法如下：

- (1) 將影像轉換成 HSV 色彩空間。
- (2) 統計顏色點的分佈情形，並找出 5 個代表整張影像的叢集 (MPEG-7 中定義最多 8 個)。
- (3) 於叢集中的顏色點之重心 (Centroid) 當做特徵值，且輸出特徵向量為 5 個主要色彩值的 DCD 向量。

假設二張影像 Img_1 和 Img_2 ，其產生二組 DCD 特徵表示如下：

$$F_1 = \{(c_{1i}, p_{1i}, v_{1i}), s_1\} \quad (i=1, 2, \dots, N_1)$$

$$F_2 = \{(c_{2i}, p_{2i}, v_{2i}), s_2\} \quad (i=1, 2, \dots, N_1)$$

其中 c_i ：第 i 個主要顏色的色彩空間向量

p_i ：顏色 c_i 的像素個數占總像素的百分比, $\sum_i p_i = 1$

v_i ：色彩的變異程度

s ：色彩空間的一致性 (Spatial Coherency)

假設二張影像 Img_1 和 Img_2 其兩組 DCD 特徵向量分別為 $F_1 = \{(c_{1i}, p_{1i}), s_1\}$ 與 $F_2 = \{(c_{2i}, p_{2i}), s_2\}$ ，根據 MPEG-7 定義在特徵相似度計算中忽略 v_i 色彩的變異程度與 s 色彩空間的一致性，其相似度計算公式定義為：

$$D^2(F_1, F_2) = \sum_{i=1}^{N_1} p_{1i}^2 + \sum_{j=1}^{N_2} p_{2j}^2 - \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} 2b_{1i,2j} p_{1i} p_{2j}$$

$b_{u,v}$ 是 c_u 與 c_v 兩種顏色之間的相似性係數，其

$$b_{u,v} = \begin{cases} 1 - disCol_{u,v} / disCol_{max}, & disCol_{u,v} \leq T_{disCol} \\ 0, & disCol_{u,v} > T_{disCol} \end{cases}$$

$disCol_{u,v} = \|c_u - c_v\|$ ：為二個顏色 c_u 與 c_v 間最大的歐基理德距離

T_{disCol} ：二個被視為相似顏色的最大距離，介於 10~20

$disCol_{max}$ ： ℓ 倍的 T_{disCol} ， ℓ ：1.0~1.5

3.2.5 Homogeneous Texture Descriptor (HTD)

同質性紋理(Homogeneous Texture)此特徵值將紋理以量化的方式來表示，MPEG-7 用來擷取紋理描述是由 30 個 Gabor 過濾器組成[15]，包含了 5 種不同的尺度(Scale)，每個尺度再各自包含六個不同的方向，當要搜尋很多看起來很多相似的影像時，紋理特徵適用於具相同紋理樣式影像的檢索。

在特徵相似度計算部份，如 Img_1 和 Img_2 兩張影像，經計算後得到兩組 HTD 特徵向量，其相似度計算公式使用歐基理德距離計算距離公式。

$$Dis(F_1, F_2) = \sum_i \left| \frac{F_1(i) - F_2(i)}{\alpha(i)} \right|$$

3.2.6 Edge Histogram Descriptor (EHD)

邊緣直方圖(Edge Histogram)在MPEG-7紋理描述定義中邊緣直方圖包含 5 種不同類型的邊緣空間分佈狀態，如圖3.6所示，其中4個為有方向邊緣(0°水平、45°、90°垂直、135°) 及1個為無方向邊緣的空間分。

其影像邊緣直方圖的特徵擷取演算法如下：

- (1) 影像分割：先將影像切割成4x4 的子影像 (Sub-Image) ，如圖3.7。
- (2) 在每一個子影像 (Sub-Image)統計五種5 種不同類型的邊緣空間分佈狀態，0°水平、45°、90°垂直、135°及無方向，因此16個子影像(Sub-Image)一共統計出 80 維的直方圖。
- (3) 將每一張Sub-Image分割成固定數個正方形的影像區塊(Image-Block)。
- (4) 每一個Image-Block以2x2的方塊進行偵測影像區塊的邊緣類型。
- (5) 統計邊緣分佈直方圖：利用邊緣濾波器(Filter) 計算影像區塊的邊緣強度。

在特徵相似度計算部份，假設二張影像 Img_1 和 Img_2 ，經過邊緣直方圖統計分別得到長度為80的邊緣直方圖係數 $h_{F_1}(80)$ 與 $h_{F_2}(80)$ ，其兩組EHD特徵向量的相似度計算公式使用L1距離計算為 $Dis(F_1, F_2)$ ，當距離越近則越相似，反之越不相似。

$$Dis(F_1, F_2) = \sum_{i=1}^N |h_{F_1}(i) - h_{F_2}(i)|$$

3.2.7 Region Shape Descriptor (RSD)

區域形狀(Region Shape) 此特徵值主要的目的利用區域內的灰階分布資訊描述物件的形狀，可以精簡及有效的描述一個物件的多個區域，包含單一封閉式區域、複雜的形狀或是不相連的區域，所以在影像切割後仍然可以保留原始影像的特性，比對上也有不錯的效果。

假設二張影像 Img_1 和 Img_2 ，經過計算分別得到兩組維度為 N 的 RSD 特徵向量係數 $F_1(N)$ 與 $F_2(N)$ ，在特徵相似度計算部份使用 L1 距離計算距離公式。

$$Dis(F_1, F_2) = \sum_{i=1}^N |(F_1(i) - F_2(i))|$$

3.2.8 Scale-Invariant Feature Transform (SIFT)

尺度不變特徵轉換(Scale-Invariant Feature Transform) 特徵是用來偵測圖像的局部特徵，對於位移、旋轉、尺度縮放、亮度變化、遮蔽和雜訊等具有很好的不變性不受其干擾，在 SIFT 特徵擷取上主要分為四個步驟：

(1) 尺度空間極值偵測(Detection of Scale-Space Extrema)

Koenderink [11] 以及Lindeberg [13]提出，將影像經過高斯函式模糊處理後，其模糊化的影像如同於做縮放(Scaling)的動作，來代表不同空間的尺度。故利用高斯模糊與影像金字塔(Image-Pyramid)，如圖3.8所示，找出在尺度空間極值，搜尋出各個尺度空間的特徵點，以達到尺度不變的目的。

首先定義一標準差 σ ，利用高斯函式與原始影像作高斯模糊，去除一定的高頻雜訊。

$$L(x, y, \sigma) = G(x, y, \sigma) \times Img_1(x, y)$$

$L(x, y, \sigma)$ 為高斯模糊後的影像， $Img_I(x, y)$ 為原始影像， $G(x, y, \sigma)$ 為高斯函式，公式如下：

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

將原始影像做五次高斯模糊，每次高斯模糊的尺度比值為 $r=\sqrt{2}$ ，做完高斯模糊後，形成一組影像集合，將尺度空間 $\sigma \times 2$ ，並再作一次影像金字塔將影像尺寸減半，並重覆以上方法，最後得到四組影像集合，接下來將每組影像集合內的相鄰高斯模糊影像，分別進行高斯差分，也就是做Difference-Of-Gaussian (DOG)，如圖3.8所示。

$$DOG(x, y, \sigma) = L(x, y, r\sigma) - L(x, y, \sigma)$$

$D(x, y, \sigma)$ 為高斯差分後的影像。找出DOG後，接下來就是極值的偵測。所謂的極值(Extrema)就是每個Octave中，若某個尺度的點的高斯差分值，為26個鄰居像素(上、下尺度各九個鄰居與本身尺度的八個鄰居)中的區域極大值(Local Maximum)或區域極小值(Local Minimum)，則稱此點為極值點，將所有點位記錄下來以供特徵點的定位使用，如圖3.9所示。

(2) 特徵點定位與篩選(Accurate Keypoint Localization)

特徵點定位與篩選目的是將上一個步驟所得到的極值點做更準確的定位，達到次像元的精度，並將對比度不佳的極值點刪除，為了達到定位極值點的目的，本步驟使用一個三維的二次方程式來完成。

首先，將DOG公式利用泰勒方程式展開取近似得到公式如下：

$$DOG(a) = DOG + \frac{\partial DOG^{Trans}}{\partial a} a + \frac{1}{2} a^{Trans} \frac{\partial^2 DOG}{\partial a^2} a$$

a 為極值點 $(x, y, \sigma)^{Trans}$ ， $Trans$ 為轉置， DOG 為高斯差分後的結果。對 a 取偏

微分並假設方程式等於0，以求取極值，可得 \hat{a} 如公式如下：

$$\hat{a} = -\frac{\partial^2 DOG^{-1}}{\partial a^2} \frac{\partial DOG}{\partial a}$$

\hat{a} 即為極值點的偏移量，若 \hat{a} 在 x, y 方向上的值大於0.5，或在 σ 方向上大於 $r/2$ ，表示此區域的極值點較靠近相鄰的點位，因此將極值點移至此相鄰點位後，將上式公式重新計算一次，以獲得較佳的次像元位置。求出偏移量 \hat{a} 後，再將 \hat{a} 代入泰勒方程式展開後的方程式，可得公式如下：

$$DOG(\hat{a}) = DOG + \frac{1}{2} \frac{\partial DOG^{Trans}}{\partial a} \hat{a}$$

若求出之絕對值小於經驗值0.03，則視此極值點稱低對比度的點，並且將此點刪除之。

(3) 計算特徵點方向性(Orientation Assignment)

此步驟主要的目的是確定特徵點的主要方位。首先，以每一個特徵點為中心，取一個 16×16 大小區塊 (Block)，計算區塊內所有像素位置的梯度強度 $m(x, y)$ 及方向 $\theta(x, y)$ ，公式如下：

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

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

其中， $m(x, y)$ 表示在座標 (x, y) 上梯度的強度， $\theta(x, y)$ 表示在座標 (x, y) 上梯度的方向，而 L 表示該方法所取的區塊內所有的像素位置。

求出梯度強度與方向後，利用方位直方圖(Orientation Histogram)的方式統計每個區間之強度值總和，找出較精確特徵點方位，則越接近特徵點位置的像素，其權值越大，最後以八個方向統計，最大的值就將此方向設定給此特徵點；有時會有些特徵點得到多個方向，則視為多個不同之特徵點，但這種狀況並不多見。

(4) 特徵點描述(Local Image Descriptor)

以特徵點的主要方位進行旋轉，並建立特徵點的描述向量，做為匹配特徵點之依據。

首先，以特徵點的方向為主軸，將以特徵點為中心的區塊旋轉到和特徵點的主方位相同，以達到旋轉不變性，並在此區塊內，將16×16的範圍切割為4×4的子區塊(Sub-Block)，接著統計每個區塊內的梯度方向 $\theta(x, y)$ 以及強度 $m(x, y)$ ，而後分別建立區塊內的方位直方圖，每個直方圖有8個區間，代表8個主要方向，如圖3.10所示。

每個特徵點共有16個方向直方圖，每個方向直方圖內有8個梯度強度值，因此共有 $16 \times 8 = 128$ 個特徵值，這些特徵值即為特徵點的描述向量。最後，將此描述向量正規化，以避免亮度差異造成的影響。

3.3 相似度正規化

特徵擷取中我們採用了八種特徵描述來描述影像的內容，得到每張影像的特徵向量後，並計算特徵向量之相似度，其計算相似度方法如3.2節所述。為使得各特徵向量計算後數值造成影響，必須對所有特徵值的相似度做正規化，以便相似值一致性。通常採用的方法有 Gaussian Normalization、Linear Normalization 等，為提高資料的準確度我們採用一般 Linear Normalization 來正規化其原始數據。求出輸入資料中的極大值和極小值，再利用以下公式便可以得出正規化值，其範圍在[0,1]中

$$Vect_i = \frac{Vect_i - \min\{Vect_i\}}{\max\{Vect_i\} - \min\{Vect_i\}} \quad \text{Range}[0,1]$$

$Vect_i$ 為一特徵值計算影像相似度後正規化之值，因本研究採用了八種特徵作為描述影像內容，故將每張影像的八種特徵值的相似度正規化後相加起來，即代表此影像的相似度。

3.4 建立 Image Proximity Network

近年來，隨著社群網路例如 FaceBook、微網誌……等興起，也開始著重於人與人之間的社群關係，因此使得社群網路的應用與研究範疇也越來越多，顯得社群關係的搜尋也顯得重要性；最初的社群網路理論是指人與人之間、或是人與組織之間，透過人際關係、社會關係等關係的連結，所形成的一個網路性的型態結構，相互之間的關係從親密到疏遠、從同樣興趣的聚落到各式各樣不同的領域，因此，以人為節點(Vertexes)，關係做為連結(Edges)來建構社會關係的模型，並藉此分析個體與個體、個體與群體、群體與群體之間互動關係及影響。然而在社會網路上的資料來源並非只受限於人與人之間的關係，也可能包含像是電影與演員合作的關係，學術論文與參考文獻之間的連結……等等，而本研究即把 Social Network 這樣的概念與影像做結合。

Social Network 的形成最重要的是將人與人之間建立關係，而要形成一個影像的 Social Network 也必須建立影像與影像之間的關係，故建立一個好的關係圖可以提升查詢的準確性。首先，根據影像低階特徵值計算出每張影像之間的相似度，接下來依照相似度去建立影像之間的關係，我們採用了二種方法來建立影像關係：

- (1) 採用 k -Nearest Neighbor 方法：每張影像連結到 k 張最相似圖片

對於每一張影像擷取特徵向量，並計算其與資料庫中其他影像的相似度，並依照相似度排序，藉此找出前 k 張相似的影像，則為每張影像至少擁有 k 張相似圖片。

- (2) Threshold Method：二張影像相似度超過門檻值 m 才建立連結關係

擷取影像特徵向量並計算其與資料庫中其他影像的相似度，若相似度超過門檻值，則視為相似影像，並建立相似影像關係。

3.5 Community Search Algorithms

事實上，社會網絡可以使用一張 Graph 來表示，組成一張 Graph 的要件為點(Node)與邊(Edge)，透過 Graph 建立的概念，將延伸出許多社會網絡的議題，以舉辦一場雞尾酒宴會的例子來說，假設今天想替三位主角 A、B、C 辦一場雞尾酒宴會，如何使得這場雞尾酒宴會辦得成功呢？與邀請來參加的朋友之間的關係有很大的因素，假設我們所邀請來參加宴會的朋友都能認識三位主角，當然是最好不過的，如果邀請來參加的朋友，他們彼此之間的熟識程度又很高，相對的宴會氣氛則較熱絡，也會使得這場宴會變得很熱鬧，像這樣一個問題以 Graph 表示時，即為 Community Search Problem[27]。

許多實際生活的問題，是可以轉換成圖形的問題來解決，故此研究將多張圖片找圖的問題看成一個Community Search Problem，影像與影像之間的相似關係構成一張Graph，User輸入多張查詢圖片為Query Node，如何從Graph中找出與Query Node相似圖片的問題則可看成一個Community Search Problem，當一個Community Search Problem以 Graph [27] 表示時，定義如下：

一個圖形 $G = (V, E)$ ， G 是由一組頂點 (Vertex) 集合 V 和一組邊 (Edge) 的集合 E 所組成，且有一組查詢影像節點(Query Nodes)集合 $Q \subseteq V$ ，利用Function f_m 從原圖 $G = (V, E)$ 中取出一個Subgraph即為 $S = (V', E')$ ，且需滿足以下條件：

(i) V' 包含 Q ($Q \subseteq V'$);

(ii) S 需為 Connected;

(iii)使用**Minimum Degree Function** $f_m(S)$ 使得Maximized the Minimum Degree

亦即，從 $G = (V, E)$ 的 V 中取出部份的Vertices V' ，並且從 E 中取出"兩點都落在 V' 之內"的 Edges E' ，而所構成的子圖 $S = (V', E')$ 稱為 G 的一個 Subgraph S ，且 $V(G') \subseteq V(G)$ 與 $E(G') \subseteq E(G)$ 。

以多張圖片找圖的例子來說，首先將影像資料庫中的影像，透過相似度計算與正規化，經過排序找出相似影像，得知每張影像與其他所有影像相似度的前 $k\%$ 來做為相似

影像，並建立影像的相似關係；每一張影像表示為一個點(Nodes)，每張影像與影像之間的相似關係為一個邊(Edge)，進而形成一張影像的Graph，而這樣一個影像之間的關係即為所有影像的Image Proximity Network，有了這個Graph，即可藉由Function f_m 找尋出多張查詢影像與其它相似影像之間的關係，即為此原Image Proximity Network的Subgraph，然而這個Subgraph S必需滿足Connected，所謂Connected是指在Graph中，每個頂點間至少存在一條邊，則互為相連通，亦即表示所找尋出來的圖片彼此之間有相似關係的。我們將利用Community Search Algorithms中的兩個演算法分述如下。

3.5.1 Maximizing the Minimum Degree

首先，從 Image Proximity Network 找出 Densely Connected Subgraph。而 Densely 有多種不同的定義方式，本研究使用了 Maximizing the Minimum Degree[27]。與點連接的邊的數量稱為 Degree，在 Image Proximity Network 中，Degree 指的是影像與影像之間的關係，當一張圖片的 Degree 數等於 3，表示與此張圖片相似的圖片有 3 張，而每張圖片的 Degree 越大表示相似關係的圖片越多，反之越少；Graph 的 Degree 越大則此 Graph 就越 Dense，為了從 Image Proximity Network 找出 Densely Connected Subgraph，因此需使得 Minimum Degree 最大。

而如何從影像資料庫中的 Image Proximity Network 找出相似的影像 Subgraph 呢？以社群網絡為例：在社群網絡中認識的人越少，表示與人群較疏遠，因此以 Image 來說，如果找出 Degree 最少的 Image，則表示此張 Image 与其它影像較不相似，將較不相似的刪除，重覆進行，即能找出 Subgraph 是 Densely 也就是 Maximizing the Minimum Degree，故從 Graph 中刪除 Minimum Degree 的點，使得 Graph 是 Densely，如以下步驟：

1. 定義 $G_0=G$ ，給予一張原始 Image Proximity Network 為 G
2. 在 t 個步驟中，每個步驟刪除一個具有 Minimum Degree 的點，則產生 G_{t-1}

3. $G_0 = \max\{f_m(G'_t) \mid t=t-1\}$ ，在 G_1, \dots, G_t 中直到符合條件(i)至少有一個

Query Nodes 為 Minimum Degree, (ii) Query Node 沒有 Connected, 則停止，亦即找出 Densest Subgraph。

故以 Greedy Algorithm，找出 Densest Subgraph 的演算法，如圖 3.11 所示。首先，針對 Image Proximity Network 裡的每個影像 Node 建立相鄰串列，並新增 Degree List 記錄每個影像 Node 的 Degree 數，從 Degree List 裡挑選出 Degree 最少的影像做為刪除，表示此張影像擁有較少的相似影像，並檢查刪除後產生的 Subgraph 是否為 Connected 且其中一個 Query Node 是否為 Minimum Degree，如符合以上條件，則終止刪除的動作，根據以上步驟重覆進行，最後找出的 Subgraph 所包含的影像 Node 即為與多張查詢影像共同相似關係的影像。此 Maximizing the Minimum Degree 方法，從 Graph 中找出 Minimum Degree 做刪除產生 Subgraph，並使得 Subgraph 之間的關係程度為最相似。

圖 3.12 所示為一個建立好的 Image Proximity Network，此 Graph 包含了 12 張 Image 與 25 條邊，當 User 選擇 A,C,D 三張影像當作查詢影像，Query Node $Q=\{A,C,D\}$ ，如何從 Graph 找出 Densely Connected Subgraph。首先計算出每張 Image 的 Degree 數，如表 3.1，統計後的 Minimum Degree=2，為 Image F，故刪除 Degree 最少的 Image F，將 Image F 刪除後即產生了第一個 Subgraph，如圖 3.13 所示。刪除後重新計算 Degree，發現 Query Node 為 Minimum Degree 故停止進行刪除動作，並產生了 Subgraph 即為 Query 結果。

3.5.2 Communities with Size Restriction

給一個 Image Proximity Network G 與 Query Node，利用 Maximizing the Minimum Degree 找出的 Densest-Subgraph 並未考慮到 Size 的問題，所以找到的 Densest-Subgraph 有可能是一個非常大 Size 的 Graph，可能包含了好幾百個 Node，就影像查詢而言，使用者並不會希望搜尋結果傳回好幾百張影像的結果，因此，我們希望能將 Densest-Subgraph 加

入限制大小條件，以致符合User的需求。

一個Community Search Problem以 Graph [27] 表示並從Graph中找出一個大小限制的Subgraph，定義如下：

一個圖形 $G = (V, E)$ ，G是由一組頂點（Vertex）的集合V和一組邊（Edge）的集合E組成，且有一組查詢影像節點(Query Nodes)的集合 $Q \subseteq V$ ，利用Function f_m 從原圖 $G = (V, E)$ 中取出一個大小限制的Subgraph即為 $S = (V', E')$ ，故需滿足以下條件：

- (i) V'包含Q ($Q \subseteq V'$);
- (ii) S需為Connected;
- (iii) $|V'| \leq \text{Num}$ ，從Subgraph中回傳Num個Node
- (iv) 使用Minimum Degree Function $f_m(S)$ 使得S為Maximized

因此，我們將Subgraph考慮了Size的限制條件，利用GreedyFAST algorithms[27]來找出符合Size=Num的Subgraph，這個Algorithm的概念在於找出與Query Node距離最近的Num個Node，當Subgraph的Node數大小限制Num時，則輸出的Num個Node為最接近Query Node的Subgraph，也就是說越接近Query Node表示與查詢影像越相似。因此，首先我們針對Subgraph計算任一個Node與Query Node之間的距離，公式如下：

$$DisNode_Q(G, v) = \sum_{q \in Q} edge_G(v, q)^2$$

$\sum_{q \in Q} edge_G(v, q)^2$ 是將每一個Node到Query Node的距離平方後加總，則 $DisNode_Q(G, v)$ 表示在Graph中任一個Node到Query Node的距離，求出距離後，即可找出與Query Node最接近的Num個Node，亦即找出的Subgraph為符合大小限制的Graph。

就影像檢索而言，圖 3.14 Image Proximity Network 中計算出每一張圖片到查詢圖片的距離，距離最小的即為找到與查詢影像最相似的圖片，如下圖 3.14 所示 Image B 為與Query Node A、C、D 為最相似的圖片。

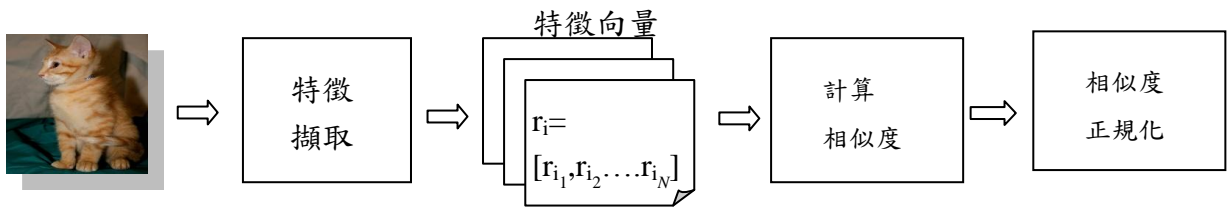


圖 3.2 影像特徵擷取流程架構。

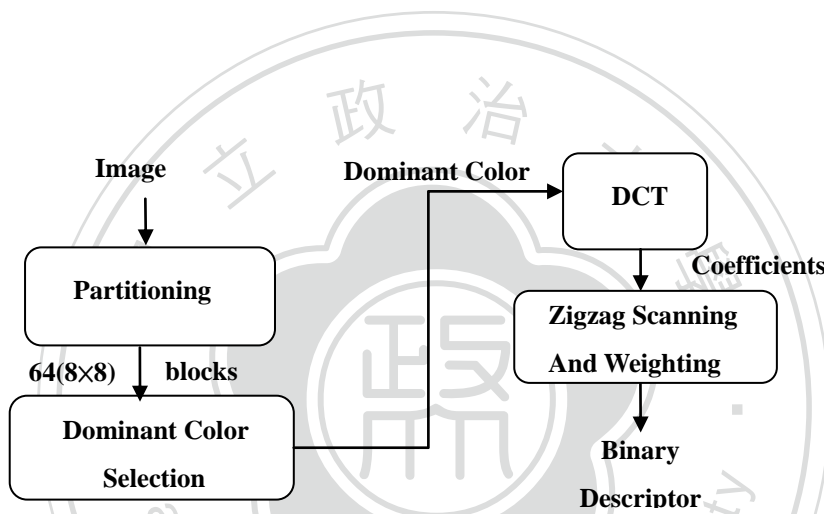


圖 3.3 CLD擷取過程[17]。

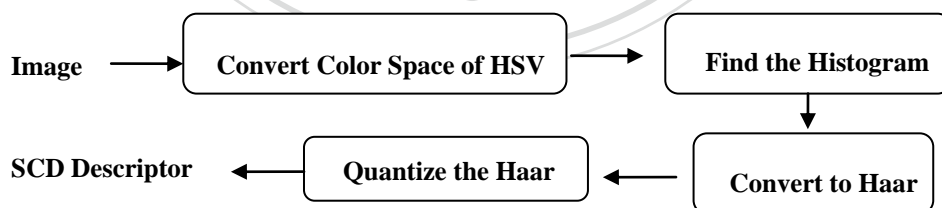


圖 3.4 SCD擷取過程[17]。

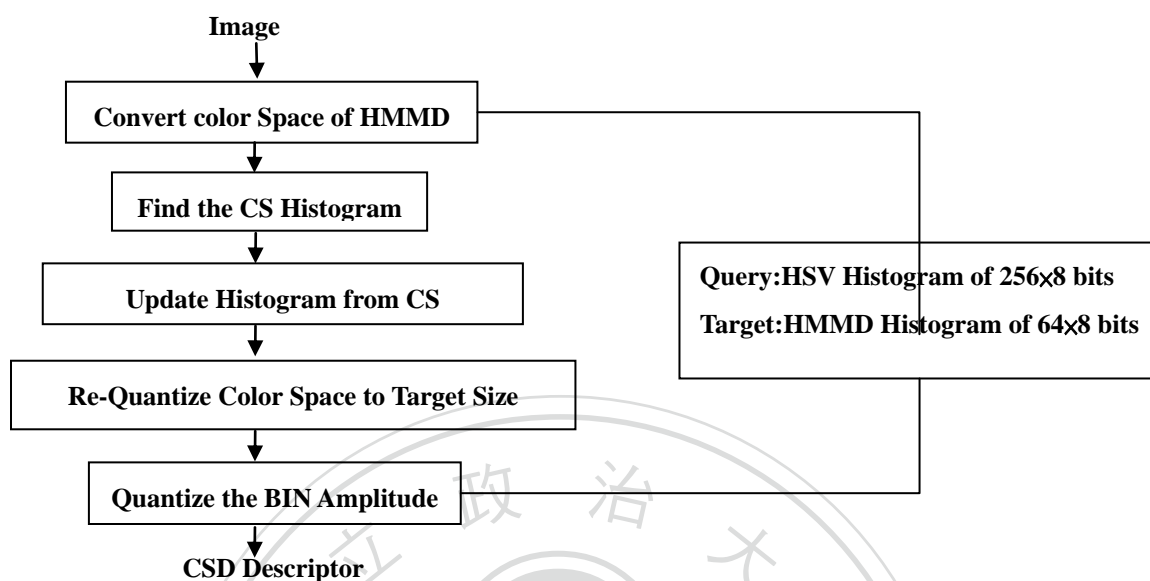


圖3.5 CSD擷取過程[17]。

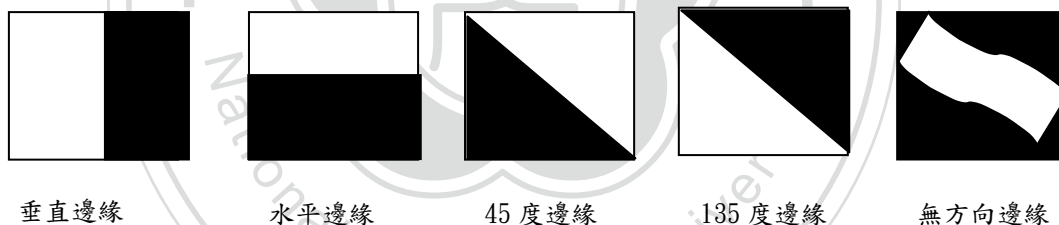


圖3.6 五種邊緣類型(Edge Type)[17]。

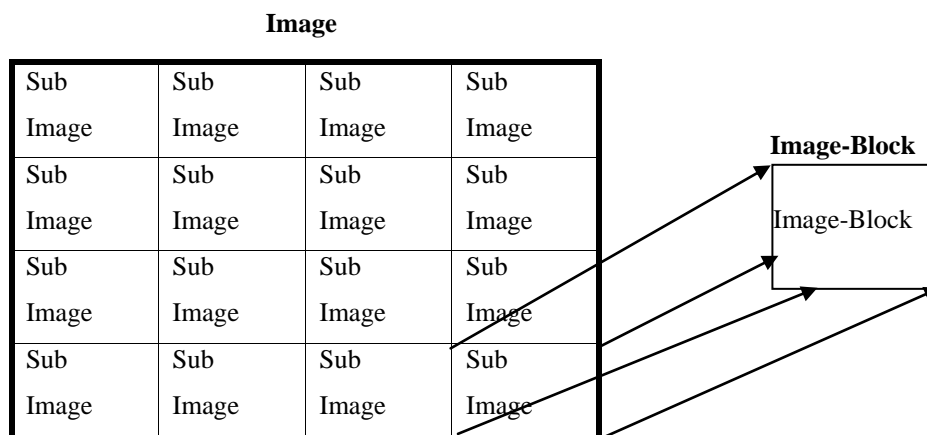


圖3.7 影像分割圖示[17]。

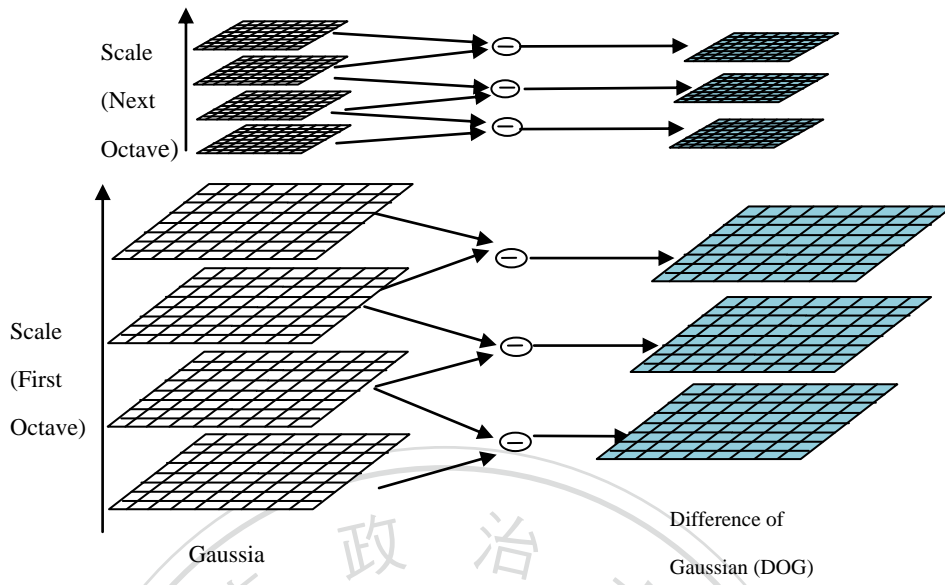


圖3.8 Difference of Gaussian (DOG)[14]。

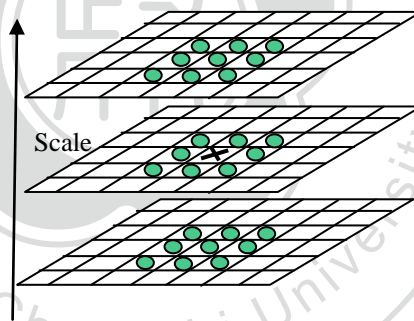


圖3.9 極值偵測示意圖 [14]。

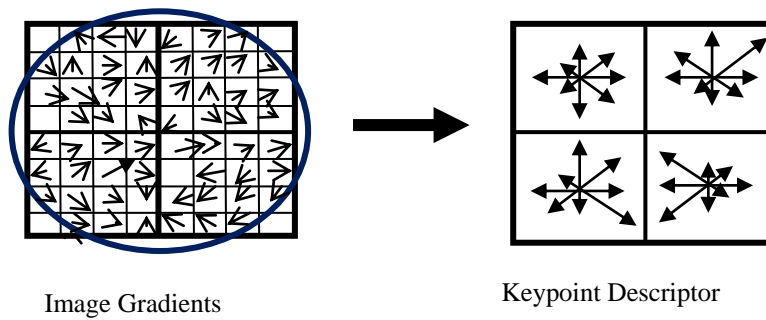


圖3.10 SIFT 演算法特徵點描述圖[14]。


```

1 Greedy algorithm for the densest subgraph
2 Input : Graph adjacent list, degree list, query node
3 output : subgraph
4 Let the first input image G
5 G'=delete minimum degree node from graph
6 While the graph G has connected
7   G=G'
8   find to minimum degree node from degree list
9   if least one of the query nodes has not minimum degree
10    delete minimum degree node from graph
11   else
12    break
13   endif
14 EndWhile
15 return subgraph

```

圖 3.11 Maximizing the Minimum Degree Algorithm。

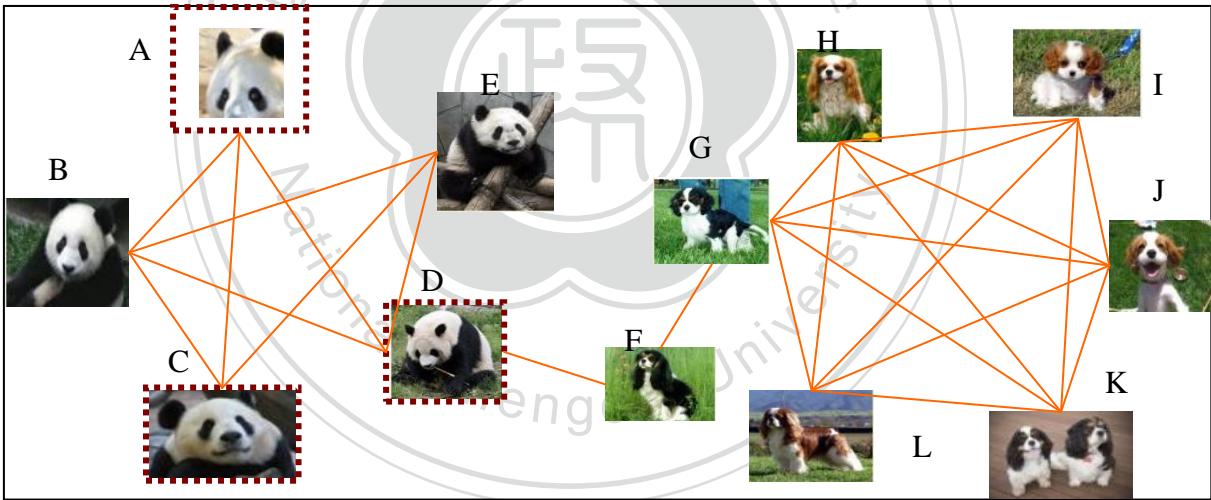


圖 3.12 Image Relationship Graph。

表 3.1 Degree List。

Degree		Degree		Degree	
<i>A</i>	3	<i>E</i>	3	<i>I</i>	5
<i>B</i>	4	<i>F</i>	2	<i>J</i>	5
<i>C</i>	3	<i>G</i>	6	<i>K</i>	5
<i>D</i>	4	<i>H</i>	5	<i>L</i>	5

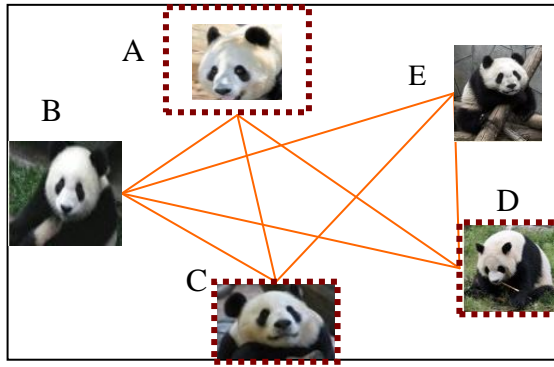


圖 3.13 Image Proximity Network。

表 3.2 刪除 Minimum Degree 後的 Degree List。

Degree	
A	3
B	4
C	3
D	3
E	3

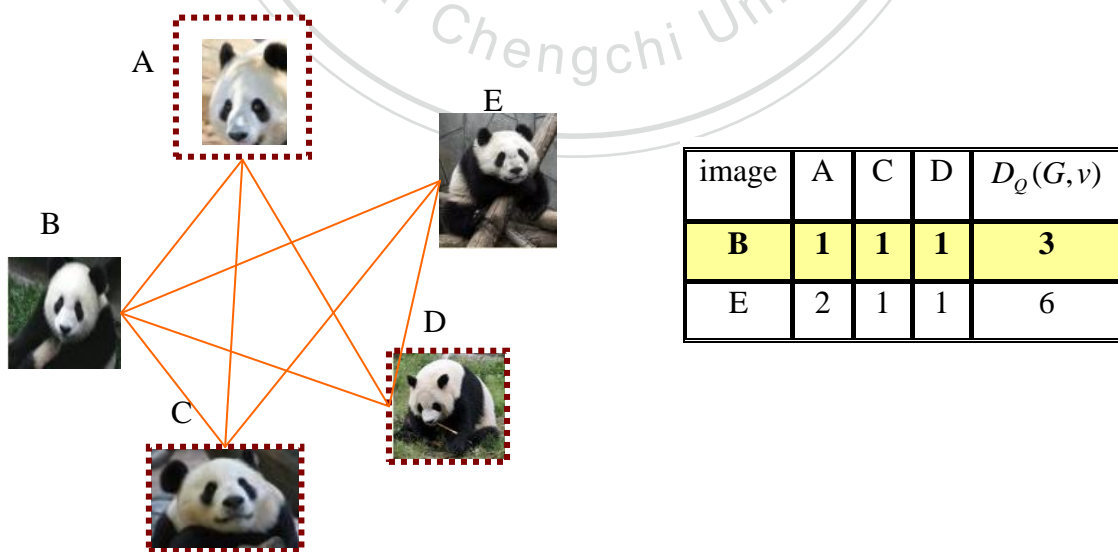


圖3.14 計算每一張圖片到查詢圖片的距離。

第四章

實驗

本章節主要針對本論文在第三章所提出將CBIR結合Community Search Algorithm的方法，使用C++語言來開發實驗程式，因此，在這章節我們將要驗證 CBIR 結合Community Search Algorithm的準確，實驗的環境如下。

實驗的環境為：

- Intel Core 2 Duo CPU
- 2G RAM
- Microsoft Windows XP
- 程式開發軟體：Code::Blocks C++

4.1 實驗資料

我們所使用的影像資料來自於ImageNet的線上圖庫，該圖庫圖片分類是根據WordNet的組織架構來為做為圖像的分類。ImageNet內容包含了一百多萬張各式圖片，我們從中取出20個類別，每個類別100張圖片來進行實驗，各類之間沒有重複的圖片，以下簡單描述各個類別的圖片，如圖4.1所示：

表4.1 圖片類別簡介。

圖片類別	張數	圖片類別	張數	圖片類別	張數	圖片類別	張數
Group	100	Panda	100	Human	100	Dog	100
Plane	100	Cat	100	Beach	100	Horse	100
Stalactite	100	Elephant	100	Pumpkin	100	Bird	100
Shell	100	Flower	100	Dish	100	Car	100
Motorcycle	100	Fish	100	Bridge	100	Building	100

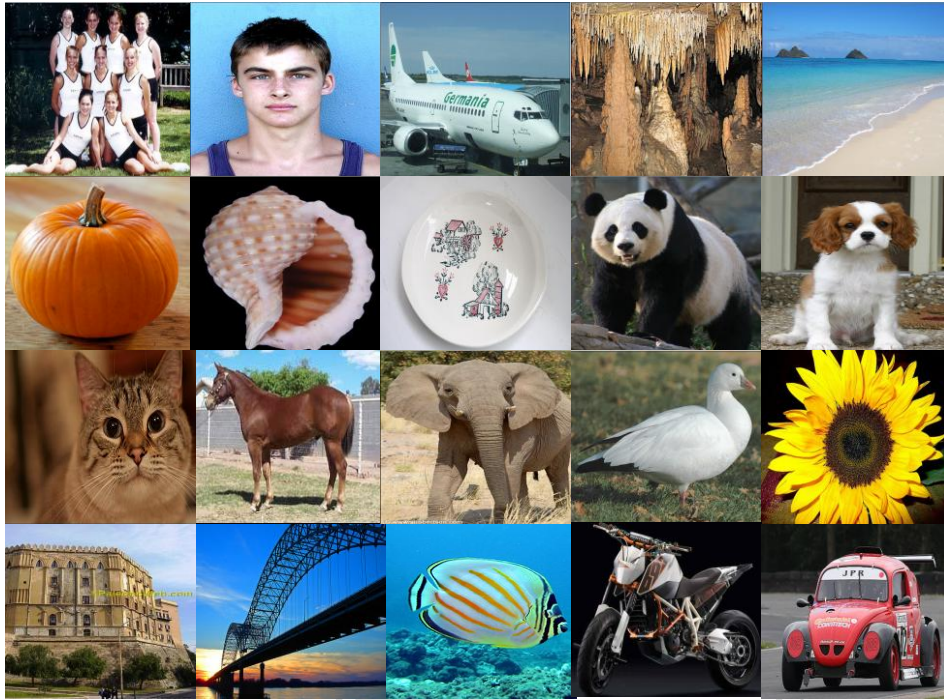


圖 4.1 影像資料庫(出處：<http://www.image-net.org/>)

4.2 實驗設計

表 4.2 各個參數說明。

符號	說明
n	系統回傳給使用者最相近TOP- n 張圖片的張數。
q	使用者給予的Query Image為一集合T，T包含了 q 張影像， q 為圖片查詢數量。
m	Thresholding Method：建立Image Proximity Network之門檻值 m 設定，依照每張影像相似度，找出前10%~30%來做為門檻值。
k	k -Nearest Neighbor：建立Image Proximity Network之 k 張最相似的連結關係。

我們實驗探討的方向主要有 5 個部分：

- (1) 相似度門檻值對於效能之影響：使用 20 類圖片來進行實驗，影像相似度之門檻值 m 設定分別為：10%、15%、20%、25%、30%。其量測的標準為當建立相似影像關係時，在哪一個門檻值 m 的設定值下有最高的準確率。
- (2) k - Nearest Neighbor 參數 k 對效能之影響：使用 20 類圖片來進行實驗，藉由 k -Nearest Neighbor 找出最相近的 k 個相似度關係的圖片張數，參數 k 設定分別為：3~15。其量測的標準為當 k 在哪一個設定值下有最高的準確率。
- (3) 相似度門檻與 k - Nearest Neighbor 方法之準確率比較
- (4) 回傳圖片張數對效能之影響：系統回傳給使用者最相近 n 張圖片的張數，TOP- n 為 10、20、30、40、50 張圖片，觀察在不同的回傳圖片張數 TOP- n 數量準確率的變化。
- (5) MPEG-7 特徵與 SIFT 查詢之效能比較：觀察使用傳統的 CBIR 方法擷取顏色、紋理、形狀等特徵查詢與僅採取 SIFT 特徵查詢的準確率變化。

準確率的評估：

在實驗中，依據查詢圖片的類別為準則，判斷系統回傳後的圖片是否與查詢圖片的類別相同，類別相同則表示回傳正確的影像，以下是我們實驗評估準確率(Precision) 的公式：

$$Precision = \frac{\text{Number of Correct}}{\text{Number of Retrieval}}$$

- Number of Correct 為回傳影像中正確的個數

-Number of Retrieval 為所回傳影像的個數

4.3 實驗結果

我們研究的目的是要在多張圖片查詢之下，系統能達到較高的準確度，因此採用 Precision 的評估方式，來比較不同的設定值，在經過了相同次數的實驗之後，所能達到的準確度。根據 4.2 章節所提到的4個實驗規劃，分別由系統自動測試100次後的測試結果詳細分析圖表與觀察心得如下：

(1) 相似度門檻值對於效能之影響

在我們建立 Image Proximity Network 的方法中，門檻值 m 做為評估相似影像的集合張數。在實驗中，門檻值 m 設定為 10、15、20、25、30，我們要觀察不同門檻值其準確率表現情況。使用的資料庫為 20 類圖片(參考表 4.1 實驗資料類別簡介)，而每類有 100 張圖片，且每張圖片都以 ImageNet 圖庫中所定義的類別為正確答案，來判斷系統回傳的圖片是否與 Query Image 為相同類別。

實驗中，使用傳統的方法做單張圖片查詢與使用本研究提出的方法做多張圖片查詢相互比較，在單張查詢部份，輸入 3 張圖片並擷取 3 張圖片的特徵產生三組特徵向量，並將三組向量平均最後得到一組特徵向量，將視為 1 張圖片的特徵向量做單張圖片查詢，在多張圖片查詢部份，分別輸入 2 張、3 張、4 張、5 張圖片做多張圖片查詢，並分別觀察查詢圖片相同類別與不同類別時的準確度變化。

觀察圖 4.2 與圖 4.3，我們可以發現，當門檻值 m 設定為 15% 時，其平均的準確率是最高的，因此我們將之後的實驗門檻值 m 設定為 15%，期望能達到最好的準確率結果。

當 m 設定為 10% 時，其準確率降低，這表示當建立相似影像關係時，如果相似影像的集合個數(m 值)過小，結果與相對應的相關圖片過少，即相似圖片過少，就會降低搜尋的準確率，而當 m 值設為定為 25% 至 30% 或更大時，我們會發現因為相似圖片過多，在使用本研究提出的方法搜尋圖片時會有所干擾，導致準確的下降。我們經過實驗測試發現無論查詢圖片是否為相同類別，其曲線跟圖的曲線走勢是相似的，因此，後面實驗

門檻值 m 將設定為15%，以達實驗之最佳準確率。

(2) k -Nearest Neighbor 參數 k 對效能之影響

在實驗中，參數 k 設定為3~15， n 設定為系統回傳20張圖片， q 設定為1~5查詢圖片張數，我們要觀察當Query Image為同類別與不同類別時，在 k -Nearest Neighbor方法中，不同 k 的設定值其準確率表現情況。

使用的資料庫為20類圖片(參考表4.1實驗資料類別簡介)，而每類有100張圖片，且每張圖片都以ImageNet圖庫中所定義的類別為正確答案，來判斷系統回傳的圖片是否與Query Image為相同類別。

實驗中，使用傳統的方法做單張圖片查詢與使用本研究提出的方法做多張圖片查詢相互比較，在單張查詢部份，輸入3張圖片並擷取3張圖片的特徵產生3組特徵向量，並將三組向量平均最後得到一組特徵向量，將視為1張圖片的特徵向量做單張圖片查詢，在多張圖片查詢部份，分別輸入2張、3張、4張、5張圖片做多張圖片查詢，並分別觀察查詢圖片相同類別與不同類別時的準確度變化。

圖4.4及4.5 我們可以觀察到：

- 當 k 參數設定為9時，其平均的準確率是較高的，當 k 值設為9跟12的準確率差異小於千分之二，因此我們將之後的實驗的門檻值 k 設定為9，期望能達到最好的準確率結果。
- 當 k 參數設定為3時，其準確率較低，這表示在建立TOP- k 張相似關係也會影響準確率的，建立的關係越少，使得Graph之間的關係程度較不Densely，則會降低搜尋的準確率。
- 不管Query Image為相同類別或不同類別時，其查詢圖片數量 q 張數越多，表示所給予的查詢條件越嚴苛，系統所回傳的結果其準確率較高。
- 從實驗中也觀察發現，多張圖片搜尋方法的準確率的确比單張的圖片搜尋方法還高。

(3) 相似度門檻與 k -Nearest Neighbor 方法之準確率比較

在相似度門檻值對於效能之影響與 k -Nearest Neighbor 參數 k 對效能之影響的實驗中，觀察到門檻值 m 設定為 15% 與 k 參數設定為 9 時，其平均的準確率是較高的，故在本實驗中將參數 m 設定為 15%， k 設定為 9， q 設定為 1~5 查詢圖片張數，參數 n 設定為 20 張圖片，我們要觀察當 Query Image 為同類別與不同類別時，在設定相似度門檻值與 k -Nearest Neighbor 方法中，其準確率表現情況。分析圖表如圖 4.6 與圖 4.7 所示，在實驗結果中發現不管使用相似度之門檻值方法或 k -Nearest Neighbor 方法，當 $q=4$ 時，都有比較高的準確度。

(4) 回傳圖片張數對效能之影響

而在方法中，我們在建立圖片關係時，使用 9 張相似圖片來建立圖片關係，而參數 n 代表為使用者每次圖片檢索時的系統回傳給使用者最相近 n 張圖片張數，在實驗中，將參數 n 設定為 10、20、30、40、50 張，我們要觀察不同 n 設定值其準確率表現情況，分析圖表如圖 4.8 與圖 4.9 所示，在實驗結果中發現不管查詢圖片張數多少及查詢圖片的類別是否相同，在 $n=20$ 時，都有比較高的準確度。

(5) MPEG-7 特徵與 SIFT 查詢之效能比較

在我們的方法中，使用了 MPEG-7 與 SIFT 的特徵來描述圖片，而在實驗中，將觀察當輸入多張同類別圖片查詢時，使用傳統的 MPEG-7 做特徵擷取與僅使用 SIFT 做特徵擷取其準確率表現情況，分析圖表如圖 4.10 所示，在實驗結果中發現當擷取 SIFT 特徵來做為多張圖片查詢依據時，其相對的準確度比 MPEG-7 的來得高。

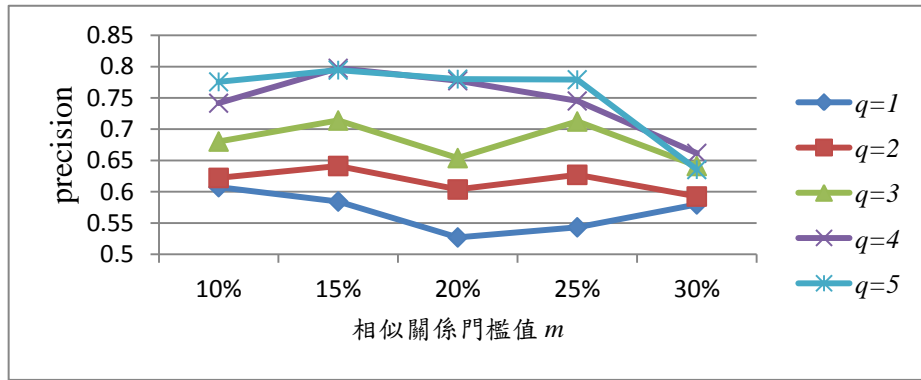


圖4.2 Query Image為同類別時，建立相似度關係門檻值 m 分析實驗結果

($m=10\% \sim 30\%$, $q=1 \sim 5$, $n=20$)。

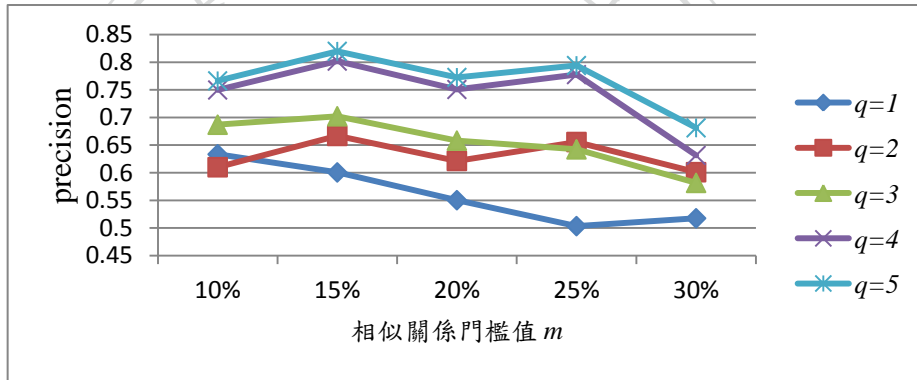


圖 4.3 Query Image 為不同類別時，建立相似度關係門檻值 m 分析實驗結果

($m=10\% \sim 30\%$, $q=1 \sim 5$, $n=20$)。

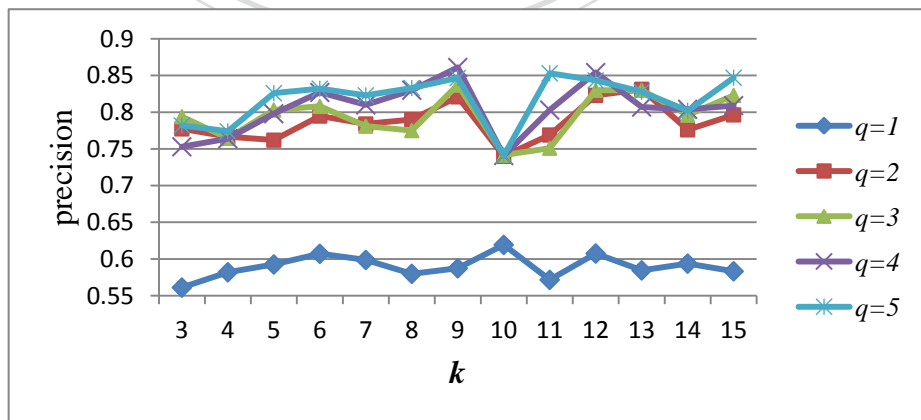


圖4.4 Query Image為同類別時，建立 k 個相似度關係分析實驗結果

($k=3 \sim 15$, $q=1 \sim 5$, $n=20$)。

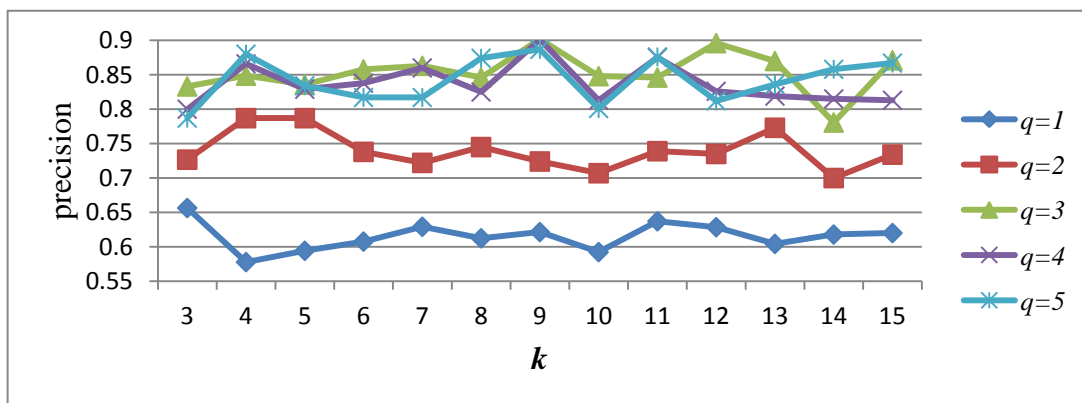


圖4.5 Query Image為不同類別時，建立 k 個相似度關係分析實驗結果
($k=3\sim 15, q=1\sim 5, n=20$)。

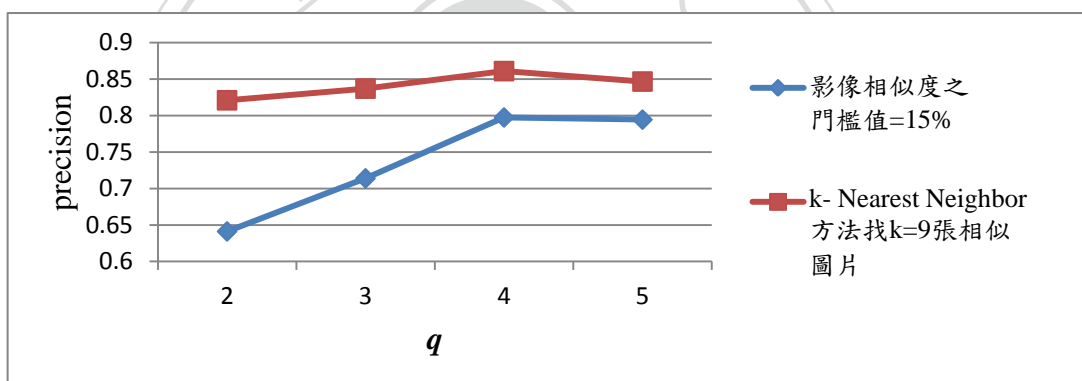


圖4.6 Query Image為相同類別時，門檻值 m vs. k 個相似度分析實驗結果
($k=9, m=15\%, q=2\sim 5, n=20$)。

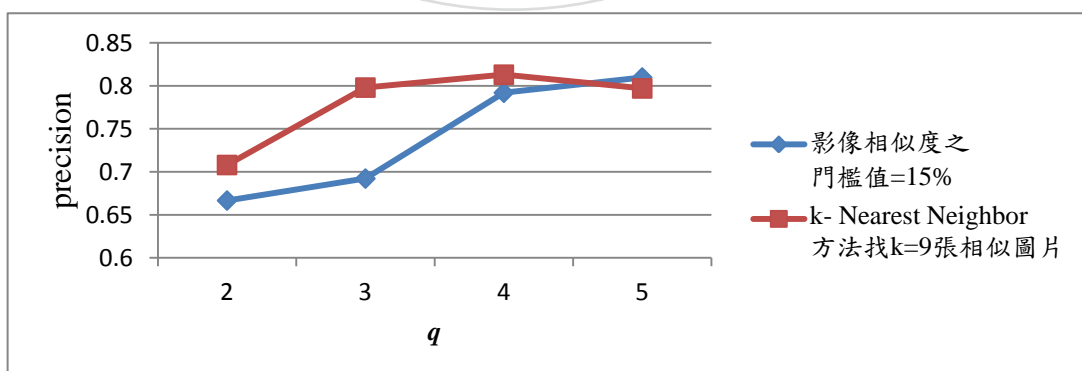


圖4.7 Query Image為不同類別時，門檻值 m vs. k 個相似度分析實驗結果
($k=9, m=15\%, q=1\sim 5, n=20$)。

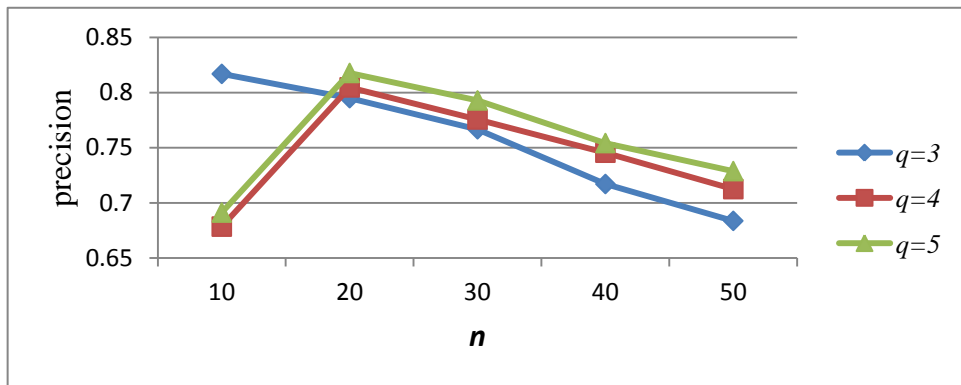


圖4.8 Query Image為相同類別時，回傳圖片 n 張之準確率分析實驗結果
($k=9, q=3\sim 5, n=10, 20, 30, 40, 50$)。

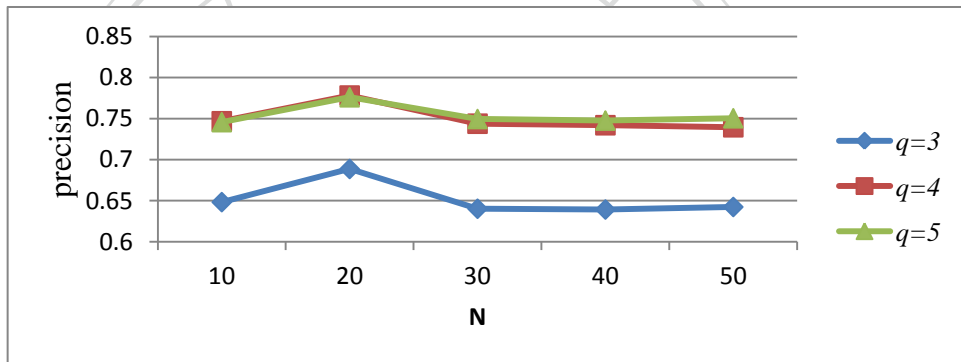


圖4.9 Query Image為不同類別時，回傳圖片 n 張之準確率分析實驗結果
($k=9, q=3\sim 5, n=10, 20, 30, 40, 50$)。

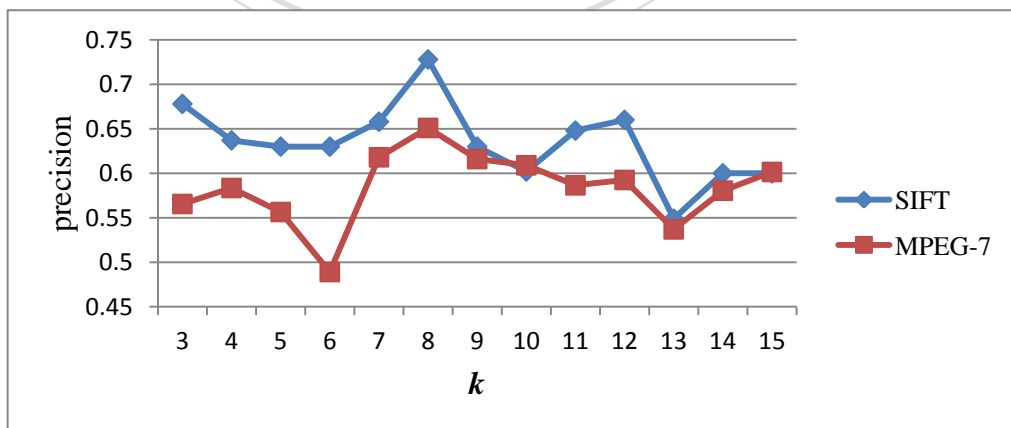


圖4.10 MPEG-7查詢與SIFT查詢之準確率分析實驗。

($k=3\sim 15, q=3, n=20$)

4.4 實驗總結

由上述所有實驗中，我們可以歸納一些特性

1. 在 k -Nearest Neighbor方法中找出最接近 k 個相似度建立關係的值實驗中，我們可以發現， k 值的設定會深深地影響到圖片的準確率變化，因此若能在建立圖片關係時，找到越相似的圖片進行關係建立，將對準確率的改進有很大的幫助。
2. 在實驗中證實了，我們提出將CBIB融入Community-Search Algorithm應用於多張影像搜尋的概念的確是有效能的，也證明這樣子的概念也較適合做多張圖片的搜尋，而實驗中也發現，在特徵擷取時，除了使用MPEG-7做特徵擷取之外，如果再加上SIFT的特徵，則對準確率有不錯的提升。

將社群的概念融入建立影像關係，從實驗中可以看出效果還不錯，主觀視覺觀察搜尋影像得到的結果，發現我們的方法在找出相關聯影像的能力是很好的，在客觀準確率圖表觀察，發現在多張影像搜尋上準確率有維持7~8成以上，穩定的準確率表現也是影像搜尋系統所須具備的特色。

第五章

結論與未來研究方向

5.1 結論

在本論文中，我們所提出的多張影像內容檢索影像內容搜尋擷取技術有不錯的成果，藉由將人際關係所產生之網路圖的概念，以此想法建立影像關係網路，距離近的影像表示其相似性越高，反之距離遠的影像表示其相似性較低，將 Community-Search Algorithm 應用於多張影像搜尋，找出相似影像出來。

我們的影像搜尋方法與一般影像搜尋不同的特色如下：

- (1) 建立影像關係網路。
- (2) 利用 Community-Search Algorithm 的技術融合社群的概念應用在圖片搜尋。
- (3) 多張影像查詢。

在多張影像搜尋上，回傳影像以人類視覺主觀觀察，發現回傳影像和查詢影像有不錯的相似度及關聯性，而以客觀數據測試結果準確度也不錯，顯示在多張影像搜尋結果效果頗佳。

5.2 未來研究方向

在本論文中所提出一個有效的多張影像內容檢索和搜尋方法，由於傳統影像搜尋都著重於影像特徵的抽取和比對，我們將影像特徵和影像關係整合為一個影像關係網路，藉由演算法將影像與影像之間相關聯的影像抓取出來且有不錯的效果。

現在是網路與數位科技蓬勃發展的世代，資料以極快的速度在成長，尤以多媒體資

訊為大宗，如何有效率且精確的搜尋出使用者所需要的影像是一個很重要的課題，傳統影像搜尋皆著重在影像關鍵字的搜尋，進階則是以影像相關特徵值找出相似的影像，不管傳統或進階影像搜尋皆是以單張圖片為搜尋來源，很少人提出多張影像搜尋的方法，本文提出一個應用在多張影像搜尋的方法，結合影像特徵值搜尋已有不錯的效果，顯示藉由影像特徵值相似度建立影像關係網路有不錯的成果，未來如果能提升影像特徵值擷取及相似度計算的效率，則應用於影像線上搜尋會更加完美。

影像關係網路目前以影像特徵值為判斷影像彼此關係的依據，未來如果能加上語意的關係，並讓使用者針對回傳的圖片勾選概念相符合的相關圖片後，回饋給系統，系統再藉由回饋資訊做配對調整，對於影像搜尋結果會更加準確且符合人類對於影像的想法及描述，使影像搜尋的使用能越來越貼近人的口語形容和人性化，長程目標未來可以進階為延伸至音樂、視訊影像搜尋。

參考文獻

- [1] M. Bober, “MPEG-7 Visual Shape Descriptors,” *IEEE Transactions on Circuits and System for Video Technology*, Vol. 11, No. 6, pp. 716-719, 2001.
- [2] C. Carson, S. Belongie, H. Greenspan, and J. Malik, “Blobworld: Image Segmentation using Expectation-Maximization and Its Application to Image Querying,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 8, pp. 1026–1038, 2002.
- [3] S. F. Chang, T. Sikora, and A. Puri, “Overview of MPEG-7 Standard,” *IEEE Transactions on Circuits Systems for Video Technology*, Vol. 11, No. 6, 2001.
- [4] M. Flickner, H. Sawhney, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, “Query By Image and Video Content : The QBIC System,” *IEEE Computer Magazine*, Vol. 28, No. 9, pp. 23–32, 1995.
- [5] A. Gupta, “Visual Information Retrieval: A Virage Perspective,” Virage, Inc. , San Mateo, Calif., 1995.
- [6] J. Han, and K. K. Ma, “Fuzzy Color Histogram and Its Use in Color Image Retrieval,” *IEEE Transactions on Image Processing*, Vol. 11, No. 8, pp. 944-952, 2002.
- [7] R. Hess, “An Open-Source SIFT Library,” In *Proc. of the ACM International Conference on Multimedia*, pp.25–29, 2010.
- [8] T. S. Huang, S. Mehrotra, and K. Ramachandran, “Multimedia Analysis and Retrieval System (MARS) Project,” In *Proc. of 33rd Annual Clinic on Library Application of Data Processing-Digital Image Access and Retrieval*, 1996.
- [9] E. Kasutani, A. Yamada, “The MPEG-7 Color Layout Descriptor: a Compact Image Feature Description for High-speed Image/Video Segment Retrieval,” In *Proc. of International Conference on Image Processing*, pp. 674-677, 2001.

- [10] H. K. Kim, J. D. Kim, D. G. Sim, and D. I. Oh, "A Modified Zernike Moment Shape Descriptor Invariant to Translation, Rotation and Scale for Similarity-Based Image Retrieval," In *Proc. of the IEEE International Conference on Multimedia and Expo*, pp. 307-310, 2000.
- [11] J. J. Koenderink, "The Structure of Images," *Biological Cybernetics*, Vol. 50, No. 5, pp. 363-396, 1984.
- [12] H. J. Lin, Y. T. Kao, S. H. Yen, and C. J. Wang, "A Study of Shape-Based Image Retrieval," In *Proc. of 24th International Conference on Distributed Computing Systems Workshops*, pp. 118-123, 2004.
- [13] T. Lindeberg, "Scale-space Theory: A Basic Tool for Analyzing Structures at Different Scales," *Journal of Applied Statistics*, Vol. 21, No. 2, pp. 224-270, 1994.
- [14] D. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.
- [15] B. S. Manjunath, G. M. Haley, and D. F. Dunn, "Efficient Gabor Filter Design for Texture Segmentation," *Pattern Recognition*, Vol. 29, No. 12, pp. 2005-2016, 1996.
- [16] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "MPEG-7 Color and Texture Descriptors," *IEEE Transactions on Circuit and System for Video Technology*, Vol. 11, No. 6, pp. 703-715, 2001.
- [17] B. S. Manjunath, P. Salembier, and T. Sikora, "Introduction to MPEG-7: Multimedia Content Description Standard," New York: Wiley, 2001.
- [18] J. M. Martinez, "Standards-MPEG-7 Overview of MPEG-7 Description tools, Part 2," *IEEE Multimedia*, Vol. 9, No. 3, pp. 83-93, 2002.
- [19] W. Niblack, R. Barber, W. Equitz, et al , "The QBIC Project: Querying Images by Content using Color, Texture, and Shape," In *Proc. of SPIE Electronic Imaging: Science*

and Technology, 1993.

- [20] J. S Payne, T. J. Stonbam, “Can Texture and Image Content Retrieval Methods Match Human Perception,” In *Proc. of Intelligent Multimedia, Video and Speech Processing*, pp.154-157, 2001.
- [21] A. Pentland, R.W. Picard, and S. Sclaroff, “Photobook : Tools for Content-Based Manipulation of image databases, ” *International Journal of Computer Vision* , Vol.18, pp. 233-254, 1996.
- [22] K. Porkaew , S. Mehrotra, and M. Ortega, “Query Reformulation for Content Based Multimedia Similarity Retrieval in Mars,” In *Proc. of IEEE Conference on Multimedia Computing and Systems*,pp.747-751, 1999.
- [23] Y. Rui, T. Huang, and S. Mehrotra, “Content-Based Image Retrieval with Relevance Feedback in MARS,” In *Proc. of IEEE International Conference on Image Processing* , pp. 815-818, 1997.
- [24] T. Sikora, “The MPEG-7 Visual Standard for Content Description-An Overview, “ *IEEE Transactions on Circuits Systems for Video Technology*, Vol. 11, No. 6, 2001.
- [25] J. R. Smith and S. F. Chang, “Visualeek: A Fully Automated Content-Based Image Query System,” In *Proc. of the ACM International Multimedia Conference*, pp.87-98, 1996.
- [26] J. R. Smith, Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis, PhD Thesis, *Graduate School of Arts and Sciences, Columbia University*, 1997.
- [27] M. Sozio and A. Gionis, “The Community-Search Problem and How To Plan A Successful Cocktail Party,” In *Proc. of 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10*, pp.939-948, 2010.

- [28] P. Y. Yin , S. H. Li, “Content-Based Image Retrieval Using Association rule Mining with Soft Relevance Feedback” *Journal of Visual Communication and Image Representation*, Vol.17, No. 5, pp.1108-1125, 2006.
- [29] Y. Zhang, M.A. Nascimento, and O.R. Zaiane, “Building Image Mosaics: An Application of Content-Based Image Retrieval,” In *Proc. of IEEE International Conference on Multimedia and Exposition*, 2003.
- [30] “MPEG-7 Visual Experimentation Model (XM) Version 10,” ISO/IEC/JTC1/SC29/WG11, Doc. N4063, 2001.
- [31] “Overview of the MPEG-7 Standard Version 5.0,” Final Committee Draft, ISO/IEC/JTC1/SC29/WG11, Doc. N4031, 2001.

