

# 智慧型數位影像內容註解及搜尋系統之設計

## A Framework for Intelligent Multimedia Contents Annotation and Retrieval

曾新穆 蘇家輝

成功大學資訊工程學系

Email: tsengsm@mail.ncku.edu.tw

### 摘要

由於資訊科技不斷地進步，愈來愈多的珍貴多媒體資料被真實地紀錄著，例如：文字、影像、視訊或音樂，這些資料反映了人類在科學、社會、人文、教育、歷史、生物等領域方面的知識累積。目前國內外各典藏機構均已累積了豐富的數位多媒體內容，然而，對這些珍貴的多媒體數位內容之搜尋仍停留在主題式瀏覽及關鍵字搜尋之階段。事實上，就使用者的搜尋需求而言，由於多媒體資料的多樣性與大量性，僅依靠主題式瀏覽及關鍵字搜尋並無法滿足使用者真正之查詢需求。因此，以內容為基礎的影像搜尋系統在近年已逐漸成為在此領域上研究發展的重點。

本研究之目標在設計一套智慧型數位影像內容註解及搜尋系統架構，研究重點可包含以下技術：1) 以叢集與分類為基礎之影像索引技術，2) 智慧型影像資料註解系統之研發，3) 互動式及群聚式影像搜尋技術，4) 運用搜尋紀錄探勘 (Log-Mining) 之高效能預測技術之研發。經由以上重點項目之研發及整合，我們預計可幫助各典藏機構產出具有創新性、高效能、高準確性之數位影像內容註解及搜尋系統。  
**關鍵詞：**影像搜尋，影像註解，資料探勘，多媒體資料庫

### 1. 緒論

電腦科技產業在莫耳定律的原則下不斷進步，不管是處理器的速度或是儲存媒體的空間都成倍速成長，加上各種數位影音技術的成熟，使得各種媒體內容轉成數位形式(即所謂數位內容)處理、儲存與傳播變得更加可行，形成隨時會有引爆一場多媒體數位影音革命的態勢。而在過去幾年，也由於資訊科技不斷地進步，愈來愈多的珍貴影音資料被真實地紀錄著，例如：文字、影像、視訊或音樂，這些資料反映了人類在科學、社會、人文、教育、歷史、生物...等領域方面的知識累積，對人類而言，正是世界進步的重要資產，故現正被妥善地保存著。

目前國內外各典藏機構均已累積了豐富的數位多媒體內容，而國內近幾年在政府兩兆雙星政策推

動下，藉由「數位典藏國家型計畫」等計畫研究，也為國內許多典藏機構建置了各種領域之寶貴數位多媒體內容。然而，目前對這些珍貴的多媒體數位內容之搜尋仍停留在主題式瀏覽 (Topic-based browsing) 及關鍵字搜尋 (keyword-based search) 之階段。如國立海洋生物博物館、國立自然科學博物館等典藏機構，提供主題式瀏覽及關鍵字搜尋方式。事實上，就使用者的搜尋需求而言，由於多媒體資料的多樣性與大量性，僅依靠主題式瀏覽及關鍵字搜尋並無法滿足使用者真正之查詢需求。因此，以內容為基礎的影像搜尋系統 (Content-Based Image Retrieval, 簡稱 CBIR) 在近年已逐漸成為在此領域上研究發展的重點 [7][8][9][12][16]。不管對使用者或研究者而言，快速而準確地搜尋到所需之影像資料為此研究領域一致的要求。目前國外已有發展一些 CBIR 之雛形系統，以多媒體資料之特徵，如顏色 (Color)、紋理 (Texture) 與形狀 (Shape) 等為基礎，有效率的分類、索引與擷取多媒體資料 (如圖 1 所示)。

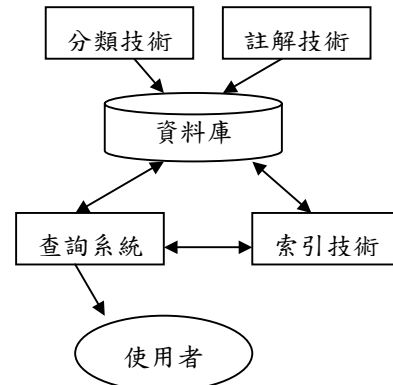


圖 1：影像處理技術流程圖。

雖然這些以內容為基礎的影像搜尋系統也許能提供不錯的擷取介面環境，但是，影像擷取的效果差強人意，搜尋的過程亦費時而無效率，在實際的應用中，仍未能達到一定的水準，其效率表現實難滿足使用者即時的要求。另外，在典藏機構端亦缺乏一套高效能之影像註解技術，以建立如 MPEG-7

等標準化之後設資料(meta-data)，以便於維護及檢索大量之數位影像內容。因此，本研究之主要目標便是要設計一套智慧型數位影像內容註解及搜尋系統架構，以便能正確而有效率地提供使用者新一代之影像搜尋功能。

本論文其餘章節的安排如下。在第二節中，我們對國內過去相關的研究作一說明。在第三節中，我們說明本系統的設計架構。最後，第四節是我們的結論與未來的研究方向。

## 2. 文獻探討

由於網路的興起，各種資料的傳遞與交換變得無遠弗屆，且因為儲存設備的進步，如數位相機，數位攝影機等的普及，帶動了多媒體資訊的快速暴增，如影像，音樂，影片等，進而需要有一個好的系統來有效率地管理這些資料。在傳統的影像搜尋系統中，以查詢的方式而言，過去有許多系統是以文字來做為查詢的索引 (QBK, Query By Keyword)，此類系統會根據使用者輸入的文字當作是搜尋索引的目標，進而找到相關的影像；但當使用者不知道如何下達正確的索引時，或是影像與文字之間多對多的關係會使得查詢的結果變得相當不穩定；故近年來以由影像做為查詢依據 (QBC, Query By Content) 的研究已成為影像查詢系統研究的主流。當一張影像加入進資料庫時，就要對這張影像擷取出多個特徵向量，而這些向量除了必須能夠精確的表示其影像的內容外，也必須遠比原影像檔案小。其中最困難的部分是要如何建立一組特徵向量可以保持影像的內容並可以快速的搜尋。經過特徵擷取後，將這些特徵向量儲存至一索引表。當使用者查詢時，則利用前面所計算過的向量索引，選出和查詢影像的特徵向量，有最短距離的索引向量。通常最短距離大多採用加權式的歐幾理德距離 (Weighting Euclidean Distance)，如下列所示。

$$|Q, T| = \sum_i w_i |q_i - t_i|$$

而這些由影像做為查詢依據可以再細分為由實際影像做為查詢，或是可以由使用者來描繪出所想要的意涵，甚至也有進一步加入使用者習性做為考量的查詢系統。這種以影像為查詢索引的影像搜尋系統便稱為以內容為基礎的影像搜尋系統。如何建立起影像低階特徵值與人類知識高階意涵，為以內容為基礎的影像搜尋系統研究的一大挑戰，現今比較有名的有 QBIC (Query By Image Content, 由 IBM Almaden Research Center 所開發) [4]，VIR Image Engine (由 Virage Inc. 所開發)，VisualSEEK (由 Department of Electrical Engineering, Columbia University 所開發) [15]，NeTra (由 Department of Electrical and Computer Engineering, University of California 所開發)，MARS (Multimedia Analysis and Retrieval System, 由 Backman Institute for Advanced Science and Technology, University of Illinois 所開

發)，及 Viper (Visual Information Processing for Enhanced Retrieval, 由 Computer Vision Group, University of Geneva 所開發)。

Li 等學者在 2000 年提出利用 2-D HMM 的方法對影像之區塊做二元分類 (Binary Classification)[11]。該研究利用影像區塊之特徵向量與影像區塊的空間關係套用在 2-D HMM 之方法建立影像分類之模組，並用建立之模組分類未標示之區塊。該研究的實驗中顯示該方法可區分高空的影像中屬於人為和屬於自然現象的區塊。Szummer 等學者在 1998 年提出一個分辨戶外與戶內影像之二元分類法。該研究將影像分成數個小區塊，透過計算擷取每個小區塊之特徵值，包括 Ohta 之色彩空間座標與離散餘旋轉換 (Discrete Cosine Transform, DCT) 之係數等特徵。接著將所有區塊，將各種不同的特徵向量分別應用 KNN 的方法，建立影像二元分類法則 (戶內或戶外) [17]。該研究的實驗中採用 1300 張影像配合各種特徵向量，比較將影像與將影像跟分割成數個區塊後之分類準確度。Aghbari 等學者在 2003 年提出利用 SVM 做影像分類之方法 [1]。該研究之目的是希望能一改過去利用影像之低階特徵向量，如顏色、形狀、紋理等做影像分類之依據，而是採用以人的概念觀點為基礎 (Concept-Based) 做影像分類。為了達到此目的，該研究重點在於如何分類影像中之每一物件。若能順利分類影像中的物件，便可利用分類完成之物件，進一步以人的概念觀點為基礎做影像分類。在該研究中利用 hill-climbing 之方法 [14] 將影像分割成數個區域，並擷取這些區域四種特徵向量 (Color Histogram、Edge Direction Histogram、Higher-Order Autocorrelation Edge Vector and Dual-Tree Complex Wavelet Transform)。套用這四種特徵向量於 SVM 方法中，建立影像物件的分類規則。Huang 等學者在 1998 年提出利用建立分類樹 (Classification Tree) 之方法建立影像分類模組 [10]。而 Carson 等學者則提出新的影像表示法 [5][6]。每張影像被分割成數個 blobs，每個 blob 包含其顏色及紋理兩特徵值。經由每個 blob 之特徵值，利用特徵距離公式計算任兩 blob 之間的距離分數。利用計算出距離分數，建立出一決策樹。該研究並利用建立出之決策樹做影像搜尋，比較其搜尋之影像正確率。資料探勘是利用分析演算法，經由自動或半自動的方式，從大量的資料中找出有意義的關係或規則。其中著名的探勘演算法 Apriori [2][3] 是 Agrawal 等學者在 1994 年所提出，其主要之概念是重複的依據候選項目集合 (Candidate Itemsets) 來讀取資料庫，做支持值的計算，產生長度相同的大項目集合，產生完後在計算出下一個長度的候選項目集合，再讀取資料庫，如此循環下去，直到沒有大項目集合被找到為止。由於影像物件在人的觀點而言，有其關係，因此尋找這些物件關係之關聯法則將有助於影像的分類 [13] 更接近人類的

觀點。

### 3.系統架構

本研究主要在設計一套智慧型數位影像內容註解及搜尋系統架構，其目的是希望能讓使用者透過互動及智慧型之介面，迅速而精確地搜尋到其所需之數位影像內容。為使本研究設計之系統架構可實際導入應用於典藏機構中，本研究將從使用者端及典藏機構端二個面向來考慮。在使用者端方面，我們之設計將以提供下述之功能為目標：

- 高彈性：可以關鍵字或影像內容來進行查詢
- 高互動性：可對搜尋結果進行篩選及回饋
- 高準確性：能由大量之影像資料庫中找到使用者真正需要之內容
- 智慧性之結果呈現：能將查詢結果歸類呈現在機構端方面，主要以提供下述之功能為目標：
- 建立影像內容之標準化後設資料，以利維護
- 自動化產生影像內容之註解資料，以減少人力之耗費

為達到上述之需求，本研究之主要設計相關技術如下：

- I. 以叢集與分類為基礎之影像索引技術：運用叢集(Clustering)或分類(Classification)技術改進現有影像查詢方法的準確度及執行效率，並進而發展智慧型影像比對搜尋技術。目前影像查詢技術中最常用的方法其準確度及效率皆有待加強，尤其缺乏對影像概念趨勢之分析。我們希望運用資料探勘中的叢集與分類技術發展新的影像索引及比對搜尋技術。
- II. 智慧型影像資料註解 (Image Annotation)：發展以 MPEG-7 為標準之影像資料註解技術及系統，以確保註解資料之正確性及標準性，除提供人工註解功能外亦藉由資料探勘技術提供自動註解功能。同時並設計適用於合作典藏機構之資料本體結構(Ontology)，整合於此影像資料註解系統中。
- III. 互動式及群聚式影像搜尋技術：我們將以概念趨勢呈現影像搜尋結果。所謂「概念趨勢」即是查詢結果不以單獨影像為單位，其中內含類別概念的觀念，將相同概念的影像群聚顯示，其優點為可同時顯示多種類別的影像。同時，我們希望以互動式及回饋式的影像查詢技術，讓使用者可篩選及強調某些影像特徵，藉以正確傳達使用者查詢影像的概念，以得到精確的結果。
- IV. 運用搜尋紀錄探勘 (Log-Mining) 之高效能預測技術：搜尋影像時，需要找出查詢影像本身與目標影像之間的關係，因此可從查詢的交易紀錄之間，找出適當的關聯規則，因此我們希望利用資料探勘技術有效的結合各項的資訊，並探勘出適當的規則資訊，預測使用者之查詢傾

向，以提高影像搜尋之精確性。

我們認為整合上述技術，可產出一套智慧型數位影像內容註解及搜尋系統，其架構如圖 2 所示，系統會先收集大量影像，經由人工智慧的方式，自我訓練並註解語意影像，存入多媒體資料庫；當使用者以影像或關鍵字進入本系統查詢時，系統會經過影像分析及辨認後，藉由搜尋元件搜尋相似語意概念之影像，並快速回傳給使用者；使用者可根據所需再回饋給系統，以獲得最佳結果；系統在使用者與系統互動回饋的同時，會將回饋的過程紀錄起來，以作為下次相同類型查詢的依據，如此，不但可以提高系統影像擷取的準確率，更能有效地縮短系統查詢的時間。以下各小節，我們將針對各技術設計做一詳細說明。

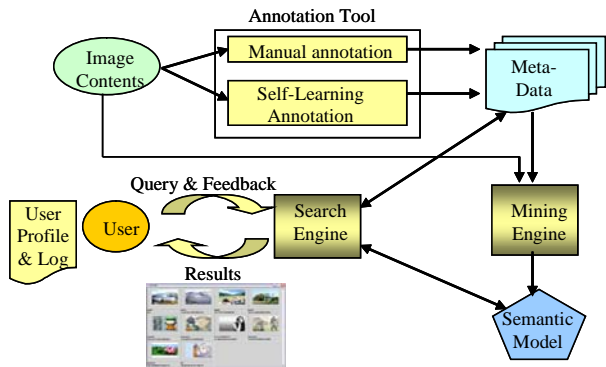


圖 2：整合系統架構。

#### 3.1 以叢集與分類為基礎之影像索引技術

多媒體資料的索引與一般文字資料有些許不同之處。首先，文字資料可以經過一些編碼壓縮，然而其他種類的資料本身就是經過編碼壓縮的，加上具有時間的特性，而且資料量龐大，因此多媒體內容之索引也格外重要。一般而言，由於多媒體資料有時間的特性，而且資料量龐大，因此在索引時，我們通常只將特徵的索引存進資料庫，而把多媒體資料本身留在檔案系統。

所以，欲提供使用者一個整合式的多媒體資料查詢系統，就必須先把數位典藏資料庫所分析出來的多媒體特徵資料，依照 MPEG-7 格式儲存到資料庫系統中，方能做有效率的管理與應用。與此同時，我們也需要制定一套有效率的儲存方式，亦即必須做好資料庫 schema 的制訂，以免在存放資料時浪費空間，或者影響到擷取資料時的效率。以下我們針對叢集與分類做簡單敘述。

- **Clustering**

叢集 (clustering;或稱聚類) 是一種將資料分群的方法，使得這些資料在群組之內的相似度較高，而群跟群之間的相似度則較低。叢集技術可以將性質相近的資料擺在一起，可以提升搜尋速度。因此叢集應用於多媒體資料庫也是相當重要的，如

- 果資料能事先經過叢集的組織，對於日後在搜尋、擷取時，能更加有效率。叢集的概念就如圖 3 所示，可以將資料以分群的方式表示之。叢集的步驟如下：
- 選擇  $k$  個點作為 cluster 的中心點，重複以下 2~5 的步驟。
  - 將每個點指定給離該點最近的中心點，形成  $k$  個 cluster。
  - 重新計算 cluster 的中心點。
  - 重複 2~3 直到達成聚類要求為止。
  - 微調聚類後結果。

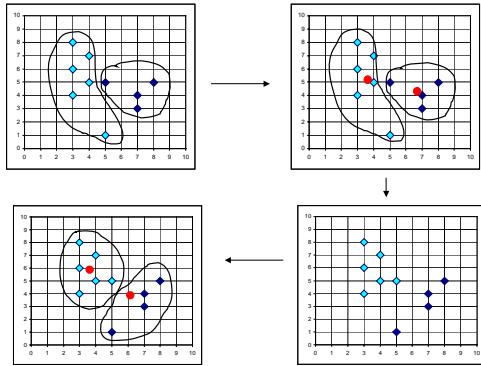


圖 3：叢集示意圖。

### ● Classification Techniques

分類 (classification) 是一種將資料分群的技术，與叢集 (clustering) 不同的地方在於，分類是根據已知的結果分析出一個能預測類別的模式，而叢集的分群結果是根據資料特徵但卻是未知答案的。因此，分類主要的用途是能夠做預測。分類的過程分兩個階段：1. 訓練過程，2. 測試過程。圖 4 代表一個分類方法中的訓練過程。根據 TENURED 欄位的分類結果，這些已知的資料被用來訓練出可預測的分類模式 (classifier model)。而圖 5 則是利用訓練好的分類模式，用測試資料進行精確度的評估。

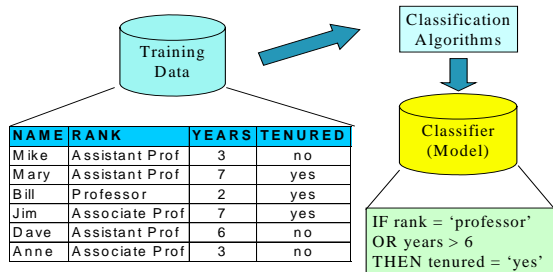


圖 4：分類的訓練過程。

與叢集不同的是，分類可以做到預測。因為分類是根據已知的結果發展出來的模式，在依據模式的規則作分類預測。所以有助於我們利用多媒體意涵特徵當作類別，以多媒體資料特徵當作要分類的欄位，這樣得到分類模式之後，只要有新的一筆多

媒體資料進來，就可以自動將它歸類。舉例而言，假設我們的資料庫裡有下列幾個分類：運動、旅遊、家庭生活等等。而我們發展出來的分類模式中，有一項規則是，當出現球就歸到運動類。於是，有一筆新的影像要存進資料庫，影像當中出現球，於是這段影像就被歸類到運動類去。如此，可以省下許多人工分類的時間。

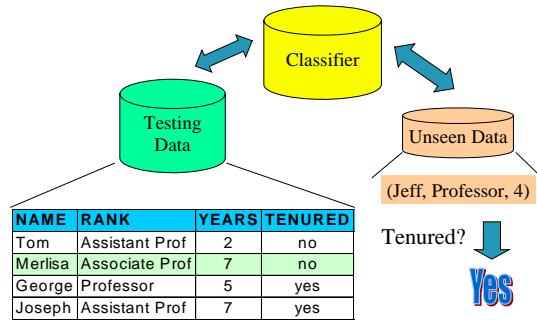


圖 5：分類的測試過程。

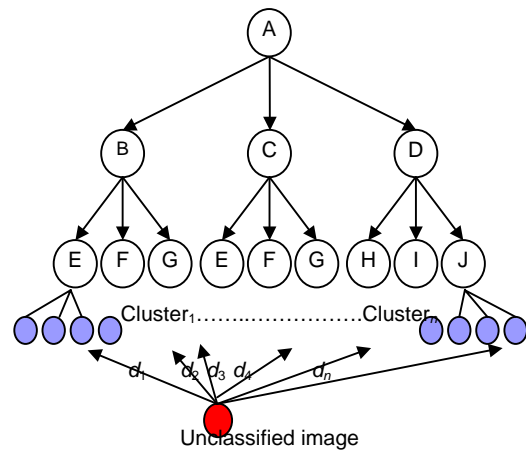


圖 6：叢集分類應用範例。

綜合以上所述，就其應用舉例而言，我們可先設計一多階層的語意實體架構，如圖 6 所示，階層愈高，代表愈高階的人類語意概念。接著，蒐集的初步的分類影像，以自動化的方式去除每類影像中的雜質 (Noise)，再將其存入語意資料庫。然後可針對每一種類別收集不等的影像，再利用分群的分法將每類影像以特徵值分群，並藉以去除雜質 (noise)，接著再從每一群中找出最密集的中心點作為範例影像，如此，每種類別因為特徵不同而分成不相等的群數，換句話說，每種類別可能出現數個不相等的範例影像。一旦範例影像被定義後，接下來的工作便是將未分類影像映射在此語意實體架構，便能得到較高階的類別 (如圖 6 所示)，甚而再經由探勘引擎並加以瀏覽可得到更高階的人類語意概念。至於如何決定一未分類影像之歸屬類別，我們採取了相似度的計算方式，也就是說，相似度愈高表示兩影

像間之距離愈近，所以找出未分類物件與各範例物件之間最短距離。

### 3.2 智慧型影像資料註解

影像註解就是賦予影像高階的語意概念，與影像分類最大的不同在於影像不再侷限於單一類別，而是著重於整張影像所隱涵的意義，其註解的關鍵字可能不只一種。如圖 7 所示，若以一般影像分類之結果應為{tiger}，但若以影像註解而言，其結果應為{tiger, water, swimming, lazy}。由圖 7 範例我們可以知道一張影像的涵義可能不侷限於影像上的物件，通常單以影像特徵值是很難找到這些高階語意，且若要以人工智慧的角度來訓練機器，其高階語意範圍太廣，訓練成本頗高，且難以一一涵蓋所有人性概念，這也是研究影像註解的最大挑戰之一。很不幸的是，此類影像在我們週遭比比皆是。最近由於網際網路的盛行，於是許多學者便利用網際網路豐富的知識來源，將網頁中的文字特徵與網頁中的影像加以連結，希望透過網頁的多元性，增加影像註解的範疇與正確性，更藉以精確抓取影像中的高階語意，這也是本研究最終的目標之一。



圖 7：高階語意概念範例。

所以在介紹完影像分類索引後，接下來是做多媒體的註解並將其以 XML 的格式儲存，同時也具備有 learning 的功能，透過事先建立好的模組自動註解多媒體檔案，利用 Annotation 的結果以及 MPEG-7 的技術，在實務上可以運用在 Content-Based Image Retrieval 的領域，以增加查詢的準確度。

如圖 8 所示，本方法是將影像以 XML 的格式加以分析，並透過特殊之註解演算法對影像加以註解，同時也具備有 learning 的功能，系統可透過人工註解的方式，事先建立好 training model，接著系統可批次大量接受影像，並自動加以註解。而利用 Annotation 的結果以及 MPEG-7 的技術，在實務上可以運用在 Content-Based Image Retrieval 的領域，以增加查詢的準確度。

### 3.3 互動式及群聚式影像搜尋技術

本研究希望使用多種演算法提供多種方法的查

詢，並且整合這些方法改進擷取的效果。當然，在這部分也必須與使用者介面作配合，使用者介面提供了使用者的多重查詢方式，譬如關鍵字加上影像內容的多重媒體方式，在結果的呈現上亦採概念趨勢的觀念，不再以單張的影像為單位，而是以類別概念為單位，讓結果更具彈性。此外，使用者可根據查詢結果，強調某些影像特徵重新查詢，使結果更加準確。

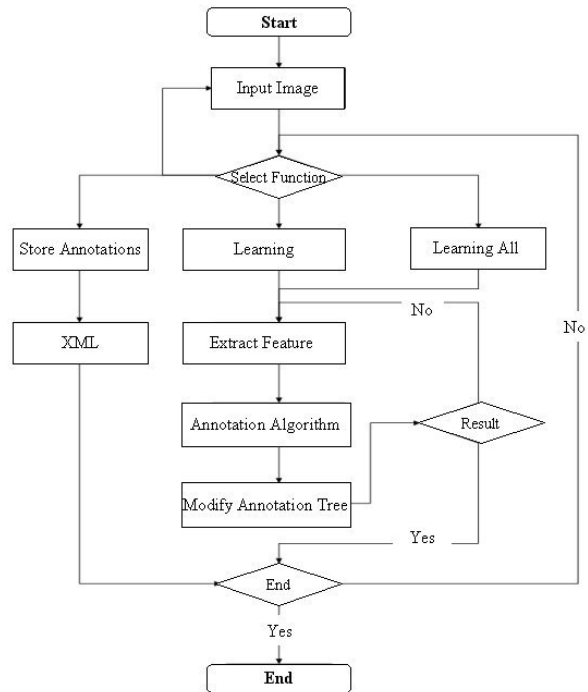


圖 8：影像註解流程圖。

#### ● 整合關鍵字與影像內容之查詢

在多媒體還不盛行的時代，電腦資料幾乎都是以文字型態來儲存，因此在檢索時，也都是以文字為基本型態。而隨著多媒體時代的來臨，資料不在只是單純的文字資料，而是多種混合型態的多媒體資料（包括文字、影像、語音、視訊與文件等）。因此，檢索方式不在單以文字為主，利用多媒體內容本身的特徵作檢索的方法也漸漸被提出來。畢竟，文字只能檢索資料的意義，並不能檢索資料本身的特徵。因此，將文字與多媒體資料內容兩種檢索方式結合是必然的趨勢。

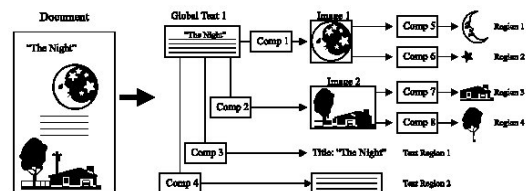


圖 9：多媒體物件模式範例。

在目前的研究報告中也顯示，多媒體檢索與擷

取有朝把文字與多媒體內容結合的趨勢，而本研究即希望對 MPEG-7 多媒體資料以 integrated keyword-and-content 來查詢之搜尋系統功能，以文字與影像為例，首先我們將影像與影像周圍的文字，依照 global 與 local 分成好幾種特徵：color histogram、color moment、concurrence texture、wavelet、title 以及 keywords 等等。接著如圖 9 所示，我們將把文件（含有多媒體資料的文件）拆成好幾個物件來看，之後在進行比對時，再各別把部分的相似度值，利用一個結構化的數學運算（如圖 10），把相似度加總起來作為最後的評量。

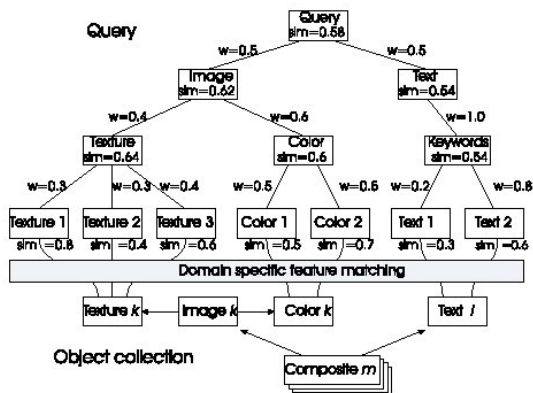


圖 10：整合文字與影像之查詢樹演算法。

## ● 互動式影像擷取

### (A) Reweighting

當使用者利用一張影像來做資料查詢或擷取時，如圖 11 所示，由系統自動擷取該影像的特徵值 (Features)，透過使用者輸入對應每一個特徵的加權值 (weight) 來比對資料庫中的資料，接著系統列出比對結果最接近的幾筆資料。而查詢系統列出搜尋比對出來的結果最接近的前幾名之後，使用者可以對這些結果進行選取，從中選擇使用者認為相關或不相關的結果，利用使用者選取的內容，系統會利用公式： $Sim_{R \rightarrow Q} = \sum_{i=1}^m w_i Sim_i$ ，透過  $m$  個特徵加權相似度  $w_i * Sim_i$  的加總，計算出查詢影像  $Q$  與搜尋影像  $R$  之間的相似度。重新比對出較適合的查詢結果，這一個程序稱之為 Relevance Feedback。透過不斷 Feedback 的方式，查詢系統會不斷更新查詢結果，而使得結果較符合使用者的需求。



圖 11：Relevance Feedback。

### (B) Query Point Movement

如圖 12，此方法則是不更動原先每個特徵的加權值，而是利用 Feedback 的結果，重新產生新的虛擬影像 (Virtual Image)，並使用該虛擬影像 (而不是使用者一開始給予的影像)，即圖中每一圓之中心點，作為系統的輸入，重新做查詢，直到中心點向正確答案逼近為止。

## ● 群聚式影像搜尋結果呈現技術

傳統的影響搜尋系統，利用特殊的相似度計算，對資料庫中的影像一一比對，最後將相似度較高的影像回傳給使用者，此比對程序不但耗時，結果亦未近人意。如圖 13 所示，其搜尋結果通常會造成某些特徵值相似卻不是使用者真正想要的影像，或是類別特徵值相似高的影像卻沒有使用者想要的概念，例如，使用者想找跳火圈的老虎，結果卻只有草地上奔跑的老虎，這也是傳統影像查詢系統所忽略的一個重點。

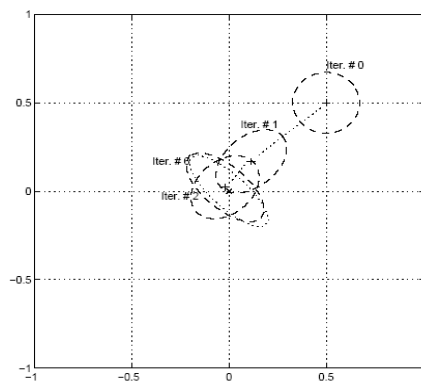


圖 12：Query Point Movement 概念圖。

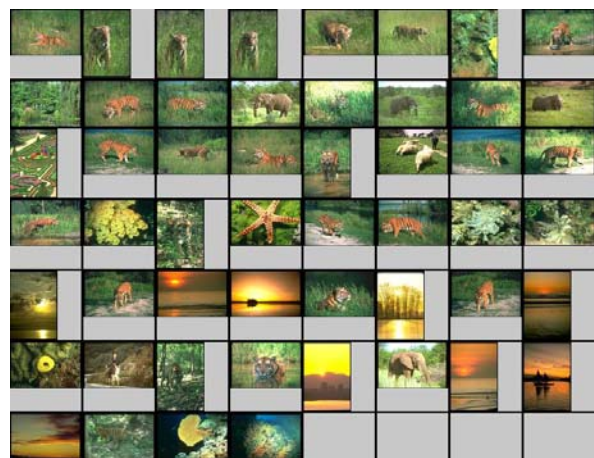


圖 13：傳統搜尋結果。

為了要使得查詢結果更具有彈性且更能接近使用者的概念意識，我們將利用上述叢集的方法將相同類別的圖根據其特徵值加以分群，如圖 14 所示，當使用者只下簡單的關鍵字「老虎」及影像查詢時，系統將以關鍵字先行比對，而找出所有老虎之分群影像，當使用者勾選類別及特徵值選項後，系統會

根據使用者下的條件，再作特徵值得進一步比對，以求取最佳結果。

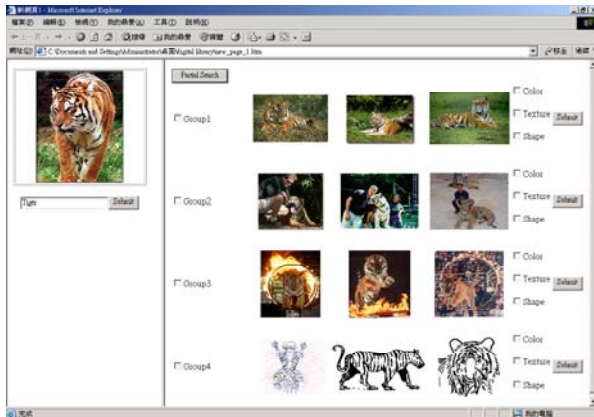


圖 14：群聚式影像搜尋結果。

### 3.4 運用搜尋紀錄探勘之高效能預測技術

為了更進一步分析多媒體資料特徵，本研究也希望利用使用者查詢紀錄探勘技術來分析這些多媒體資料特徵，資料探勘技術可以從大量資料中挖掘出有用的資訊或樣本 (pattern)，這些知識可以幫助更快且更精準地搜尋目標，提供使用者個人化的服務。如圖 15 所示，從查詢的交易紀錄之間，找出適當的關聯規則，因此我們利用資料探勘技術，能夠有效的結合各項的資訊，並產生出適當的規則資訊，在使用者進行影像查詢時能立刻透過搜尋使用規則庫的資料，進行比對後能順利的作出更精確的搜尋。

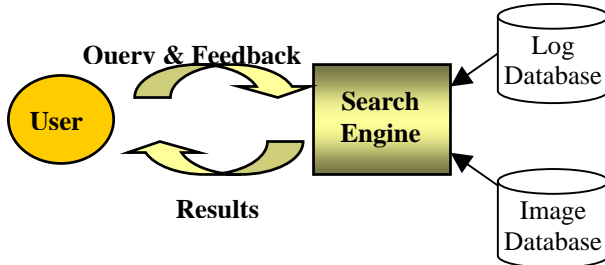


圖 15：查詢紀錄探勘之高效能預測技術。

表 1：影像查詢紀錄。

	Relevance Feedback	Relevant Images
1	(I <sub>1</sub> , I <sub>5</sub> ) (I <sub>1</sub> , I <sub>5</sub> , I <sub>6</sub> ) (I <sub>1</sub> , I <sub>5</sub> , I <sub>6</sub> , I <sub>8</sub> )	I <sub>1</sub> , I <sub>5</sub> , I <sub>6</sub> , I <sub>8</sub>
2	(I <sub>0</sub> ) (I <sub>2</sub> , I <sub>3</sub> ) (I <sub>2</sub> , I <sub>3</sub> , I <sub>6</sub> ) (I <sub>2</sub> , I <sub>3</sub> , I <sub>5</sub> , I <sub>6</sub> )	I <sub>0</sub> , I <sub>2</sub> , I <sub>3</sub> , I <sub>5</sub> , I <sub>6</sub>
3	(I <sub>6</sub> , I <sub>8</sub> ) (I <sub>2</sub> , I <sub>6</sub> , I <sub>8</sub> ) (I <sub>2</sub> , I <sub>6</sub> , I <sub>7</sub> , I <sub>8</sub> )	I <sub>2</sub> , I <sub>6</sub> , I <sub>7</sub> , I <sub>8</sub>
4	(I <sub>2</sub> ) (I <sub>5</sub> , I <sub>6</sub> ) (I <sub>2</sub> , I <sub>5</sub> , I <sub>6</sub> ) (I <sub>2</sub> , I <sub>5</sub> , I <sub>6</sub> , I <sub>7</sub> , I <sub>8</sub> )	I <sub>2</sub> , I <sub>5</sub> , I <sub>6</sub> , I <sub>7</sub> , I <sub>8</sub>
5	(I <sub>3</sub> , I <sub>4</sub> ) (I <sub>2</sub> , I <sub>3</sub> , I <sub>4</sub> ) (I <sub>2</sub> , I <sub>3</sub> , I <sub>5</sub> , I <sub>6</sub> , I <sub>8</sub> , I <sub>9</sub> )	I <sub>2</sub> , I <sub>3</sub> , I <sub>4</sub> , I <sub>5</sub> , I <sub>6</sub> , I <sub>8</sub> , I <sub>9</sub>

接下來我們舉一例子，如表 1 所示，在每一筆查詢時，系統會將使用者每次回饋 (Feedback) 時所選取的影像記錄下來。以第三次查詢為例，使用者第一次選取了 I<sub>6</sub>, I<sub>8</sub> 兩張圖，第二及第三次則分別選取了 I<sub>2</sub>, I<sub>6</sub>, I<sub>8</sub>, I<sub>2</sub>, I<sub>6</sub>, I<sub>7</sub> 與 I<sub>8</sub> 這幾張圖，最後我們將這些圖作聯集，產生相關影像 (Relevant Images) 的集合，表示這些影像具有相同的語義。由這些紀錄我們可以產生影像間的關聯規則 (Association Rule)。在這裡我們所需要的關聯規則則是 I<sub>i</sub> → I<sub>j</sub> 的格式，也就是說當 I<sub>i</sub> 被選取時，I<sub>j</sub> 這張圖也很可能同時被選取，藉由此類查詢影像之關聯規則，我們可以縮短查詢的次數，進而更快速地找到使用者語意概念影像。

### 4. 結論與未來研究方向

在本論文中，我們提出了一套智慧型的語意概念式多媒體影像搜尋系統之架構，透過此系統，機器能學習並分析影像之特性。此系統雖然仍以影像內容為基礎，卻有別於一般傳統影像內容為基礎的搜尋系統。其目的不但兼顧了一般影像搜尋系統的特色，更兼具了人類語意的概念。為了達成此概念存取之目的，我們結合了影像內容與語意概念，並發展出以下幾項特色：

- I. 索引技術能準確地將影像叢集或分類，幫助系統更有效率地找到使用者想要的影像。
- II. 系統透過人工智慧的方法，能夠自我學習分類註解，使低階影像特徵能更切近人類語意概念。
- III. 互動式的搜尋方法透過較佳的資料探勘技術，使得系統能了解使用者的深層概念，更能準確且快速地回答使用者的問題。

在未來，隨著數位電視的普遍，本研究可實際應用於數位影像及視訊資料庫自動註解之環境，大幅降低人力需求及改善多媒體資料搜尋結果精確度之品質，並逐漸發展至家庭個人化隨選視訊系統。此外，本研究所提出之方法亦可應用於各種典藏機構中，我們除了設計具高準確性極高效能之技術元件外，也將同時考慮其使用者介面，希望亦能提供一個具高度自動化與親和力之介面。未來之研發成果將與國內相關之典藏機構及業界進行整合合作，預期將可提昇我國在多媒體內容之技術層次及國際化能力。

### 致謝

本研究由國科會數位典藏科技計畫所補助，計畫編號 NSC 95-2422-H-006 -001

### 參考文獻

- [1] Zaher Aghbari and Akifumi Makinouchi, "Semantic Approach to Image Database Classification and Retrieval," NII Journal, no. 7, September, 2003.
- [2] R. Agrawal, T. Imielinski, and A. Swami,

- “Mining Association Rules between Sets of Items in Large Databases,” Proc. of ACM SIGMOD, pages 207-216, May 1993.
- [3] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules,” Proc. 20<sup>th</sup> Very Large Databases (VLDB) Conference, pp 487-499, Santiago, Chile. 1994.
- [4] R. Agrawal, M. Mehta, J. Shafer, and R. Srikant, “The Quest Data Mining System,” Proc. 2<sup>nd</sup> Int. Conf. on Knowledge Discovery and Data Mining, 1996.
- [5] Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik, "Color- and Texture-based Image Segmentation Using the Expectation-Maximization Algorithm and Its Application to Content-Based Image Retrieval," Int. Conf. Computer Vision, Bombay, India, Jan 1998.
- [6] Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik, “Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying,” IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(8), 1026-1038, August 2002.
- [7] [11] S. F. Chang, W. Chen, H. J. Meng, H. Sundaram, and D. Zhong, “A Fully Automated Content-Based Video Search Engine Supporting Spatiotemporal Queries,” IEEE Transactions on Circuit and Systems for Video Technology, Vol. 8, No. 5, September 1998.
- [8] Chabane Djeraba, “Association and Content-Based Retrieval,” IEEE Trans. Knowl. Data Eng., 2003.
- [9] 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, Sep. 1995.
- [10] Jing Huang, Ravi Kumar, and Ramin Zabih, “An Automatic Hierarchical Image Classification Scheme,” ACM Multimedia, 1998.
- [11] J.Li, A. Najmi, and R.M.Gray, “Image classification by a two-dimension hidden Markov model,” IEEE Transaction on Signal Processing, vol. 48, no. 2, pp. 517-533, February 2000.
- [12] Ze-Nian Li, Osmar R. Zaiane, and Zinovi Tauber, “Illumination Invariance and Object Model in Content-Based Image and Video Retrieval,” Journal of Visual Communication and Image Representation, Academic Press. Vol. 10, No. 3, pp. 219-244, 1999.
- [13] Bing Liu, Wynne Hsu, Yiming Ma, "Integrating Classification and Association Rule Mining," Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98, full paper), New York, USA, 1998.
- [14] T.Ohashi, Z.Aghbari, A.Makinouchi, “Hill-Climbing Algorithm for Efficient Color-Based Image Segmentation,” IASTED Int’l Conf. On Signal Processing, Pattern Recognition, and Applications (SPPRA 2003), Rhodes, Greece, June, 2003.
- [15] J.R.Smith and S-F.Chang, “VisualSEEK: A fully automated content-based image query system,” ACM Multimedia Conf., Nov. 1996.
- [16] M.V. Srinivasan, S. Venkatesh, and R. Hosie, “Qualitative estimation of camera motion parameters from video sequences,” Pattern Recognition, Vol. 30, No 4, pp. 593-606, 1997.
- [17] M. Szummer and R.W. Picard, “Indoor-Outdoor Image Classification,” IEEE International Workshop on Content-based Access of Image and Video Databases, in conjunction with ICCV’98. Bombay, India, 1998.