

# 台灣二〇〇五年國際科學展覽會

科 別：電腦科學

作品名稱：灰階影像內容之檢索技術

得獎獎項：大會獎第二名

美國團隊正選代表:參加美國第 56 屆國際科技展

覽會

英特爾電腦科學獎:第一名

學 校：國立臺中第一高級中學

作 者：吳冠龍、尤孝庭

評語與建議事項：

作品之主要貢獻為能將影像加以作量化分析，然後利用分析之結果進行影像搜尋。研究動機，方法，步驟與原理均佳。若能將其作品與網路聯結成爲一功能強大之搜尋引擎更具功能性。另外亦建議可在提出更具智慧之比對影像演算法，來增加其程式之效能。

## 作者簡介

姓名：尤孝庭

興趣：程式語言、網頁設計、音樂

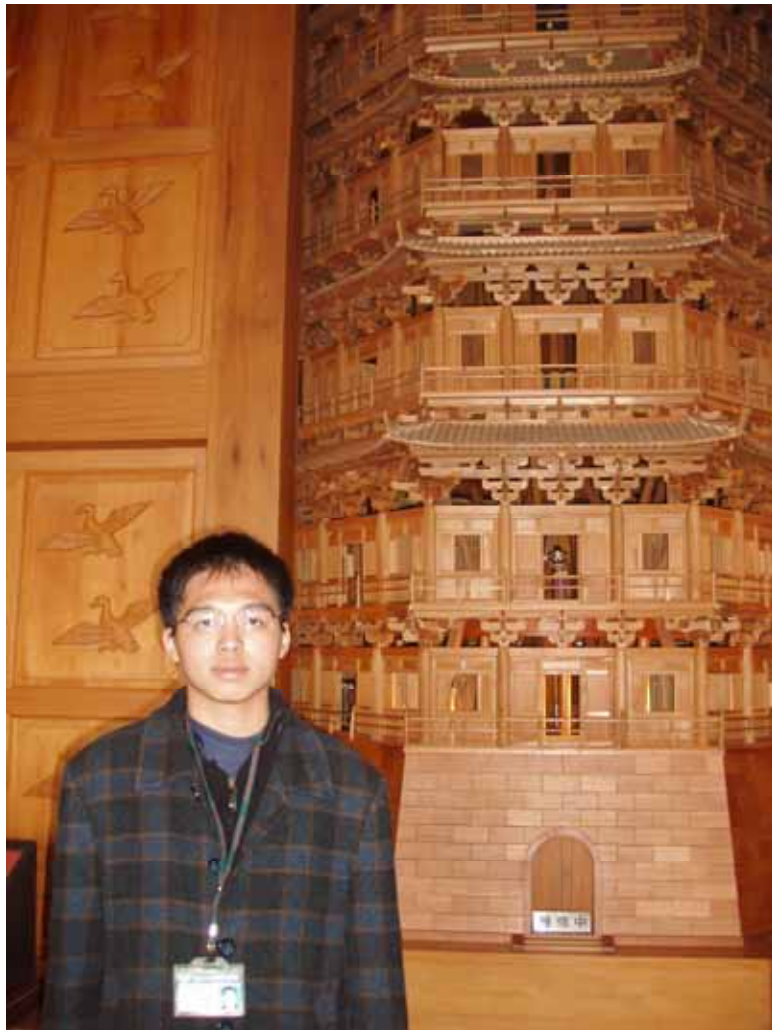
簡介：小時後幾乎把電腦當作玩具的我，對於這種工具已經非常的熟悉，幾乎成爲了我的第二生命。自從小四開始寫 C 語言之後，對於程式設計的熱誠更是發揮無疑。現在主要以撰寫網頁應用類的程式爲主。參加電研社和人才培育班，對於寫程式已經十分熟悉。在本回的實驗研究之中，在程式的最佳化和執行效率上做了很大的努力，也使得我在這次的學習中有很大的進步。



姓名：吳冠龍

興趣：關於電腦的所有事物、騎腳踏車、看書…

簡介：自從小學四年級擁有第一台個人電腦，我對於電腦的狂熱就不曾間斷。在中一中時因緣際會參加學校所舉辦的資訊科人才培育班，讓我的程式功力大幅增進，此次的研究也能順利的完成。在修習學業的同時，我也是國立自然科學博物館的高中導覽義工，在科博館執勤，除了使我對於自然科學的觀點更有廣度與深度，也在替觀眾講解的同時提昇自己的台風與說話技巧。



## 摘要

資訊時代的來臨，促使我們的社會型態大幅改變，無一不朝數位化的方向邁進。網際網路與資訊科技的快速發展，近年來影像資料庫和數位圖書館大量的成立，關於影像資料檢索的研究已漸漸成爲一門極爲重要的研究議題。在本研究報告中，我們提出一種植基於向量量化編碼法的灰階影像檢索技術。向量量化編碼法是一種極爲簡單的影像壓縮技術。我們應用這個編碼法，來萃取出灰階影像內容的特徵值。同時我們也計算出整張影像與影像中間位置的像素平均值，作爲日後檢索影像時過濾掉影像資料庫中不需要比對特徵值的依據。

我們所提出的方法能有效地萃取影像內容的特徵值並且讓使用者可以快速且正確地查詢到所需要的影像。當影像資料庫中存在與查詢影像完全相同的影像時，我們的檢索技術都能在第一時間第一順位檢索出這張影像。即便影像資料庫中不存在與查詢影像完全相同的影像時，我們的檢索技術平均 74.3%也能在第一時間前五個順位檢索出最相似的影像。

## ABSTRACT

With the coming of the information age, our sociological system change extensively, and everything has moves toward digitization. Due to the rapid development of Internet, information technology, the rapid growth of image databases and digital libraries recently, the related researches of image retrieval have become a very important issue. In this memoir, we propose an image retrieval scheme based on the vector quantization to retrieve similar images from the image database according to the pre-collected image features. Vector quantization is a very simple image compression scheme. We have applied vector quantization to extract features from grayscale image. In order to speed up the retrieval process, we also calculated the mean pixel values of whole images and the central part of the image to filter the images, which are significantly dissimilar to the query image.

The experimental results show that our proposed approaches can effectively extract features from the image and enable users to retrieve images from image database quickly and accurately. When images stored in image database match query image, the proposed scheme can instantly retrieve the stored image at the first rank. Even though images stored in image database query image exactly, the proposed scheme can instantly retrieve the stored image over 74.3% at the first five ranks.

# 壹、前言

## 一、研究動機

近年來，隨著網際網路與電腦科技快速地發展，政府、公司企業、醫院、博物館、圖書館等公家或私人機構，經常需要將大量的影像資料數位化，一來可節省影像儲藏的空間，二來也便於日後的維護與管理。然而面對龐大數量的影像資料，如何建構一個有效率處理影像的機制，並且設計一種快速又簡單檢索影像的方法，已經成為當前最受矚目的研究議題。

在早期影像檢索系統大都依賴人工處理影像資料的編號、名稱與註解等，之後才運用這些標記資料當作關鍵字來檢索影像。這種檢索影像的方式，無法切合實際的使用需求，原因有二：

1. 以人工處理影像資料的註記相當費時費力，而且每個人對影像資訊的解讀不盡相同。
2. 文字無法精確地描述影像資料。

目前市面上已經出現好幾種影像檢索系統，廣泛地應用在博物館文物典藏、醫療影像、地理影像、人臉指紋辨識、數位圖書館等各種領域。因此近年來發展的重點是如何從影像資料的本身擷取出適當的特徵，如色彩、紋理、形狀與空間關係等來描述此影像，進而運用這些特徵作為日後檢索影像的依據，這便是目前市場上漸漸發展出的以影像本身的資料擷取出特徵值的影像檢索系統(Image Retrieval System)。

## 二、研究目的

本研究的主要目的，在於研究並設計出一種有效處理影像本身資料特徵的演算法與快速查詢所需要影像的檢索技術。

## 三、研究問題

面對龐大數量的影像資料庫，如何幫助使用者非常便利又快速地查詢到所需的影像資料，並且瞭解當前市場上所發展以影像內容為基礎的查詢系統，進而設計出處理影像資料特徵的方法和檢索影像的技術，使其更具正確性與快速性。為達成上述的目標，本研究的問題計有下列三個項目：

1. 設計一個有效處理影像本身資料特徵的演算法。
2. 如何設計一個使用較小記憶體空間儲存影像本身資料特徵值的方法。
3. 如何設計一種加速影像檢索速度及提昇檢索正確率的方法。

## 四、研究範圍與限制

1. 本研究所提到的「影像」，主要是指一般二度空間，靜態的圖像資料，並不包含動態的視訊資料(Video)。實驗的影像皆為 512x512 的灰階影像，並沒有包含彩色影像。
2. 本研究所提到的影像檢索系統，主要是限定在使用以影像本身資料擷取出的特徵值作為檢索所需要影像方式的系統，並沒有包含使用以文字作為關鍵字來進行影像檢索的系統。

## 貳、研究過程－文獻探討

### 一、影像檢索系統

影像檢索系統(Image Retrieval System)，主要是以影像本身資料擷取的特徵值或利用介面的輔助工具所繪出的圖形當作輸入資料來進行檢索影像的一種系統。這種檢索系統是藉著擷取影像資料中某些特徵值或向量，與資料庫中存著事先經由某種演算法所得到的特徵值來進行相似性的比對，然後將最相似的數張影像以及一些相關資訊回傳給檢索者。

影像資訊的檢索技術目前發展出四種查詢介面的方法，茲簡要說明如下：

#### 1. 文字查詢(Query by Text)

這種查詢是以關鍵字來查詢影像，用文字來表達影像是高階的表達方式，可以描述影像的語意(Semantic)部分。提供以關鍵字為主的介面(Keyword-based Interface)，使用者不一定要對影像領域有足夠的了解。

#### 2. 特徵查詢(Query by Feature)：

是用描繪的方式查詢，這種方式具有較友善的查詢介面，例如以顏色分布或調色盤選擇顏色、在畫布畫出顏色分布、形狀，但是通常精確度較低，因為使用者用繪圖的方式不夠精確。特徵查詢(Query by Feature)是提供讓使用者點選特徵的介面，從固定的特徵集裡，拼出一個要查詢的影像出來。

#### 3. 範例圖片查詢(Query by Image Example)：

是指系統首先隨機顯示一些樣本影像，使用者從中選擇與目標相似的影像，作為查詢影像，系統再根據這些查詢影像，找出相似的影像。使用者也可透過瀏覽的方式或由查詢結果，從中選擇目標。

#### 4. 瀏覽(Browsing)：

是指系統先經過叢聚(Clustering)技術，依據影像的相似度將影像分類。使用者因此根據分類，瀏覽找尋影像。

### 二、傳統以文字檢索影像的缺點

當我們比較以文字檢索影像的方式與傳統資訊檢索(Information Retrieval)的技術，可以說兩者並沒有什麼不同。差別之處只是在於最後取得的前者是影像而後者是文字。

影像資料建構的過程中，首先建構者自己事先依據自身的知識分析影像，決定使用何種方法的索引或分類，再將影像鍵入資料庫中。然後讓影像資料庫的使用者利用文字或分類來查詢，擷取所需要的影像。相對於其他種類的數位資料，影像資料最大的不同點，在於其非結構性的特色。由於影像資料的高度非結構性的特質，使得影像資料之間的關聯性低，因此對於系統建構者而言，很難制定一個有效的標準來找到影像間的關聯性，使得影像資料的分類組織變得非常困難。傳統這種以文字檢索影像資料庫的方式有下列幾項缺點：

1. 一張影像都需要經過人工標註分類的程序。當影像資料量日趨龐大的時候。這些標註分類的人工成本將成為龐大的負擔。
2. 標註分類的標準完全靠資料建構者本身的知識作判斷，但不同的建構者對同一張影像可能會有不同的詮釋的方法，所以很難達成標註分類的一致性。
3. 資料庫的使用者與資料建構者對同一張影像的認知可能也有某種程度的差異，因此會影

響到資料檢索的效率。

4. 使用者無法以影像內容所呈現的特徵，如顏色、紋路、形狀、空間關係等，來作為查詢影像的條件。

還有如果影像全由人工加以註解以及使用者查詢時依照自己的語意作為輸入條件，由於人類個人主觀的因素，往往會造成極大的出入。基於上述種種的缺點，造就了能夠自動編碼、分類以及自動檢索技術的需求因應而生，這種技術就是所謂的以影像本身資料擷取各種特徵值的影像檢索技術。

### 三、現有國內外影像檢索系統簡介

現有國內外影像檢索系統已經發展較為成熟的有國內的台灣魚類檢索系統、台灣蝴蝶外觀檢索系統；國外的有美國哥倫比亞大學的 VisualSEEK、美國加州大學柏克萊分校的 Blobworld、美國哥倫比亞大學的 WebSEEK、美國波士頓大學的 ImageRover、美國麻省理工學院的 PhotoBook...等。以下針對這些較具代表性的影像檢索系統作簡要式的介紹：

#### 1. 國內的台灣魚類檢索系統

台灣魚類檢索系統[1]提供的檢索方式可分為兩大類：第一大類是經由使用者從系統本身內建的魚類資料庫中挑選出一張所需影像後，分析其頭身比例、顏色、長寬比例等等象徵性資料，再從資料庫選出符合或相似以上資料之影像。第二大類則是讓使用者自行上傳所需進行比對的影像，再依照上述比對步驟篩選出相似影像。在搜尋結果中，系統會將每張影像對於原圖的精確率(相似度)以百分長條圖表示，具有相當程度以上的效率。

#### 2. 國內的台灣蝴蝶外觀檢索系統

台灣蝴蝶外觀檢索系統[2]是自民國 87 年 8 月 8 日至 88 年 8 月，國立自然科學博物館與暨南大學共同合作開發的國科會數位博物館先導計畫---蝴蝶生態面面觀。蝴蝶檢索系統以 QBF/QBE 查詢模式，讓使用者以交叉查詢的互動模式，按照蝴蝶的特徵和相似度來進行查詢。QBF(Query by feature)特徵查詢，是讓使用者指定不同的特徵做查詢，藉此解決使用者認知不同的問題；QBE(Query by example)相似查詢，讓使用者指定相似的蝴蝶進行查詢，在過程中系統根據查詢的結果提供使用者有用的資訊，使用者根據這些資訊再做更一步精確的查詢。

網站提供兩種查詢方式，分別是關鍵字查詢和外觀查詢兩種，讓使用者除了使用一般常用的關鍵字檢索之外，還可以用蝴蝶本身的特徵來進行查詢，共有 18 種顏色特徵、7 個形狀特徵和 17 個花紋特徵，並且可以進一步獲得蝴蝶的詳細資料。如圖一所示：

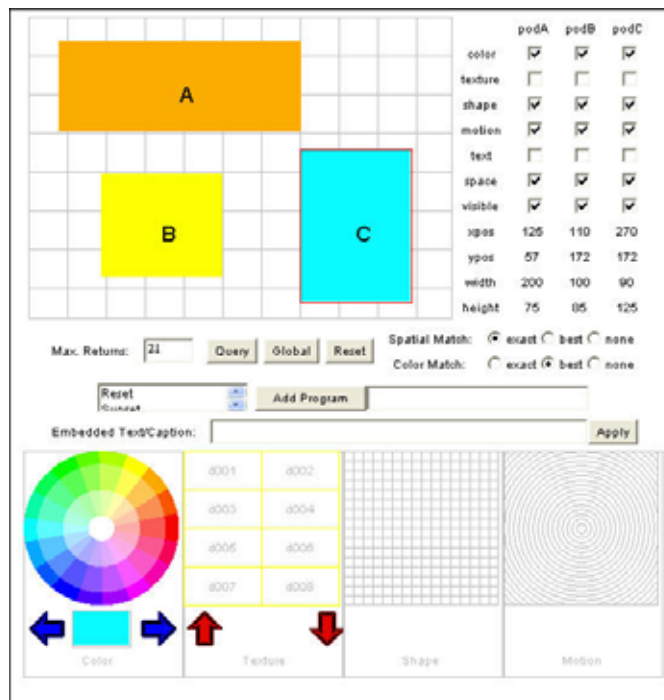




圖一 台灣蝴蝶影像檢索介面

3. 美國哥倫比亞大學的 VisualSEEK

VisualSEEK[3]是由哥倫比亞的新媒體技術中心(Image and ATV LAB of Columbia University)所開發的，VisualSEEK 是一個用來查詢影像和視訊資料的網路工具，讓使用者能夠對全球資訊網上的影像及視訊來進行搜尋。VisualSEEK 提供檢索的影像特徵包括顏色、形狀、紋理和空間資訊，如圖二所示。其主要特色是將物件之間的空間關係亦列入查詢時的考量，以提昇搜尋的能力，使用者可以結合影像的特徵及物件的空間關係來進行更精準的查詢。此外系統還有提供 12,000 張彩色影像的資料庫，並讓使用者利用影像資料的網址 (URL)來進行查詢。

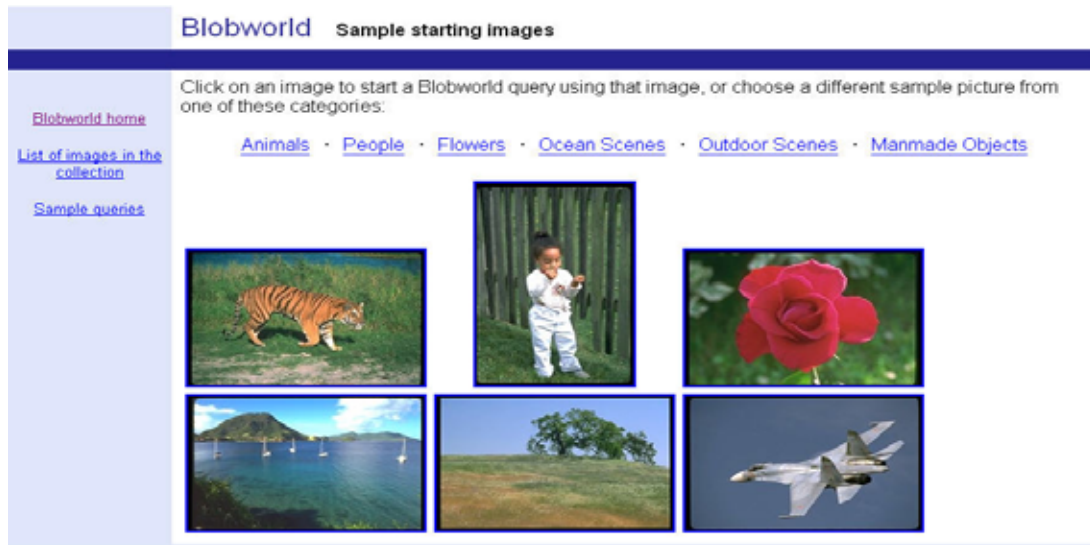


圖二 VisualSEEK 檢索介面

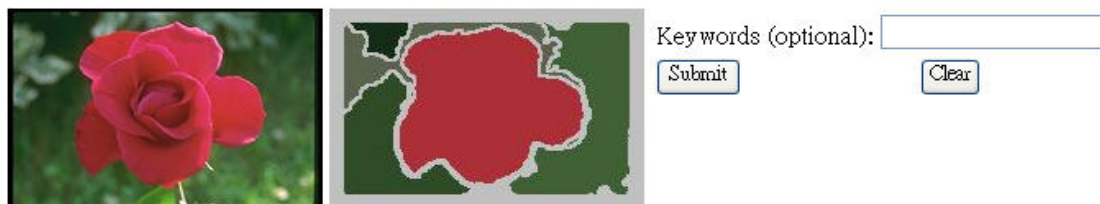
4. 美國加州大學柏克萊分校的 Blobworld

Blobworld[4]是加州大學柏克萊分校所開發的影像資料庫專案，系統提供四種檢索方式，包括顏色、外型、紋理和範例圖片，其進行的方式是先讓使用者依照圖像的主題來選擇範例圖片，當使用者選擇了某個範例圖片之後，系統接著會顯示出代表該圖像的氣泡圖，使用者從其中選擇接近查詢目標的外型，然後再指定氣泡的權值，包括色彩、紋理、位置、外型及大小的權值。最後系統根據這些氣泡權值的設定來進行檢索比對，從資料庫中檢索出最相近的影像資料。Blobworld 檢索介面的操作步驟如以下圖三所示：

Step 1 先選擇範例圖形



Step 2 指定代表的氣泡進行檢索



Step 3 選擇色彩、紋理、位置、外型及大小的權值

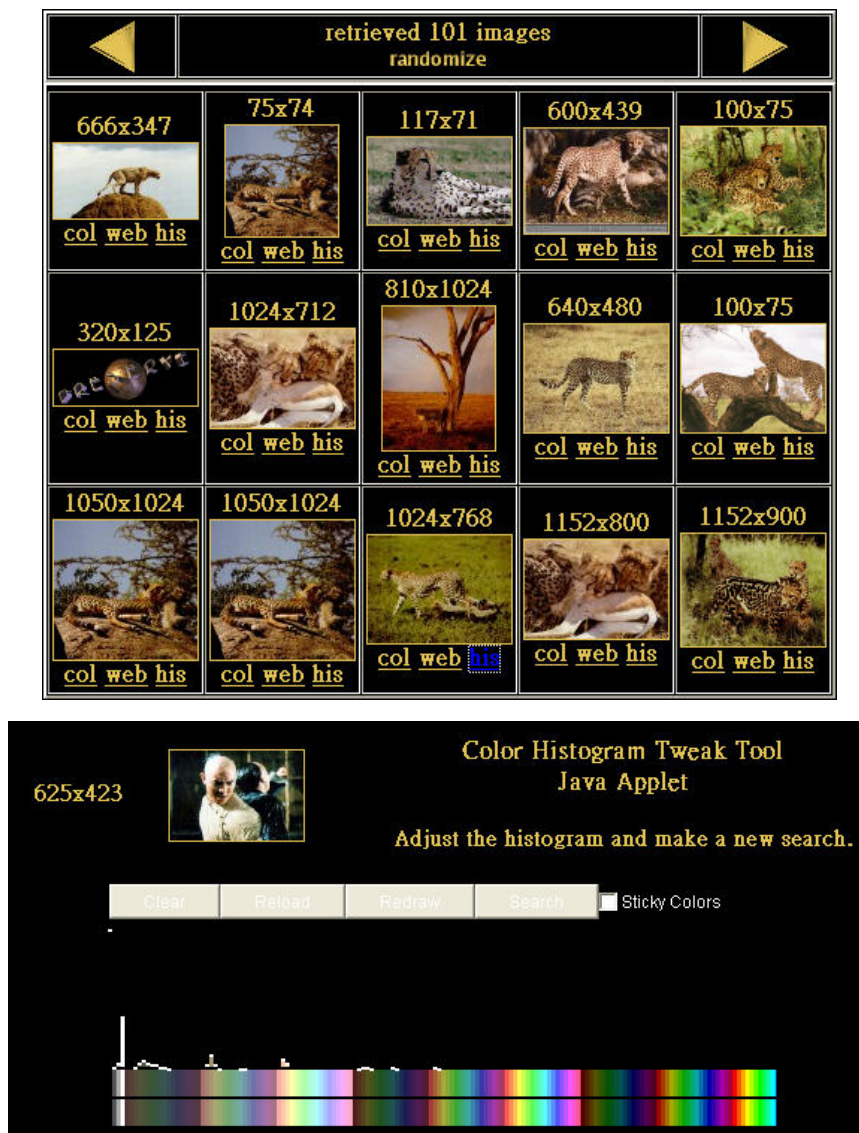
How important is the selected region?	<input type="radio"/> Not	<input type="radio"/> Somewhat	<input checked="" type="radio"/> Very
How important are the features of this region?			
Color	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Texture	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Location	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Shape/Size	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

How important is the background (everything outside the region)?	<input type="radio"/> Not	<input checked="" type="radio"/> Somewhat	<input type="radio"/> Very
--	---------------------------	---	----------------------------

圖三 Blobworld 檢索介面

## 5. 美國哥倫比亞大學的 WebSEEK

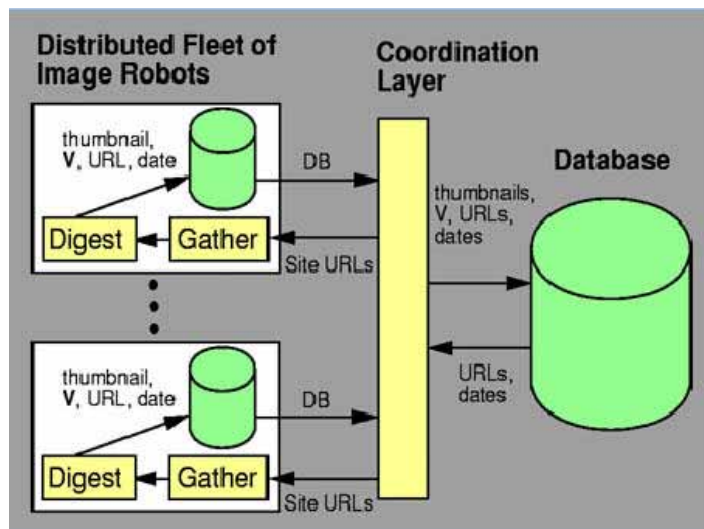
WebSEEK[5]是由哥倫比亞大學開發的一種實驗性系統。採用代理自動搜索可視信息並對其進行分類，目前已分類的圖像有 66 萬多幅，形成了一個極富創新性的圖像目錄，主題分類是它的主要優點。WebSEEK的檢索途徑有三種，一是使用關鍵詞進行自由全文檢索，但不支持短語檢索；二是利用不同的類目等級進行主題瀏覽；三是在檢索出圖像的基礎上利用其可視屬性進一步檢索。其檢索結果顯示為沒有文本的簡圖及圖像大小。圖四即為美國哥倫比亞大學的WebSEEK的檢索介面。



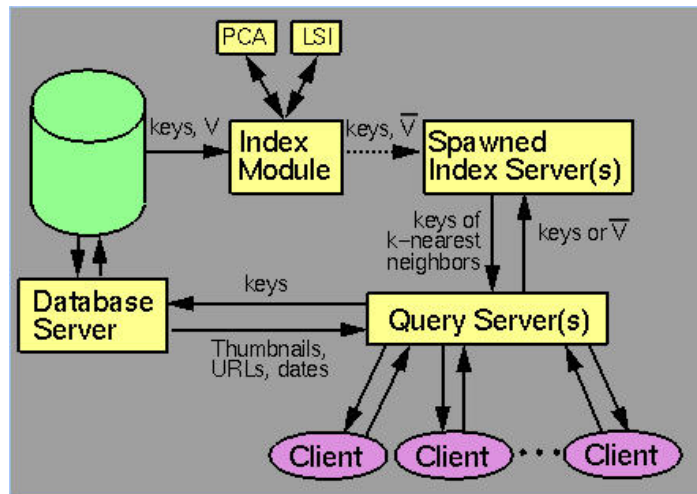
圖四 WebSEEK 檢索介面

## 6. 美國波士頓大學的 ImageRover

ImageRover[6]影像資料庫可送出軟體機器人(圖五)遊走於網際網路，將所找到的影像傳回伺服器並作成索引。特徵以色彩和紋理為主，其中尤重紋理的處理，還另將紋理做方向性、週期性與亂度的分析。其他的特色還有關聯性回饋 (Relevance Feedback )，就是由使用者從伺服器所找到的影像中選取數張作為關聯影像，伺服器則根據這幾張關聯影像將比對結果更細緻化。圖六即為美國波士頓大學的 ImageRover 的檢索介面。



圖五 軟體機器人



圖六 ImageRover 的檢索介面

## 7. 美國麻省理工學院的 PhotoBook

PhotoBook[7]是由麻省理工學院所開發的影像檢索系統。提供顏色、形狀和紋理等影像特徵讓使用這進行查詢，使用這可依據學科領域的不同來選擇不同的特徵和比對之演算法，並提供 FourEyes 互動學習代理(Interactive Learning Agent)作為使用者介面。該系統可依據使用者的使用模式，並提供 Texture Modeling、Face Recognition、Shape Matching、Brain Matching 等四個範例資料庫供使用者使用。如圖七所示：





圖七 PhotoBook 檢索介面

歸納上述各種影像內容檢索系統，經過綜合比較分析，可以歸納出影像特徵檢索的方式共有七種，分別是關鍵字(Keyword)查詢、顏色(Color)查詢、紋理(Texture)查詢、形狀(Shape)查詢、範例(Example)查詢、影像空間關係(Spatial relationship)和權重數值(Weighting)。其中以顏色查詢、紋理查詢、形狀查詢這三種檢索方式最常使用到，每一種影像內容檢索系統都有提供這些檢索功能。

接著我們比較國內外現有的影像內容檢索系統，我們嘗試將它們影像檢索方式的運用情形，歸納分析作成表一。

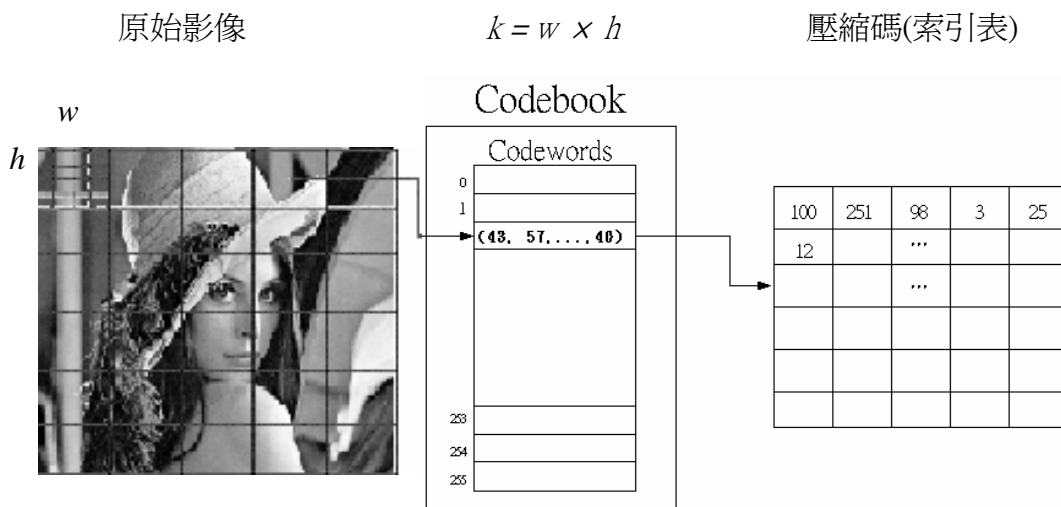
表一 國內外各種影像檢索系統—特徵檢索方式比較

影像檢索方式 影像檢索系統	關鍵字 (Keyword)	顏色查詢 (Color)	紋理查詢 (Texture)	形狀查詢 (Shape)	範例查詢 (Example)	影像空間關係 (Spatial relationship)	權重 (Weighting)
魚類檢索系統		✓	✓	✓	✓		
蝴蝶檢索系統	✓	✓	✓	✓	✓		
VisualSEEK		✓	✓	✓	✓	✓	✓
Blobworld		✓	✓	✓	✓		✓
WebSEEK	✓	✓			✓		✓
ImageRover		✓	✓		✓	✓	✓
PhotoBook		✓	✓	✓	✓		

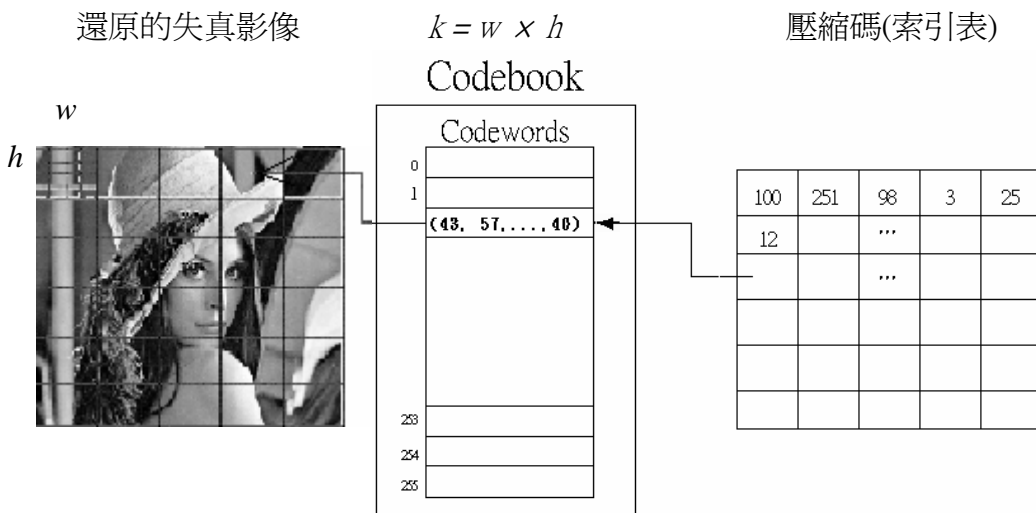
註：✓代表檢索系統有此項檢索方式

#### 四、向量量化編碼法

向量量化編碼法(Vector Quantization, 簡稱 VQ)[8], 顧名思義是將一向量值以一個量化值來取而代之, 進而達到資料壓縮的目的。向量編碼法應用在數位影像中作法可以想成像是在貼磁磚一樣, 如圖八, 其中係將原始影像切割成一塊塊大小相等的區塊(磁磚), 每個區塊包含了許多的像素且通常將這些像素由左至右、由上而下地串成一個向量。每塊磁磚都有其花色(由像素值所形成的向量), 編碼時只要到事先已收集完成, 且具有代表性的花色集合, 稱之為編碼簿(codebook)中找出與其最相似的花色(編碼字, codeword)出來, 再用此一相似花色位在編碼簿中的索引值取代此一整塊資料即完成壓縮。解壓縮時, 如圖九所示只需利用該索引值到編碼簿中取出該區塊的花色資料還原即可。



圖八 向量量化編碼法編碼示意圖



圖九 向量量化編碼法解碼示意圖

## 叁、研究過程－研究方法與設計

### 一、研究流程與架構

本研究主要先瞭解以影像本身資料為主的檢索系統的發展情形，進而設計一種有效萃取影像資料特徵的演算法，並且研發一種快速查詢影像資料的檢索技術。研究的進行分成初期、中期和後期，總計花費九個月的時間完成。以下簡述每個時期所進行的研究方法與步驟：

#### 1. 初期 (93.03-93.05)：確定研究主題與文獻探討

初期主要分析、閱讀目前有關以影像本身資料為主的檢索技術、論文、期刊、圖書及網路資料等，藉此過程引發研究的動機，進而確定研究主題、範圍與方向。接著蒐集與研究主題相關的文獻資料，內容包括影像資料庫的使用與開發、影像檢索系統與技術、數位影像處理技術、資料壓縮、電子影像技術等，從中篩選出與主題相關之文獻，進行閱讀與分析，並撰寫文獻分析。

#### 2. 中期 (93.06-93.09)：研究設計、資料蒐集

在中期的階段主要進行研究設計與資料蒐集。首先設計影像搜尋模式，並依據這個模式設計一個有效萃取影像資料特徵的演算法。進而將萃取出來的特徵值存放在一個特徵值比對資料庫內。接著將查詢影像，依相同的演算法萃取出特徵值，再與特徵值比對資料庫進行比對，最後將搜尋結果的資料蒐集起來以便作為日後修改整個搜尋演算法的參考。

#### 3. 後期 (93.10-93.12)：資料分析與整理、結論與建議

將中期蒐集的實驗資料加以歸納、分析、討論，針對影像搜尋演算法的缺點加以改進，目的是為了使影像檢索系統達到最佳化的情形。最後根據實驗結果將本研究寫成研究報告、結論與建議，作為未來影像檢索系統開發者在設計或改良系統時的參考依據。

### 二、萃取灰階影像本身資料特徵值的方法

在我們的實驗中所使用的影像稱為 Grayscale(灰階影像)亦稱為 Graylevel，通常用 8 bit 來記錄顏色，因此可以記錄  $2^8 = 256$  種色階；在這種類型的影像中，除了純黑(灰階值為 0)與純白(灰階值為 255)之外的 254 種顏色(灰階值介於 1~254 之間)，都被用來表示不同程度的灰色，因此稱為灰階影像[9]。

在本研究中，我們使用 C 語言(gcc 編譯器)來撰寫與整合此研究利用到的演算法，使得我們設計出的程式能準確且快速的處理大量實驗影像。下列是本次研究所使用的研究設備及實驗資料簡介：

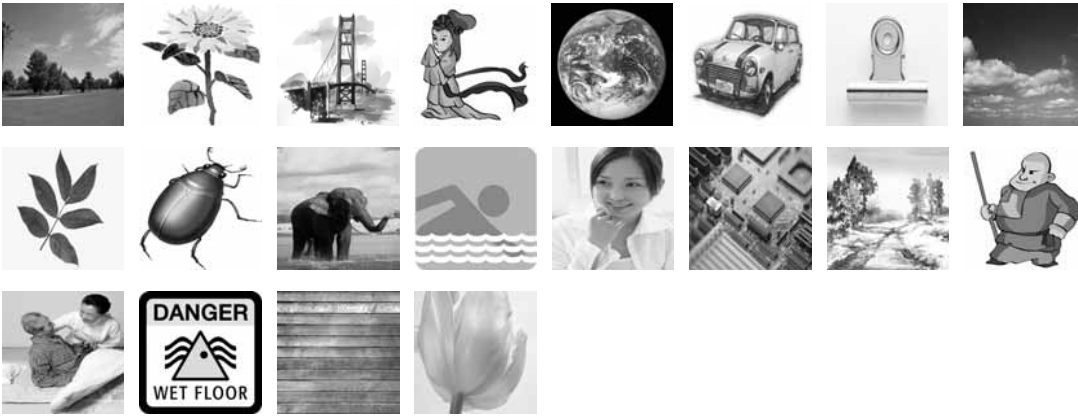
研究設備：

CPU：Intel Pentium 4 3.0GHz (HT 技術)

RAM：512 MB

實驗資料:共 3100 張，每個類群各有 100-300 張影像以進行實驗。

接下來我們將對本次實驗過程以淺顯易懂的方式作詳細的敘述。圖十為 3100 張影像 20 個類群中的代表影像。



圖十 3100 張影像 20 個類群中的代表影像

## 1. 訓練一本含有 256 個編碼字的編碼簿

在萃取影像特徵值之前，我們先要訓練出一本具有 256 個編碼字的編碼簿[10]。其步驟如下：

步驟 1: 選出五張具代表性的影像。將每一張影像切割成不重疊  $4 \times 4$  的區塊。每張影像大小為  $512 \times 512$ ，其中每個不重疊的區塊，我們可視之為一個 16 維的向量。

步驟 2: 從這些眾多的區塊中隨意挑出 256 個區塊當作初始編碼簿。

步驟 3: 以這 256 個初始向量為重心，將其餘區塊分別歸群到這 256 個重心，也就是每個區塊去尋找與其距離最近的的重心，形成 256 個群。

步驟 4: 重新計算這 256 個群的重心，得到新的一本編碼簿。如此重複步驟 3 及步驟 4，直到每個群重心的變化不再很大時(意謂收斂)，即完成了編碼簿的訓練工作。

## 2. 標準 VQ 編碼法-Full Search

萃取影像特徵值乃是應用向量量化編碼法 Full Search 的技巧。將使用編碼字次數最多的前 10 個索引值當作特徵值。萃取特徵值的步驟如下：

步驟 1: 針對每一張輸入的灰階影像先切割成一堆不重疊  $4 \times 4$  的影像區塊。

步驟 2: 接著每一個影像區塊 16 維的向量一一算出與每個編碼字歐基里德的距離。

步驟 3: 找出距離最小的編碼字的索引值(index)。

步驟 4: 計算出所有影像區塊使用索引值次數最多的前十名，其 10 個索引值依序成為每張輸入影像的特徵值。

步驟 3 及 4 實際程式運作的程式片段如下：

```

for(i=0;i<256;i++)
{
    // 利用陣列作出平方的效果，以便計算歐幾里德距離
    squar[i]=(unsigned long)(i*i);
}
for(i=0;i<vectornum;i++)
{
    found=0;
    min=21474847;
    for(j=0;j<codenum;j++)

```



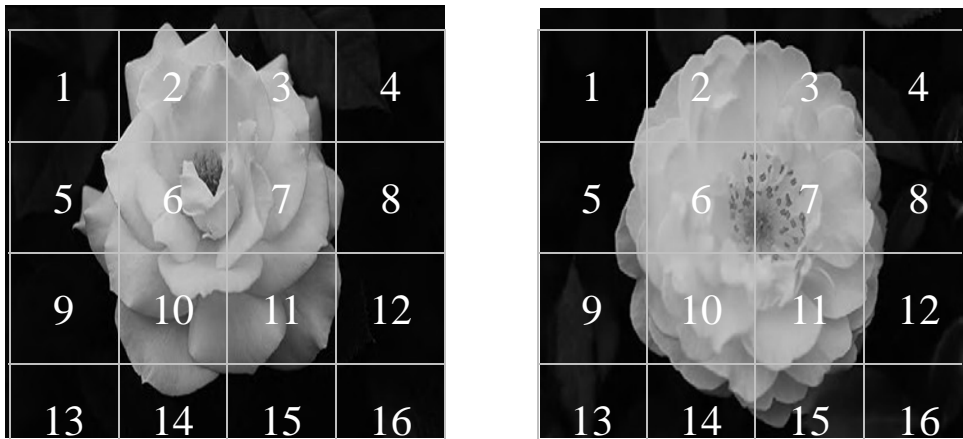
```

{
  dist=0;
  for(k=0;k<vectorsize;k++)
  {
    // 計算每個編碼字歐基里德的距離
    dist+=suar[abs(vectors[i][k]-codebook[j][k])];
  }
  if ( dist<min )
  {
    // 找到了距離最小的編碼字，寫入索引值 index
    min=dist;
    index=j;
  }
}

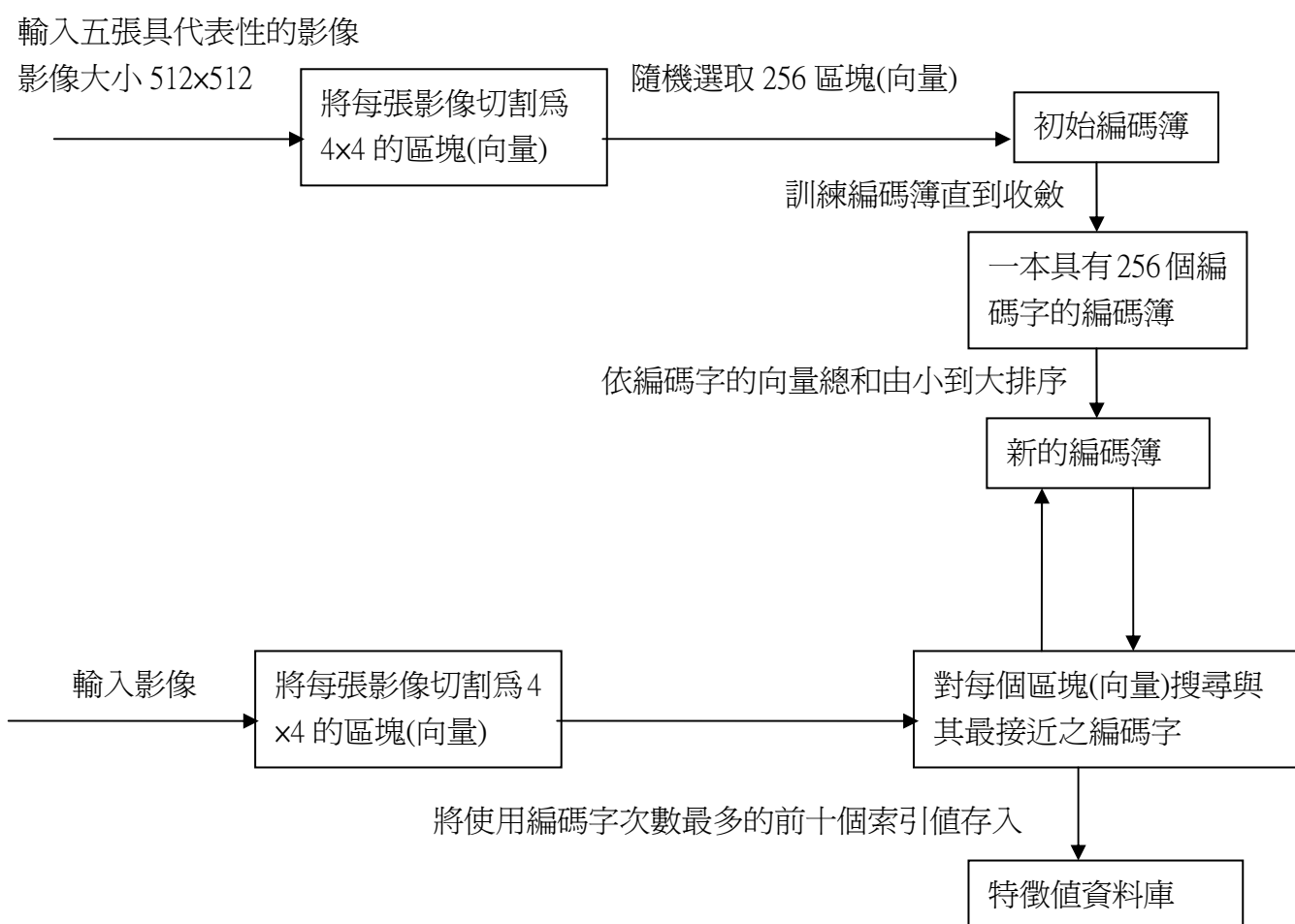
```

步驟 5: 計算出整張影像的像素平均值與中間位置影像的像素平均值。

一般而言，當我們拍攝一張影像時總是將拍攝的主要對象放在中間的位置。在我們實驗中所使用的灰階影像，我們將它等分成 16 個方格。依序由左至右、由上至下給他編號為 1、2、3...16，如圖十一所示。編號 6、7、10、11 的位置便是本研究報告中所定義的中間位置。從計算影像資料庫中每張影像的像素平均值與中間位置影像的像素平均值的過程中，我們發現到一個有趣的現象。兩張相似性很大的影像，它們的整張影像的像素平均值與中間位置影像的像素平均值的差距大多是介於 16 至 20 之間，這也是我們在進行檢索影像實驗時以 16 至 20 之間的數當作門檻值來過濾掉不需要比對特徵值的影像，以節省檢索影像的時間。圖十二是萃取灰階影像特徵值的流程。



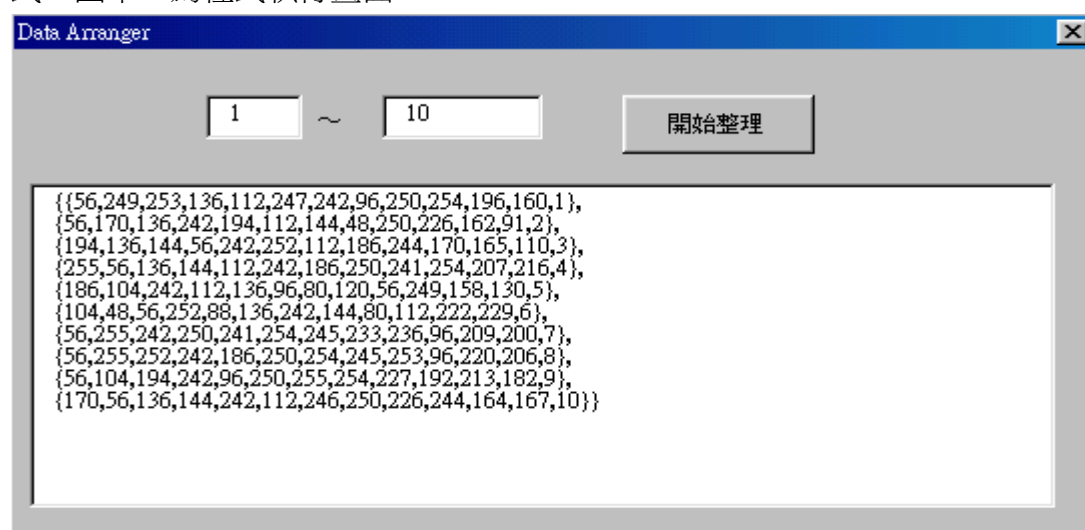
圖十一 編號 6、7、10、11 的位置便是本論文中所定義的中間位置



圖十二 萃取灰階影像特徵值的流程

### 3. 利用 Microsoft Excel 巨集程式將萃取出來的特徵值資料轉換成特徵值資料庫所需的陣列形式

我們將程式資料轉換為陣列形式，因為程式在檔案處理的撰寫上出現困難，所以不使用 C 語言程式。因此我們使用 Microsoft Excel 將資料匯入之後進行處理，轉換為我們需要的陣列形式。圖十三為程式執行畫面



圖十三 巨集程式執行畫面

### (三)檢索灰階影像的方法

當進行影像檢索時，首先計算查詢影像與影像資料庫中的影像其整張影像的像素平均值與中間位置影像的像素平均值，如果兩者差距的絕對值皆小於等於 20 者，才進行特徵值比對的程序。檢索影像的步驟如圖十四，簡要說明如下：

步驟 1:將查詢影像依照萃取輸入影像內容特徵值的方法，找出使用編碼字次數最多前 10 個索引值當作特徵值。

步驟 2:計算出查詢影像的像素平均值與中間位置影像的像素平均值。

步驟 3:計算查詢影像與影像資料庫的影像兩者之間整張影像像素平均值與中間位置影像像素平均值差距絕對值皆小於等於 20 者，才進行特徵值相似度比對的程序。

步驟 4:進行特徵值相似度比對是依照特徵值比對加權計分模型來計算其得分，得分愈高者表示相似度愈大。表二是我們依據每個特徵值使用編碼字次數的比例所設計出來的加權計分模型。

步驟 3 及 4 實際的程式設計如下：

特徵值資料庫陣列格式介紹： $a[3][10]=\{27,159,155,139,245,151,3,94,28,205,153,154,1\}$

a 放置的是所有的特徵值平均值資料，為兩維陣列，放置每一張圖片的特徵值和平均值資料。第一維的項目數即為圖片總數，每項包含 13 個項目，前 10 項為內容特徵值，第 11 項為整張影像像素平均值，第 12 項為中間位置像素平均值，第 13 項為圖片的編號。

(例如： $a[3][10]$  就是  $a[3]$  的整張像素平均值。第 11 項的編號是 153)

check 放置的是欲比較的特徵值和平均值資料，其中包含 13 個項目，如前所提。)

```
for(h=0;h<amount;h++)
{
// 整張影像像素平均值與中間位置影像像素平均值差距的絕對值小於等於 20
if(abs(check[10]-a[h][10])<=20 && abs(check[11]-a[h][11])<=20)
{
for(i=0;i<10;i++)
{
for(j=0;j<10;j++)
{
if(check[i]==a[h][j])
{
// 依照特徵值比對加權計分模型計算得分
if (i>j) {count[h]+=100-10*(9-j)+110*(9-i);}
if (i<j) {count[h]+=100-10*(9-i)+110*(9-j);}
}
```

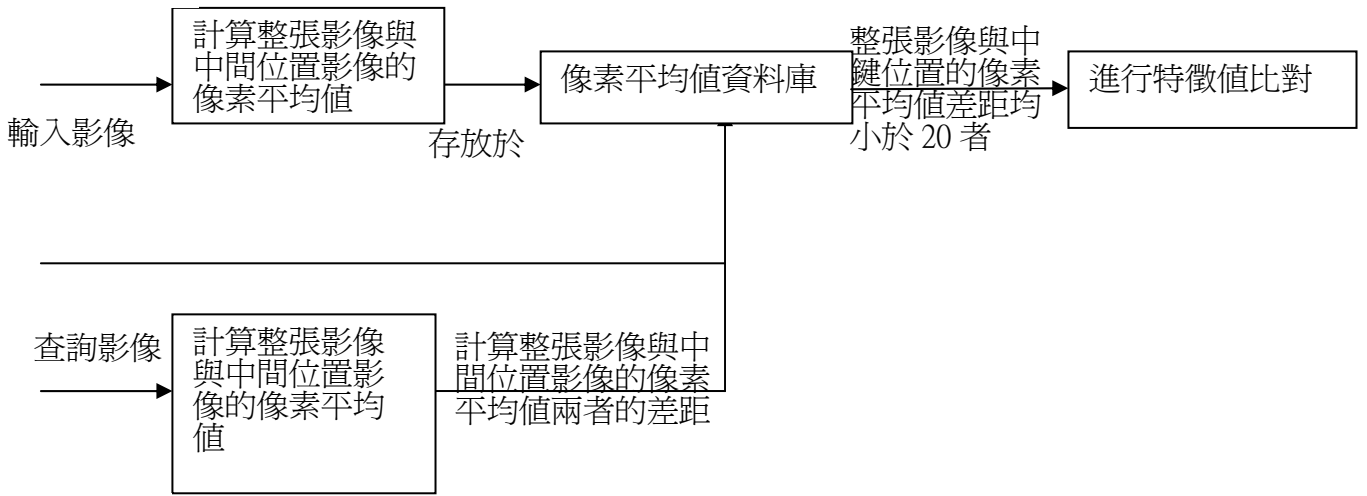
```

    if (i==j) {count[h]+=2000-200*i;}
  }
}
}
b[h] = a[h][12];
c[h] = count[h];
}
}

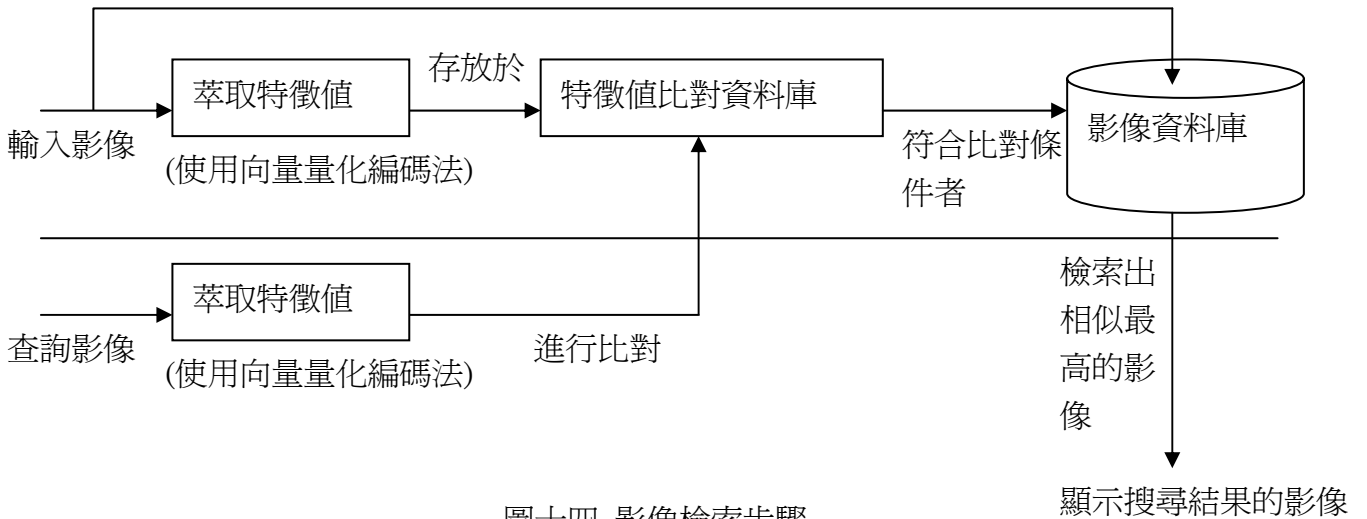
```

步驟 5:按照使用者需要將相似度最大的影像檢索出來。

Step 1 :



Step 2 :



圖十四 影像檢索步驟

表二 特徵值比對加權計分模型

資料庫影像 查詢影像	第 1 特徵值	第 2 特徵值	第 3 特徵值	第 4 特徵值	第 5 特徵值	第 6 特徵值	第 7 特徵值	第 8 特徵值	第 9 特徵值	第 10 特徵值
第 1 特徵值	2000	1880	1760	1640	1520	1400	1280	1160	1040	920
第 2 特徵值	1880	1800	1680	1560	1440	1320	1200	1080	960	840
第 3 特徵值	1760	1680	1600	1480	1360	1240	1120	1000	880	760
第 4 特徵值	1640	1560	1480	1400	1280	1160	1040	920	800	680
第 5 特徵值	1520	1440	1360	1280	1200	1080	960	840	720	600
第 6 特徵值	1400	1320	1240	1160	1080	1000	880	760	640	520
第 7 特徵值	1280	1200	1120	1040	960	880	800	680	560	440
第 8 特徵值	1160	1080	1000	920	840	760	680	600	480	360
第 9 特徵值	1040	960	880	800	720	640	560	480	400	280
第 10 特徵值	920	840	760	680	600	520	440	360	280	200

## 肆、研究結果與分析

我們利用設計出來的程式和處理流程，經由多次的實驗和偵錯後，實驗的結果所得數據如表三所示：

表三 檢索系統對於各種類群的檢索成功率

類群	檢索出同一張張數百分率	前一張檢索出本類群張數百分率	前二張檢索出本類群張數百分率	前三張檢索出本類群張數百分率	前四張檢索出本類群張數百分率	前五張檢索出本類群張數百分率
多種材質圖	93.0%	40.0%	58.0%	66.0%	69.0%	72.0%
標誌圖	97.0%	79.0%	80.0%	80.0%	81.0%	81.0%
女孩生活圖	100.0%	32.0%	47.0%	65.0%	69.0%	76.0%
中國神話人物漫畫	100.0%	48.0%	67.7%	79.7%	84.7%	89.3%
水彩繪圖	100.0%	26.0%	38.0%	51.0%	62.0%	71.0%
水彩花卉	100.0%	37.0%	54.0%	66.0%	76.0%	82.0%
四季大自然景象	100.0%	26.0%	43.0%	53.5%	63.0%	67.0%
真實花卉	100.0%	28.0%	39.0%	51.0%	59.5%	62.5%
老人寫真&家庭一瞥	99.5%	32.5%	43.0%	55.0%	63.5%	70.0%
中國民俗漫畫	100.0%	59.3%	76.0%	83.7%	89.0%	91.7%
大自然水彩畫	100.0%	40.0%	54.0%	59.0%	67.0%	70.0%
資訊電科技	99.0%	33.5%	49.5%	56.5%	63.0%	70.5%
柔和標語	99.5%	83.0%	85.0%	91.5%	92.0%	93.5%
大象	100.0%	47.0%	62.0%	68.0%	72.0%	80.0%
昆蟲	100.0%	32.0%	44.0%	57.0%	62.0%	67.0%
樹葉	99.0%	49.0%	65.0%	70.0%	74.5%	76.5%
藍天白雲	100.0%	36.0%	51.5%	63.0%	69.0%	73.5%
辦公生活用品	100.0%	27.0%	39.0%	44.0%	50.0%	59.0%
素描	100.0%	50.0%	61.0%	73.0%	76.0%	79.0%
水相宇宙	99.0%	27.0%	34.0%	44.0%	50.0%	55.0%
平均	99.3%	41.6%	54.5%	63.8%	69.6%	74.3%

註：詳細實驗資料與數據請查閱附件－實驗記錄

### 一、各類群檢索出同一張影像的實驗分析

表三「檢索出同一張影像的百分率」欄位是我們在實驗過程中逐一地將每張影像當作查詢影像進行檢索的結果，由欄中的數據可以明確的得知，對於本研究所實驗的影像類群，利

用我們所設計出的影像檢索技術進行比對，結果各類群檢索出同一張張數百分率平均都在 99.3% 以上。其中在本實驗所使用的 20 個類群，有 13 個類群檢索出同一張張數的比率達到 100%，表示我們所設計出的影像檢索技術能在資料庫擁有與查詢影像相同的影像時確實找出與查詢影像相同的影像。

對於其他 7 個類群檢索出同一張張數的比率未達 100%，在查閱過未比對出來的實驗影像資料及數據後，發現未檢索出同一張影像的原因如下：

影像的像素值的種類太少：相對於檢索率達到 100% 的 13 個類群，其他 7 個類群的影像所使用的像素值數目就減少了許多，例如樹葉、材質與標誌…等。又我們所使用的標準 VQ 編碼法－Full Search 取用的像素值是使用最多像素值數目的前十名，因此有極少部分的影像在萃取出特徵值時會出現類似下面的陣列格式：{158,166,174,134,219,143,150,0,0,62,62,76}

當初我們在設計程式時我們並未考慮到此種情形，導致在使用特徵值加權計模型作影像相似度的排序時出現無法比對到同一張影像的情形，這是我們在未來還可以再做改進的地方。

## 二、有關檢索出相似度最大之影像的實驗分析

接下來的實驗爲了檢驗檢索系統能否檢索出相似度最大的影像，我們逐一地將每張影像當作查詢影像並且同時從資料庫中抽走這張影像，然後進行檢索影像的程序，記錄檢索出相似度最大影像的情形。如表二所示，當影像資料庫存有 20 個類群合計 3100 張影像。實驗結果顯示前一張、前二張、前三張、前四張、前五張檢索出相同類群的張數比率依序是 41.6%、54.5%、63.8%、69.6、74.3%。即使影像資料庫不存在與查詢影像完全相同的影像時，我們的檢索技術也可以達到 74.3% 的正確率，把同類群相似度最大的前五張影像即刻回傳給使用者。

## 伍、結論與應用

在研究報告中，我們提出一種植基於向量量化編碼法的灰階影像檢索技術。向量量化編碼法是一種極為簡單的影像壓縮技術。我們應用這個編碼法，來萃取出灰階影像內容的特徵值。同時我們也計算出整張影像與影像中間位置的像素平均值，作為日後檢索影像時過濾掉影像資料庫中不需要比對特徵值的依據。

面對龐大數量的影像資料庫，如何幫助使用者非常便利又快速地查詢到所需的影像資料，並且瞭解當前市場上所發展以影像內容為基礎的查詢系統，進而設計出處理影像資料特徵的方法和檢索影像的技術，使其更具正確性與快速性。本論文乃為達成上述的目標而努力，本報告的貢獻計有下列四個項目：

一、設計出一個有效處理影像資料特徵的演算法：

我們所提出的方法能有效地萃取出影像內容的特徵值並且讓使用者可以快速且正確地查詢到所需要的影像。當影像資料庫中存在與查詢影像完全相同的影像時，我們的檢索技術都能在第一時間第一順位檢索出這張影像。即便影像資料中不存在與查詢影像完全相同的影像時，我們的檢索技術平均約 74.3%也能在第一時間前五順位檢索出最像的影像。

二、設計出一個使用較小記憶體空間儲存影像資料特徵值的方法：

依據處理影像資料特徵值的演算法我們可以發現，我們所提出的方法儲存每張影像特徵值我們僅需  $10 \times 8\text{bits} = 10 \text{ bytes}$  的記憶體空間，遠低於每張灰階影像  $512 \times 512 \times 8\text{bits} = 262144\text{bytes}$ 。我們所提出的方法大大節省儲存影像資料記憶體的空間。

三、設計出一種可以加速影像檢索速度及提升檢索正確率的方法：

我們所提出的方法在檢索影像的過程中，應用查詢影像與資料庫影像整張影像的像素平均值與中間位置的像素平均值的差距的絕對值不大於 20 的技巧，不僅可以過濾掉大多與查詢影像相似性不大的資料庫影像，如此就節省了許多不必要比對特徵值的時間，而且當檢索影像時，也可以提昇檢索出相似度最大影像的機率。

(四)、整合本研究中的程式處理流程，使我們所設計出來的影像檢索技術能更快速的萃取出影像的特徵值，以達到快速建立影像特徵值資料庫的目的：

在本次研究中，我們使用了 Linux C(gcc 編譯器)與 Microsoft Excel 巨集來整合我們所設計出來的程式，經由實際實驗後，我們所整合的程式萃取出 3100 張  $512 \times 512$  的影像特徵值所花費的時間約 1.1 小時，相對於早期以人工方式建立同樣影像張數的影像特徵值資料庫所花費的時間約 52 小時明顯快了許多。



## 陸、展望

在本研究中，我們提出了一個植基於向量量化編碼法的灰階影像檢索技術。在影像檢索的方法中，我們主要應用了向量量化編碼法的壓縮技術來萃取出影像本身資料有效的特徵值。同時我們也應用整張影像的像素平均值與中間位置的像素平均值，作為過濾掉影像資料庫不需要比對的依據。雖然影像在處理的時間、儲存記憶體的空間…等，我們均做了相當多的考量。但不可諱言地在未來研究的方向上，還是存在許多可以改進的空間：

### 一、改進訓練編碼簿的演算法：

一本好的編碼簿的產生，會提昇整個檢索系統萃取出影像本身資料特徵值的代表性。因此如何選出具有代表性的影像與訓練出一本好的編碼簿是未來相關研究者可以思考改進的一個方向。

### 二、改進向量量化編碼法萃取出影像本身資料特徵值的技術：

使用向量量化編碼法萃取出影像本身資料特徵值的技術，未來可以在縮短處理的時間與萃取出更有效之特徵值的方向加以改進，使其影像檢索系統達到讓使用者快速且正確地獲得所需要的影像。

### 三、增加實驗影像的類群與張數：

每個類群的影像都各自擁有它們的特性，未來研究的方向，如果能針對更多不同的類群的影像進行相關的實驗，一定會有一些新的發現。深信這些發現將進一步改進現有影像檢索系統的缺點，使其更臻完美的境界。

### 四、將技術應用在彩色的影像上面：

本篇研究報告我們所提出的方法是應用在檢索灰階影像上面，相信未來我們也可以將此方法應用在檢索彩色影像上面，以因應數位時代的潮流。

## 柒、參考文獻

- [1] [http://smart.iis.sinica.edu.tw/~fish/eupload\\_main.html](http://smart.iis.sinica.edu.tw/~fish/eupload_main.html)
- [2] <http://turing.csie.ntu.edu.tw/ncnudlm/index.html>
- [3] <http://www.ctr.columbia.edu/VisualSEEk/>
- [4] <http://elib.cs.berkeley.edu/photos/blobworld/>
- [5] <http://disney.ctr.columbia.edu/webseek/>
- [6] <http://cs-www.bu.edu/groups/ivc/ImageRover/Home.html>
- [7] <http://www-white.media.mit.edu/~tpminka/photoboos/>
- [8] 陳同孝、張真誠、黃國峰，數位影像處理技術，初版，臺北市，松崗圖書，2001
- [9] 戴顯權，資料壓縮，第二版，高雄市，紳藍出版社，2002
- [10] 陳同孝、張真誠、黃國峰，電子影像技術，初版，臺北市，旗標，2003

# **A Content-Based Grayscale Image Retrieval System Based on Vector Quantization**

## **Project Book**

**Guan-Long Wu and Hsiao-Ting Yu**

National Taichung First Senior High School, Taichung, Taiwan

Fair ID: 680101

Project ID: CS304

# 1. Introduction

Content-Based Image Retrieval (CBIR), the technique to retrieve images based on automatically derived image features, is developed due to exponential growth of Web pages [1, 5, 6]. However, current CBIR systems have their drawbacks: one way to solve it is to find a new representative feature to improve efficiency. A simple compression scheme named Vector Quantization is used in this project.

Vector quantization (VQ) is used in many applications such as image and voice compression.

According to common definitions, VQ is composed of two operations: the first is an encoder, and the second is a decoder. The encoder takes an input vector and outputs the index of the codeword that offers the lowest distortion. In this case the lowest distortion is found by evaluating the Euclidean distance between the input vector and each codeword in the codebook. Once the closest codeword is found, the index of that codeword is sent through a channel (the channel could be computer storage, communications channel, and so on). When the encoder receives the index of the codeword, it replaces the index with the associated codeword.

We hypothesize that a small part (vectors) of a compressed image, which consists of the most used indices of codeword, can significantly represent the whole image, and so can be compared to other images efficiently. An image retrieval system which records indices as features is proposed.

## 2. Materials & Methods

### 2.1 Programming Environment

Images: 2,363 images classified into 15 categories      Type: 512x512 raw format (Grayscale)  
Compiler: Borland C++ Builder 6 Professional      Platform: Windows XP with Intel Pentium 4  
3.0GHz

### 2.2 The Experimental Method

#### 2.2.1 Design a Codebook with LBG Algorithm

We manually select five images for training, and create a codebook by LBG (Linde-Buzo-Gray) algorithm. This codebook is used in VQ while compressing images.

#### 2.2.2 Encode All Images with Full Search

Full search is a method to search the image pixels (Fig. 2), where each vector is compared to all codewords, and the codewords, which are nearest to the vectors, are regarded as indices.

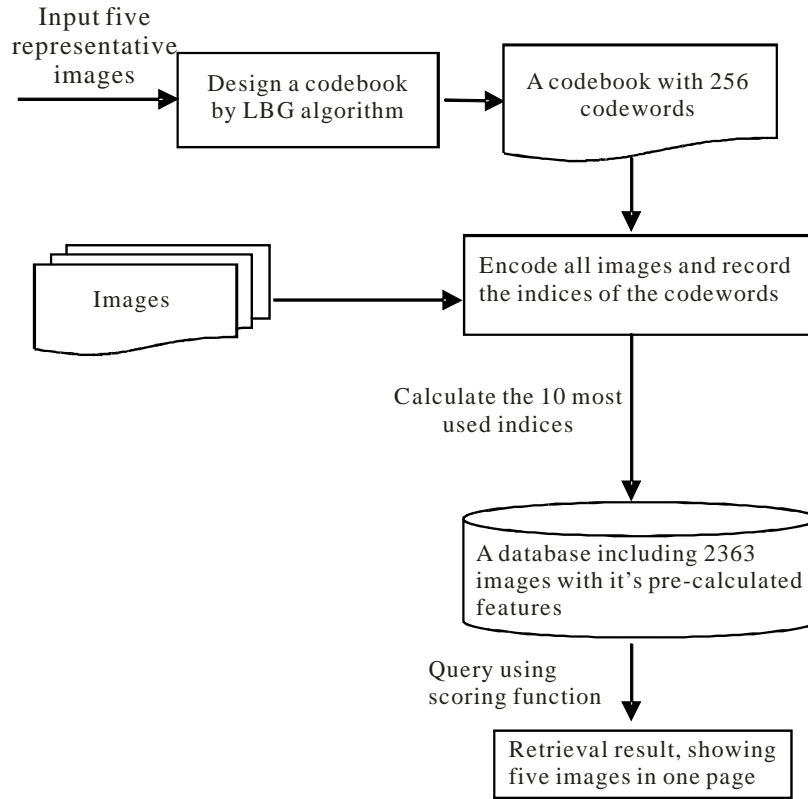
We developed a program to extract features of all images with full search, and then the most-used 10 indices of each image were recorded in the feature database.

#### 2.2.3 Query Images

We query image (Fig. 4-8) with the above feature database. The scoring function in Fig.2 is used when we calculate the image similarity. Finally, a result with images of best 5 matches in similarity order for every image will be shown. We also provided a precision-recall diagram, however, without early rejection.

### 2.2.4 Early Rejection

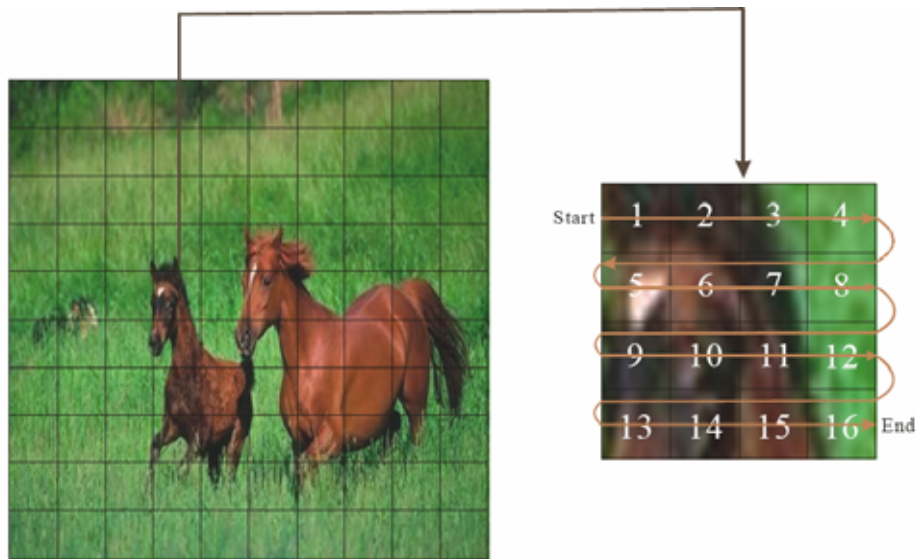
For Computational efficiency, we calculate the mean intensity of the central block of the query image, and compare it to the mean of all images within the same class. If the difference is bigger than a given threshold, early rejection is applied.



**Fig.1** The experimental method.

**Table 1** Scoring function as a table, where query image features are compared to database image features using the weighting factors listed above.

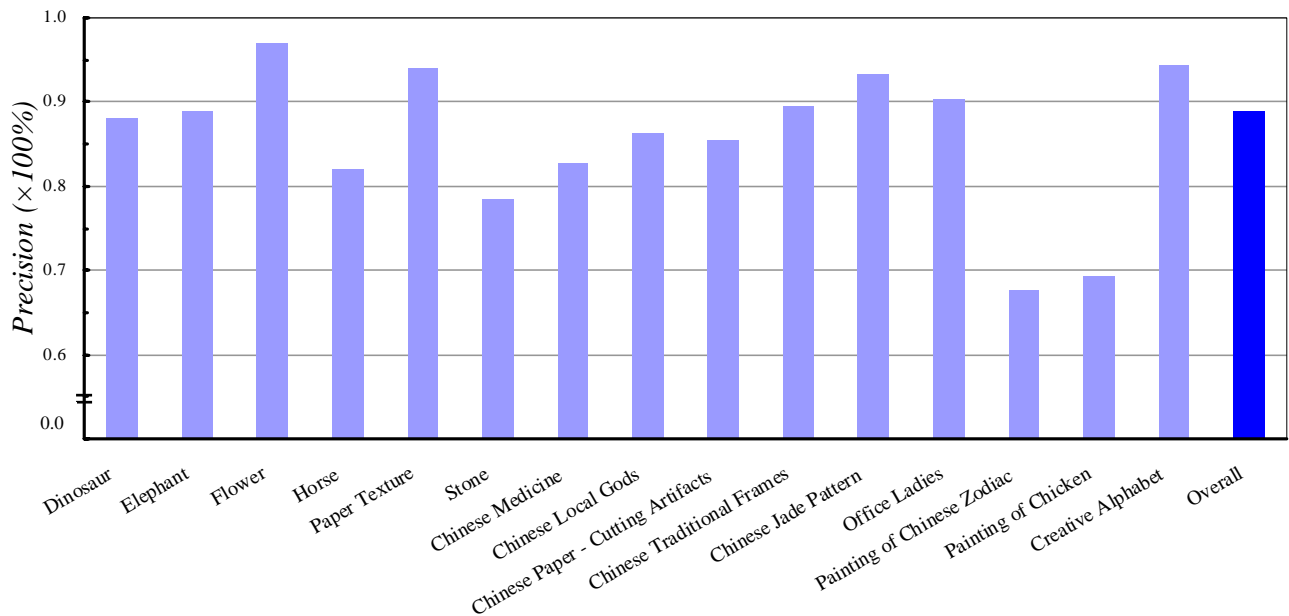
Features	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>	6 <sup>th</sup>	7 <sup>th</sup>	8 <sup>th</sup>	9 <sup>th</sup>	10 <sup>th</sup>
1 <sup>st</sup>	0.200	0.188	0.176	0.164	0.152	0.140	0.128	0.116	0.104	0.092
2 <sup>nd</sup>	0.188	0.180	0.168	0.156	0.144	0.132	0.120	0.108	0.096	0.084
3 <sup>rd</sup>	0.176	0.168	0.160	0.148	0.136	0.124	0.112	0.100	0.088	0.076
4 <sup>th</sup>	0.164	0.156	0.148	0.140	0.128	0.116	0.104	0.092	0.080	0.068
5 <sup>th</sup>	0.152	0.144	0.136	0.128	0.120	0.108	0.096	0.084	0.072	0.060
6 <sup>th</sup>	0.140	0.132	0.124	0.116	0.108	0.100	0.088	0.076	0.064	0.052
7 <sup>th</sup>	0.128	0.120	0.112	0.104	0.096	0.088	0.080	0.068	0.056	0.044
8 <sup>th</sup>	0.116	0.108	0.100	0.092	0.084	0.076	0.068	0.060	0.048	0.036
9 <sup>th</sup>	0.104	0.096	0.088	0.080	0.072	0.064	0.056	0.048	0.040	0.028
10 <sup>th</sup>	0.092	0.084	0.076	0.068	0.060	0.052	0.044	0.036	0.028	0.020



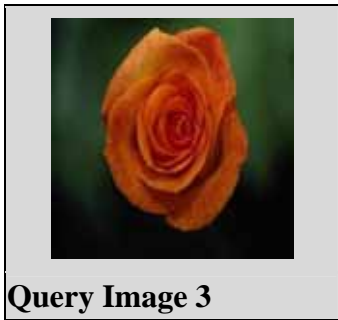
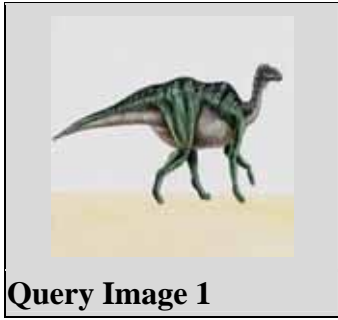
**Fig. 2** Microblock and encoding sequences. Each microblock contains 16 pixels, and there are 16384 microblocks in an image (512×512 pixels).

### 3. Result

The experimental results show that our proposed approach can effectively extract features from the images and enable users to retrieve them from an image database of 2363 images with 15 classes quickly and accurately. When images stored in image database are matched to the query image, the proposed scheme can instantly retrieve the stored image in a proper ranking order in terms of similarity. In short, the proposed scheme can retrieve the stored image at a precision of over 87.9% for the first five ranks (Fig. 4). If the query image is also in the database, it can be retrieved in the top place with a precision of 99.7%. In the following, a precision-recall diagram [1, 3, 4] (Fig.6) is provided, and so can be compared to other existing approaches.



**Fig. 3** Precision of 15 image categories with at least one correct image in the top five ranks.



**Fig. 4** An example query interface, where the left most is the query image, and the right five images are the query results, ranked in terms of similarity defined in Table 1.



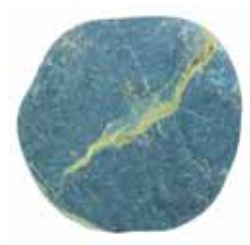
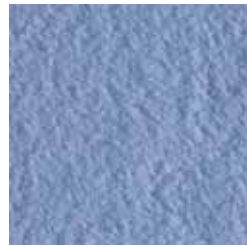
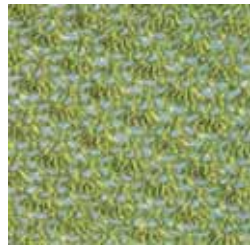


Fig. 5 More results



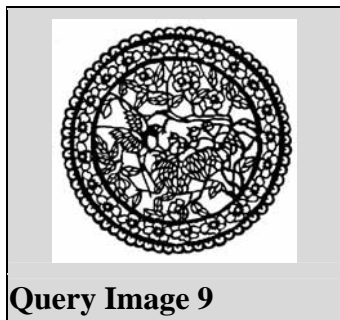
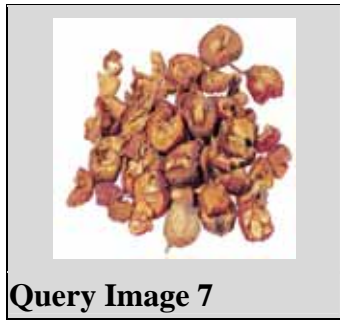


Fig. 6 More results

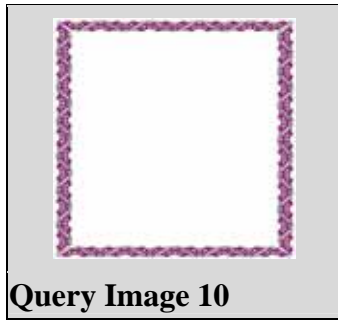


Fig. 7 More results

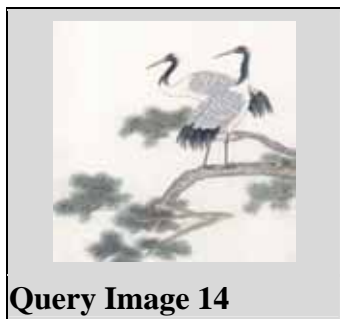
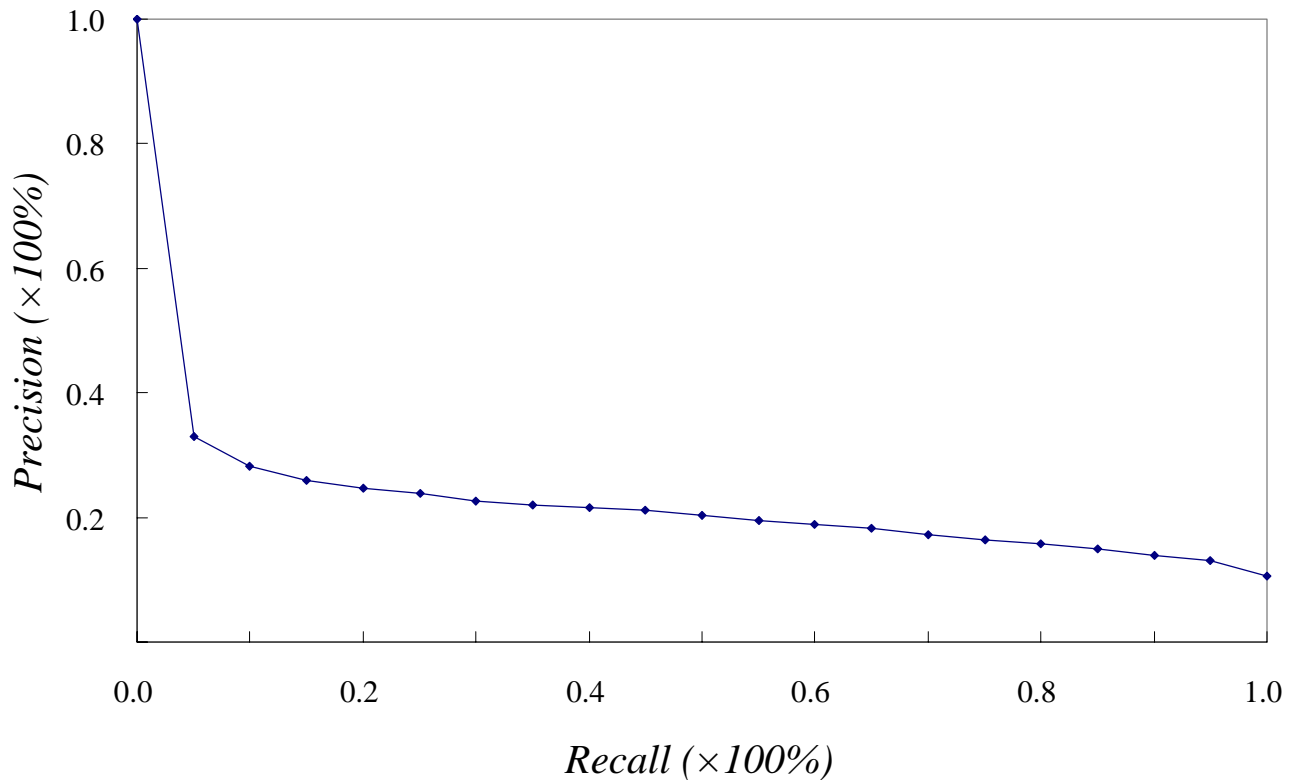


Fig. 8 More results

## 4. Discussion

Although the current system (Fig. 4-8) is efficient, limitations of our current implementation include:

1. The proposed feature is not rotation invariant, and it means that the same image can not be retrieved if rotated 90 degrees, for instance.
2. It does not deal with color, although the proposed method can be modified to include color as a feature. The simplest method is to treat color as three independent components such as (R, G, B) values.
3. Advanced features such as texture are not given special treatment here, although frequency analysis as an independent feature can be easily applied.



**Fig. 9** Precision-recall diagram of 2363 images in 15 categories, where the precision is given as the correct retrieval rate for an average of 15 categories, and the recall rate is defined as the ratio of the images that are correct images retrieved over the number of all retrieved images.



## 5. Conclusions

1. Using VQ in image retrieval is efficient and useful.
2. User feedback and manual filtering can potentially increase the retrieval precision, since our system can retrieve top five candidates in 0.7 seconds.
3. The proposed scheme can retrieve the stored image at a precision of over 87.9% for the first five ranks. If the query image is also in the database, it can be retrieved in the top place with a precision of 99.7%.

## 6. Acknowledgments

First of all, we would like to offer our deep thanks to our Consultants Mr. Ming-Cheng Liang and Professor Ming Ouhyoung to come to our assistance in our project. Second, our great thanks go to English teachers Mr. Chia-Shun Li and Miss Jia-Jen Luo to assist us in our posters and the project book. In addition, we are grateful to our teacher Mr. Szu-Ping Yueh and Ms. Tisng-Hsiang Chung for equipment supply.

Source of the images: four categories of images are provided by James Z. Wang's Research Group Website (<http://wang.ist.psu.edu/>) for research comparison [6], the rest are from Light Box which is authorized for usage by IMAGEMORE Co., Ltd.

## 7. References

- [1] D. Y. Chen, X. P. Tian, Y. T. Shen and M. Ouhyoung, "On Visual Similarity Based 3D Model Retrieval," Computer Graphics Forum (EUROGRAPHICS'03), Vol. 22, No. 3, pp. 223-232, Sept. 2003
- [2] H. Muller, W. Muller, D. M. Squire, S. Marchand-Maillet, and T. Pun, "Performance Evaluation in Content-based Image Retrieval: Overview and Proposals," Pattern Recognition Letters, Vol. 22, No. 5, pp. 593-601, 2001.
- [3] Q. Iqbal and J. K. Aggarwal, "CIRES: A System for Content-based Retrieval in Digital Image Libraries," International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, pp. 205-210, December 2-5, 2002.
- [4] Q. Iqbal and J. K. Aggarwal, "Combining Structure, Color and Texture for Image Retrieval: A Performance Evaluation," International Conference on Pattern Recognition (ICPR) Vol. 2, pp. 438-443, Quebec City, Canada, August 2002.
- [5] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton Shape Benchmark," pp. 167-178, Proc. of Shape Modeling International conference, Genova, Italy, June 2004.
- [6] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics Sensitive Integrated Matching for Picture Libraries," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, pp. 947-963, September 2001.