

以四元樹為基礎抽取影像特徵之影像檢索技術

曾修宜 黃文宣 楊智宇 劉濬毅
東吳大學 資訊科學系
tseng@cis.scu.edu.tw

摘要

本研究提出以四元樹(quadtree)分解之結果為基礎來抽取影像特徵值的技術，以改善影像資料庫檢索之效果。在我們提出的方法之中，因為四元樹分解可以強調出影像的細微部份，所以可以抽取出現覺上重要物件的特徵資訊。實驗結果顯示此方法使得影像中微小但視覺上重要物件的特徵值可以有效地被抽取出來且明顯地增進了影像檢索的效能。

關鍵詞：內涵式影像檢索 (CBIR)，四元樹 (quadtree)，向量量化編碼(Vector Quantization)。

1. 簡介

近幾年來，隨著數位影像領域快速地成長，內涵式影像檢索(CBIR)開始受人關注。一般來說，影像檢索的結果通常都存在人類感官與電腦認知上差別的問題。採用以影像裡重要物件為基礎的影像檢索技術是近年來常用被討論的議題之一。

在 Jin 等人提出的影像檢索架構裡[3]，他們發表了一種採用向量量化編碼(Vector Quantization, 以下簡稱 VQ) 分群型式作為區域的表達方式。在 Yoo 等人提出的影像檢索技術裡[6]則是在相同大小的區塊中抽取物件的顏色與紋路特徵(texture feature)，然後使用 VQ 分群(clustering)演算法來求出數個影像的主要物件，透過主要物件的特徵，可以有效地檢索出相似的影像。

為了要實作以物件為基礎的影像檢索，在抽取區域特徵值前要先執行影像分割。目前有許多影像分割的方法可以應用[2]。最近，有人提出了以適當的局部門檻值為基礎的自動彩色影像分割型式[4]。他們的這種分割型式有兩個階段，第一階段是分水嶺演算法，第二階段則是重覆迴圈的結合過程。雖然他們的演算法結果相當完善，但缺點是其計算成本相當高。四元樹分解是一種既簡單又有效率的影像分割方式[5]。使用這種方式可以在影像或紋路簡單的區域產生較少較大的區塊，而影像較複雜的區域，產生較小又較多的區塊。由 Chung 與 Tseng[1] 所提出的影像漸近傳輸方法中已証實，四元樹分解能夠保留住影像裡視覺上的重要物件之特徵。

在本研究中，我們著重於如何抽取視覺上重要物件的特徵，且能夠與電腦所認知的互相謀合。視

覺上重要的物件通常是影像中的前景物件，且這種物件的特徵在影像擷取時比背景的特徵來得重要。在我們提出的方法中，我們在抽取特徵的過程之前，先執行四元樹分解來產生大小不一的區塊，而這些區塊即是抽取特徵值的單位。在計算完這些區塊的特徵向量之後，則應用 VQ 分群演算法來找出視覺上重要的物件。實驗結果顯示，形狀細小但視覺上重要的物件可以有效地被抽取出來，因而顯著地改善了影像檢索的效能。

這篇文章的其他章節內容如下：第二節說明四元樹分解的方法，第三節說明影像特徵值抽取的方式，第四節列出詳細的實作步驟，第五節展示實驗結果，第六節則是本研究的結論。

2. 四元樹分解

四元樹分解是依據區塊的複雜程度連續地將影像分割成四分之一區塊。如果分割後的子影像不是同質區塊(homogeneous block)，它又會再被分割成四塊相同大小的子影像，持續分割下去直到所有子影像都是同質區塊為止。

所謂的同質區塊是指如果一個區塊中所有像素(pixel)的灰階值(graylevel)變化量小於某一特定常數則稱此區塊稱為同質區塊。假設 $f(x+i, y+j)$ 代表一個左上角座標位於 (x,y) 的區塊上每一個像素的灰階值， $0 \leq i, j < n$ ， n 代表區塊的大小。這個區塊的平均灰階值記作 $avg(x,y)$ ，其被定義如下：

$$avg(x, y) = \frac{\sum_{j=0}^{n-1} \sum_{i=0}^{n-1} f(x+i, y+j)}{n^2}$$

已知有一錯誤容忍值 ε ，如果某區塊中的所有像素都滿足下列條件：

$$|avg(x, y) - f(x+i, y+j)| \leq \varepsilon$$

則此區塊被定義為同質區塊。我們可以選擇錯誤容忍值 ε 的大小來控制影像分割後區塊的個數， ε 值高者影像被分割出的區塊較少較大塊，反之， ε 值低者區塊可以被分割到很小塊，甚至到 1×1 。如何選擇一個合適的容忍值，為本研究值得探討之議題。圖一為四元樹分解結果的例圖，在圖 1 中，每一個不同大小的區塊都是同質區塊且每一個同質區塊將被視為接下來抽取特徵值的單位。

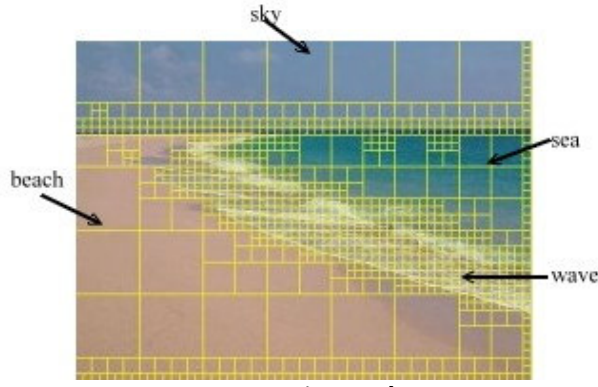


圖 1 四元樹分解範例

3. 影像特徵值抽取

為了要實作內涵式影像檢索，顏色、紋路、形狀和影像中物件的空間關係都應該被抽取。在本研究之中，我們沿用了 Hun-Woo Yoo 等人所提出的技術[6]。我們所考量的影像特徵包括顏色特徵及紋路特徵，在顏色特徵方面，每一點像素的 RGB 值將被轉換成色調、純度、亮度(HSV)色彩空間來表示。而在紋路特徵方面，先將影像分割成許多 7x7 相同大小的區塊，再用各像素的灰階值計算出每個區塊在四個角度(0°, 45°, 90°, 135°)正規化後的共生矩陣(co-occurrence matrix)[2]。依照下列公式在每一個共生矩陣計算出五個紋路特徵。

$$\text{Angular Second Moment} = \sum_i \sum_j c_{ij}^2, \quad (1)$$

$$\text{Inverse Element Different Moment} = \sum_i \sum_j \frac{c_{ij}}{1 + (i - j)^2}, \quad (2)$$

$$\text{Correlation} = \frac{\sum_i \sum_j (ijc_{ij}) - \mu_x \mu_y}{\sigma_x \sigma_y}, \quad (3)$$

$$\text{Variance} = \frac{1}{n_g} \sum_i \sum_j (i - j)^2 c_{ij}, \quad (4)$$

$$\text{Entropy} = - \sum_i \sum_j c_{ij} \log_{N_g} c_{ij}, \quad (5)$$

其中 c_{ij} 是正規化後共生矩陣的元素， n_g 是量化影像中不同灰階值的個數； $\mu_x, \mu_y, \sigma_x, \sigma_y$ 是邊際機率 p_x 與 p_y 的平均值及標準差。在計算完一個區塊中的五個紋路特徵值之後，賦予這個區塊中的每一點像素這五個數值當作它的紋路特徵。一旦抽取完所有特徵值之後，影像中的每一點像素即可用一個八個特徵值的向量 $\langle H, S, V, \text{Angular Second Moment}, \text{Inverse Element Different Moment}, \text{Correlation}, \text{Variance}, \text{Entropy} \rangle$ 來表達。

為了要實作內涵式影像檢索，我們應用 VQ 分

群演算法[6]來計算出影像的重要物件。分群的演算法是根據一已知的門檻值，如果加入的特徵向量與所有已存在群集的圓心向量(centroid)之間距離都大於門檻值，即會產生以此特徵向量為圓心的新群集。分群後向量個數多的前幾群即可分別代表影像的幾個主要物件。舉例來說，VQ 分群過後前四個群組之四個圓心向量就可以表達如圖 1 的四個主要物件：天空，海洋，沙灘和波浪。

上述所謂的已知門檻值高低是會直接影響到分群結果的群數。對同一張圖而言，門檻值高則群數會多，門檻值低則群數會變少。如何取得一個好的門檻值，得以讓分群的結果能剛好代表該影像的重要物件，亦是本研究所探討的一個問題。

根據我們的觀察，影像內容較複雜的區域需要較多的特徵資訊，反之，較簡單的區域則不需要太多的特徵資訊。舉例來說，圖 1 的波浪相較於天空或沙灘，就需要較多細微的特徵描述。然而，以相同大小的區塊處理後的前幾群當作主要物件很少與視覺上的主要物件互相謀合，為了要抽取影像中不同等級的細微部份，就要使用四元樹分割來分割影像。

4. 實作

以每張影像為單位，將彩色圖轉換成灰階圖，然後再用前述四元樹分解方法，容忍值 ϵ 定為 30，將其分割成大小不同的同質區塊。本研究中考量灰階值範圍在 0 到 255 之間，以及以同質區塊為單位計算特徵值之考量，同質區塊不宜太小，因此依實驗經驗結果設定容忍值 ϵ 30 是個合適的值。且限制區塊最小為 7x7，也就是說不論其是否已達到同質區塊的標準，7x7 的區塊不會再被切割下去。因此四元數分解結果，所有區塊的大小都是 7×2^n 的倍數，例如：14x14, 28x28, 56x56, ... 等等。

對於每一個同質區塊，將其所有的像素抽樣成 7x7 大小的區塊。舉例來說，如果某個區塊的大小為 14x14，我們將每 4 個像素取樣 1 個。對於取樣過後 7x7 的區塊，我們採用用第三節所描述的方式求得各像素的 HSV，加上採用共生矩陣的方式找出此區塊的紋路特徵向量值，給予該區塊每個抽樣的像素相同的紋路特徵值，組成八個值的特徵向量。為了減少計算量，在計算共生矩陣時，我們將原先 0 到 255 的灰階值以線性量化為 0 到 31 之間的整數值。然後將該影像所產生的全部特徵向量當作 VQ 分群演算法的輸入，以求出此影像的主要物件。

就像[6]所提到的，VQ 分群演算法有個缺點就是分群的結果容易受到輸入特徵向量的順序影響，尤其是初始的幾個輸入。為了減少此問題，在分群前，我們會先給定四個固定的初始群組，該四個初始群組之距離皆大於預設門檻值。在適當地選擇初始群組之後，VQ 分群演算法則可以降低因輸入順序之不同所造成的差異而順利地分群。在分群

之後，我們依各群所包含的向量個數多寡排序，找出最大的六群。這六群的圓心向量即被拿來代表這個影像的六個重要物件，且被儲存到影像資料庫中。

在檢索的過程中，要檢索的影像一樣要透過相同的方法來抽取六個特徵值向量。我們給予使用者在檢索影像及檢索過程之間一些空間，使用者可以在六個主要的物件中選取兩個當作檢索的條件。然後利用這兩個條件進入影像資料庫比對相似度，相似度的計算是透過 Euclidean 距離測量的方式，最後，會依照相似度排序列出影像的檢索結果。

5. 實驗

我們用影像分析及檢索的結果來展示本研究之實驗結果。實驗的資料庫由 719 張圖片所組成，所有圖片都是由數位相機所拍攝，每一張圖片的大小都是 300x400 像素。我們故意設計在不同地點所拍攝的各類圖片來組成影像資料庫，總共有 10 個系列的圖片，第一至第九系列圖各包含有 25 張至 39 張不等的圖片，第十個系列圖則為沒有分類的圖片，圖 3 即為十張代表不同系列的圖片。我們的系統在 IBM 相容個人用電腦上用 Java 程式語言實作，特徵資料值則用微軟 MS ACCESS 建構儲存之。

我們的系統實作了三個版本，版本一即為以像素為基礎的方式，即為 Yoo 等人[6]所提之方法。版本二及版本三為本研究採以四元數分解實驗過程之結果。其中版本二是以區塊為單位，每個區塊取樣的 49 個像素皆分別算出其顏色特徵值以及區塊共同的紋路特徵值，每個像素皆有一個八個元素的特徵向量作為 VQ 分群之輸入。版本三則改進版本二，考量既為同質區塊，顏色差異一定很小，用該區塊像素之平均顏色代表該方塊之顏色特徵，因此每個區塊用一個特徵向量代表即可。版本內容比較如下表：

表 2 實驗版本比較表

版本	四元樹分解	輸入 VQ 分群之特徵向量個數	VQ 群組初始化
一	無	每 4 個像素取樣 1 個	隨機產生 4 個
二	有	每個區塊 49 個	固定 4 個
三	有	每個區塊 1 個	固定 4 個

這三個版本之間存在密切的關係，版本二因為使用四元樹分解，實際反應影像的重要特徵，所以明顯地改善版本一中分群時受門檻值嚴重影響分群結果的問題。但版本二必須要處理的資料量很大。有賴版本三減少資料量，且能得到與版本二相近的檢索結果，且計算速度增快非常多，效果亦相

當理想。

我們設定各版本一個門檻值 (版本一 =1.5, 版本二 =1.5, 版本三=1.3)，然後執行批次程式，以求得所有圖片的特徵值，並存入資料庫。然後，開始執行檢索程式。圖 2 是我們系統的檢索查詢介面，在查詢介面中，位於左側的是所要查詢的圖片，中央的六個彩色區塊則是所查詢圖片之六個主要物件的色彩特徵。我們讓使用者從六個彩色區塊之中點選二個他們認為這張圖的主要特徵，然後，當使用者按下執行鍵時，即開始至影像資料庫比對，檢索的結果則在介面右側依相似度由大至小排序顯示。

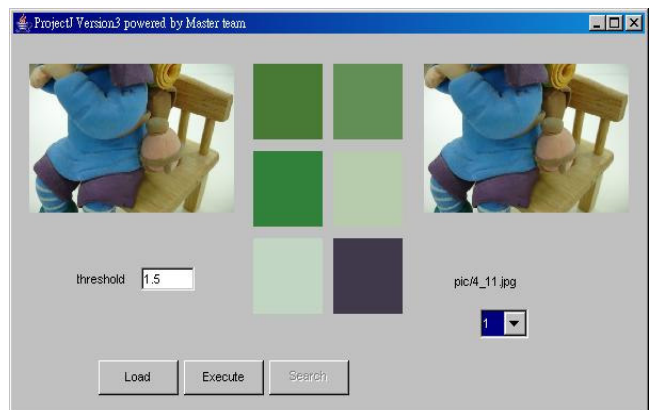


圖 2 影像檢索查詢介面

圖 4 展示各版本檢索的結果，左上角是查詢的圖片，左下角則是該圖片重要物件色彩特徵。其右側為檢索結果依相似度排序，由左至右，由上至下。從圖 4 的版本三可以看出，可在六個彩色區塊中找到牛、遠山、草地的色彩特徵，且檢索結果都為相關影像。圖 5 展示另一個例子。

最後，我們比較各版本的檢索結果數據。表 2 即為各版本的檢索比率，其值顯示在各系列在下二十次不同查詢的平均結果，檢索結果所得到的 5 張、10 張、15 張以及 20 張當中是屬於該系列的影像張數。我們可以很清楚地發現數值從版本一至版本三呈現上揚的趨勢，這表示版本二及版本三因為增加了四元樹分解，使得分析結果可以與人類視覺互相謀合。且版本三相較於版本二，因為資料量少，效能更高。

圖 5 為各實驗版本之精確率與回收率比較圖。回收率(recall rate)與精確率(precise rate)是評估檢索系統效能的兩大重要指標，公式如下

$$\text{回收率} = \text{相關影像張數} / \text{該系列圖片總數}$$

$$\text{精確率} = \text{相關影像張數} / \text{檢索所得張數}$$

由圖 5 中我們可以看出，版本一至版本三的曲線由左下至右上分佈，代表版本三具有最高精確率，這與我們預期的相同，顯示在四元樹分解的幫助下，可以更精確地檢索出圖片。

6. 結論

在本研究中，我們發表了以四元樹分解的方式為基礎來抽取影像特徵的方法。在提出的方法之中，使用四元樹分解可以強調出影像的細微部份，較之影像中較單調的部份，視覺上重要的物件可以抽取較多的特徵資訊。實驗結果顯示，相較於以像素為單位抽取特徵的方式[6]，我們所提的方法可以有效地將視覺上重要的物件特徵抽取出來，使得檢索的效能明顯地提升。

參考文獻

- [1] K.L. Chung, S.Y. Tseng, "New progressive image transmission based on quadtree and shading approach with resolution control," *Pattern Recognition Letters*, vol. 22, pp. 1545-1555, 2001.
- [2] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, Sec. 8.3, Addison-Wesley Publishing Company, Reading, MA, 1993.
- [3] F. Jin, M. Li, H-J Zhang, and B. Zhang, "An efficient and effective region-based image retrieval framework," *IEEE Trans. on Image Processing*, vol. 13, no. 5, pp. 699-709, 2004.
- [4] E. Navon, O. Miller, and A. Averbuch, "Color image segmentation based on adaptive local thresholds," *Image and Vision Computing*, vol. 23, pp. 69-85, 2005.
- [5] H. Samet, *Applications of Spatial Data Structures*, Addison-Wesley, New York, 1990.
- [6] H.W. Yoo, S.H. Jung, D.S. Jang, and Y.K. Na, "Extraction of major object features using VQ clustering for content-based image retrieval," *Pattern Recognition*, vol. 35, pp. 1115-1126, 2002.



圖 3 十個不同系列的圖片

版本一



版本二



版本三

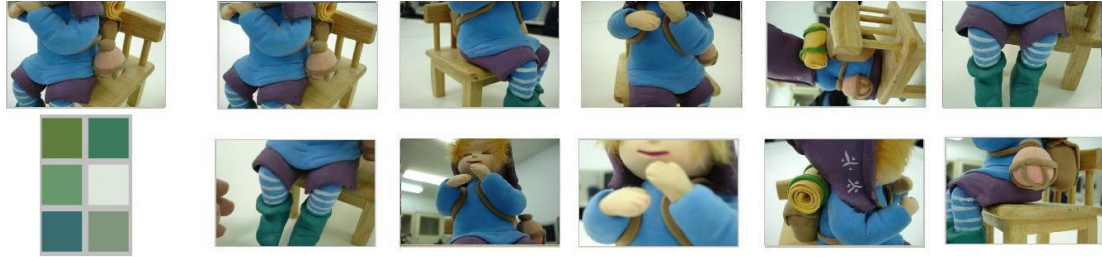


圖 4 草地上的牛檢索結果

版本一



版本二



版本三



圖 5 娃娃的檢索結果

表 2 各系列各版本的檢索比率
相關影像張數

系列	水平線			草地上的牛			沙灘			娃娃			海濱			荷花			電腦教室			籃球場			五大樓		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
檢 5	29	37	38	38	43	43	22	36	37	42	45	46	21	29	33	33	43	45	23	33	35	33	37	41	21	39	40
索 10	44	57	61	64	80	79	31	62	63	76	86	87	28	45	51	52	82	84	32	55	57	52	63	73	28	68	70
張 15	55	72	76	86	113	110	37	81	86	107	123	127	34	58	62	66	112	117	38	72	74	66	81	98	34	88	92
數 20	63	83	86	105	143	139	43	98	104	135	156	161	37	68	72	73	136	141	43	86	89	76	99	118	40	103	107

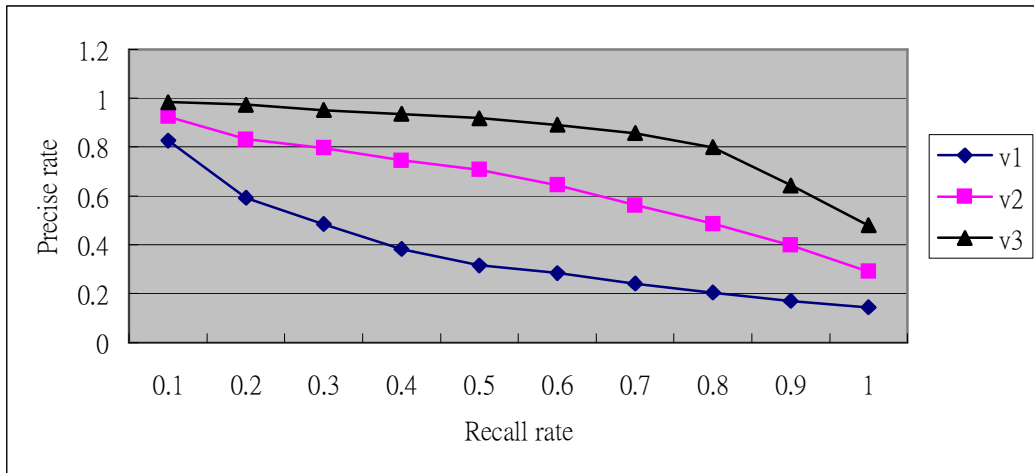


圖 5 精確率與回收率比較圖