

## 利用區塊相似比對法則設計一個適用於網路環境之 階層式視訊內容查詢系統

### Hierarchical Video Retrieval on the Internet Using Block Similarity Measures

鄭卜壬  
Pu-Jien Cheng

李曉玲  
Hsiao-Ling Lee

楊維邦  
Wei-Pang Yan

國立交通大學資訊科學研究所  
Institute of Computer and Information Science  
National Chiao Tung University  
{pjcheng, gis87502, wpyang}@cis.nctu.edu.tw

#### 摘要

在本篇論文中，我們提出一個適用於網路環境的階層式視訊內容查詢系統，本系統分為線上(on-line)與離線(off-line)兩個子系統，線上子系統採用多層解析度(multi-resolution)的比對方法，離線子系統則採用關鍵畫面(key-frame)的視覺特徵以建立多維空間的索引架構，此兩種方法的結合可以有效地改進在網路上搜尋視訊內容的效能。由於目前大部份視訊壓縮技術基於區塊編碼的方式，本系統採取區塊相似(block similarity)比對法則，在壓縮條件下提供部分範例查詢的功 (query by partial example)，以有效地擷取相似的視訊物件。

關鍵詞：視訊查詢系統、視訊內容擷取、視訊索引以及相似物件查詢。

#### Abstract

In this paper we propose a hierarchical video content search system on the Internet. The search system is comprised of on-line and off-line subsystems. The on-line subsystem uses the multi-resolution-comparison method while the off-line subsystem builds a multi-dimensional index on the visual features of key-frames. Combining these two approaches efficiently improves the performance of content-based video retrieval on the Internet. For the advent of video compression standards relying on block-based encoding, the system, which aims at working on compressed video, adopts block-based similarity measures to efficiently retrieve similar video objects by partial example.

Keyword: video search system, content-based retrieval, video indexing, and query by partial example.

#### 1 簡介

由於全球資訊網(www)採取一種分散式、異質性、合作式之開放式資訊架構，導致多媒體資料廣泛地散佈在各種不同應用領域的資訊網路系統，在各種展示媒體中，以視訊(video)的內容最為豐富，視訊資料由數以千計個連續的視訊畫面(frame)所組成，傳統文字型態的儲存與

索引技術僅適用於視訊內文字或語意內容的擷取[4,14]，無法滿足數位圖書館整合的需求。現存的視訊資訊擷取系統普遍發展於獨立的電腦環境，缺乏考慮整體網路的因素，由於儲存空間與網路頻寬的限制，傳送大量的視訊資料需要較多的時間與空間，為有效地使用有限的資源，數位視訊資料經過壓縮的方法來減少資料的儲存與傳輸量，利用較快的運算能力來改進傳輸的速度。傳統的視訊內容查詢系統[2,14]是在離線(off-line)的狀態下針對擷取出的關鍵畫面(key-frame)索引，例如：QBIC系統[9]提供使用者以顏色、紋理、形狀和輪廓查詢視訊資料，Photobook系統[10]以2D形狀和紋理的特徵進行視訊資料的搜尋，VisualSeek系統[11]可自動由影像中擷取出顯著的顏色區域，提供以顏色區域和區域空間關係進行查詢的功能。此類系統可應用在影像或視訊資料庫以提供多媒體資料內容的查詢，其自動擷取視訊中每個場景的關鍵畫面，並利用影像處理技術分析影像內容特徵(如顏色、紋理等)以作為索引。離線系統適用一般的多媒體資料庫系統，針對已存在的網路視訊進行索引，如新聞片段的查詢、電影的編輯與剪接等，但由於此類系統必須在資料庫處於離線狀態時完成關鍵畫面選取、內容特徵擷取和索引的運算，因此無法適用於即時(real-time)的應用環境。

為了滿足網路上即時性的需求(如視訊會議、遠距教學、監視保全系統、即時衛星影像分析、網路節目直播等)，視訊內容查詢系統必須提供線上查詢的功能，和傳統(離線)架構的差異在於線上系統無法事先選取視訊資料片段中的關鍵畫面完成索引，而是當網路視訊在播放的同時，進行即時比對的工作。為了降低傳輸資料量，線上系統通常採用壓縮和多層解析度(multi-resolution)兩種方法來避免因為網路頻寬限制造成傳輸延遲的情形發生。為了避免因解壓縮計算而遮蔽壓縮在傳輸上的優點，此類系統必須要能有效地在壓縮領域(compressed domain)下進行視訊內容特徵的擷取與比對，以加快特徵擷取和查詢比對的速度。目前已存在許多相關的研究[6,7,12]。VideoZoom[12]是一個在網路環境中以多層解析度方式瀏覽視訊的瀏覽系統，使用者可先瀏覽較粗糙、解析度較低的視訊資料，再根據需要來選擇有興趣的片段，在空間與時間的座標系統瀏覽較高解析度的視訊內容，產生焦距調整(zoom in/out)的效果，其目的在減少網路的資料傳輸量；[7]說明如何從MPEG的壓縮視訊中擷取區塊的DCT係數和運動向量方向作為查詢比對

的依據；在[6]這篇論文中，整理數種直接在壓縮領域下擷取空間和時間上特徵點以及視訊資料比對的方法。

以網路而言，不同的網路環境(如 ATM、ISDN 與區域網路)有不同的頻寬限制，而適用的應用系統也隨之不同，因此為了因應不同的網路傳輸需求，以及非即時和即時性的應用環境，本系統同時結合線上和離線兩個子系統的特性，採用可適性(adaptive)的開放式架構，以增進各種網路環境的執行效益。以查詢而言，傳統的查詢系統都有提供範例查詢的功能，有些利用影像或視訊處理技術(如運動向量、顏色、形狀等)允許查詢範例為視訊畫面的一個部份影像[9,11]，有些應用 DCT 區塊比對兩張影像相似程度[7]，但上述的系統皆無法以區塊相似法則來提供部分範例查詢的功能，就我們所知，目前尚無研究分析如何在壓縮的情形下，同時在離線與線上進行自動化部份區塊相似性比對的分析。

本論文的目的是提出一個有效且具可適性的網路視訊查詢系統，在不解壓縮的前提下，同時在離線與線上模式提供部份區塊相似比對的功能，使用者可輸入一個比視訊畫面相同或較小的影像，查詢有哪些視訊畫面中有局部相似於此一查詢影像，並且指出在該畫面的哪個位置，查詢影像可分散在視訊畫面中一個或多個區域，為了增進查詢效率，整個查詢與索引過程皆自動化且直接在 DCT 頻率空間(frequency domain)內運算。

本論文第 2 章介紹整個系統的概觀，第 3 章描述離線子系統的索引架構與查詢處理，第 4 章說明線上子系統的多階層比對法則，第 5 章是結論與未來研究的方向。

## 2 系統概觀

本系統為一可適性(adaptive)架構，包含離線、線上以及同時具有離線和線上功能。

### 2.1 離線(Off-line)子系統

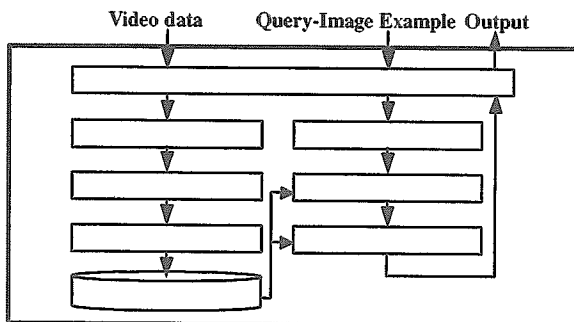


圖 1. 離線子系統之系統架構圖

離線子系統的系統架構如圖 1 所示，網路上的視訊資料先經過場景(scene/shot)偵測、關鍵畫面選取及特徵擷取的處理，再將擷取出的特徵以多維索引架構完成索引與儲存的工作，如 R\*-tree[1]，查詢相似的視訊場景時，由使用者輸入指定的查詢影像(query image)，查詢影像同樣必須先擷取其特徵，根據這些特徵在之前已建構的索引中搜尋具有相似內容特徵的關鍵畫面，計算其相似的程度，並根據相似度以排序所找到的關鍵畫面，查詢的結果輸出即為排序好的關鍵畫面、相關場景在視訊中的起始與終止位置及其所屬視訊在網路上的連結位置(URL)。類似文字搜尋引擎的結果，其輸出為相關視訊的超連結(hyperlink)排序，使用者或系統必須透過該連結以擷取視訊實際內容。

離線系統適用於一般的傳統的多媒體資料庫，例如影像資料庫之索引與擷取。在查詢之前事先建立索引，查詢處理(query processing)直接根據查詢影像的特徵和索引執行相似比對，以減少回應時間(response time)。然而離線系統無法應用在即時多媒體環境中，此外，由於索引範圍不包含視訊內的所有畫面，只根據關鍵畫面找出可能的場景，因此離線系統無法完全找出場景或視訊中有哪些畫面和查詢影像相似。

### 2.2 線上(On-line)子系統

線上子系統之系統架構如圖 2 所示。視訊輸入為視訊物件或是離線子系統的場景輸出，對於使用者輸入的查詢影像，系統將所有的視訊連續畫面依序和查詢影像做比較，每張畫面先傳送低解析度的部分進行比對，若相似度太低則此畫面視為和查詢影像高度不相似，不需要再繼續傳送高解析度的部份與之比對；若相似度高，則進一步傳送較高解析度的部分做比對，以此方法一直遞迴下去，直到相似度低於臨界值(threshold)或是比對到最高解析度，輸出的結果為視訊或場景中的哪幾個畫面和查詢影像相似以及相似的位置，其可依照畫面在視訊中的順序或相似程度高低排序列出。

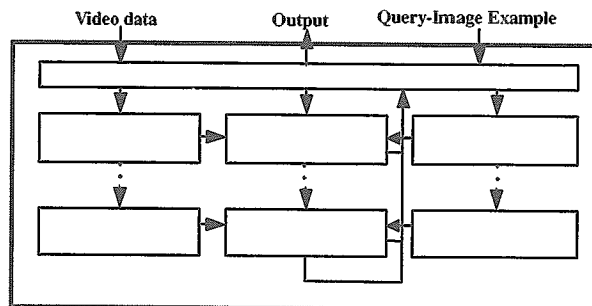


圖 2. 線上子系統之系統架構

線上子系統不需事先完成資料索引的工作，以線上的方式達到視訊內容搜尋和查詢的功能，因此符合即時視訊分析與擷取的應用環境，由於此系統針對連續畫面進行比對，資料量較離線子系統大，為了盡量減少在網路上的傳輸量，所以利用壓縮和多層解析度的方式，直接在壓縮領域下進行比對，只有相關性高的影像資料需要傳送高解析度的部分，其餘只需傳送資料量較少的低解析度部分，以避免解壓縮大量資料造成計算量過大，或是因網路頻寬的限制導致傳輸時間過長的情形。線上子系統亦適用於即時影像之分析與擷取，但不適合於非即時性質的應用環境。

### 2.3 混合型(Hybrid)系統

結合離線和線上兩個子系統以查詢網路視訊時，首先離線子系統找出關鍵畫面和查詢畫面相似的視訊場景集合，再將離線子系統的查詢結果輸出到線上子系統進行連續畫面的比對，利用多層解析度(multi-resolution)的方法依序比對場景中每個畫面，找出視訊畫面中和查詢影像相似的位置。離線索引的目的是選出所有可能含查詢影像的場景，縮小網路查詢的範圍，而線上多層比對的目的是利用不同解析度依序過濾不相似於查詢影像的視訊畫面，減少網路所需頻寬，避免傳送不相似的影像。根據不同的網路與應用環境，如圖 3 所示，線上和離線兩種系統架構分為三種不同模式：

- (1) 線上子系統的比對工作全部在伺服器 (server) 端進行。
- (2) 線上子系統的比對工作部分在伺服器端進行, 部分在客戶端(client)進行。
- (3) 線上子系統的比對工作全部在客戶端進行。

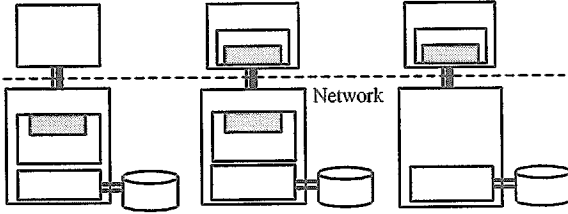


圖 3. 混合型系統的網路架構

圖 3. 混合型系統的網路架構

離線子系統需要一個儲存索引檔案的資料庫, 因此通常放置在一個或多個伺服器上, 而由於線上子系統以即時比對的方式進行, 需要一個緩衝區(buffer pool)暫時存放要播放的視訊資料。圖 3 的第(1)種情形表示線上比對在伺服器端執行, 伺服器端直接根據離線子系統的結果(超連結集合)從網路上傳回視訊場景資料進行線上比對, 再將結果傳送給客戶端, 如此客戶端在計算、儲存與網路傳輸上的負擔較輕, 此架構適用於伺服器端速度較快或是客戶端無足夠的設備環境; 相對地, 第(3)種情形伺服器端將離線子系統的結果傳給客戶端, 再由客戶端自網路上傳回視訊進行線上比對, 如此可減輕伺服器端的負擔, 避免伺服器成為執行效率上的瓶頸, 此架構適用於當客戶端機器可進行較複雜的計算與擁有較多資源; 第(2)種情形則介於上述兩者之間, 低解析度的線上比對在伺服器端而高解析度比對在客戶端進行。目前視訊擷取系統大部份只包含離線子系統, 線上子系統已應用於網路的影像漸層(progressive)傳輸上, 本系統同時具備兩者特性, 實際應用時可根據網路當時的環境資源以選擇三者中最適合的模式。

### 3 離線子系統

本章將分別說明離線子系統中索引的詳細步驟以及查詢處理的演算法則, 最後以電視新聞播報的視訊分析為例。

#### 3.1 索引結構

首先視訊資料被切割成多個場景, 場景變換的偵測即是找出連續兩個畫面差異度很大的位置, 本系統利用運動向量的方法偵測場景轉換[15], 在 MPEG 壓縮格式的視訊, 每張畫面中包含向前及向後參考(forward/backward reference)的區塊, 當一畫面與前一個畫面的差異度很大時, 其畫面中的區塊向前參考的數目會大量降低, 系統計算每張畫面中向前及向後參考的區塊數目, 當數目小於一臨界值時, 此畫面視為場景變換的位置。視訊資料切割成場景後, 再由每個場景中選取一個或多個畫面做為關鍵畫面, 並記錄其為場景中的第幾個畫面, 而關鍵畫面的選取方法, 可利用計算每張畫面和場景中其他畫面的差異度, 選擇和其他畫面差異最小的畫面作為場景的關鍵畫面, 本系統採用分群(clustering)法則, 根據同一場景內每張影像的顏色分佈(color histogram 自動分類, 距離每一群中心點最近的一張或多張影像視為關鍵畫面, 此方法可找出包含最多場景特性的關鍵畫面, 之

後依序對每張關鍵畫面進行區塊特徵擷取及索引。

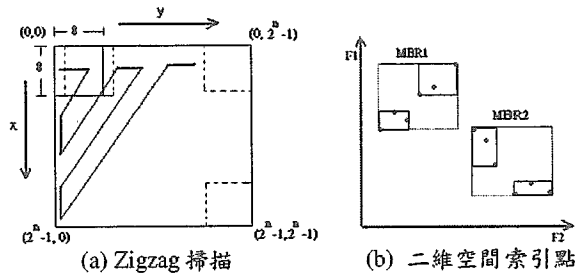


圖 4. 視訊畫面掃描方式以及其對應的二維空間表示

由於許多視訊或影像標準(如 MPEG 和 JPEG)採用 DCT 轉換函數壓縮影像中的顏色區塊, 本系統選擇擷取關鍵畫面的 DCT 係數作為索引和相似比對的特徵向量, 為了允許查詢範例對應到視訊畫面中的任意位置, 每張關鍵畫面以  $8 \times 8$  大小的滑動視窗(sliding window)做 zigzag 掃描以記錄每個位置的 DCT 係數, 如圖 4(a)所示。每移動一個像素(pixel), 計算在滑動視窗位置區塊的 DCT 係數, 換言之, 在每個位置都可算得 64 個 DCT 係數, 為了降低多維索引架構中的維數, 本系統利用 DCT 係數具有左上角部分的係數包含較多影像資訊的特性, 只取 64 個係數的前  $k$  個係數, 因此, 每個滑動視窗位置的 DCT 係數可視為在  $k$  維索引空間中的一個點, 座標為  $(c_1, c_2, \dots, c_k)$ , 圖 4(b)以二維空間( $k=2$ )為例, 顯示各個區塊的 DCT 係數對應到空間上的點。

本系統應用  $R^*$ -tree 多維索引架構, 將上述的索引點依序插入到  $R^*$ -tree 中, 同時紀錄這些點所對應的關鍵畫面和視窗區塊位置, 以最小矩形區域(MBR)將鄰近的點包圍起來[5], 方便在查詢時快速找尋與查詢點(DCT 係數)相近的點。

#### 3.2 查詢處理

使用者輸入的查詢影像大小為  $8 \times 8$  區塊(block)大小的倍數( $8 \times 8, 16 \times 16, \dots$ ), 系統利用範例部分查詢, 搜尋視訊中包含全部或局部查詢影像區塊的關鍵畫面。

**演算法:** 離線系統查詢處理

輸入: 查詢影像(影像大小為  $8 \times 8$  倍數的矩形)。

輸出: 與查詢影像相似或是包含查詢影像區塊的關鍵畫面、相似區塊的位置、其所代表的場景以及所屬的視訊資料鏈結位置, 依照相似程度排序。

- 步驟1. 將使用者輸入的查詢影像切割為多個  $8 \times 8$  大小的區塊, 假設區塊的數目為  $N \times N$  個  $(b_{11} \dots b_{1N}, b_{21} \dots b_{2N}, \dots, b_{N1} \dots b_{NN})$ 。
- 步驟2. 依序針對每個區塊  $(b_{11} \dots b_{NN})$  計算區塊  $b_{ij}$  的 DCT 係數, 並取前  $k$  個係數, 視為  $R^*$ -tree 多維空間中的一個點, 訂定一個臨界值  $\sigma$ , 依據  $R^*$ -tree 的架構和最鄰近查詢處理方法(nearest neighbor query)[1], 找出和  $b_{ij}$  查詢點距離小於臨界值  $\sigma$  的索引點, 這些點即為關鍵畫面中和查詢區塊  $b_{ij}$  的相似區塊  $b'_{ij}$ 。
- 步驟3. 檢查由  $R^*$ -tree 中找出的相似點各屬於哪些關鍵畫面, 選取包含了  $b_{11} \dots b_{NN}$  每個區塊相似點的關鍵畫面  $I_1 \dots I_M$ 。
- 步驟4. 依序對選取出的每個關鍵畫面  $I_1 \dots I_M$  檢查相似的區塊  $b'_{11} \dots b'_{NN}$  在畫面中的空間關係是否與查詢區塊  $b_{11} \dots b_{NN}$  的空間關係相同。此步驟中, 同一關鍵畫面中對於任一個查詢區塊  $b_{ij}$

可能包含一個以上的相似區塊  $b'_{ij}$ ，而且  $N \times N$  個相似區塊  $b'_{11} \dots b'_{NN}$  在空間上的關係會有多種的組合，在此應用 2D B-string [8] 的觀念篩選出所有和查詢區塊空間關係相同的組合，2D B-string 包含 type-0、type-1 與 type-2 三種相似對應，其中 type-1 同時考慮拓撲 (topological) 與方向 (directional) 上二維空間關係的特性，為了允許更彈性的相似比對，本系統允許查詢區塊間 meet 的拓撲關係可對應到關鍵畫面中相似區塊間的 meet、partly overlap 與 disjoint 關係，以擴充 type-1 規則來選擇相似的空間組合。假設使用者的查詢影像如圖 5 所示分割為九個  $8 \times 8$  大小的區塊，則選出區塊  $b_{11}, b_{12}, b_{13}$  的相似區塊  $b'_{11}, b'_{12}, b'_{13}$  在 y 軸上符合  $b'_{11} < b'_{12} < b'_{13}$  的方向關係與  $b'_{11}$  meets/partly overlaps/disjoints  $b'_{12}$ 、 $b'_{12}$  meets/partly overlaps/disjoints  $b'_{13}$  及  $b'_{11}$  disjoints  $b'_{13}$  的拓撲關係， $b_{21}, b_{22}, b_{23}$  和  $b_{31}, b_{32}, b_{33}$  的相似區塊亦類似，同理，在 x 軸上的所相似區塊亦符合上述 type-1 的空間與方向關係。

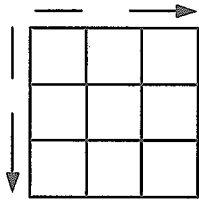
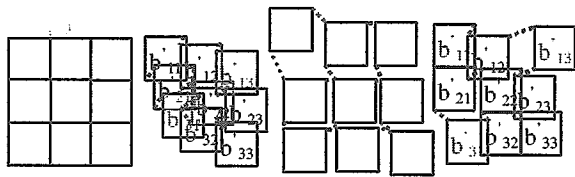


圖 5. 查詢影像切割成多個  $8 \times 8$  大小的區塊



(a) 完全相鄰 (b) 完全重疊 (c) 完全分離 (d) 部份相鄰、重疊及分離  
圖 6. 相似區塊間的空间關係與距離

步驟 5. 計算步驟 4 所選擇的關鍵畫面中相似區塊組合與查詢影像中區塊組合在距離上的相似程度，以作為輸出結果排序的依據。考慮圖 5 中的查詢影像，步驟 4 在一關鍵畫面中選出一由  $b'_{ij}$  相似區塊組成的組合，則相似區塊之間的 type-1 空間關係可能有四種情形：(1) 如圖 6(a) 表示相似區塊間互相相鄰，與查詢影像中區塊間的距離完全相同；(2) 如圖 6(b) 表示相似區塊間互相重疊；(3) 如圖 6(c) 表示相似區塊間是分離的關係及(4) 如圖 6(d) 表示相似區塊間是部份相鄰、重疊與分離的關係。在此以計算區塊之間距離的方式來辨別上述 4 種情形之  $b_{11} \dots b_{NN}$  與  $b'_{11} \dots b'_{NN}$  的相似程度。區塊之間的距離即為圖 6 中虛線所示，圖 6(a) 中區塊間的距離為 0，與查詢影像的相似度最高，圖 6(b)、(c) 和 (d) 中當區塊間的距離愈小時，則視為與查詢影像(圖 5)空間上的排列方式愈相似。本系統中相似程度的計算方法是同時考慮查詢區塊  $b_{ij}$  與相似區塊  $b'_{ij}$  在 DCT 係數的差異以及區塊間的距離， $b_{11} \dots b_{NN}$  和  $b'_{11} \dots b'_{NN}$  的差異

度(Diff)計算方法定義為

$$Diff = \frac{1}{N^2} \left( \sum_{i=1}^N \sum_{j=1}^N b_{ij} \text{ 和 } b'_{ij} \text{ 前 } k \text{ 個 DCT 係數的差異} \right) + \frac{1}{2N(N-1)} \left( \sum_{i=1}^N \sum_{j=1}^{N-1} b'_{ij} \text{ 和 } b'_{i,j+1} \text{ 的距離} + \sum_{i=1}^{N-1} \sum_{j=1}^N b'_{ij} \text{ 和 } b'_{i+1,j} \text{ 的距離} \right), \dots \dots \dots \text{式 (1)}$$

其中區塊的前  $k$  個 DCT 係數差異就相當於查詢點與相似點在  $R^*$ -tree 中的距離，兩點距離愈遠表示差異愈大。系統以滑動視窗區塊最左上角的點在視訊畫面中的位置表示該視窗區塊的位置，例如：(0,0) 為第一個視窗區塊位置，若視訊畫面的解析度為  $256 \times 256$ ，則 (248,248) 為最後一個視窗位置，而相似區塊間的 Euclidean 距離可利用區塊位置座標算出，

$$\sum_{i=1}^N \sum_{j=1}^{N-1} b'_{ij} \text{ 和 } b'_{i,j+1} \text{ 間的距離} = \sum_{i=1}^N \sum_{j=1}^{N-1} \text{sqrt} \left( (b'_{ij} \text{ 的 } y \text{ 座標} + 8 - b'_{i,j+1} \text{ 的 } y \text{ 座標})^2 + (b'_{ij} \text{ 的 } x \text{ 座標} - b'_{i,j+1} \text{ 的 } x \text{ 座標})^2 \right),$$

$$\sum_{i=1}^{N-1} \sum_{j=1}^N b'_{ij} \text{ 和 } b'_{i+1,j} \text{ 間的距離} = \sum_{i=1}^{N-1} \sum_{j=1}^N \text{sqrt} \left( (b'_{ij} \text{ 的 } y \text{ 座標} - b'_{i+1,j} \text{ 的 } y \text{ 座標})^2 + (b'_{ij} \text{ 的 } x \text{ 座標} + 8 - b'_{i+1,j} \text{ 的 } x \text{ 座標})^2 \right),$$

由上面的式子即可看出當相似區塊間和查詢區塊間一樣是相鄰時，區塊間的距離計算結果等於 0，當距離越遠時，則和查詢影像的差異越大。

此外，系統另外加入一個臨界值  $\rho$  的條件，當相似區塊間的距離超過臨界值  $\rho$  時則這個組合視為與查詢影像不相似而去除，亦或在式(1)中加入權重(weight)，使其可調整 DCT 差異以及區塊間距離在計算整體差異度上所佔的權重比例。當式(1)計算總和愈小即表示與查詢影像相似度愈高。

步驟 6. 根據式(1)的結果排序，並依序輸出關鍵畫面、相似區塊的位置、相對應的場景以及所屬的視訊資料。

### 3.3. 離線子系統之範例

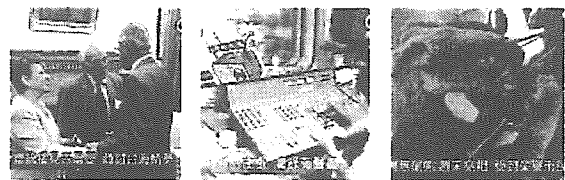


圖 7. 三段新聞報導場景的關鍵畫

本節以新聞報導的視訊資料作為例子以說明離線子系統的運作流程。假設有三段新聞報導的視訊場景，其關鍵畫面  $k_1, k_2, k_3$  皆為  $64 \times 64$  像素大小的影像，如圖 7 中所示。根據 3.1 節介紹建立索引的方式，分別計算這三張關鍵畫面各個滑動視窗區塊的 DCT 係數，取前  $k$  個係數，在此例中設  $k$  為 4，將每個區塊 DCT 係數點  $(c_1, c_2, c_3, c_4)$  插入  $R^*$ -tree 的索引架構中完成索引的動作。

<sup>1</sup> 圖片資料來源民視新聞資訊網，版權屬民視所有，經同意於本論文中使用。

當使用者給定一個查詢影像，例如輸入一個16×16大小的影像，系統搜尋所有視訊中有出現與查詢影像相似區塊的關鍵畫面，以下根據3.2節查詢處理的步驟：

1. 首先將查詢影像切割成4個8×8大小的區塊，分別為 $b_{11}, b_{12}, b_{21}, b_{22}$ 。
2. 計算 $b_{11}, b_{12}, b_{21}, b_{22}$ 的DCT係數，同樣取前 $k$ 個係數( $k=4$ )，將查詢鄰近範圍之臨界值 $\sigma$ 定為45，在多維索引架構中找出和查詢區塊DCT係數點距離小於45的點。
3. 檢查由步驟2找出來的相似點各屬於哪個關鍵畫面以及在關鍵畫面中的位置為何，在本例子中找出來的相似區塊如下所示。

$b_{11}$ 的相似區塊： $k_2$ 中(12, 4)位置的區塊，其DCT係數點與 $b_{11}$ 查詢點的距離為23.7。

$b_{12}$ 的相似區塊： $k_2$ 中(6, 42), (7, 32), (12, 12)以及 $k_3$ 中(7, 3), (7, 4)位置的區塊，其DCT係數點與 $b_{12}$ 查詢點的距離各為44.5, 29.0, 11.32和33.5, 17.9。

$b_{21}$ 的相似區塊： $k_2$ 中(20, 4)位置的區塊，其DCT係數點與 $b_{21}$ 查詢點的距離為33.3。

$b_{22}$ 的相似區塊： $k_2$ 中(13, 11), (20, 12)以及 $k_3$ 中(4, 16)位置的區塊，其DCT係數點與 $b_{22}$ 查詢點的距離各為44.3, 13.9和42.9。

其中 $k_1$ 不包含和查詢影像相似的區塊， $k_3$ 只包含 $b_{12}$ 和 $b_{22}$ 的相似區塊， $k_2$ 包含 $b_{11}, b_{12}, b_{21}$ 和 $b_{22}$ 的相似區塊，因此只需針對 $k_2$ 進行下一步驟處理。

4. 在 $k_2$ 關鍵畫面中 $b_{11}, b_{12}, b_{21}, b_{22}$ 的相似區塊各有1, 3, 1, 2個，所以4個區塊的組合共有6種，利用每個區塊在畫面中的位置座標，檢查每種組合中區塊的空間關係是否與查詢影像相同，其中 $b_{11}, b_{12}$ 和 $b_{21}, b_{22}$ 相似區塊在 $y$ 軸上要符合type-1的方向與空間關係，而 $b_{11}, b_{21}$ 和 $b_{12}, b_{22}$ 相似區塊在 $x$ 軸亦相同，刪除不符合type-1的區塊組合，在本例中6種組合的相似區塊空間關係則皆符合這些條件，例如 $b_{11}$ 和 $b_{12}$ 的相似區塊在 $y$ 軸上為(12, 4) < (6, 42)或(7, 32)或(12, 12)，且(12, 4)disjoints(6, 42)或(7, 32, (12, 4)meets(12, 12)。
5. 計算與查詢影像的相似程度，區塊與查詢區塊DCT係數的差異即為步驟2中兩個點的距離，區塊之間的距離的臨界值 $\rho=25$ ，例如(12,4)和(6,42)兩個區塊在 $y$ 軸的距離超過25，則我們可將6種組合中同時包含這兩個區塊的兩種組合去除，將區塊與查詢區塊DCT係數的差異以及區塊間的距離相加以估算與查詢影像的差異度，每兩個區塊之間的距離的權重設為 $1/\rho=0.04$ ，每兩個DCT係數的差異的權重設為 $1/\sigma=0.02$ ，本例子以{(12,4), (12,12), (20,4), (20,12)}組合的差異度數值(Diff=0.457)最小，其餘三組依序為{(12,4), (12,12), (20,4), (13,11)}(Diff=0.767)，{(12,4), (7,32), (20,4), (20,12)}(Diff=0.967)以及{(12,4), (7,32), (20,4), (13,11)}(Diff=1.212)。
6. 由以上步驟可推得本例的結果如圖8所示， $k_2$ 關鍵畫面內有4種區塊組合相似於查詢影像，使用者可以調整臨界值 $\sigma$ 與 $\rho$ 來縮小查詢結果集合。

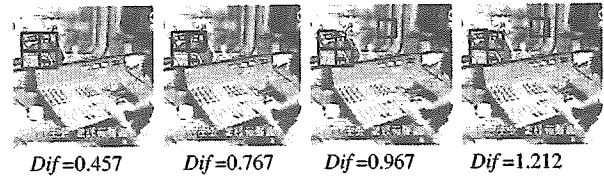


圖8. 關鍵畫面中相似區塊的有效組合

#### 4 線上子系統

本章將說明線上子系統中多層解析度(multi-resolution)的階層式比對方法(hierarchical comparison)，以及查詢影像和關鍵畫面在壓縮領域(compressed domain)下的相似比對，最後同樣地以電視新聞播報的視訊分析為例。

##### 4.1 多層解析度的網路傳遞模式

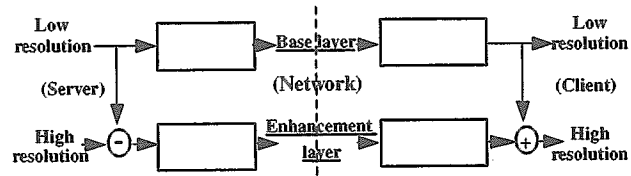


圖9. 多層解析度的網路傳送方式

視訊資料可以將原始資料流分解成多層解析度的格式分別傳送，一般方式如圖9的流程圖所示，低解析度的資料為原圖模糊或低頻的資料部分，傳送過程中首先計算原圖(即高解析度)和低解析度資料的差值，再分別將低解析度和差值的資料經過某種(壓縮)轉換，由不同的分層格式傳送。頻寬較大的客戶端在接收後，分別將各層資料解壓縮的處理，以還原高解析度的影像，對於頻寬較小客戶端可選擇性的只解低頻的部份，此時差值的部份不需傳送，以滿足同步(synchronization)與即時的需求。

線上子系統如圖2所示，先將使用者輸入的查詢影像和視訊畫面在低解析度的情形下比對，若相似度則再進一步作高解析度的比對，此方法可以在低解析度時利用臨界值先過濾不相似的影像，如圖10<sup>2</sup>所示，此例中影像1為查詢影像，用以比對另三張影像，每張影像被切割成數個8×8大小的區塊，假設 $c_{ij}$ 與 $q_{ij}$ 分別表示被比對影像與查詢影像中第 $i$ 個區塊的第 $j$ 個係數，考慮每個區塊中的前 $2^k$ 個係數， $k$ 愈小表示解析度較低，則兩張影像的相似距離(Dist)可由式(2)計算得知

$$Dist = \frac{1}{\# \text{ of blocks}} \sum_{i=0}^{\# \text{ of blocks}-1} \sqrt{\sum_{j=0}^{2^k-1} (c_{ij} - q_{ij})^2} \dots \dots \dots \text{式}$$

(2)

Dist值愈小表愈相似，執行結果如表1所示，若臨界值設為200，當 $k=0$ 時，影像4為不相似而不再繼續比對， $k=2$ 時影像3不相似，若臨界值設為290，在 $k=1$ 時，影像4為不相似， $k=5$ 時影像3不相似，因此利用此方法，可以有效地提早減少搜尋的範圍。

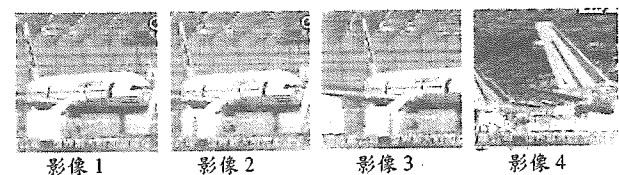


圖10. 利用多層解析度法比較四張影像的不同

<sup>2</sup> 圖片資料來源民視新聞資訊網，版權屬民視所有，經同意於本論文中使用。

表 1. 不同係數個數時, Dist 值的計算結果

臨界值 $k$	影像	影像 2	影像 3	影像 4
$k = 6$ 64 coefficients		64.72	303.09	434.22
$k = 5$ 32 coefficients		58.69	297.52	428.66
$k = 4$ 16 coefficients		48.12	281.92	414.66
$k = 3$ 8 coefficients		37.84	259.09	389.17
$k = 2$ 4 coefficients		27.88	241.86	367.97
$k = 1$ 2 coefficients		22.34	197.49	313.92
$k = 0$ 1 coefficient(DC)		16.55	173.33	289.38

由於線上子系統無法先進行建立索引的工作, 在視訊播放的過程中, 即時地分析緩衝區內的資料, 對同一個場景而言, 需要處理的畫面數目較離線子系統多, 若是做解壓縮的動作, 則會耗費太多的時間, 因此系統在接收到資料後必須在壓縮領域下直接進行相似比對。圖 9 顯示低解析度的資料在壓縮領域下可在客戶端直接獲得, 但高解析度傳送至客戶端是差值的壓縮結果, 必須將差值和低解析度相加才能得到高解析度的資料, 高解析度 = 低解析度 + 差值, 由於 DCT 是線性的轉換,  $DCT(\text{高解析度}) = DCT(\text{低解析度}) + DCT(\text{差值})$ , 所以可以直接將接收到的差值 DCT 係數和低解析度資料的 DCT 係數相加得到高解析度資料在壓縮(DCT)領域下的資料, 換言之, 圖 9 的流程中仍不需經過解壓縮的步驟即可獲得高解析度的 DCT 係數。

但在壓縮格式的限制下, 例如 MPEG 將視訊畫面切割為  $8 \times 8$  大小區塊的 DCT 係數, 若使用離線子系統以滑動視窗做 zigzag 掃描的方法來得到所有位置的 DCT 係數, 在連續影像分析的情形下, 需要的計算量可能會影響線上系統即時播放的功能, 所以系統在此考慮採用一種預測的方式來進行特徵擷取和比對的運算。

#### 4.2 階層式比對

在[3]這篇論文中, 利用 DCT 線性轉換的特性, 計算介於 4 個區塊中間位置的 DCT 值, 如圖 11 所示, 區塊 P 的 DCT 係數為與它重疊的 4 個區塊之 DCT 係數的線性組合(linear combination), 任意位置的 DCT 值可在不反轉 DCT 的運算下獲得, 因此圖 4(a)的 zigzag 掃描亦可在壓縮領域下計算。

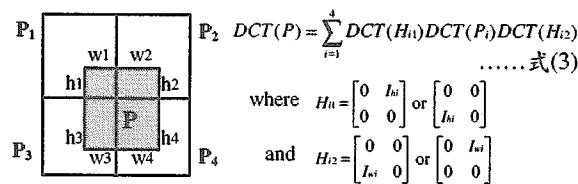


圖 11. 區塊 P 與 P<sub>1</sub>, P<sub>2</sub>, P<sub>3</sub>, P<sub>4</sub> 重疊

式(3)計算過程需要多個乘法, 考慮影像之 DCT 轉換的特性, 即每個區塊左上角位置的 DCT 係數(DC 值)包含了影像大部分的資訊, 因此在低解析度的比對上, 系統選擇只比對 DC 係數, 同時應用 DC 值的近似值推算法[13], 加快比對的速度, 在[13]這篇論文中提出兩種區塊 DC 係數值的近似值, 一是 zero-order 近似值,  $DC(P)^0 = DC(P_i)$ , 其中  $h_i w_i \geq h_j w_j$ , 即區塊 P 的 DC 值以重疊範圍最大的區塊 DC 值表示, 另一是 first-order 近似值,  $DC(P)^1 = \dots \text{式(4)}$ 將有重疊

$$\sum_{i=1}^4 \frac{h_i w_i}{64} DC(P_i)$$

的區塊 DC 值各自乘上重疊的面積比例, 其誤差的最大值為影像像素最大值的四分之三, 例如 8 位元影像的最大誤差是  $191(255 \times 3/4)$ 。

線上子系統使用階層式的比對方式, 壓縮視訊畫面與查詢影像被表示成多個  $8 \times 8$  的區塊(如圖 5), 對每個視訊畫面, 利用區塊 DCT 係數的多寡, 由低解析度至高解析度依序比對, 並根據圖 9、[13]以及[3]提出的方式來推算所有位置區塊的 DC 及 DCT 係數, 茲分別詳述比對法則如下:

#### • 低解析度的比對(Low-resolution comparison)

在低解析度的部分, 系統比對查詢影像區塊和視訊畫面區塊的 DC 值, 當兩者的差異愈大表示兩個區塊的相似度愈低, 並設定兩個臨界值( $\theta_1$  和  $\theta_2$ )其中  $\theta_1 > \theta_2$ 。第一個步驟應用 zero-order 近似值的觀念, 先將查詢影像區塊的 DC 值直接和壓縮視訊畫面中所有區塊的 DC 值比較, 當兩者差異小於  $\theta_1$ , 則將視訊畫面中該區塊視為候選區塊(candidate block)。第二個步驟進一步檢查這些候選區塊附近範圍內是否有更相似的區塊, 系統利用一個滑動視窗, 檢查和候選區塊重疊範圍超過區塊面積四分之一的所有區塊, 計算這些區塊和查詢影像區塊的 DC 值差異, 當兩者的差異小於  $\theta_2$  時, 則視為相似區塊, 並紀錄下它們在視訊畫面中的位置。

第二步驟中統考慮以下兩種方式來估算和候選區塊有重疊的區塊的 DC 值, 第一種方式是利用 first-order 式(4)的方法估算區塊 DC 係數的近似值, 其中  $h_i$  和  $w_i$  各介於 0 到 8 之間,  $h_2 = 8 - h_4$ ,  $h_3 = h_4$ ,  $h_1 = h_2$ ,  $w_2 = 8 - w_1$ ,  $w_3 = w_1$ ,  $w_4 = w_2$ , 這種方法雖然計算較為簡單, 但由於誤差的最大值為影像像素最大值的四分之三, 為了提高召回值(recall)必須把與查詢影像區塊的差異的臨界值  $\theta_2$  增大, 但相對的精確值(precision)就會降低。第二種方式是應用式(3)的計算方式, 實際計算出區塊的 DC 值, 使用這種方式的正確率較高, 但計算量較大。

#### • 高解析度的比對(High-resolution comparison)

經過低解析度的比對後, 若視訊畫面中包含有查詢影像的相似區塊, 則必須再由網路傳送高解析度的資料, 進行進一步的比對。高解析度的比對範圍是針對畫面中在低解析度比對步驟後的結果, 利用式(3)計算介於區塊中間位置的區塊 DCT 係數, 注意式(3)中  $H_{i1}$  和  $H_{i2}$  的 DCT 值可事先計算儲存, 並利用稀疏矩陣相乘技術(sparse matrix multiplication technique) [3]以減少計算時間。計算後的 DCT 係數和查詢影像比較, 差異小於臨界值  $\theta_3$  的區塊即為和查詢影像相似的區塊。

經過低解析度和高解析度比對找出視訊資料畫面中相似區塊的位置後, 相同於離線子系統查詢處理(參考 3.2 節)的第四及第五步驟, 檢查視訊畫面中是否包含了每個查詢影像區塊的相似區塊, 並檢查這些相似區塊在視訊畫面中的空間關係是否和查詢影像的區塊相同, 並考慮相似區塊之間的距離, 以此判斷視訊畫面和查詢影像是否相似, 同樣地, 當相似區塊之間的距離超過一臨界值  $\rho$  時, 即使區塊與查詢影像區塊的 DCT 係數相似, 但仍將這區塊的組合視為不相似。輸出的結果為包含查詢影像之相似區塊的視訊畫面、區塊位置、視訊畫面所屬的場景以及其視訊資料來源。

#### 4.3 線上子系統之範例

本節繼續 3.3 節中以新聞報導節目的視訊資料為例子, 說明線上子系統的處理流程。假設離線子系統輸出包含



查詢影像區塊的視訊場景中有三張畫面如圖 12<sup>3</sup> 所示，皆是 64×64 像素大小的畫面，輸入的查詢範例仍為 16×16 大小的影像，同樣地先將查詢影像切割成 4 個 8×8 大小的區塊 ( $b_{11}, b_{12}, b_{21}, b_{22}$ )，計算每個區塊的 DCT 係數，再根據 4.2 節中介紹的階層式比對方式進行查詢處理。兩個 DCT 向量的相似程度定義為 Euclidean 距離函數，距離愈小表示相似程度愈高。

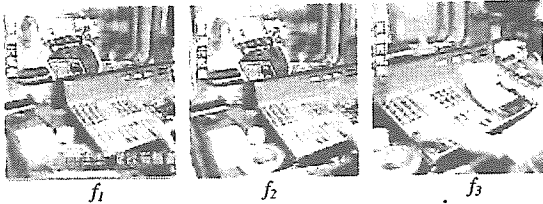


圖 12. 新聞報導場景中的四張畫面

### • 低解析度的比對

在低解析度的部分只比對區塊的第一個 DCT 係數即 DC 值，依照 4.2 節描述的步驟，第一步將查詢區塊的 DC 值直接和三張畫面中每個 8×8 大小分割區塊比較，在此例中  $\theta_1$  設為 175，以選擇每個查詢區塊在視訊畫面中的候選區塊。第二個步驟採用下述任一種方式估算畫面中和候選區塊重疊範圍超過區塊面積四分之一的區塊 DC 值。

方法一：以式(4)估算區塊的 DC 係數近似值，估算出的近似值和查詢影像區塊的 DC 值比較，設定臨界值  $\theta_2$  為 155。此方法每計算一個區塊的 DC 值需要 4 個乘法和 3 個加法。

方法二：以式(3)計算出區塊實際的 DC 值，設定臨界值  $\theta_2$  為 155。此方法每計算一個區塊的 DC 值至少需要 72 個乘法和 63 個加法。

當區塊 DC 值和查詢影像區塊的 DC 值差異小於  $\theta_2$  時，即視為與查詢影像區塊相似的區塊。此例將分別使用這兩種方式計算圖 12 三張畫面中區塊的 DC 值，並比較這兩種方法在相似區塊比對效益上的召回值和精確值，畫面  $f_i$  中區塊  $b_j$  的召回值和精確值定義為

$$\text{recall}(f_i, b_j) = \frac{\# \text{ of } b_j \text{ relevant and retrieved blocks in } f_i}{\# \text{ of } b_j \text{ relevant blocks in } f_i} \text{ 和}$$

$$\text{precision}(f_i, b_j) = \frac{\# \text{ of } b_j \text{ relevant and retrieved blocks in } f_i}{\# \text{ of } b_j \text{ retrieved blocks in } f_i}$$

其中的相關區塊(relevant block)是利用 DCT 中全部的係數計算而決定，將畫面中區塊的 64 個 DCT 係數和查詢區塊比對，取兩者差異小於 475 的區塊視為畫面的相關區塊。每個視訊畫面對於查詢影像區塊的召回值和精確值，則是計算視訊畫面對 4 個查詢區塊召回值和精確值的平均， $\text{recall}(f_i) = \frac{1}{4} \sum_{j=0}^4 \text{recall}(f_i, b_j)$  和

$$\text{precision}(f_i) = \frac{1}{4} \sum_{j=0}^4 \text{precision}(f_i, b_j)$$

表 2. 方法一與方法二的效益比較

視訊畫面	召回值和精確值		精確值 (%)	
	方法一	方法二	方法一	方法二
$f_1$	76.2	78.9	30.3	31.2
$f_2$	81.3	84.3	31.4	32.3
$f_3$	68.1	70.2	27.8	29.6

表 2 資料來源：視訊圖象分割與查詢影像相似性比對於單詞的使用。

較，其結果顯示實際計算出區塊 DC 值的方法二在平均召回值和精確值上都略優於估算 DC 的方法一，但值得注意的是方法一計算區塊 DC 值所需的計算量遠小於方法二。

### • 高解析度的比對

經過低解析度資料的比對，本例中三張畫面皆包含相似區塊，於是進行較高解析度的比對，此例採用 4 個 DCT 係數計算和查詢影像區塊的差異，介於視訊畫面分割區塊中間位置的區塊之 DCT 係數以式(3)計算，臨界值  $\theta_3$  設為 48，意即找出視訊畫面中和查詢區塊(在 4 個 DCT 係數時)差異小於 48 的相似區塊，距離的臨界值  $\rho$  設為 20，符合條件的區塊組合再以式(1)計算其相似程度，每兩個區塊之間的距離的權重設為  $1/\rho = 0.05$ ，每兩個 DCT 係數差異的權重設為  $1/\theta_3 = 0.02$ ，比對結果三張畫面中  $f_3$  不包含  $b_{22}$  的相似區塊， $f_1$  和  $f_2$  皆包含四個查詢區塊的相似區塊。本例查詢的結果為， $f_1$  中有 4 組相似區塊組合，排序後順序為

{(10,16), (10,24), (18,16), (18,24)} (Diff=0.412),  
 {(10,16), (10,24), (18,16), (11,23)} (Diff=0.747),  
 {(10,16), (17,25), (18,16), (18,24)} (Diff=0.779),  
 {(10,16), (17,25), (18,16), (11,23)} (Diff=1.113),

畫面  $f_1$  中 4 組相似區塊組合如圖 13 所示。 $f_2$  中有 12 組相似區塊組合，排序後順序為

{(13,15), (13,23), (21,15), (21,23)} (Diff=0.412),  
 {(12,16), (13,23), (21,15), (21,23)} (Diff=0.563),  
 {(13,15), (13,24), (21,15), (21,23)} (Diff=0.591),  
 {(13,15), (12,22), (21,15), (21,23)} (Diff=0.610),  
 {(12,16), (13,24), (21,15), (21,23)} (Diff=0.725),  
 {(13,15), (13,23), (21,15), (14,22)} (Diff=0.746),  
 {(12,16), (12,22), (21,15), (21,23)} (Diff=0.751),  
 {(12,16), (13,23), (21,15), (14,22)} (Diff=0.897),  
 {(13,15), (12,22), (21,15), (14,22)} (Diff=0.913),  
 {(13,15), (13,24), (21,15), (14,22)} (Diff=0.915),  
 {(12,16), (13,24), (21,15), (14,22)} (Diff=1.049),  
 {(12,16), (12,22), (21,15), (14,22)} (Diff=1.054)。

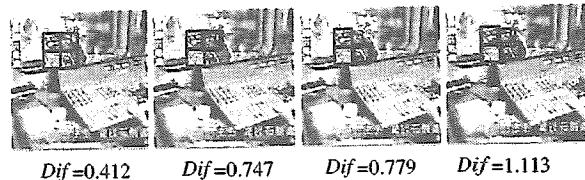


圖 13. 畫面  $f_1$  中相似區塊的有效組合

在檢查區塊比對的效益方面，採用相同於 DC 值的方法，設定差異小於 265 的區塊為相關區塊，如此在  $f_3$  中同樣找不到  $b_{22}$  的相關區塊， $f_1$  和  $f_2$  比對效益上的召回值和精確值則如表 3 所示，此處相關區塊的臨界值(265)小於低解析度時的臨界值(475)是為了說明高解析度可計算更精確的相似比對，如表 2 和表 3 的結果顯示。本例以兩層比對進行查詢處理，實際應用亦可根據需要選擇進行更多層的比對。

表 3. 高解析度採用 4 個 DCT 係數的效益評估

視訊畫面	召回值和精確值		精確值 (%)	
	方法一	方法二	方法一	方法二
$f_1$	78.1	79.2	68.8	55.0
$f_2$	79.2	70.2	55.0	29.6

## 5 結論

本篇論文提出一個結合離線和線上功能的整合系統，離線子系統利用多維空間索引，根據關鍵畫面先找出與查詢影像相似的場景，而線上子系統即時比對場景的每個視訊畫面，針對使用者輸入的查詢影像找出相似視訊畫面的部份內容。此系統具有下列五大特性：(1)可適性(adaptability)，系統可視不同的網路頻寬限制與應用環境，彈性選擇最適當模式以增加整體效益；(2)擴充性(scalability)，主從式的開放架構的使其易於建構在網路中；(3)有效性(efficiency)，無論離線或線上處理皆可在壓縮領域下運算且完全自動化；(4)可行性(effectiveness)，線上多解析度的擷取方式符合視訊壓縮標準的規格，例如 MPEG 提供 SNR、spatial 及 temporal scalability 的功能；(5)高彈性(flexibility)，系統提供區塊的近似與部份比對法則，同時兼顧查詢的效能和品質。本系統提供網路上搜尋視訊資料的設計架構，但目前尚有些限制，例如：區塊比對方法無法查詢物件的放大縮小與旋轉等情形，且無法分離出前景與背景物件。除了適用於網路環境上的視訊資料內容查詢系統，本系統同樣地可應用於文字、影像、聲音等多媒體資料庫查詢系統。未來的研究方向考慮在網路查詢系統中再加入相關性回饋(relevant feedback)的功能，以自動學習使用者在視覺感官上相似的規則，更有效地找出使用者真正想要的視訊資料。

國科會計劃編號 NSC89-2213-E-009-015

## 參考文獻

- [1] N. Beckmann, H. -P. Kriegel, R. Schneider and B. Seeger, "The R\*-tree: An Efficient and Robust Access Method for Points and Rectangles," in *Proc. of ACM SIGMOD*, Atlantic City, USA, pp.322-331, May 1990.
- [2] M. L. Cascia and E. Ardizzone, "JACOB: just a content-based query system for video databases," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 7-10, 1996.
- [3] S. F. Chang and D. G. Messerschmitt, "Manipulation and compositing of MC -DCT compressed video," *IEEE Journal on Selected Areas in Communications: Special Issue on Intelligent Signal Processing*, vol. 13, no. 1, pp.1-11, Jan. 1995.
- [4] P. J. Cheng and W. P. Yang, "A new content-based video retrieval system - data model and query processing," in *Proc. of Workshop on Software Engineering and database Systems, International Computer Symposium*, pp. 1 -122, 1998.
- [5] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast subsequence matching in time-series databases," in *Proc. of ACM SIGMOD*, Minnesota, USA, 1994.
- [6] F. Idris and S. Panchanathan, "Review of image and video indexing technique," *Journal of Visual Communication and Image Representation*, vol. 8, no. 2, pp.146-166, 1997.
- [7] V. Kobla and D. Doermann, "Indexing and retrieval of the MPEG compressed video," *Journal of Electronic Imaging*, vol. 7, no. 2, pp.294-307, April 1998.
- [8] S. Lee, M. Yang and J. Chen, "Signature file as a spatial filter for iconic image databases," *Journal of Visual Languages and Computing*, vol. 3, pp. 373-397, 1992.
- [9] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker and C. Faloutsos, "The QBIC project: Querying images by content using color, texture, and shape," in *Storage and Retrieval for Image and Video Databases*, SPIE, vol. 1908, pp.173-187, February 1993.
- [10] A. P. Pentland, R. W. Picard and S. Sclaroff "Photobook: Content -based manipulation of image databases," *International Journal of Computer Vision*, vol. 18, no. 3, pp.233-254, 1996.
- [11] J. R. Smith and S. F. Chang, "VisualSEEK: a fully automated content-based image query system," *ACM Multimedia*, Boston, MA, November 1996.
- [12] J. R. Smith, "VideoZoom Spati -temporal Video Browser," *IEEE Transactions on Multimedia*, vol. 1, no. 2, pp. 157-171, June 1999.
- [13] B. L. Yeo and B. Liu, "On the extraction of DC sequence from MPEG compressed video," *IEEE International Conference on Image Processing*, vol. 2, pp. 260-263, 1995.
- [14] A. Yoshitaka and T. Ichikawa, "A survey on content-based retrieval for multimedia databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, no. 1, pp. 81-93, 1999.
- [15] H. J. Zhang, C. Y. Low, Y. Gong, S. W. Smoliar and S. Y. Tan, "Video parsing compressed data," in *Proc. Of SPIE: Image Video Processing II* 2182, pp.142-149, 1994.