國　立　交　通　大　學

資訊工程學系

碩士論文

電視新聞及廣告分段與視訊短片搜尋之研究

The Study of TV News and Commercial Segmentation and Video Clip Search

研　究　生：　黃子源

指導教授: 傅心家 教授

中華民國 九 十 三　年 七 月

電視新聞及廣告分段與視訊短片搜尋之研究
The Study of TV News and Commercial Segmentation and Video Clip Search

研 究 生: 黃子源　　Student: Tzu-Yang Huang

指導教授: 傅心家　教授　　Advisor: Prof. Hsin-Chia Fu

國 立 交 通 大 學
資 訊 工 程 學 系
碩 士 論 文

A Thesis
Submitted to
Institute of Computer Science and Information Engineering
College of Electrical Engineering and Computer Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master
in
Computer Science and Information Engineering
July 2004
Hsinchu, Taiwan, Republic of China.

中 華 民 國 九十三 年 七 月

電視新聞及廣告分段與視訊短片搜尋之研究

研究生:黃子源　　指導教授: 傅心家　教授

國立交通大學資訊工程學系

摘要

本論文旨在發展自動區分電視新聞中的廣告段落, 以及能準確搜尋目標短片之技術. 在論文中, 我們以聲音及視訊之特徵差異區分新聞及廣告, 並以新聞節目結構上之特性, 提升分辨新聞及廣告的準確性; 另外, 我們發展出以鏡頭爲準 (shot-based) 之短片搜尋技術, 以利加強於實際的廣播節目錄影中搜尋特定視訊短片之準確性.

The Study of TV News and Commercial Segmentation and Video Clip Search

student:Tzu-Yang Huang Advisor:Prof. Hsin-Chia Fu

Institute of Computer Science and Information Engineering
National Chiao Tung University

Abstract

This paper mainly develops the technique of dividing commercials from TV news programs, and the technique of video clip searching. We use the audio and video features to distinguish between news and commercials, and then some properties of the structure of TV news programs are applied to improve the accuracy of news and commercial segmentation. And, we also develop a shot-based video clip search method to enhance the accuracy of video clip searching in the practical broadcasts.

誌　謝

感謝老師在過去兩年中對我在課業研究上的指導與教誨, 使我在碩士班這段日子受益良多.

同時也要感謝徐永煜學長, 陳岳宏學長, 曾政龍學長及賴柏伸學長經常在課業上無私的給予我幫助, 尤其是賴柏伸學長, 謝謝這些日子來你們對我的照顧; 還有可敬的學弟妹：宗儒, 揚智及宜玲, 在課業操勞之餘, 尚要幫忙處理雜事, 眞是辛苦你們了. 再次感謝老師, 學長及學弟妹們對我的幫助.

另外, 當然也不能忘了已畢業的三位學長姊：博傑, 信憲及留圓, 感謝你們分享給的我們寶貴經驗.

喔, 對了! 還有我們實驗室的榮譽成員, Chris (抱歉, 我還是記不住你的全名) , 雖然只有兩個月的相處, 但也謝謝你這兩個月帶給我們很特別的回憶.

最後是與我同甘共苦過這兩年的同學, 俊銘及聖育, 謝謝你們; 因爲有你們的陪同, 讓我在碩士班這段日子的回憶更加豐富.

目　　錄

表 目 錄

圖 目 錄

1 緒論

1.1 研究動機

電視新聞爲人們最普遍的資訊來源之一; 如何對其進行自動化整理, 以利人們更有效率的接收與搜尋人們感興趣的資訊, 已成爲一重要課題.

然則, 若不先行分開新聞節目中的新聞及廣告, 便欲對其中一部分 (新聞或廣告) 內容作進一步處理, 將會對系統整體的效能及準確率造成不良影響; 因此, 先行區分開新聞及廣告的段落, 將有助於改善此現象.

另外, 若人們想自一堆長時間的影片中找到自己感興趣的視訊短片 (video clip) 所在, 目前已有一些有效率的視訊短片搜尋方法可供使用[7][11][12]; 但這些搜尋方法在實際的視訊廣播節目, 例如, 電視新聞中, 卻可能產生搜尋準確度不佳的情況, 因爲某些視訊短片 (例如 : 廣告) 在不同時間出現時的播放速度或長度, 畫面亮度會因各種原因而有些變化, 這些變化會使得現存的一些視訊短片搜尋法失效. 因此, 發展一套能容忍上述各項變化的視訊短片搜尋方法, 才可以在實際的狀況中, 取得較佳的搜尋準確率度.

1.2 研究目標

本論文所著重的研究目標, 基於上述的原因, 主要有兩項：

1. 依據新聞及廣告的特徵及差異性, 將兩者區分開來, 以利之後對新聞或廣告的其他處理動作.
2. 發展一短片搜尋技術, 以利使用者能於實際的廣播節目錄影中搜尋到其感興趣的特定短片, 且能容忍短片發生長度變化或亮度變化等情況之影響.

1.3 章節介紹

第二章先簡介有關新聞及廣告分段 (或分類) , 以及視訊短片搜尋, 這兩項技術的相關研究. 在第三章中, 我們介紹用以區分新聞節目中新聞及廣告的一些特性, 並且提出我們的新聞及廣告分段系統. 第四章將介紹我們所發展出的, 一套能容忍視訊短片發生長度變化及亮度變化等情形, 適合用於實際視訊廣播節目中的視訊短片搜尋技術. 第五章將實作我們在第三章及第四章所提出的方法, 並列出實驗與結果分析. 最後, 於第六章總結本論文所得結論.

2 相關研究

在本章中, 我們將簡介有關新聞及廣告分段 (或分類) , 以及視訊短片搜尋, 這兩項技術的相關研究成果.

2.1 新聞及廣告分段

對於視訊影片內容的分類及分段, 可以從該影片的視訊（ video ）訊號或聲音（ audio ）訊號上開始著手分析. 藉著統計一些影像（ image ）上的物理性質, 例如畫面（ frame ）中的色彩, 紋理（ texture ）, 輪廓（ contour ）, 或是視訊上的物理性質如物體的動作（ motion ）, 攝影鏡頭的切換（ shot change ）, 乃至於聲音上的物理性質等, 配合不同影片內容如新聞或廣告等各自的特性, 我們便可依此對影片內容進行自動分類的動作, 而後連續一段時間同一類別的視訊內容, 即可被分爲同一段落, 或是進行其他後續的分析工作. 利用不同的統計方式, 可以達成分析各種視訊內容的效用. 例如, 使用 motion vector 來偵測物體移動量大的連續畫面, 用以判斷體育相關的影片內容.

S. Srinivasan 等人[2]使用 ZCR (Zero Crossing Rate) 及 STE (Short Time Energy) 等特徵 (features) 爲基礎, 能將人聲 (speech) 及樂音 (music) , 或兩者的混合段落區分出來, 這將有助於我們區分人聲爲主的新聞片段以及有樂音爲背景的廣告片段; 而 L. Lu 等人[10][5]更進一步使更多種的聲音特徵來區分樂音, 環境音, 靜音, 以及人聲, 並能做到線上 (online) 建立語者模型 (用來分辨該語者) 以及即時分段等功能.

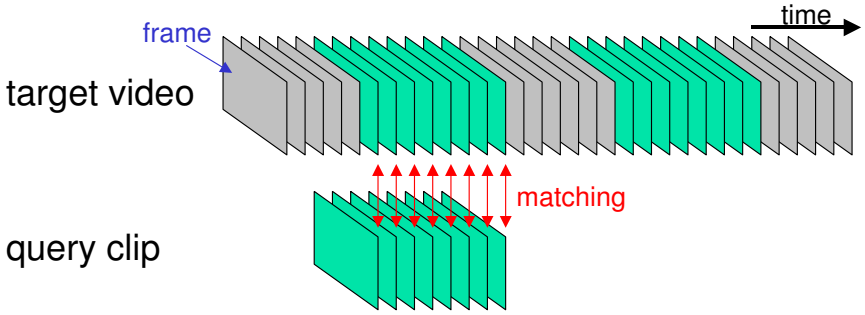
Z. Liu 等人[1]則利用聲音上的一些特徵, 建立新聞及廣告等片段各自的 HMM (Hidden Markov Model) , 直接爲這些不同類型的視訊影片片段建立模型, 用建立好的模型來分辨各視訊影片片段的類別; 而 C. Lu 等人[6]則是以視訊上的特徵爲基礎來建立各類型視訊影片的 HMM .

但是, 由於單獨使用聲音 (或視訊) 的特徵, 雖可有效的判別出如樂音及人聲, 但若用於判別新聞及廣告等無固定特性的視訊內容, 便會有其不足之處; 且要以事先建立好的模型來辨別各視訊內容的類別, 則此模型是否能夠全面性適用於各種狀況, 仍是個問題.

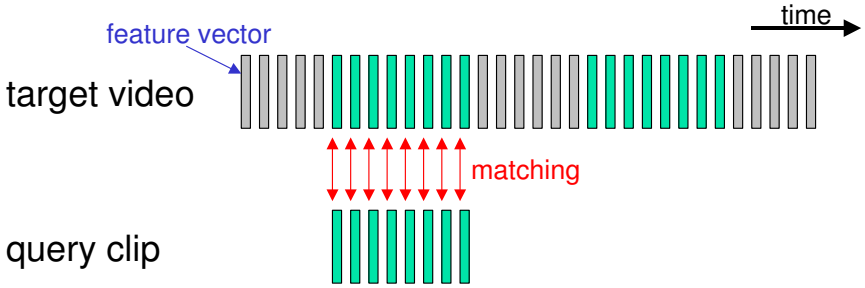
因此, 同時使用聲音及視訊上的特徵, 便有其必要性, 且最好不要事先便存在一些用以判定類別的模型.

2.2 視訊短片搜尋

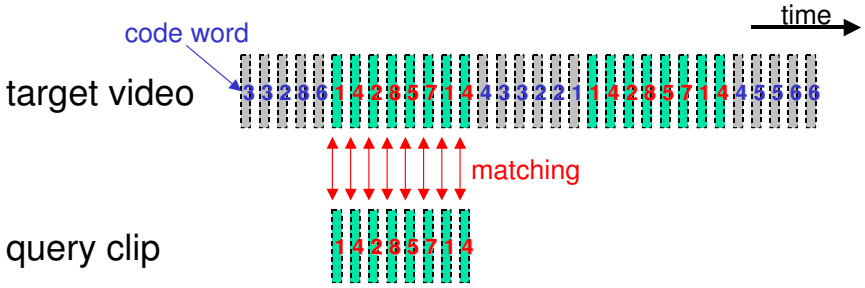
搜尋視訊短片最直接的方法, 即是以待尋視訊短片 (query video clip , 或簡稱 query clip) 每個畫面 (frame) 逐一與目標視訊影片 (stored target video , 或簡稱 target video) 中的畫面 (frame-by-frame) 進



(a) frame-by-frame 的比對方式



(b) 改以每張畫面取一特徵向量來進行比對



(c) 進一步對特徵向量做 VQ 以加速比對過程

圖 2.1: frame-by-frame 的比對方式及其加速之法.

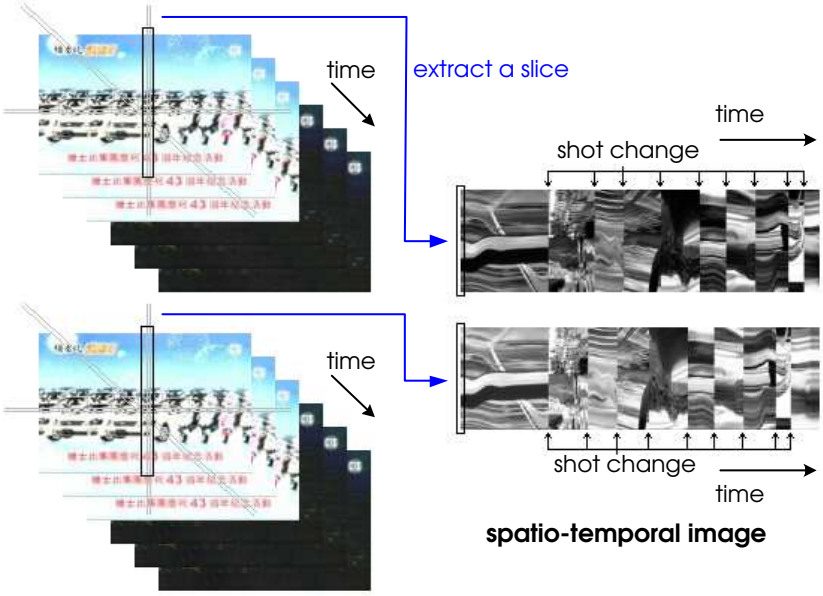


圖 2.2: 這是兩則短片, 以及分別由這兩則短片抽取出的兩段 spatial-temporal images. spatial-temporal images 的每個切片（ slice ）是由短片的每張畫面的中央橫軸, 縱軸或斜軸上的像素（ pixels ）依序取出的.

行比對, 如圖2.1 (a) 所示. 然而, 如此的方法顯然過於緩慢, 因此, 爲提升搜尋速度, 自視訊影片中的畫面中抽取各項特徵（ features ）列成一特徵向量（ feature vector ）以代表整張畫面, 便成爲了一項通用的做法, 如圖2.1 (b) 所示. 常見的一些特徵值的取法有：取畫面中各區塊各自的平均色, 取畫面中亮度或色彩分佈的統計圖（ histogram ）, 取畫面中的紋理（ texture ）, 取畫面中物件的相對位置, 或是取一張畫面的部分切片（ slices ）爲特徵 [4] (如圖2.2所示) .

若是要更進一步減少計算量, 以加快搜尋速度, 可以對特徵向量作 VQ（ Vector Quantization ）, 再對 VQ 後的 codewords 進行字串比對（ string matching ）, 如圖2.1 (c) 所示.

另外, K. Kashino 等人[12]更進一步, 對 VQ 後的 codewords 做

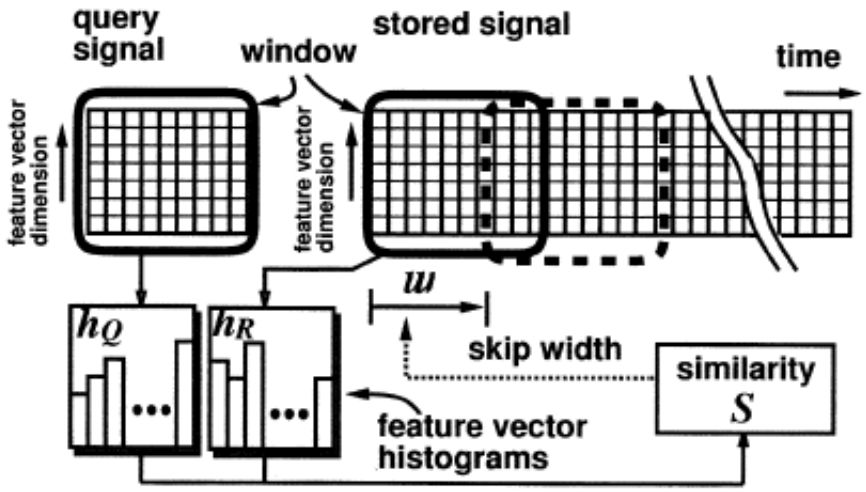


圖 2.3: 對 VQ 後的 codewords 做統計圖, 比對雙方的統計圖形越不相似, 則可跳過比對的次數 (skip width) 越大.

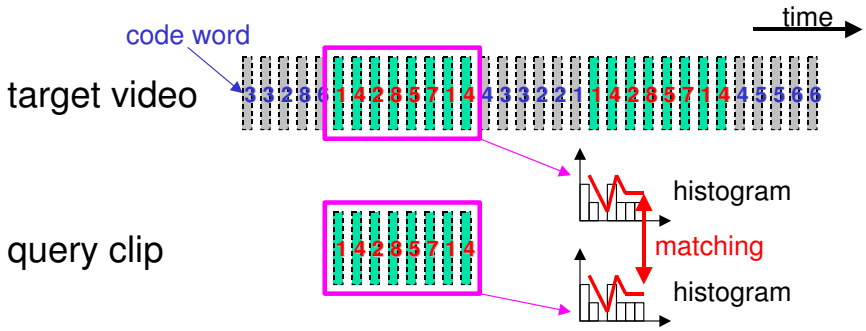


圖 2.4: 對 VQ 的 codewords 統計圖, 做分段式 PCA , 減少比對次數.

統計圖 , 並用其統計圖的分佈狀況在時間上的延續性, 減少比對次數; 如圖2.3所示, 比對雙方的統計圖形越不相似, 則可跳過比對的次數 (skip width) 越大.

而 A. Kimura 等人[11]則是再對 VQ-histogram 的軌道 (trajectory) 做分段式 PCA (segment-based PCA) , 以其分段 PCA 後的線段比對之, 再次減少比對次數, 如圖2.4所示.

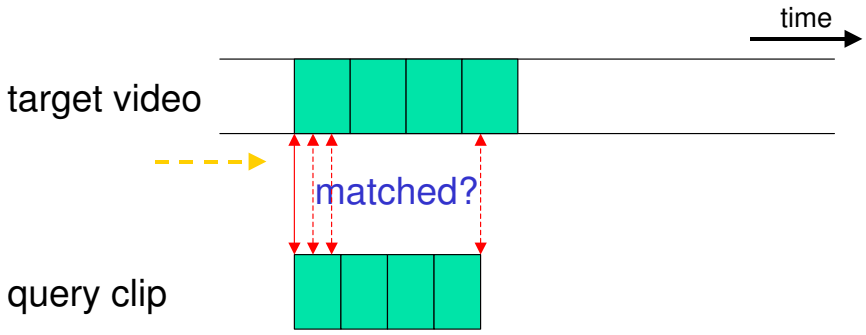


圖 2.5: 若待尋短片在 target video 中發生長度變化, frame-by-frame 的比對方式將不可行.

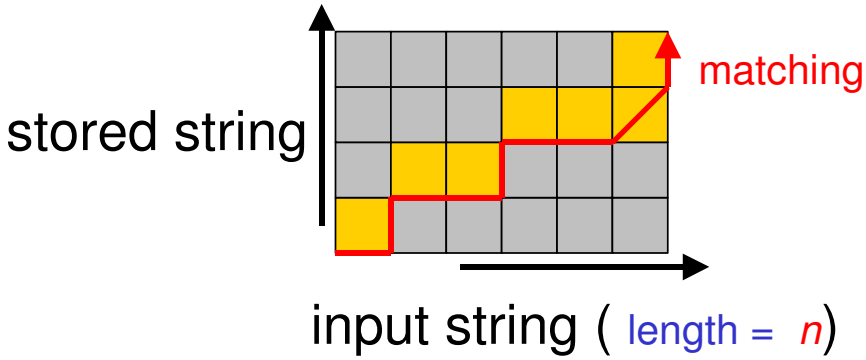


圖 2.6: 用 Dynamic-Programming Matching 的方法解決速度變化後無法正確比對的問題.

上述的方法, 可加速比對搜尋的速度; 但這些方法, 基本上仍是 frame-by-frame 的比對方法. 而在實際的視訊廣播節目中, 某些相同情節的短片 (例如：廣告) 在不同時間的播放速度或長度, 畫面亮度會因各種原因而有些變化, 甚至同一個鏡頭 (shot) 亦可能有長短變化; 因此, 若將這些 frame-by-frame 的比對方法, 應用於這些實際狀況中時, 其準確率並不一定能令人滿意 (如圖2.5 所示) .

當然, 也有人提出過解決類似問題的方法, 例如, H. Nagano 等人[9] 利用 Dynamic-Programming Matching 的方法解決類似 (速度變化) 的問題, 如圖2.6 所示; 但此法太過耗時, 以圖2.6 爲例, 比對一回即需時

$O(n^2)$.

因此, 我們便希望能找出一種視訊短片搜尋方法, 能兼顧搜尋準確度, 短片長度變化等容錯能力, 以及搜尋速度這三者.

3 新聞及廣告分段系統

在我們希望進一步處理電視新聞節目的內容之前, 有必要先將新聞節目的新聞部分及廣告部分作區隔; 因此, 在本章中, 我們將會提出一套方法, 嘗試將電視新聞節目中的新聞段落及廣告段落劃分開來.

在3.1節中, 我們會介紹我們所提出的新聞及廣告分段系統. 而3.2節, 則會對於我們提出的這套方法進行進一步的討論.

3.1 新聞及廣告分段系統

我們利用新聞及廣告的聲音及視訊特徵上的差異性來區分新聞及廣告的段落（ section ）, 並以新聞節目及廣告的結構性改善分辨兩者的準確度.

我們將區分新聞及廣告段落的流程分爲五個步驟, 每個步驟皆採用新聞及廣告的一部分特性來區分兩者, 再由後一個步驟來調整之前幾個步驟的分辨結果; 其流程如圖3.1 所示, 各流程的文字簡述如下 :

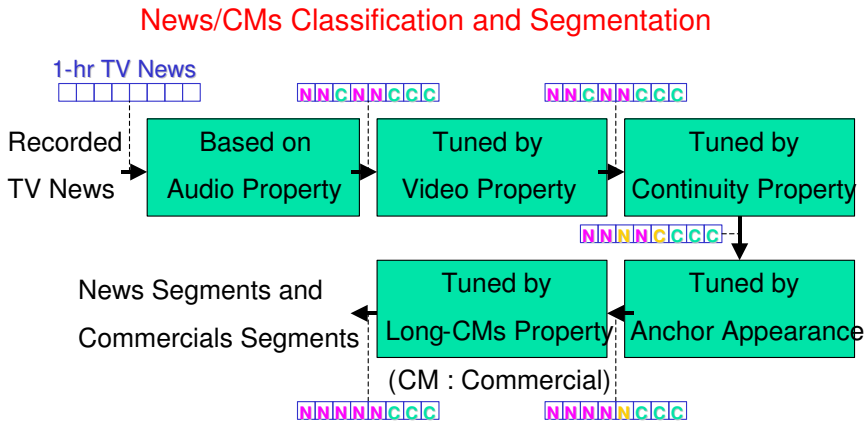


圖 3.1: 新聞及廣告分段系統流程圖.

1. 先以兩者聲音上的差異性做初步判斷依據.

2. 再輔以視訊上的差異性, 調整初步判斷.

3. 考慮時間上的連續性, 調整可能誤判部分.

4. 考慮主播出現處爲新聞, 調整判斷結果.

5. 考慮廣告段落的長度, 調整判斷結果.

另外, 由於新聞及廣告的部分特性, 需要有一段時間 (10秒以上) 爲基本單位 - 我們稱之爲一個片段 (piece) - 來做統計, 連續的新聞或廣告片段的序列, 即爲一新聞或廣告的段落; 因此, 上述的五個步驟的區分結果, 只能精確至此基本單位的時間. 所以在完成以上流程後, 我們還需要補上一個步驟, 將區分出來的各新聞或廣告段落的起始及結束時間作適當的調整.

以上這些步驟, 將在以下各小節中, 依序詳述.

3.1.1 新聞及廣告的聲音特性

觀察新聞及廣告在聲音 (audio) 上的差異之處, 不難發現, 廣告由於希望加強觀衆對它的印象, 常常會有背景樂音來襯托其主題; 相對地, 新聞多半爲主播或外景記者等人聲播報爲主, 且由於新聞有其時效性, 須在短時間內編輯完成, 故較少有機會加入背景樂音.

因此, 我們便打算利用此一特性, 初步區分新聞及廣告兩者.

我們實際採用的聲音特徵 (audio features) 爲 ZCR (Zero Crossing Rate) 的變化率 (variation rate) 及 STE (Short Time Energy) . 根據上述的觀察結果, 配合我們所採用的聲音特徵, 可以整理出以下幾點特性 :

1. 人聲 (新聞) 的 ZCR 變化率較高, 樂音 (廣告) 則反之.
2. 極高的 ZCR 變化率 + 極低的 STE = 靜音 (silence) , 這是用來輔助的特性.

人聲中的氣音或濁音, 其 ZCR 值甚高, 而清音部分的 ZCR 值則較低, 故相對於樂音而言, 人聲的 ZCR 值變化甚大.

在這裡, 我們只採用了兩種聲音特徵來協助判斷人聲及樂音, 主要是因爲, 繼續加入其他聲音特徵, 並不能加強改善判斷結果, 反而可能使結果變差; 我們曾試過其他的聲音特徵來判斷人聲及樂音, 例如 LPC (Linear Predictive coefficients) 及 LSP (Linear Spectral Pairs) [10], 但結果並不理想 (判斷正確率小於60%) .

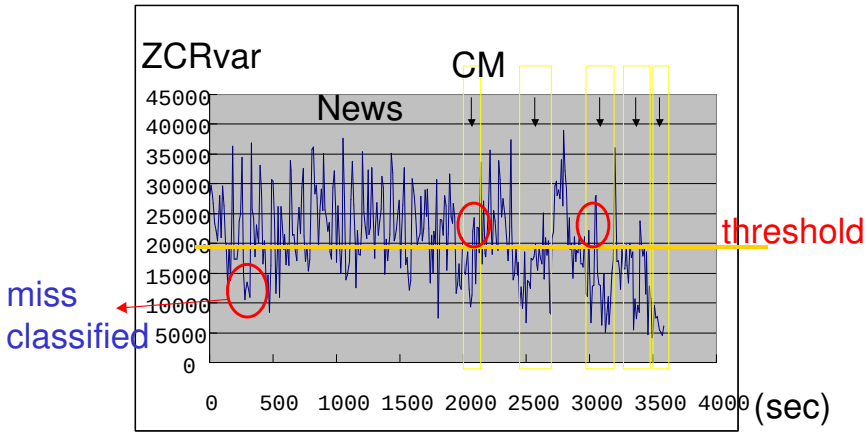


圖 3.2: 一小時新聞的 ZCR 變化率統計圖.

由上述的特性, 我們可以初步假設, ZCR 變化率高於某個門檻值 (threshold) 的部分即爲新聞, 反之則爲廣告.

(另外, 我們濾出靜音的部分, 是爲了避免靜音部分之數據影響對新聞及廣告的判斷, 且這裡所找出的靜音處, 將在3.1.6小節有所用途.)

如圖3.2 所示, 這是我們以約10秒爲一個基本統計單位 (往後幾個小節若未特別提及, 則亦是沿用此基本單位) , 對一個小時 (3600秒) 的電視新聞節目所做的 ZCR 變化率統計圖, 以及我們用以劃分新聞及廣告的 threshold . 圖中共有五段廣告段落 (以箭頭所指的段落) , 其餘部分則爲新聞.

由圖3.2 可以觀察到, 在大部分的情況下, 新聞部分的 ZCR 變化率會高於 threshold , 而廣告部分的 ZCR 變化率則會低於 threshold ; 但仍會有一小部分的情況不同, 這部分將會導致誤判. 因此尚需之後的步驟來輔助更正.

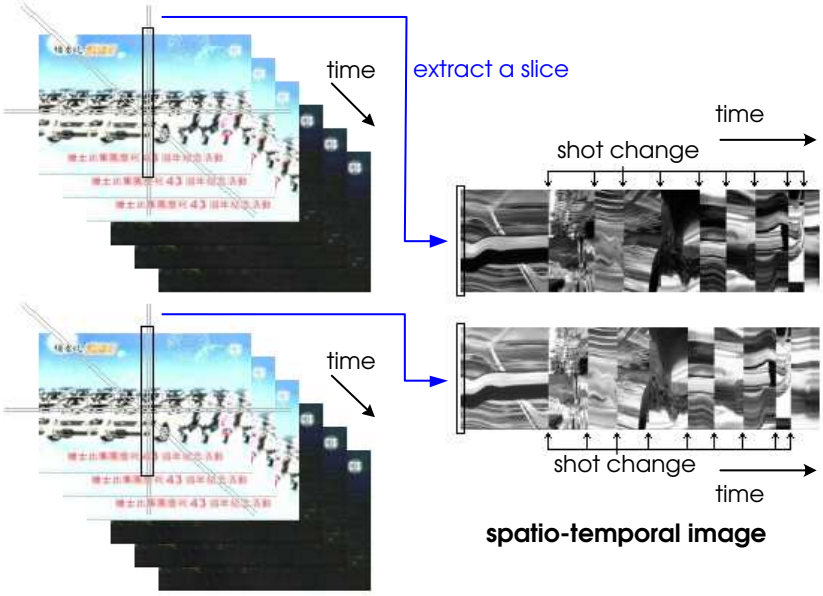


圖 3.3: 這是由這兩則短片抽取出的兩段 spatial-temporal images. spatial-temporal image 內紋理的垂直不連續面爲出現 shot change 之處.

3.1.2 新聞及廣告的視訊特性

對廣告而言, 由於它需要在短時間內盡可能的將其商品的特點展示出來, 因此廣告的整體節奏通常較快, 其鏡頭切換 (shot change) 的速度也較快.

故而, 我們可以假設, Shot Change Rate (單位時間內的鏡頭切換次數) 較高的部分, 較有可能是廣告. 我們可藉此調整前一個步驟所得之判斷結果.

我們用以找出 shot change 的方法, 是利用 Ngo 等人[4]所提出的 spatial-temporal image 其特性找出來的.

如圖3.3所示, 在 spatio-temporal images 內紋理 (texture) 的垂

直不連續面 (意即相鄰兩 slices 中, 水平方向相鄰的各點亮度差異過大) , 皆可能爲出現 shot change 之處. 我們判斷這種不連續面的方式, 是以橫軸, 縱軸及斜軸三個 spatio-temporal images 中各相鄰 slices 之間的差異大小來決定. 而在這裡, 由於我們希望取得較有可能爲 shot change 之處, 故而我們會對 slices 間的差異度訂定一個較高的 threshold TSC , 若相臨兩 slices 的差異度超出此 threshold 時, 即認定此處有 shot change , 我們稱它爲一個 "tight-shot change" ; 以此法所切分的各個 shots , 我們稱之爲 "tight-shots" .

(我們計算 spatio-temporal images 內兩相鄰 slices 的差異度, 是計算兩 slices 相鄰像素 (pixels) 亮度灰階值的差值總和.)

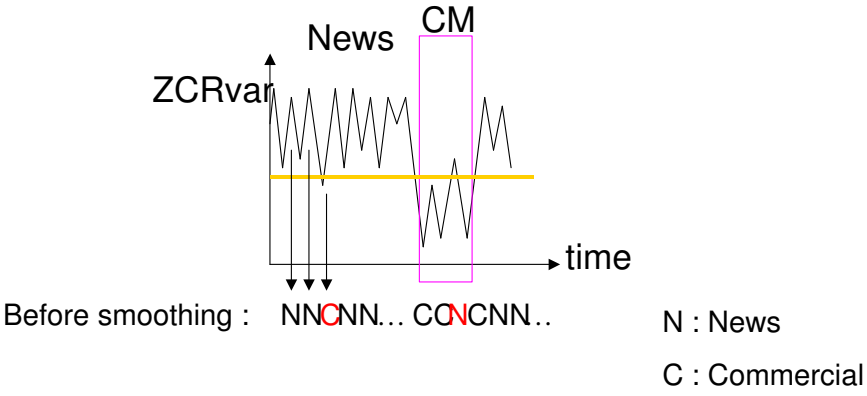
而 shot change rate 的統計方法, 是每隔10秒做一次統計, 每次統計所觀察的時間單位爲30秒 (所以會有重複統計處) , 其時間單位內的 "tight-shot change" 發生次數.

我們依 shot change rate 的結果來調整前一步驟中的判斷準則.

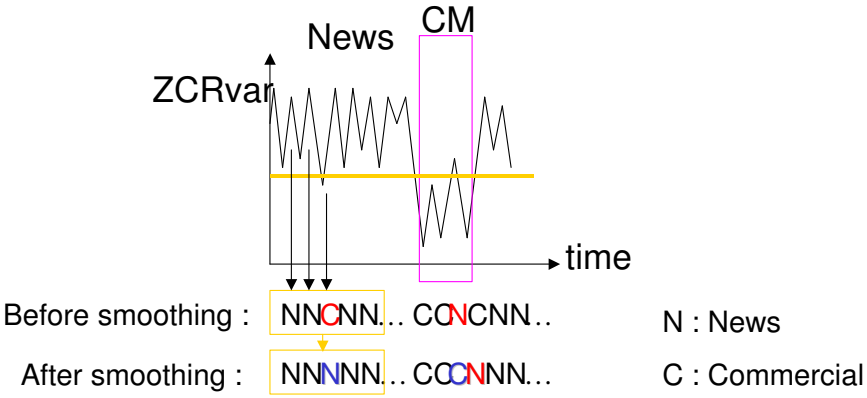
上一個步驟中, 我們爲 ZCR 變化率設定一 threshold , 低於此值者爲廣告; 現在, 我們認爲 shot change rate 高處較可能爲廣告出現處, 故而, 我們可以在 shot change rate 較高處, 將 ZCR 變化率的 threshold 調高 (使得此處較容易被判斷爲廣告) , 達到調整前一步驟所得結果的目的.

如圖3.4 所示, 在 shot change rate 較高處, 將 ZCR 變化率的 threshold 調高, 相當於將該處的 ZCR 變化率調低; 如此將可更正部分在前一步驟誤判的結果.

圖 3.4: 在 shot change rate 較高處, 將 ZCR 變化率的 threshold 調高, 相當於將該處的 ZCR 變化率調低; 如此將可更正部分在前一步驟誤判的結果.



(a) 以新聞及廣告連續性的特性修正判斷結果前



(b) 以新聞及廣告連續性的特性修正判斷結果後

圖 3.5: 以新聞及廣告連續性的特性修正判斷結果.

當然, 在此一步驟中, 仍有一些誤判的部分沒有更正; 且有一小部分 shot change rate 較高處, 其實並非廣告, 在經過此步驟調整後, 反而可能導致誤判.

3.1.3 新聞及廣告的時間連續性

在經過上述兩個步驟的處理之後, 我們已可以將新聞節目中的各個片段 (例如：一個小時的新聞節目, 以10秒爲一片段, 共360個片段) , 個別分類爲新聞或廣告, 但仍可能有一些誤判的情況存在, 如圖3.5 (a) 所示.

由圖3.4 觀察可知, 被誤判的片段通常是單獨出現或是僅連續出現極短的時間, 因此可能發生如圖3.5 (a) 中, 在一整段新聞中突然冒出一個廣告片段的情況; 所以, 如果我們能將圖3.5 (a) 中的某片段判斷結果, 再參考該片段前後數個片段的判斷結果, 選取當中佔多數的類別 (新聞或廣告) 做爲該片所屬類別, 如圖3.5 (b) 所示, 則應該可以消除此種被誤判的情形.

不過, 此步修正的動作, 對於新聞及廣告段落交界的部分, 會有較差的影響.

3.1.4 主播出現處應爲新聞

通常而言, 新聞主播出現處, 即爲新聞, 而非廣告; 因此, 若我們能知道主播出現的時間點, 便可判斷該處的片段爲新聞片段.

根據鄭士賢的論文所述[8], 在新聞節目的時段中, 新聞主播出現的部分, 通常也是整個新聞節目中出現的人物裡, 所佔時間最長的. 因此, 在其論文中, 便將新聞節目時段的聲音切分爲許多極短 (數秒) 的聲音片段, 並將這些聲音片段分群; 之後, 只要挑選分群之後的最大群, 應當就是主播出現的片段.

我們利用其結果來輔助判斷, 令主播出現處的片段爲新聞片段, 如圖3.6 所示. 如此可修正部分被誤判的新聞片段.

3.1.5 廣告段落的長度

廣告段落通常會持續一段時間以上 (例如 : 一分鐘以上) , 而不會一

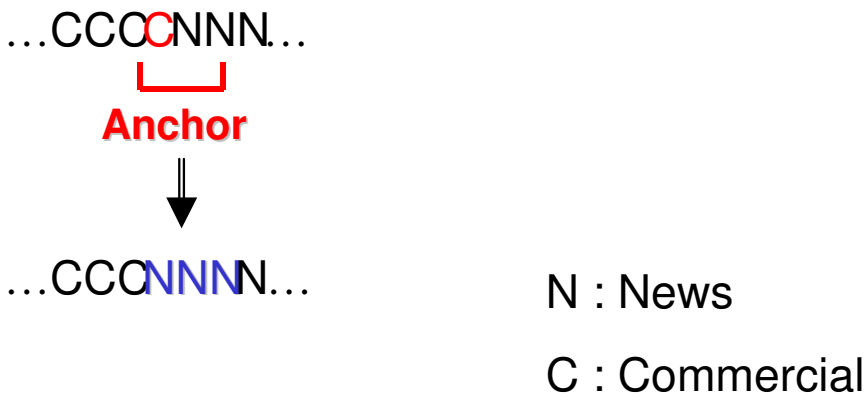


圖 3.6: 將前一步驟各時間片段的判斷結果, 以主播出現處爲新聞的特性修正之.

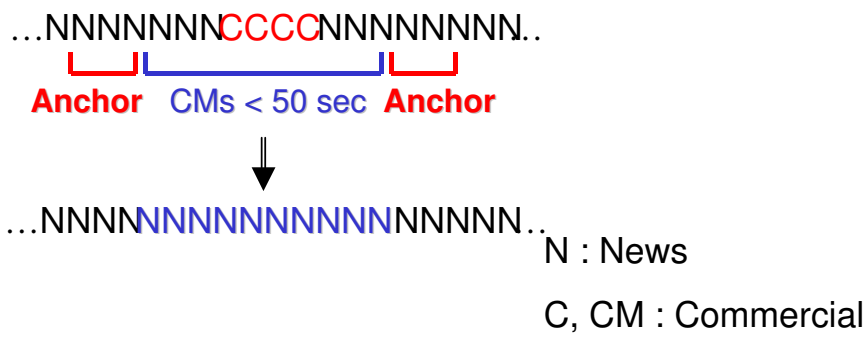
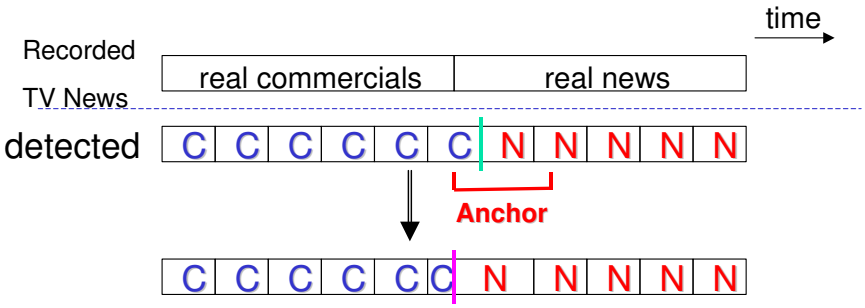


圖 3.7: 在同一節的節目中出現的廣告總長不夠長, 通常爲誤判, 應予修正.

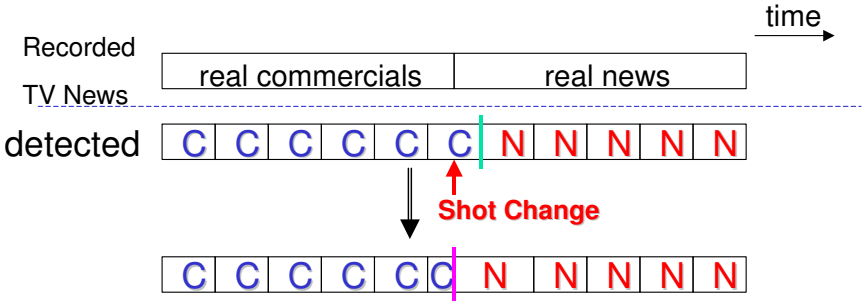
次僅播放一小段時間.

因此, 我們再利用此特性, 對我們的判斷結果做一次修正. 我們利用前一步驟所找到的主播片段來將新聞節目分段, 如果一段節目中的廣告片段出現的時間總和, 沒有超過一定時間 (如：一分鐘) 以上, 則令該段中的所有片段皆判斷爲新聞片段, 如圖3.7 所示.

至此, 我們已將新聞節目中的片段分類完成, 且有不錯的準確度.



(a) 以鄰近的主播出現時間點爲新聞及廣告段落分段點



(b) 以鄰近的鏡頭切換處 (或靜音處) 爲新聞及廣告段落分段點

圖 3.8: 修正新聞及廣告段落分段點.

3.1.6 新聞及廣告段落的斷點

上述之聲音及視訊之特徵, 均需以一小段時間 (例如 : 10秒) 爲基本的統計單位, 才能具有較可靠的準確度; 因此, 之前的五個步驟, 對新聞及廣告段落的分界點, 僅能精確到一個片段的時間單位.

在上述五步驟將新聞及廣告作粗略分段之後, 我們需再找適當的時間點, 做爲各分段的開頭及結尾.

而我們選擇各分段開頭及結尾適當的時間點之方式如下述 :

1. 若該分段開頭或結尾處, 接近主播結束或開始談話處, 則爲最合適之開頭或結尾時間點, 如圖3.8 (a) 所示.

2. 其次, 則以分段開頭或結尾處附近之靜音處, 或是鏡頭切換 (shot change) 處爲較佳的分段時間點, 如圖3.8 (b) 所示.

3.2 討論

我們將處理流程如此排序的原因如下 :

處理流程應以新聞節目 ”全面性” 的新聞或廣告特性, 做爲一開始粗略區分新聞及廣告的依據; 之後再以 ”局部性” 的特性, 對前一步驟的判斷結果做調整或細部修正. 將 ”局部特性” 提前處理, 較容易造成 ”顧此失彼” 的情形. 因此主播的特性及廣告長度的特性, 應當在聲音及視訊等特性的流程之後處理.

另外, 新聞及廣告的連續性, 此特性並無法單獨用來判斷新聞及廣告, 因此也需排在聲音及視訊特性的流程之後處理; 而由於主播出現處的特性, 可以用來修正 ”新聞及廣告的連續性” 此步驟所造成的, 在新聞及廣告段落分界處的錯誤, 因此被排在其後處理.

4 以鏡頭爲準的視訊短片搜尋技術

此章介紹我們所發展出來的一套以鏡頭爲準的視訊短片搜尋（ shot-based video clip search ）技術. 在第二章的相關研究中, 我們已介紹過了一些常用於視訊短片搜尋的方法, 但這些搜尋方法, 主要皆是以 frame-by-frame 的比對搜尋爲基礎, 這使得它們在面對視訊影片中的短片播放速度或長度可能有變化的實際情況時, 無法準確的得出正確的搜尋結果; 因此, 我們所設計的搜尋法便以此爲切入點, 試著發展出一套同時兼顧搜尋準確率, 容錯能力 (能容忍電視台播放速度, 亮度及剪接長度上的些許變化) 及搜尋速度的搜尋方法, 以期能在有限的時間內, 利用使用者給定的待尋視訊短片, 快速且準確的在電視節目的錄影中, 找到使用者所要的相同視訊短片.

我們在4.1節介紹在實際的視訊廣播節目中可能發生的短片播放速度, 亮度及剪接長度變化情形. 然後在4.2節介紹我們所提搜尋方法的原理及演算法.

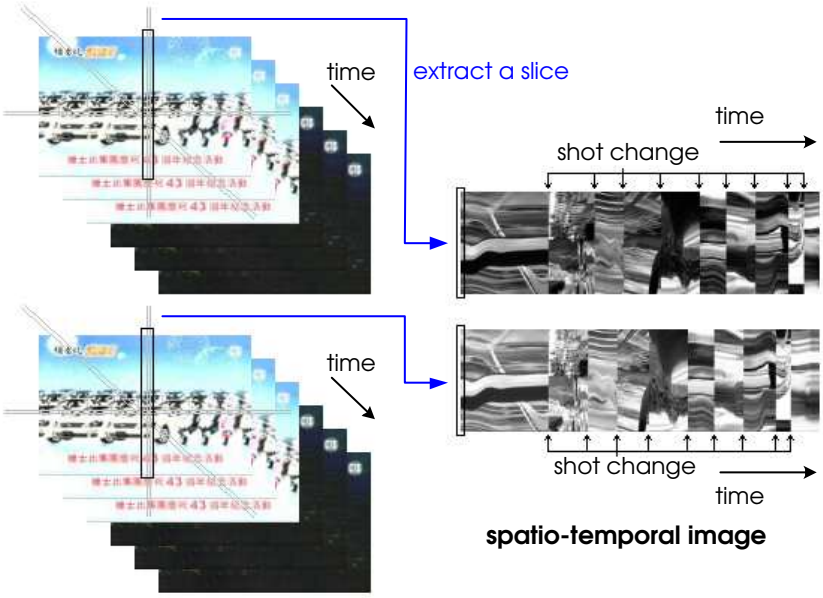


圖 4.1: 這是兩則具有相同情節的短片, 以及分別由這兩則短片抽取出的兩段 spatial-temporal images. 由兩則短片的 spatial-temporal images 相互對照可知, 這兩則短片的亮度及各鏡頭的播放長度, 有些微的差異.

4.1 視訊廣播節目中的短片播放變化情形

在一般的視訊廣播節目中, 偶爾會有一些特定的短片內容會一再重複出現, 例如 : 廣告, 節目片頭片尾, 節目預告, 政令宣導等. 而在某些重複出現的相同短片中, 其播放速度, 畫面亮度或播放長度, 可能會有些許不同; 甚至, 在兩則相同的短片中, 某些鏡頭 (shots) 的剪輯長度, 也會有不同; 就如圖4.1 所示, 兩段 spatial-temporal images , 分屬兩段不同時段播出的相同廣告, 具有相同的情節, 但亮度及播放長度有少許變化. 這些變化會令原本內容應當完全一致的兩則短片變得不甚相同, 但在人類 (觀衆) 的認知上, 這兩則短片仍是 ”相同” 的.

在上述的這種情形下, 若使用者有一則待尋視訊短片, 欲利用第二章

所述的一些搜尋方法, 找到於目標視訊節目中曾出現的 ”相同” 視訊短片, 極有可能會找不到, 因爲該則短片於目標視訊節目中可能出現如上述的一些變化.

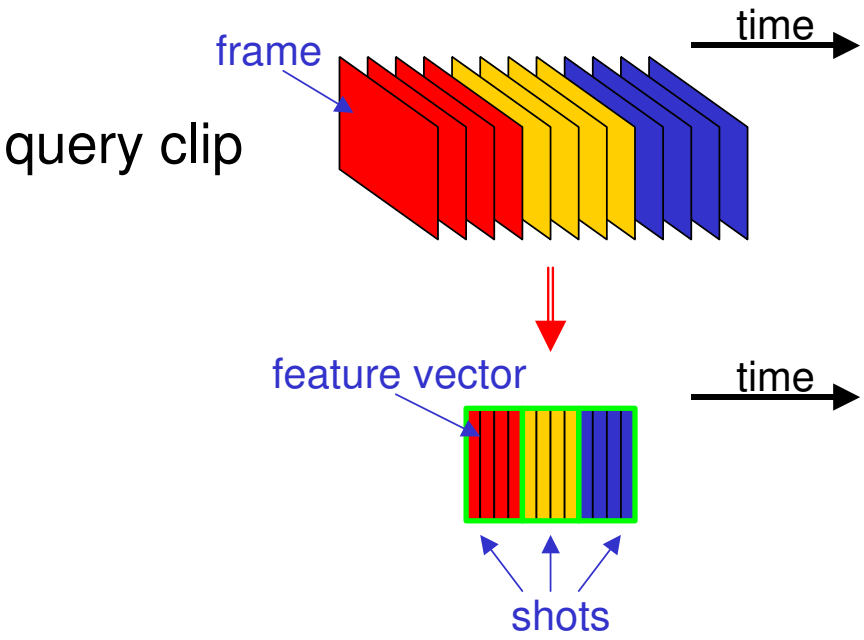
4.2 以鏡頭爲準的視訊短片搜尋技術

由於在實際的視訊廣播節目中, 視訊短片的長度變化等情況難免, 以 frame-by-frame 的比對搜尋不可行, 故此, 我們改用一種以鏡頭爲準的 (shot-based) 比對搜尋法. 以下, 我們在4.2.1小節解釋我們爲何使用 ”shot-based” 的方法, 以及該搜尋法的基本概念; 之後, 在4.2.2小節, 我們將介紹我們所使用的特徵向量; 4.2.3小節說明各項前處理步驟; 4.2.4小節則詳述搜尋階段的比對流程; 最後在4.2.5小節討論此搜尋法的時間複雜度.

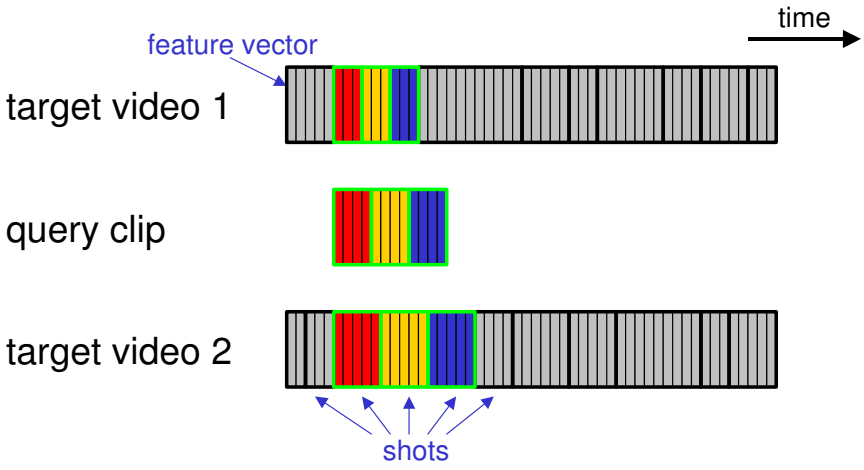
4.2.1 基本概念

假設有使用者希望於已錄影存檔的目標視訊影片 (stored target video, 之後將簡稱爲 target video) 中, 找到使用者欲搜尋的待尋視訊短片 (query video clip, 之後將簡稱爲 query clip) ; 但在 target video 當中的該則短片可能已有長度變化等情形下, 使用 frame-by-frame 的搜尋比對法並不適合, 如圖4.2 所示.

然而, 若 query clip 及 target video 中的 shots 已知, 則可依 query clip 中的各 shots 及其排列順序, 做爲搜尋標的; 這是因爲, 即使視訊短片的長度有變 (或是某些 shots 的長度有變) , 但短片中的 shots 順序不變.



(a) 待尋視訊短片



(b) 視訊短片於視訊影片中可能會有長度變化

圖 4.2: 如圖, frame-by-frame 的搜尋比對法於 (b) 的情況下並不適用.

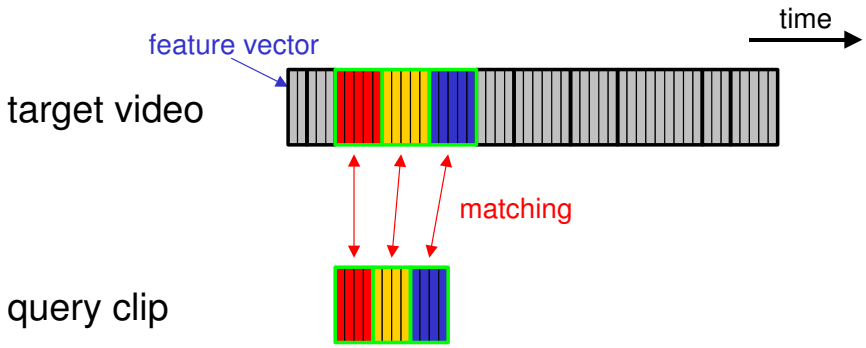


圖 4.3: Shot-based 比對法示意圖.

因此, 我們若能將 query clip 及 target video 中的 shots 一一比對, 如圖4.3 所示, 則不難找出正確的結果 (當然, 這是指在理想狀況下的結果) .

4.2.2 特徵向量

我們會對 query clip 及 target video 等, 各別取出它們的 spatio-temporal images, 包括橫軸, 縱軸及斜軸.

之後, 將橫軸, 縱軸和斜軸的 spatio-temporal images 並排, 合併成一張包含三軸的 spatio-temporal image; 我們將從這個合併後的 spatio-temporal image 取出我們所需的特徵向量.

我們取此 spatio-temporal image 中的一行 (slice) 或數行爲一個 segment , 針對每個 segment 取出一個特徵向量.

假設在一個 segment 中的像素 (pixels) 有 $t \times n$ 個, 其中 t 爲一個 segment 所含的行數 (slices) , 而 n 則爲每一行 (slice) 所含的像素個數. 然後, 我們將該 segment 切成 $n/4$ 個 $t \times 4$ 像素的方塊,

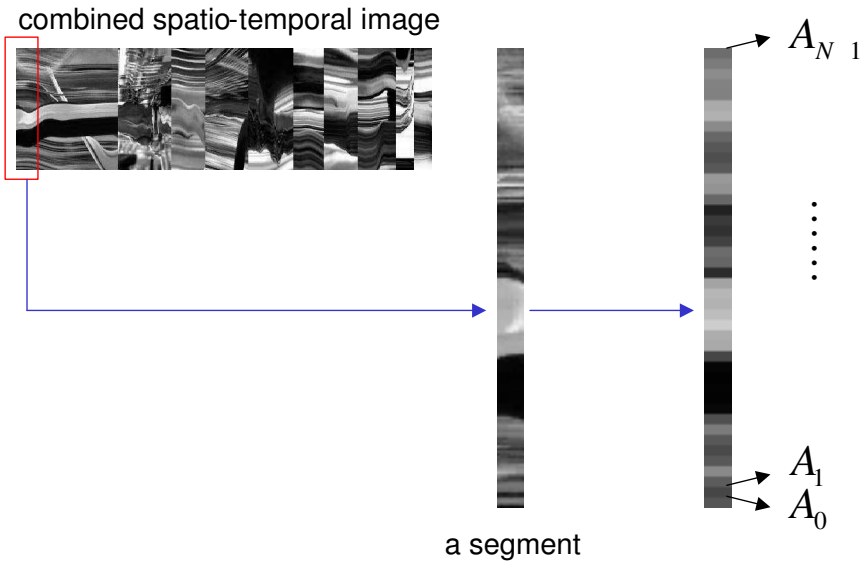


圖 4.4: 對 combined spatio-temporal image 的每個 segment 取出 A_0 到 A_{N-1} 的數值.

每個方塊取一平均值, 將這 $1 \times (n/4)$ 個方塊的平均值依序命名爲 A_0 到 A_{N-1} , $N = n/4$, 如圖4.4 所示. 最後, 對數列 (A_1, A_2, ... , A_{N-1}) 做 1D-DCT (one-dimensional Discrete Cosine Transform) , 取得 :

$$B_k = \sum_{i=0}^{N-1} 4 \cdot A_i \cdot cos(\frac{\pi k}{2N}(2i+1))$$

而 (B_1, B_2, ... , B_{10}) 即爲我們所取的特徵向量.

以此特徵向量來代表整個 segment , 可以有效減少比對時的計算量.

另外, 我們所取的特徵向量, 也具有能容忍發生亮度變化的能力, 其說明如下 :

首先, B_0 爲 DCT 過程的 DC coefficient , 它可代表 DCT 轉換前, 該 segment 的亮度總量, 故而 B_0 的值會隨亮度變化而改變, 因此, 我們

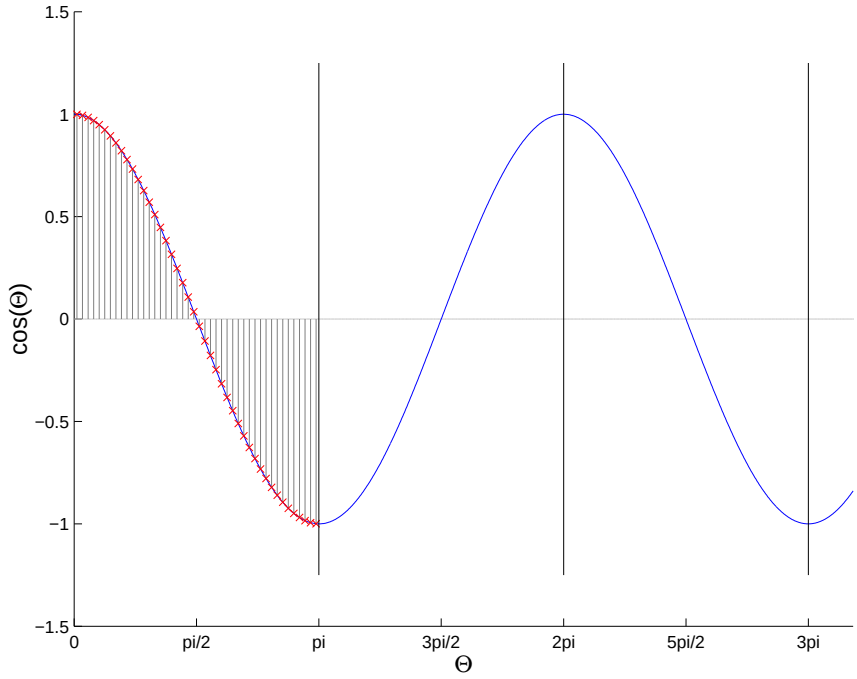


圖 4.5: 當 k 爲自然數, 且 N 夠大時, $\sum_{i=0}^{N-1} cos(\frac{\pi k}{2N}(2i+1))$ 其值爲零.

不將 B_0 列入我們的特徵向量中.

其次, 假設我們在另一則相同故事, 但亮度不同的短片中取得與上述 segment 對應位置上的數列 (A'_1, A'_2, ... , A'_{N-1}) , 以及其特徵向量 (B'_1, B'_2, ... , B'_{10}) , 其中 $A'_i = A_i + C$ (亮度變化) , 則可得 :

$$\begin{aligned} B'_k &= \sum_{i=0}^{N-1} 4 \cdot A'_i \cdot cos(\frac{\pi k}{2N}(2i+1)) \\ &= \sum_{i=0}^{N-1} 4 \cdot (A_i + C) \cdot cos(\frac{\pi k}{2N}(2i+1)) \\ &= B_k + 4 \cdot C \cdot \sum_{i=0}^{N-1} cos(\frac{\pi k}{2N}(2i+1)) \\ &= B_k \\ &(k = 1, 2, ..., 10) \end{aligned}$$

由圖4.5 可知, 當 k 爲自然數, 且 N 夠大時, 上式中的 $\sum_{i=0}^{N-1} cos(\frac{\pi k}{2N}(2i+1))$ 其值爲零.

故可知, 若該 segment 的各像素亮度變化同爲一定值 C (在我們的先導實驗中, 我們發現在實際的視訊廣播節目中所發生的亮度變化, 多爲此類情形) , 則我們所取的特徵向量並不會受到影響.

4.2.3 前處理

在開始比對 query clip 及 target video 前, 我們必須先做一些前處理的動作, 包括抽取特徵向量, 以及判斷 shots 等工作; 前處理完成後, 其所得之特徵向量及 shots 等資訊, 皆可在往後的搜尋中重複使用.

我們的前處理主要分爲三部分, 其一爲針對 query clip 及 target video 共同的前處理步驟, 其二爲針對 target video 所做的前處理, 其三則爲針對 query clip 所做的前處理.

query clip 及 target video 共同的前處理步驟, 即是將該影片或短片中的 spatio-temporal images 擷取出來, 並同時抽取我們所需的特徵向量.

而針對 target video 所做的前處理, 則是依靠前一步驟中所得之 spatio-temporal images 的特性, 判斷 target video 中的所有可能的 shot change 之處. 如圖4.1 所示, 在 spatio-temporal images 內紋理 (texture) 的垂直不連續面 (意即相鄰兩 slices 中, 水平方向相鄰的各點亮度差異過大) , 皆可能爲出現 shot change 之處. 我們判斷這種不連續面的方式, 是以橫軸, 縱軸及斜軸三個 spatio-temporal images 中各相鄰 slices 之間的差異大小來決定. 而在這裡, 由於我們希望取得任何可能爲 shot change 之處, 故而我們會對 slices 間的差異度訂定一個較低的

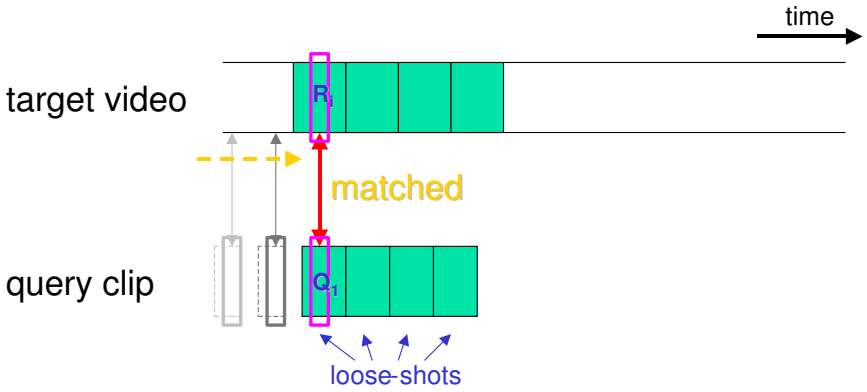
threshold LSC , 若相臨兩 slices 的差異度超出此 threshold 時, 即認定此處有 shot change , 我們稱它爲一個 "loose-shot change" ; 以此法所切分的各個 shots , 我們稱之爲 "loose-shots" .

然後, 在針對 query clip 所做的前處理中, 我們也以同樣的方法, 找出其 "loose-shots" .

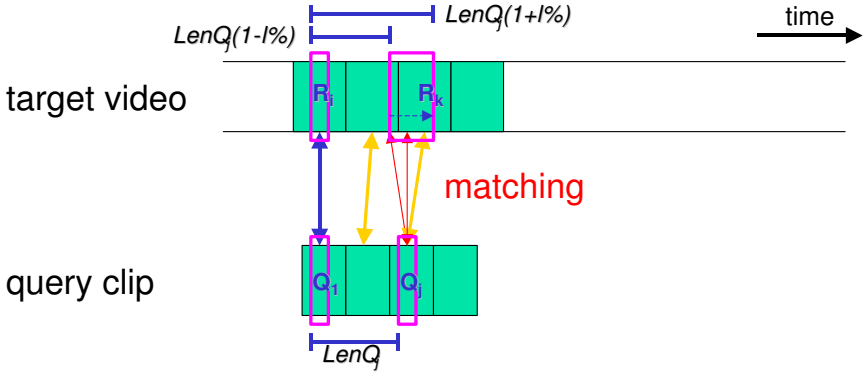
4.2.4 比對流程

我們所提出的視訊短片搜尋技術, 其比對流程示意圖如圖4.6 所示, 詳細演算法如下 :

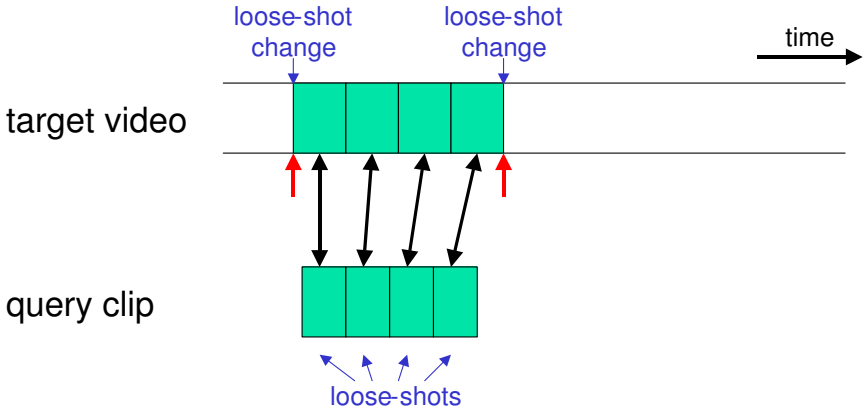
1. 允許每個 shot 的最大長度變化率爲 $l\%$.
2. 自 query clip 的第一個 loose-shot , 取其中央位置之 segment (的特徵向量) Q_1 , 依時間上的順序分別與 target video 的各 segment R_i 比對其相似度 $sim(Q_1, R_i)$. $sim(Q_1, R_i) := 1 - \frac{dist(Q_1,R_i)}{MaxDist}$, $dist(Q_1, R_i)$ 爲 Q_1 及 R_i 之間的歐氏距離, $MaxDist$ 爲兩向量之間可能的最大距離. 若 $\exists R_i$, 能使得 $sim(Q_1, R_i) \geq SSim$ ($SSim$ 爲一 threshold) , 則跳至下一步驟; 否則跳至最後一步驟.
3. 令 Q_j 爲 query clip 第 j 個 loose-shot 其中央 segment 的特徵向量; 令 $LenQ_j$ 爲 Q_j 與 Q_1 所代表的 segments 在時間軸上的距離.
4. 令 R_k 爲 target video 某 segment 的特徵向量; 令 $LenR_k$ 爲 R_k 與 R_i 所代表的 segments 在時間軸上的距離.



(a) 比對 query clip 第一個 loose-shot 居中的 segment



(b) 比對 query clip 其他 loose-shots 居中的 segments



(c) 爲搜尋到的短片定出頭尾時間

圖 4.6: 以鏡頭爲準的視訊短片搜尋比對法.

5. 令 $j := 2$.

6. $\forall R_k, LenQ_j \cdot (1 - l\%) \leq LenR_k \leq LenQ_j \cdot (1 + l\%)$, 比對 Q_j 與 R_k 的相似度 $sim(Q_j, R_k)$. 若 $\exists R_k$, 能使得 $sim(Q_j, R_k) \geq SSim$, 則跳至下一步驟; 否則 $i := i + 1$, 跳至第 2 步.

7. 若 query clip 的最後一個 loose-shot 爲第 j 個 loose-shot , 則跳至下一個步驟; 否則 $j := j + 1$, 跳至上一步.

8. 至此步驟, 表示有一則比對結果吻合 (matched) 的短片被找到. 令此短片的開頭爲目前 R_i 所在之 loose-shot 的開頭處; 令此短片的結尾爲目前 R_k 所在之 loose-shot 的結尾處. 記錄其位置.

9. $i := i + 1$, 跳至第2步.

10. end

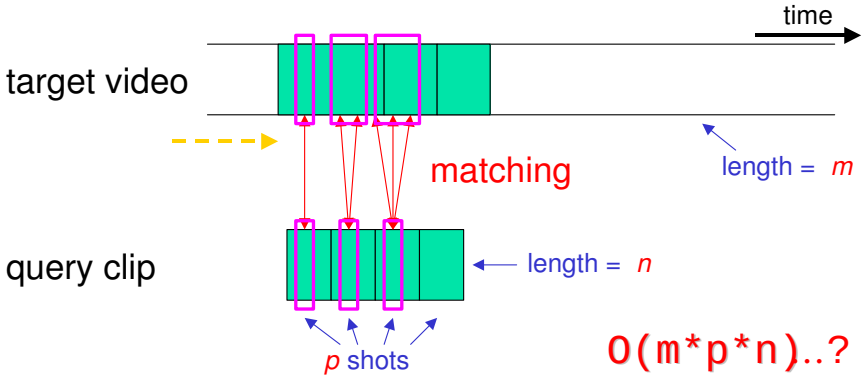
以這種 shot-based 的比對方式進行搜尋, 便可有效的避免大部分因電視節目播放速度不同而造成的問題.

4.2.5 時間複雜度

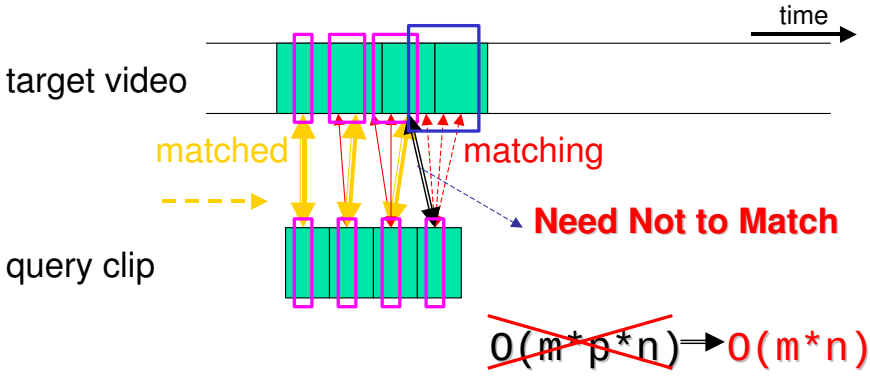
假設 target video 的長度爲 m , query clip 的長度爲 n , 而 query clip 中共有 p 個 shots (loose-shots) .

則以 frame-by-frame 的搜尋方式進行比對, 其完成搜尋比對所需的時間複雜度應爲 $O(m \cdot n)$.

而以我們所提出的搜尋方法進行比對, 如圖4.7 (a) 所示, 則其時間複雜度似爲 $O(m \cdot p \cdot n)$; 但若考慮各 shots 的順序不會倒置, 則如圖4.7



(a) 越後面的 shot 需比對範圍越大



(b) 但前後兩個 shots 的比對範圍不需重疊

圖 4.7: 我們所提出的搜尋方法, 其時間複雜度爲 $O(m \cdot n)$.

(b) 所示, 前後兩個 shots 重疊比對區域的情形也不會發生, 故而我們所提出的搜尋方法, 其所需的時間複雜度最多也只需 $O(m \cdot n)$.

5 系統實作與實驗結果分析

5.1 新聞及廣告分段系統實作與實驗結果

我們實作了新聞及廣告分段系統, 並以93年5月底到6月份的華視午間新聞及晚間新聞, 做爲實驗新聞及廣告分段系統的測試資料. 根據實驗結果, 我們分別統計了針對廣告部分的 recall rate 及 precision rate , 以及整體的準確度（ accuracy ）. 其中：

1. 對於廣告的 $recall\,rate = \frac{\#\,of\,Correctly\,Detected\,Commercial\,Pieces}{\#\,of\,Real\,Commercial\,Pieces}$.
2. 對於廣告的 $precision\,rate = \frac{\#\,of\,Correctly\,Detected\,Commercial\,Pieces}{\#\,of\,Detected\,Commercial\,Pieces}$.
3. 新聞整體的 $accuracy = \frac{\#\,of\,Correctly\,Detected\,Commercial\,and\,News\,Pieces}{\#\,of\,All\,Pieces}$.

統計結果如表5.1至表5.3所列.

表 5.1: 對於廣告部分的 recall rate

	廣告的 recall rate
以聲音特性判斷	76.5951%
加入視訊特性輔助	87.5643%
加入連續特性輔助	92.4160%
加入主播特性輔助	91.3648%
加入廣告長度特性	91.2911%

表 5.2: 對於廣告部分的 precision rate

	廣告的 precision rate
以聲音特性判斷	51.3849%
加入視訊特性輔助	53.2382%
加入連續特性輔助	60.3746%
加入主播特性輔助	64.6911%
加入廣告長度特性	80.1268%

由實驗結果可知, 我們所提出的方法中的各步驟, 確實有其效用; 且整體而言, 對於新聞及廣告分段有足夠的準確度, 達到將近92%的判斷準確度.

5.2 以鏡頭爲準的視訊短片搜尋實驗與結果分析

我們將所提的方法, 在 PC (Personal Computer) 的系統上實作了一個原型. 並實驗測試此搜尋方法的容錯能力, 準確度, 以及效率.

我們用以測試的資料, 是由我們錄製的, 以一個小時爲單位的, 不同時

表 5.3: 新聞及廣告分段整體的準確度

	新聞節目整體的準確度
以聲音特性判斷	75.4543%
加入視訊特性輔助	77.0493%
加入連續特性輔助	82.3061%
加入主播特性輔助	84.9547%
加入廣告長度特性	91.9360%

段, 不同電視台播出的電視節目; 其格式爲 NTSC, 畫面解析度 352×240 , 以 mpeg-1 格式儲存. 所有的錄製節目, 皆爲 target video 的測試資料. 而 query clips , 則是由錄影中節選出來的不同內容的節目片頭, 廣告, 政令宣導等, 這些短片在我們的錄製節目中重複出現, 且由於亮度, 播放速度或剪接略有不同, 其內容及長度也有些許變化. 例如, 在我們錄製的電視節目的其中12個小時 (92年11月13日中午12時至午夜12時, 華視) 的資料中, 我們選出的20則不同內容的 query clips , 共出現了31次, 而且有部分的短片, 存在些許亮度變化或長度變化的現象 (即與原 query clips 略有不同) .

我們的 target video 測試資料來源爲92年底, 以及93年4月至6月的華視及 TVBS 各時段共300小時的節目錄影; 而 query clips 的選取也是由這些錄影中挑選出的, 不同內容的短片, 主要爲各時段出現的廣告.

5.2.1 對於亮度變化的容忍力

在這個小節中, 我們在12個小時的 target video 中隨機選取360段時間長度10秒的短片, 以人工加工的方式, 將所有畫面 (frames) 調亮或

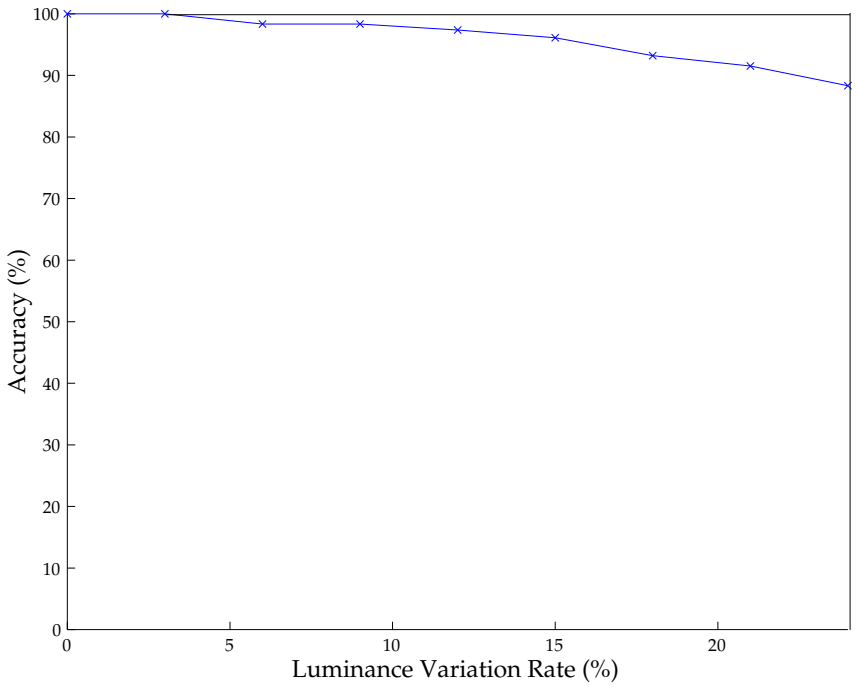


圖 5.1: 實驗結果顯示在亮度變化不超過15%的情形下, 我們的方法其搜尋準確度皆能保持在96%以上.

調暗若干個百分比, 再以所提的方法去搜尋已經過調變亮度的短片, 觀察其搜尋準確度, 以測試本演算法對於欲搜尋的短片產生亮度變化時的容忍度. 實驗結果如圖5.1 所示, 在亮度變化不超過15%的情形下, 其搜尋準確度皆能保持在96%以上.

在亮度變化超過15%時, 搜尋準確度會大幅下降的原因, 主要是因爲亮度變化超過一定範圍後, 視訊影片中的畫面所保留的原始資訊將會大幅減少 (例如, 將一段平均亮度偏暗的影片片段再調暗, 則該段影片的畫面內容將變成一片漆黑) , 因而導致其與未發生亮度變化前的畫面內容之間的相似度亦大幅降低.

我們另外也實驗了本演算法對電視節目中發生的實際亮度變化之容忍度; 我們在24個小時的 target video 中共有50則與我們選出的20則 query

clips 內容相同, 但亮度有些變化的短片 (其平均亮度變化率大部分在12% $\sim$ 18%之間) , 以之作測試, 所得的搜尋準確度爲92% (46/50) .

由此節的實驗結果可知, 我們所採用的方法, 的確對於視訊影片的亮度變化有足夠的容忍度.

5.2.2 對於長度變化的容忍力

在這個小節的第一個實驗中, 我們在12個小時的 target video 中有30則與我們選出的20則 query clips 內容相同, 但時間長度有些變化的短片, 作爲我們搜尋的標的; 然後, 再訂定在4.2.4小節所提的 "允許每個 shot 的最大長度變化率" ($l\%$) 爲10%, 進行搜尋, 所得的搜尋準確度爲 73.3% (22/30) .

我們所做的第二個實驗, 主要是希望了解在4.2.4小節所提到的 $l\%$ 的大小, 對實際上的搜尋準確度的影響. 實驗數據, 是以我們在總共12個小時的電視節目中, 分別搜尋20則 query clips (共出現過84次, 其中30次有長度變化) 的數據作平均所得. 結果如圖5.2 所示, 搜尋準確度在 $l\% := 10\%$ 附近較高, 約爲91%左右, 相較於不容許短片長度變化 ($l\% := 0\%$) 的72%, 高出近兩成, 顯示容許短片長度變化, 確實有助於提升實際情況下的搜尋準確度; 另外, 亦可由圖5.2 的結果觀察出, 在 $l\%$ 繼續加大後, 搜尋準確度不升反降的情形, 這是由於 $l\%$ 過高易產生 false-positive 現象的緣故.

對於部分未能被成功搜尋到的短片, 其原因可能爲 : 該待尋短片的 shots 甚少 (少於2個 shots) , 且其 shots 在長度變化後, 由各 segments 所

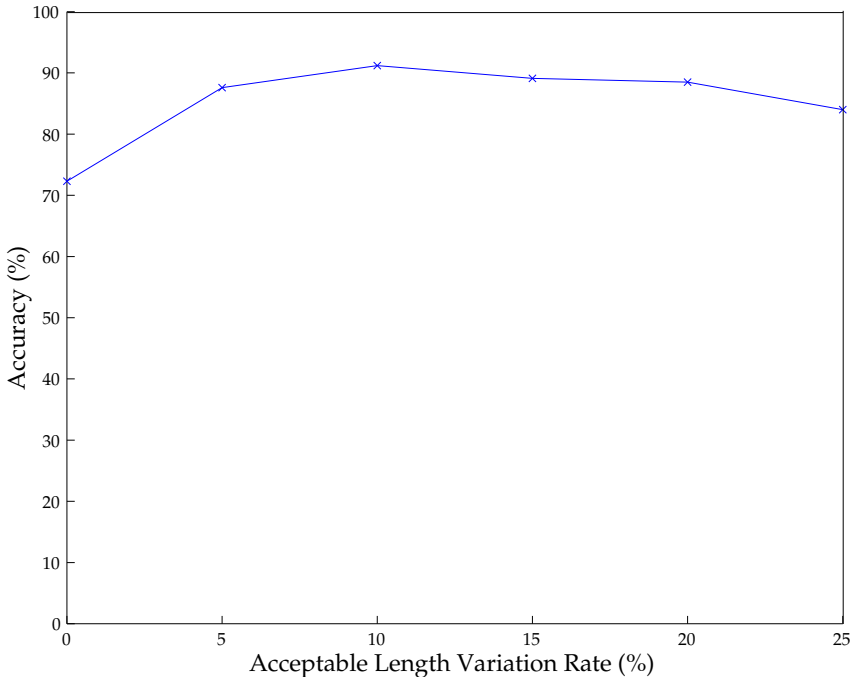


圖 5.2: 此圖顯示在 "允許最大長度變化率" $l\% := 10\%$ 的情況下, 可以有91%最佳的搜尋準確度.

取的特徵向量亦產生較大的變化.

特徵向量在短片長度變化後亦產生較大變化的情形, 在我們取較多 frames 爲一個 segment 的單位時, 較容易發生.

5.2.3 搜尋準確度

本演算法在實際情況下的搜尋準確度, 以圖5.2 來說, 在令 $l\% := 10\%$ 時, 可以有最佳的結果, 搜尋準確度爲91%.

我們再以93年4月中至6月份的華視及 TVBS 的共300小時電視節目錄影爲 target video 的測試資料, 並將這300小時中出現的所有內容不同的廣告 (共325則) 爲 query clips 的測試資料, 用較大量的資料量測試

我們的方法其搜尋準確度.

在300小時的 target video 中, 325則內容相異的 query clips (即廣告) 共出現了7323次, 其中2858次出現長度變化 (與325則 query clips 相較) ; 而我們最終的搜尋準確度爲84.8% (6210/7323) . (其中對於發生長度變化的短片, 其搜尋準確度爲65.5% (1871/2858) .)

而此實驗的搜尋速度, 在共300小時的 target video , 共325則 query clips (即所有內容相異的廣告, 平均每則長約18.5秒) 的數據爲依據, 可推知在一個小時的 target video 中, 搜尋一則18.5秒的 query clip , 需時約0.1秒.

此結果證明我們的搜尋法, 確實能有效地在實際的視訊廣播節目中, 搜尋到所需的待尋短片.

6 結論與未來方向

在本論文中, 我們提出了一個新聞及廣告分段系統. 在第5章的實驗結果中, 其爲新聞及廣告段落分類及分段的準確度, 已達到將近92%, 顯示此新聞及廣告分段系統, 的確能有效的區分電視新聞節目中的新聞及廣告.

然而, 我們所採用的各種特性及特徵是否能通用於大部分的新聞節目? 其流程是否合理? 或者, 是否有更爲合適的特性, 特徵或流程可以使得對於新聞節目中的新聞及廣告的分類分段更爲準確? 這些問題都是我們尙待進一步去解答的.

另外, 我們也發展出了一套以鏡頭爲準的視訊短片搜尋技術. 在發展本篇論文所提出的演算法之前, 我們曾經嘗試使用一般 frame-by-frame 的比對及搜尋方式, 來搜尋一些電視節目中重複出現的短片 (例如, 廣告) , 但卻發現, 這些在電視節目中重複出現的短片, 有時會有長度變化等現象, 使得原本預期會被搜尋出來的短片, 會有部分被判定爲 "不相似" ; 這些問題並無法用 frame-by-frame 的比對方式來解決 (其搜尋準確度甚低) , 而我們所提出的方法, 則爲這種現象提供了一種解決之道. 同樣的, 我們的演算法所取的特徵向量, 也能有效的抵抗因影片發生亮度變化而對搜尋準

確度產生的負面影響. 目前, 我們的演算法, 在實際狀況下的搜尋準確度, 可達到約85%; 而搜尋效率亦不差, 在一個小時的 target video 中, 搜尋一則18.5秒的 query clip , 約需時0.1秒. 由實驗結果來看, 我們提出的方法, 確實能有效地在實際的視訊廣播節目中, 搜尋到所需的待尋短片.

但於本文所提之演算法, 仍有一些有待加強之處. 我們的方法雖然增加了對亮度變化的容忍度, 但亦可能產生 false-positive 的現象, 尤其是在待尋短片中的內容單調且鏡頭變化少的時候; 而其搜尋速度, 顯然也比不上 K. Kashino 等人[12]所提出的方法, 也有繼續改善的空間. 這將是我們未來改進的方向.

參 考 文 獻

[1] Z. Liu, J. Huang and Y. Wang, "Classification of TV programs based on audio information using hidden Markov model," *IEEE Workshop on Multimedia Signal Processing*, pp. 27–32, Dec. 1998.

[2] S. Srinivasan, D. Petkovic and D. Ponceleon, "Toward robust features for classifying audio in the CueVideo system," *Proc. 7th Int. Conf. ACM Multimedia*, pp. 393–400, Nov. 1999.

[3] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[4] C. W. Ngo, T. C. Pong and H. J. Zhang, "On clustering and retrieval of video shots," *Proc. 9th Int. Conf. ACM Multimedia*, pp. 51–60, Sept. 2001.

[5] L. Lu, H. Jiang and H. J Zhang, "A robust audio classification and segmentation method," *Proc. 9th Int. Conf. ACM Multimedia*, pp. 103–122, Sept. 2001.

[6] C. Lu, M. S. Drew and J. Au, "Classification of summarized videos using hidden Markov models on compressed chromaticity signatures," *Proc. 9th Int. Conf. ACM Multimedia*, pp. 479–482, Sept. 2001.

[7] T. Kurozumi, K. Kashino and H. Murase, "A method for robust and quick video searching using probabilistic dither-voting," *Proc. of ICIP 2001*, vol. 2, pp. 653–656, Oct. 2001.

[8] 鄭士賢, "高斯混合模型的學習與其在語者識別上的應用," 國立交通大學資訊工程學系碩士論文, Jul. 2002.

[9] H. Nagano, K. Kashino and H. Murase, "Fast music retrieval using polyphonic binary feature vectors," *IEEE ICME*, vol. 1, pp. 101–104, Aug. 2002.

[10] L. Lu, H. J. Zhang and H. Jiang, "Content analysis for audio classification and segmentaion," *IEEE Trans. Speech and Audio Processing*, vol. 10, pp. 504–516, Oct. 2002.

[11] A. Kimura, K. Kashino, T. Kurozumi and H. Murase, "Dynamic-segmentation-based feature dimension reduction

for quick audio/video searching," *Proc. of ICASSP 2003*, vol. 3, pp. 357–360, Apr. 2003.

[12] K. Kashino, T. Kurozumi and H. Murase, "A quick search method for audio and video signals based on histogram pruning," *IEEE Trans. Multimedia*, vol. 5, no. 3, pp. 348–357, Sept. 2003.