運動節目語意分析與重點事件的自動搜尋系統

Semantic Analysis of Sports Video Programs and Highlight Retrieval System

施皇嘉、黃仲陵
清華大學電機工程系

摘要

由於視訊資料大量數位化，以致於數位資料暴增，雖然結合各種最新之壓縮及視訊處理標準，可以讓數位資料編碼以最精簡的容量來儲存，所以資料的表示與儲存已不是問題，現階段最急於解決的是一個能夠提供使用者在最短的時間內找到最符合他們需求的數位內容(Digital Content)，也就是說，如何快速地、有效率地存取我們所要的數位資料成為一項重要的課題。因此，我的研究主要是去建立一個能夠認知、分類及總結視訊資料的系統，我們以運動節目做為我們實驗的應用範疇，因為運動節目具有高重覆性，高相關性的鏡頭特性，對我在做分析時較有利。

一、 前言

未來的寬頻應用，主要可以分為三種主要應用媒體，分別是視訊(Video)、語音(Audio) 、資料(Data)，使用者可以存取他們想要的多媒體資料經由數位網路電視、網路電話及寬頻網路等介面，只要透過一個簡單的 set-top-box 及 DSL system 就能夠輕鬆得到所有服務。當所有的硬體設備完成建置後，最後要解決的就是數位內容的問題了。因為電腦是無法自行處理人類高語意的意含。比如拍攝鏡頭內的影片類別，鏡頭內所含的人物角色，場地資訊等，為了讓使用者能夠透過網路，直接存取他們所想要的數位資訊（Video, Audio and Data)。這所有的「知識」必須由人類親自去安排與分類。這無形中耗費了許多人力資源，所以我們為了解決這個問題，我們針對最受歡迎的視訊－運動節目來當做我們的研究領域，我們提出了一個多階層多機能的視訊索引系統，來使數位資料達到最快速、最接近人類的感觀特性的分類與管理，便於數位內容發行者來提供觀眾視聽上的需求。

人對於多媒體資料最能直接充分瞭解的是當中的高階語意(Semantics)資料，因為高階的語意是最接近人類的思考模式，無疑地，人在接收各種外在資訊時，最能夠讓人類充分瞭解資訊提供者所想要表達的意含就是高階語意，但相反地，電腦是一個沒有生命的物體，而它若沒有經過人類的訓練及推演，它只能接收及判斷一些最低階的視訊特徵，每一個高階的語意資訊與低階視訊特徵，若沒有經過人類的相關性鏈結部署，電腦是無法自行辨識高階語意資訊的意含，而多媒體資料是十分雜亂的，若要建立一

個共通的解決方案似乎不太可能，我們能做到的就選擇一個視訊片斷重覆較高的運動節目來做解析。近年來有相當多的文章以運動節目為例子來做摘要分析(Summarization)[1,2]、學習認知(Understanding)[3,4]、視訊索引及搜尋 (Video Indexing and Retrieval)[5,6]。而未來的視訊語意分析系統不僅要適用在異質性的平台(Heterogeneous Platforms)上，更應結合以下的網路功能：

1) 內容之可調適性(Content Scalable)，系統能夠適應當下的網路頻寬而供應不同劇性細節的多媒體資訊。
2) 存取上的可調適性(Adaptive Access)，提供使用者多功能地存取不同使用環境下的多媒體。
3) 有興趣的影片(Video-of-Interesting)，籍由視訊分析後，提供使用者能夠得到他們真正感興趣影片。

在此，我們為了完全解析輸入的運動影片的內含，推演出不易直接觀測到的高階語意特徵，我們利用動態貝氏網路(Dynamic Bayesian Network)的架構，推論出視訊的高階特性來做來做運動節目的語意認知。動態貝氏網路可以建構起高低特徵的橋梁和結合不同的特徵資訊，進而達到 semantic analysis 的目的。測試視訊先經過我們所設計的各種特徵分析器求得一些低階的資訊，我們把這些資訊視為動態貝氏網路的輸入，籍由在訓練程序所求得的一些機率分佈，向上推演得到我們所要的類別資訊。除了單純考慮一個時間點的貝氏網路外，我也希望利用貝氏網路時間上的相關性(time dependency)，也就是動態貝氏網路來判斷視訊的特性，如此可以得到較好的結果。

在此篇論文中，我們會舉例足球影片做語意上的分析及重點事件的偵測，在本文的第二部份對低階的視訊特徵及足球影片可能出現的高階語意及物件的偵測做介紹，而在第三部份針對重點事件的偵測提出我們的研究架構。

二、 低階視訊特徵的分析

近年來，有愈來愈多的研究是關於籍由資訊分析的視訊瀏覽(Video Browsing)、摘要(Summarization)及搜尋(Retrieval)的題目，而大部份的作者都是仰賴鏡頭的場景移動，畫面色調及物件的形狀來做對視訊節目做摘要分析及認知。在此節中，我們將對低階的視訊特徵及足球影片可能出現的高階語意及物件的偵測做介紹，我們偵測了九種的視訊特徵及物件，如特寫(Close-up)、偏移(Panning)、觀眾席(Audience)、重播(Replay)、球門(Gate)、告示版(Board)、裁判(Referee)、音訊(Audio)、靜止畫面(Static Camera)。這些特徵都是一些必須的證據(Evidence)來提供後端的DBN做分析。

2.1 特寫鏡頭偵測

特寫鏡頭在一個運動節目中占了一個十分舉足輕重的角色。我們使用了主色偵測法(dominant color detection)來判斷某個鏡頭是屬於特寫鏡頭(close-up view)或是全域的鏡頭(global view)，我們將每一個畫面轉換到HSI顏色坐標上，取得一個主色，設定一個範圍，只是與這個主色的距離小於一個範圍之內的點，就屬於主色域的領域，在足球節目中，主色也就是草地的顏色-綠色，圖就是我們偵測的結果，下圖白色為非主色區域，黑色區域為主色區域，利用主色區域的大小就可以分析此鏡頭為特寫或全域的鏡頭。

圖十六 (a) 全域鏡頭,(b) 特寫鏡頭.

2.2 偏移偵測

.

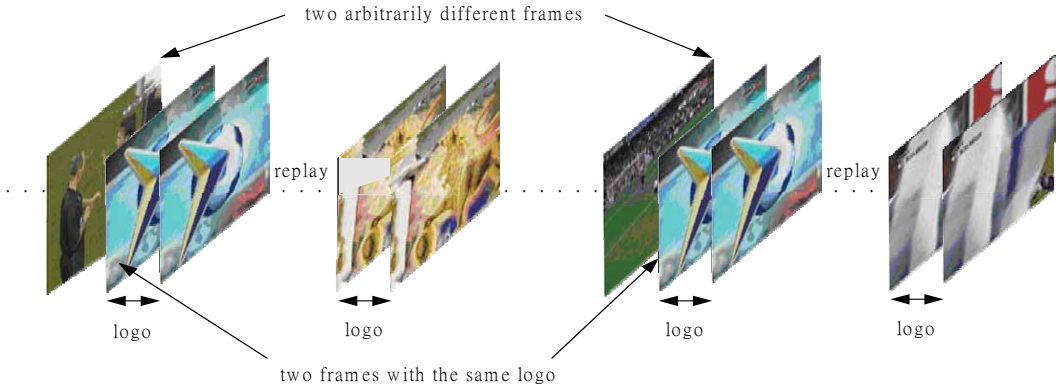
3-1.B 重覆播放鏡頭(replay) 偵測

在足球節目中,慢動作重播鏡頭常常被使用來讓觀眾再一次清楚地收看精采事件發生時的全程時錄，有時常會用不同的角度，不同的攝影機來呈現它。針對這個重播鏡頭的偵測顯得非常重要,因為這些重要的事件的發生正是我們所以去抓取的。而電視公司不約而同都會利用一個 logo 來做為轉場的中接鏡頭。也就

是說我們可以輕易地去偵測到重播鏡頭的播放，只要我們找一個方法來去偵測 logo 的出現。

圖十七 轉場效應中的 logo.

我們所使用的方法是利用連續兩張畫面的色彩濃度(hue)和亮度(intensity)對比差值若大於某個程度就代表是某個 logo 出現的起始點，我們假設轉場時間大約都在一到兩秒鐘之間。下一個大差值即為 logo 轉場的終止點。

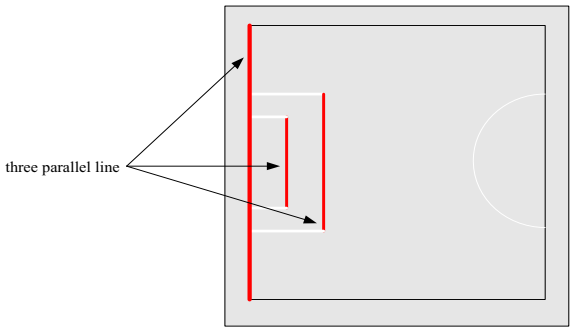


圖十八 同一個轉場效應中有相同的 logo.

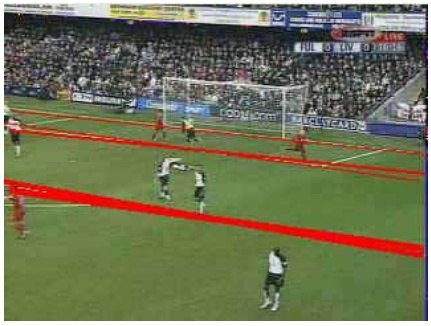
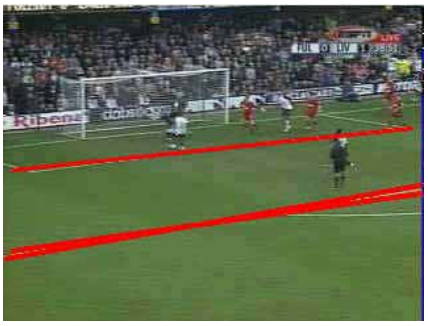
3-1.C 球門(gate) 偵測

在鏡頭為遠距離拍攝時，常常會有球門的出現，我們利用球門前方的白色線條來取代直接去偵測球門。我們可以很容易地去找出三條的平行白色場地線。由圖十九可以看出來，有三條明顯的白線在球門的前方。當球員進入這個罰球區時，鏡頭將會被帶到這三條白線。這是十分有用的偵測球門的資訊。我們利用找線十分出名的瞿夫轉換(Hough transform)來找出有用的線條。

在此我們先從裡面找出有用的邊緣(edge)資訊再用瞿夫轉換來找平行線，圖二十可以看出來平行線被偵測出來，並且他們的角度大都在 140o 到 170o(圖 a), 圖 b 都在 10o to 40o. 我們更可以利用角度的資訊來判斷現在的球門是在左方還是右方，更可以來判斷是哪一隊球隊得分。



圖十九　三條明顯的線在球門區域

(a) 
(b) 

圖二十　球門的偵測

3-1. D 裁判的偵測

在足球比賽中，裁判也是一樣十分重要的資訊，對於我們在偵測事件發生時常被用到，比如黃／紅牌的事件，也就是重大犯規時的事件，通常都會有特寫裁判的鏡頭。而且裁判的球衣顏色一定會不同於兩隊球員。在此我們就是利用物體的顏色若是黑色或為黃色時，並且是異於兩隊球員時。此物體即為裁判的出現。

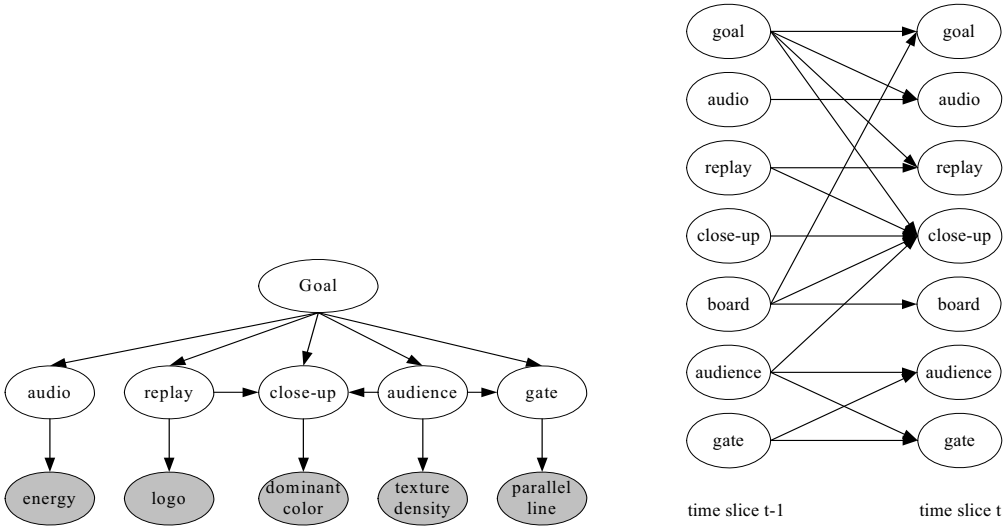
圖二十一　裁判的偵測

3-2 足球的精采鏡頭偵測

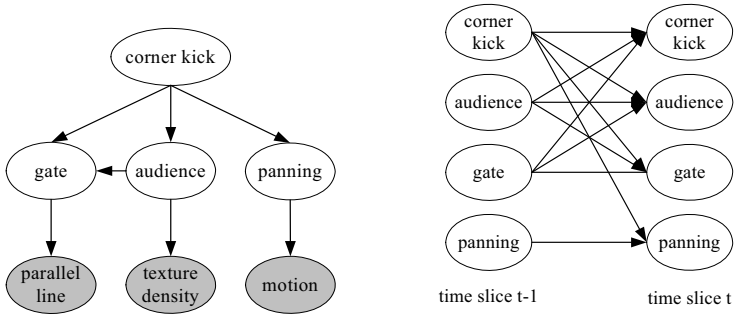
動態貝氏信度網路對於來做不確定或無法接觀測的特徵推演是項極為有用的工具。傳統的視訊認知方法皆是由視訊資料的低階特徵做比對，對於某些抽象或是劇情太過複雜的視訊影片卻無法確實表達，而動態貝氏信度網路的角色就是在連接低階特徵與高階語意特徵間的橋樑。在這裡我們將介紹一個多階層式的動貝氏信網路架構來對運動節目做視訊認知，對特定領域的節目做深入的探討，如

足球節目。我們將整個描述足球節目的動態貝氏信度網路分成數個子網路來討論，包含了射門鏡頭(goal kick)、角球(corner kick)、十二碼罰球(penalty kick)、黃牌紅牌(card)。

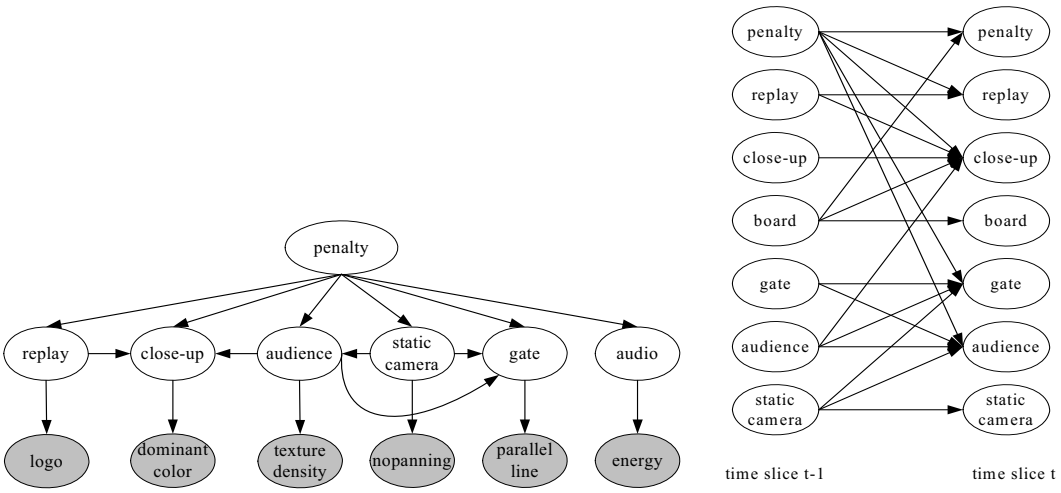
我們設計了四個動態貝氏信度網路來偵測以上四種精采鏡頭。



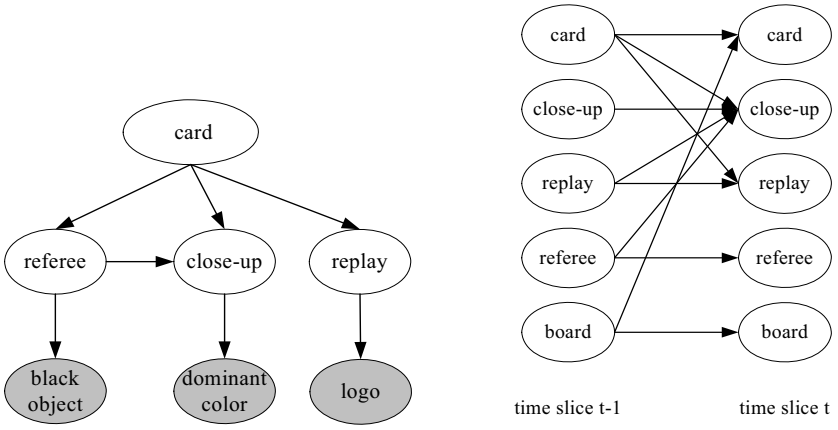
(a)



(b)



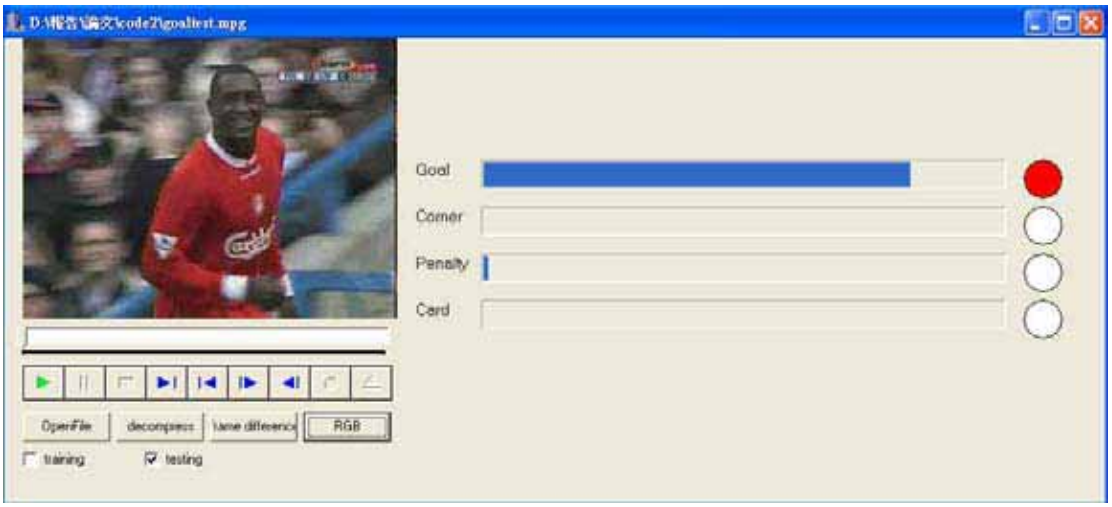
(c)



(d)

圖二十二　四種事件的動態貝氏信度網路

以上的動態貝氏信度網路一定都需要經過訓練(training)的過程。並能夠找到每一個點(node)的事前機率(priori probability)和每一個連結(link)的條件機率(conditional probability)然後在測試階段再利用動態貝氏信度網路的機率推演(probability inference)才能夠得到我們最後所要的結果，事後機率(posterior probability)。最後我們就能求得輸入的畫面屬於某個事件的機率有多高。近而來偵測以上四種精采鏡頭的出現。圖二十三即是我的一個使用界面。當影片都被分析好後，往後的應用將更多。例如數位化後的資料存取，索引。



圖二十三　四種事件偵測的界面

4. 結論

在我們的計畫當中，最值得一提的是對於運動節目的分析的重要性，因為科

技的進步，大量的影視資訊被數位化後，被儲存在數位儲存設備當中，雖然資料能夠被永久保存，但現在卻面臨一個十分嚴重的問題，也就是搜尋的問題，每一個運動節目，每一個鏡頭出現在什麼事件，什麼人物，我們不可能只利用人工去指定輸入，這樣對於人力的浪費過於龐大。電腦不像人腦這麼的聰明，不能夠一眼就看出這個畫面是什麼事件什麼人物，但人類還有許多事情等著我們去實現，電腦必須取代人腦去做一些時間消耗大，重覆性高的事情。有鑒於此，我們這個計畫就是要去發揮我們的想像力，去追求一個最適合人類的夢想，最適合人類的使用習慣的收看模式。觀眾能夠隨心所欲地去找出他們想看的。不需要再浪費太多時間在等待無意義的暫停時段，如此，觀眾收視將更有效率，也將會花更多時間在收視。運動節目是我們第一步，將來我們將朝向電視新聞分析，各種電影的分析領域邁進。

五、 參考文獻

[1] A. Ekin, M. Tekalp, and R. Mehrotra, “Automatic Soccer Video Analysis and Summarization,” *IEEE Trans. on Image Processing*, Vol. 12, No. 7, July 2003.

[2] A. Ekin, “Sports Video Processing for Description, Summarization, and Search,” University of Rochester, New York, 2003.

[3] H. C. Shih and C. L. Huang, “MSN: Statistical Understanding of Broadcasted Sports Video Using Multi-level Semantic Network,” to be appear in *IEEE Trans. on Broadcasting*, Dec. 2005.

[4] H. C. Shih and C. L. Huang, “Detection of The Highlights in Baseball video Program,” IEEE-ICME 2004, Taipei, Taiwan. 2004.

[5] W. Zhou, A. Vellaikal, and C.-C. J. Kuo, “Rule-based video classification system for basketball video indexing,” in *ACM Mult. Conf.*, Los, Angeles, USA, Nov. 2000.

[6] V. Mihajlovic, M. Petkovic, “Automatic Annotation of Formula 1 Races for Content-Based Video Retrieval,” *Technical Report*, *TR-CTIT-01-41*, 2001.

[7]