

行政院國家科學委員會專題研究計畫 成果報告

子計畫二：視訊串流伺服器及視訊資料庫整合(III)

計畫類別：整合型計畫

計畫編號：NSC94-2219-E-009-021-

執行期間：94年08月01日至95年07月31日

執行單位：國立交通大學資訊工程學系(所)

計畫主持人：李素瑛

計畫參與人員：蕭銘和 陳#28175；紋 陳華總 陳宣勝 田敏君 李卓皓

報告類型：完整報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 95 年 10 月 30 日

行政院國家科學委員會專題研究計畫成果報告

視訊串流伺服器及視訊資料庫整合(III)

計畫編號：94-2219-E-009-021-

執行期限：94年8月1日至95年7月31日

主持人：李素瑛教授 交通大學資訊工程系

計畫參與人員：蕭銘和 陳漪紋 陳華總 陳宣勝 田敏君 李卓皓

一、中文摘要

本計畫主要目的在於研究與實踐視訊資料庫以及視訊伺服器。在視訊資料庫方面，除了視訊資料內容之分析與高階與語意特徵擷取之外，並結合音訊資料之特徵加上影片內容字幕之偵測及擷取，以內容分析之結果為依據的影片調整方法，讓資源能夠有效地利用，並提高影片的視覺品質。並且利用分析所得之語意資訊加以解譯及推測，以偵測影片資料內容之事件，藉此達成高階特徵之擷取並以此為基礎產生既簡短又具語意意義之視訊資料摘要提供視訊資料預覽之用，除此之外在產生影片內容摘要之過程中，同時產生符合MPEG-7之低階至高階特徵之描述(descriptor)以及描述結構(description scheme)，以利資料庫索引之建立並藉此建立互動式影片瀏覽系統。

在視訊伺服器方面，我們希望設計支援單點傳送與多點傳送的視訊串流伺服器。在單點傳送伺服器設計方面，主要有兩項任務，一項是傳輸速率控制(rate control)，另一項則是錯誤控制(error control)。在多點傳送視訊串流伺服器方面，由於多點傳送有效率地使用頻寬，因此我們的重心在於錯誤控制。在多點傳送的錯誤控制中，當封包遺失發生時，有兩個主要的問題需要解決，一個是區域性地錯誤恢復(Local Error Recovery)，一個是NACK爆炸(NACK Implosion)的議題，前者是希望能盡量縮短錯誤恢復所需要的時間，後者則避免區域性網路壅塞的問題。

最後，我們分別實作出支援單點傳送與多點傳送的視訊串流伺服器以及視訊資料庫，最後再將它們整合以達到快速地從資

料庫中找到最佳的位元配置與影片調整方法，來做視訊串流的調整。從實驗結果可以看出所提出的方法相當實用並可達到較好的品質。

關鍵詞：視訊內容分析；音訊內容分析；互動式影片導覽系統；視訊串流；視訊資料庫

Abstract

The purpose of this project is to develop an efficient and effective video-streaming server. In video database, we will analyze visual, audio features and localize captions in order to support content-based video retrieval based on acquired high-level/semantic features. Furthermore, we will detect video events by inferring from analyzed information of video and audio streams. Besides, video summaries can be obtained by merging semantic video clips that are identified according to the video events. The description/descriptors corresponding to low-level and high-level features and the description schemes can be automatically generated during the process of video summaries generation and will be used for indexing.

In video streaming server, we will implement unicast video streaming server and multicast video streaming server. There are two major tasks for unicast video streaming server: one is rate control and the other is error control. In rate control, unicast video streaming server optimizes the usage of the network bandwidth according to not only the estimated available bandwidth but also the importance of the video stream. In error control, we will adopt a method called

FEC (Forward error correction) with UEP (Unequal Error Protection) to reduce the time spent to recover packet loss and waste of the network bandwidth because traditional retransmission mechanism needs to spend more time to complete the packet loss error recovery.

In multicast video streaming server, we will focus on the error control because multicast is an efficient technique to deliver video content over the Internet. In error control, there are two major issues needed to be resolved: one is local error recovery and the other is NACK implosion. The first issue is to recover the packet loss as soon as possible and the second is to avoid local area network congestion. Therefore, in order to realize the multicast video-streaming server, we will investigate in the achievement of local error recovery while avoiding local area network congestion.

Keywords: Content-Based Video and Audio Analysis, Interactive Video Browsing System, Video Streaming, Video Database

二、緣由與目的

隨著網際網路規模的不斷擴大以及資訊科技進步快速，愈來愈多的多媒體資料以數位的形式藉網路為媒介廣泛地傳播著。在傳輸多媒體資料的過程中，有網路頻寬、封包遺失以及封包延遲的限制，但是現今的網路並沒有提供 QoS (Quality of service) 的能力，來保證多媒體資料的傳輸，因此，如何設計一些機制來確保多媒體資料的傳輸便面臨很多的挑戰。

資訊種類之多元以及資訊量之豐富對於使用者來說哪些是有用的資訊及哪些是可忽略的資料自然成為一個重要的議題。因此對於使用者而言，在眾多資料之中如何有效且快速地獲得真正所要的內容便是我們所要努力的目標。然而除了提供快速有效的查詢方式之外，還必須讓使用者同時享有自然且友善的查詢環境，而利用資料內容為索引及查詢(Indexing & Retrieval)為依據的方式(Content-Based Access)正能夠滿足此需求。因為所謂 Content-Based Access 即是以資料內容之特徵當成索引來

建立資料庫，因此特徵描述之充分與否則直接影響查詢結果之優劣。而傳統資料內容之描述可為某些關鍵字、屬性的集合、或是某些以文字為主的抽象表示法。

對於大量的音樂資料、新聞、紀錄片、電影以及影像提供充分適當的描述對於不同領域及不同應用的使用者而言相當地重要。在應用上比如，數位圖書館、隨選視訊以及網路多媒體傳播等等。因此 MPEG-7 試著定義多媒體特徵的描述方法使得以內容為基礎的資料索引及資料搜尋更為容易。此外 MPEG-7 包含更多的資料型態並且將內容的描述(descriptor)和描述法(description scheme)標準化藉以達成能夠描述更多種類的多媒體資料。然而特徵產生之方式並非 MPEG-7 欲標準化的部份，除此之外由於多媒體資料型態及內容的多樣性使得更有效、更具高階意義的特徵亟待產生。另一方面為了提供系統間之相互作用(interoperability)，特徵抽取以及搜尋引擎兩大議題皆非 MPEG-7 欲標準化之部分。

本計畫主要目的為視訊資料內容之分析與擷取，並利用分析所得之資訊加以推理(inference)，以偵測視訊資料內容之事件，藉此達成高階特徵之擷取並以此為基礎產生既簡短又具語意意義之視訊資料摘要(summary)，除此之外在產生視訊內容摘要之過程中，同時產生低階至高階特徵之描述(descriptor)以及描述結構(description scheme)以符合 MPEG-7 之標準。最後，我們希望分別實作出支援單點傳送與多點傳送的視訊串流伺服器以及視訊資料庫，最後再將它們整合。

三、實驗結果與討論

◆ 單點傳送與多點傳送視訊串流伺服器之整合

由於單點傳送視訊串流伺服器與多點傳送視訊串流伺服器屬於不同的傳輸模式，因此在實際運作時，其負責的模組也不相同，在單點傳送視訊串流伺服器中包含傳輸速率控制模組、傳輸錯誤更正模組以及單點傳送網路模組；在多點傳送視訊串流伺服器中包含傳輸頻道分割模組、應用層協定模組以及多點傳

送網路模組。另外，由於視訊串流伺服器傳輸模式的不同，使得接收端不管在接收資料緩衝區的設計以及錯誤恢復的機制皆不盡相同，所以要將這兩種視訊串流伺服器整合起來時，必須先判斷是否是單點傳送要求再進入單點或多點傳送模組，方可確保不論在單點傳送或是多點傳送的運作下，皆可正確無誤地接收資料並且執行錯誤恢復的能力。整合之視訊串流伺服器如下圖一所示。



圖一、視訊串流伺服器介面

◆ 視訊影片資料庫

隨著無線網路的發展與可攜式隨身裝置能力的提升，在這樣的環境中傳遞影片的應用越來越廣泛。在有限的資源和與生俱來的限制下，如何適當地調整影片長度以及品質成為在無線多媒體應用相關領域中的一個最重要且富挑戰性的議題。

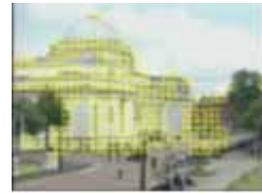
由於不同種類之視訊資料則有不同之特性，如運動類影片與新聞類及一般電影類影片其特性則相當不同，視訊片段之分類及其影片長度及品質調整之設定則有不同之標準，在本研究中，我們提出了一個以內容為依據的影片調整方法，讓資源能夠有效地利用，並提高影片的視覺品質。藉由分析影片中物件的亮度、位置、移動向量與能量等屬性，配合客戶端裝置能力與關聯性統計模型取得場景中較吸引注意的部分如圖二所示。利用區域加權速度失真模型、位元配置與影片調整方法，動態地從物件、畫面與 GOP 三個層面來做視訊串流的調整，如圖三所示。



(a) Class 1



(b) Class 2

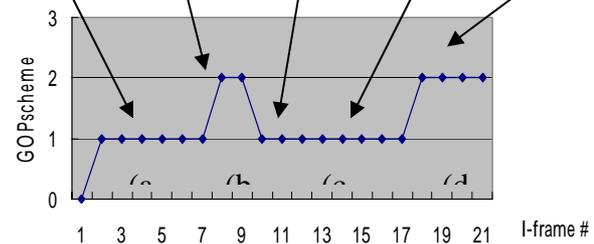


(c) Class 3



(d) Class 4

圖二、不同影片類型中，感興趣區域的實驗結果



圖三、以影片特性做視訊串流的調整結果

另外，我們利用物件軌跡作為基礎，再加上 domain knowledge 作為推測之依據，並利用研發之事件推測演算法判斷視訊片段是否重要，是否隱含高階語意，之後即可將重要之片段擷取並合併成為更具意義之視訊摘要，相信這將是在結構化視訊計算範疇當中相當重要之進程。我們亦結合音訊內容以增進視訊摘要之正確性。另外 MPEG-7 亦定義了 Summary DS，我們也將產生符合其架構及格式之 descriptors 及 description schemes 來建立索引並藉此建立互動式影片瀏覽系統(如圖四所示)，使得使用者做視訊資料擷取時能快速有效地比對出正確之資料。



圖四、互動式影片瀏覽系統

四、參考文獻

1. W3 Consortium, XML 1.0 Specification. <http://www.w3.org/TR/REC-xml>
2. MPEG-7: Evaluation Process Document, ISO/IEC JTC1/SC29/WG11 N2463, Atlantic City, October 1998.
3. H. Wang, A. Divakaran, A. Vetro, S. F. Chang and H. Sun, "Survey of Compressed-Domain Features Used in Audio-Visual Indexing and Analysis," Journal of Visual Communication and Image Representation, Vol. 14, Issue 2, pp. 150-183, June 2003.
4. ISO/IEC /JTC1/SC29/WG11: "MPEG-7 context and objectives," N2460, Atlantic City, October 1998.
5. ISO/IEC JTC1/SC29/WG11: "MPEG-7 Generic AV Description Schemes (V0.7)," N2966, Melbourne, October 1999.
6. ISO/IEC JTC1/SC29/WG11: "MPEG-7 Visual Part of eXperimentation Model Version 3.0," N2931, Melbourne, October 1999.