

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(II)

計畫類別：整合型計畫

計畫編號：NSC93-2219-E-009-018-

執行期間：93年08月01日至94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 劉祺昱 周冠宏 馬兆慶

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 94 年 10 月 25 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(II)

計畫編號：NSC 93-2219-E-009-018

執行期限：93年8月1日至94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 劉祺昱 周冠宏 馬兆慶

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，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。在單點傳送視訊串流伺服器方面，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)。在傳輸速率控制中，不僅要根據網路狀況，更要根據視訊資料的重要性來做網路頻寬的最佳使用；在錯誤控制中，傳統的重送機制必須花費較多的時間才能達成錯誤恢復，因此我們希望透過傳輸錯誤更正(forward error correction)機制，並且搭配不平等的錯誤保護機制(Unequal Error Protection)以降低錯誤恢復所需的時間並且減低頻寬的浪費。

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，後者則避免造成區域性網路壅塞的問題。因此如何達到區域性地錯誤恢復以及避免 NACK 爆炸所造成的區域性網路壅塞便是我們的目標。

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

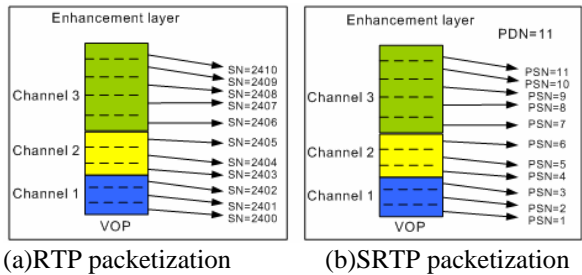
◆ 支援多點傳送的視訊串流伺服器

1. 應用層通訊協定設計

(Application-layer Protocol Design)

在多點傳送的視訊串流伺服器中，由於結合 FGS 適合在網路上傳送的特性，伺服器只需要將多媒體資料分成數個 channel 並把它們送到網路上即可，而且只需送一份資料，然而在這樣的模式下，一樣會發生封包遺失的情況，因此，必須加

入錯誤恢復的機制，在這個計劃中，我們設計一應用層通訊協定稱 SRTP (Scalable Real-time transport protocol)以提供接收端錯誤恢復的能力。圖一所示為 RTP 與 SRTP 將某畫面切割封包的差異。



圖一、RTP 與 SRTP 之封包化

2. 接收端緩衝區管理模組 (Buffer Management Module)

為了降低接收端封包延遲以及封包延遲變動的影響，必須在接收端配置緩衝區以讓接收端的多媒體品質保持平順的狀態，因此，如何設計一個緩衝區，使得它具有快速存取、有效的空間利用以及支援封包遺失的恢復能力，便是一個很重要的研究議題。因此我們提出一個新的接收端緩衝區結構，這個結構包含緩衝區索引表、指標以及資料緩衝區三個部分。緩衝區索引表屬於一個雜湊表(Hash Table)，被用來紀錄每張畫面在資料緩衝區的起始位置以及快速地將資料存放至資料緩衝區中；資料緩衝區為多媒體資料真正儲存的地方，而且還儲存有關多媒體資料封包的資訊，例如，我們設計的 SRTP 中的標頭資訊，以讓該緩衝區的設計支援快速的錯誤偵測。

3. 錯誤恢復模組 (Packet-loss Recovery Module)

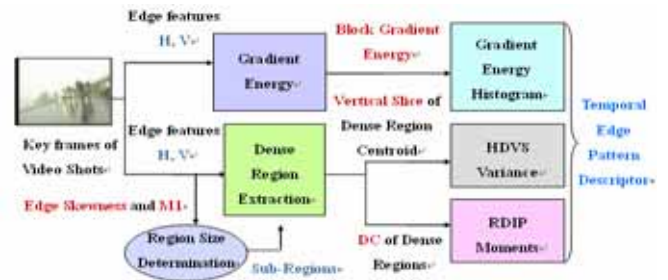
在多點傳送的模式下，當封包遺失發生時，有兩個主要的問題需要解決，一個是區域性地錯誤恢復 (Local Error Recovery)，一個是 NACK 爆炸(NACK Implosion)的議題，前者是希望能盡量縮短錯誤恢復所需要的時間，後者則避免造成區域性網路壅塞的問題。因此，我們提出的錯誤恢復模組也正朝這兩個方向努力，其中包含(1)錯誤恢復啟動管理(Activating Loss Recovery Manager)、(2)服務品質監視器(QoS Monitor)、(3)虛擬群組管理(Virtual

Group Manager)、(4)錯誤恢復管理 (Recovery Manager)四個部分來監控封包發生遺失時以使服務品質獲的確保。

◆ 視訊影片資料庫

視訊影片資料庫的有效性依靠於是否能夠容易找到有興趣的視訊影片。為了提供能根據視覺的內容來探索、瀏覽和擷取影片的能力，在視訊影片資料庫中有效的偵測相似影片的技術成為在多媒體與資料庫相關領域中的一個最重要與富有挑戰性的議題。我們在視覺特徵方面，提出了擷取較特別之特徵如 **edge**，來建構一個新的時間化之邊緣型態表示法以用來偵測影片的相似度。

為了減少計算複雜度，我們使用影片中的關鍵畫面來做比對。利用一個以 GOP 為基礎的方法來做場景變換偵測，並將影片切割為許多片段。從已切割好的影片片段中選出 I-frame 作為其關鍵畫面。最後，使用一個結合梯度能量直方圖、最高邊緣密度垂直切片變化與區域逆機率差異矩的混合邊緣特徵表示法來計算影片間的相似度。圖二為我們所設計創新的 temporal edge pattern descriptor 產生過程。產生步驟包含(a) Edge Detection; (b) 計算 Gradient Energy Histogram; (c) Dense Region Extraction; (d) 計算 HDVS Variance。



圖二、temporal edge pattern descriptor 產生



圖三、query results for a FV shot

Query Shot



Results



圖四、query results for an API shot

從影片查詢實驗結果可以看出不管全景影片或特寫主播片段(如圖三、圖四所示)，我們所提出用來查詢相似影片的特 temporal edge pattern descriptor，以及設計的視訊資料庫查詢架構相當實用。表一分別為查詢各種片段所得到相似影片的結果，可以看效果非常好，分別接近 90%，可達到很高的準確率。

表一、視訊片段查詢整體效能

Clips	Full-Court View (FV)	Pitching View (PV)	Anchor Person (API)	Bicycle Racing (BR)	Close-Up Tracking (CUT)
Ground-Truth Video Shots	48	32	150	50	95
Recall	89%	81%	88.8%	96%	89.6%
Precision	89%	81%	88.8%	96%	80.7%
Average Recall		Average Precision			
88.8%		87%			

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.
7. ISO/IEC JTC1/SC29/WG11: "Supporting Information for the Generic AV Description Schemes," N2967, Melbourne, October 1999.
8. Chu-Hong Hoi, Wei Wang, and Michael R. Lyu, "A Novel Scheme for Video Similarity Detection," In International Conference on Image and Video Retrieval (CIVR2003), Vol. 2728, pp. 373-382. LNCS, USA, Springer, 2003.
9. Sang Hyun Kim and Rae-Hong Park, "An Efficient Algorithm for Video Sequence Matching Using the Modified Hausdorff Distance and the Directed Divergence," IEEE Transactions on Circuit and Systems for Video Technology, Vol. 12, No. 7, July 2002.
10. Arun Hampapur and Ruud M. Bolle, "Comparison of Distance Measures for Video Copy Detection," in Proceeding of International Conferences on Multimedia and Expo (ICME2001), Aug. 2001.
11. Sang Hyun Kim and Rae-Hong Park, "An Efficient Video Sequence Matching Using the Cauchy Function and the Modified Hausdorff Distance," in Proceeding of Storage and Retrieval for Media Databases (SPIE2002), Jan. 2002.