

行政院國家科學委員會專題研究計畫 成果報告

視訊與音訊資料的自動內涵分析及其應用

計畫類別：個別型計畫

計畫編號：NSC92-2213-E-216-015-

執行期間：92年08月01日至93年07月31日

執行單位：中華大學資訊工程學系

計畫主持人：劉志俊

計畫參與人員：范世鎮、吳智偉、鄭煒平、許肇、凌；、葉；億真、
陳建志、古鎬華、賴明楷、劉；韋志、陳毅修、盧；
偉智、楊志強、黃鈞蔚、林；怡伸、歐家銘、陳竹晟

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 9 月 27 日

行政院國家科學委員會補助專題研究計畫成果報告

※

視訊與音訊資料的自動內涵分析及其應用

※

計畫類別：個別型計畫 整合型計畫

計畫編號：NSC 92-2213-E-216-015

執行期間：92年8月1日至93年7月31日

計畫主持人：劉志俊

本成果報告包括以下應繳交之附件：

- 赴國外出差或研習心得報告一份
- 赴大陸地區出差或研習心得報告一份
- 出席國際學術會議心得報告及發表之論文各一份
- 國際合作研究計畫國外研究報告書一份

執行單位：中華大學資訊工程學系

中華民國 93 年 8 月 1 日

行政院國家科學委員會專題研究計畫成果報告

視訊與音訊資料的自動內涵分析及其應用

Automatic Content-Based Analysis of Video and Audio Data

計畫編號：NSC 92-2213-E-216-015

執行期限：92年8月1日至93年7月31日

主持人：劉志俊 中華大學資訊工程學系

計畫參與人員：范世鎮、吳智偉、鄭煒平、許肇凌、葉億真、陳建志、古鎬華、賴明楷、劉韋志、陳毅修、盧偉智、楊志強、黃鈞蔚、林怡伸、歐家銘、陳竹晟 中華大學資訊工程學系

一、中文摘要

多媒體資料的內涵式檢索技術 (content-based retrieval) 為目前多媒體、資料庫與資訊檢索等領域的主要研究主題之一。『數位內容』產業也是國家推動所謂『兩兆雙星』國家發展計劃的兩個明星產業之一，而數位內容的許多應用需要利用多媒體內涵式分析 (content-based analysis) 的技術。因此，多媒體內涵式分析的技術為目前學界與業界的重點研究項目。

在我們以往的相關研究計劃中，我們對 MPEG 視訊、MIDI 音樂以及 MP3 音樂的分段 (segmentation)、分類 (classification)、索引 (indexing)、查詢 (retrieval) 等內涵分析技術進行一系列的研究，在理論與系統實作方面皆有相當不錯的成果。在本計劃中，我們以這些研究成果為基礎，對數位多媒體內容的分析技術進行更為深入的探討。

關鍵詞：內涵分析、視訊資料庫、音訊資料庫、內涵式查詢、內涵式檢索

Abstract

Content-based retrieval of multimedia data is a major research issue in the area of multimedia computing, databases, and information retrieval. On the other hand, the industrialization of "digital content" related products or services plays a major role in our national development, in which many technologies about content analysis of

multimedia data are highly demanded. Thus, multimedia content analysis is an active and important research issue in both research and industrial communities.

In our previous project, techniques about the segmentation, classification, indexing, retrieval of MPEG video, MIDI music, and MP3 music data were investigated in depth. We have also developed some experimental prototypes to show the feasibility of our proposals. In this project, the research on multimedia content analysis will be extended in three directions. First, core techniques on the content-based analysis of the AC-3 audio data will be studied. Second, we will study how to summarize MPEG-4 video objects and MP3 music objects based on the cinematic and music theories. Finally, since the illegal downloads of the MPEG-4 (or divx) movies and the MP3 songs can be found everywhere on the Internet, we will try to develop an automatically illegal download detecting system for these two kinds of medias.

Keywords: Content Analysis, Video Databases, Audio Databases, Content-Based Retrieval

二、緣由與目的

多媒體內涵式分析的技術為目前學界與業界的重點研究項目。本計劃的主要重心有三：首先，我們將對 AC-3 音訊此一廣泛應用於 DVD / HDTV 等媒體系統，然而 AC-3 音訊卻在內涵分析方面沒有學者對其研究的媒體格式，在 AC-3 音訊媒體特

性、環場模式建立(surrounding mode modeling)、特徵值擷取(feature extraction)、索引、事件偵測(event detection)、內涵式查詢等相關技術展開研究。其次，我們將針對MPEG-4 視訊的視訊摘要以及MP3 音訊格式的音樂摘要等研究主題，以電影理論以及樂理進行探討並實作一多媒體摘要系統。最後，我們將內涵式查詢的技術結合多媒體封包捕捉技術，用於MP3 音樂以及MPEG-4 (divx) 電影的非法下載偵測之上，試著解決目前網際網路上隨處可見的多媒體資料遭到非法下載的問題，保障多媒體內容產業業者的智慧財產權。

三、研究成果

(一) AC-3 音訊內涵式分析

DVD 為目前最主要的電影儲存媒介之一。DVD 與以往錄影帶或 VCD 相比，除了視訊的品質大幅提升之外，其音訊資料亦由立體聲改進為多聲道環場音效，而 Dolby AC-3(簡稱 AC-3) 為最主要之環場音效格式。在本計劃中，我們提出了一種針對 AC-3 電影音效與電影劇情關聯之資料探勘模型。我們透過對 AC-3 電影音效進行內涵分析，由 AC-3 解碼過程中擷取出其特徵係數，再根據 AC-3 的環繞音場特性以及人類聽覺感知性質，計算出環場音量字串，再透過一組樣板規則，找到隱藏在環境音量字串中的音效事件，並利用關聯法則分析演算法，進而對電影劇情與音效事件進行關聯分析，找到其間的關聯法則。最後，透過對電影劇情的關聯法則，便可以對未知電影進行自動摘要與劇情自動分析。

當我們在欣賞一部電影時，往往因為它的劇情，心情也會隨之起伏不定，例如驚悚的、感性的、緊張的劇情。在這些精采的片段中，往往都會配合許多的音效，例如戰爭片的爆破效果、喜劇片的笑聲、文藝片的抒情音樂等等，透過這些音效，讓觀眾的心境也不自覺的投入這部影片。

當導演想要表達出不同的情境時，他會利用適合的音訊來進行搭配，使得整體的效果更為彰顯。換言之，我們可以利用對音訊的分析，進而了解導演的內心世

界。如圖 1 所示，導演會將所要詮釋的情境傳達給攝影師以及剪輯的人員，也因為導演情境的不同，視訊的內容也會隨之改變。導演是電影的靈魂人物，換句話說，電影是導演內心情境的一種衍生物。透過電影，可以展現出導演內心的構思與想法。因此，我們認為電影視訊摘要必須以導演的中心思想為出發點，使得電影摘要能夠有效且快速的表達出整部電影的菁華內容。上述電影理論為本篇論文尋找 AC-3 環場音效與電影劇情關聯之基礎理論根源。



圖 1 導演情境與音訊的關聯圖

(二) MPEG-4 視訊摘要技術

由於多媒體應用與網際網路的快速蓬勃發展，使用者因此可以藉由網際網路去獲得許多的聲音、視訊、影像等多媒體資料，使得多媒體資訊變的隨手可得。多媒體資料的摘要(multimedia summarization)為目前多媒體資料庫領域的重要研究主題之一。在各種多媒體資料中，視訊資料的資料量最為龐大。因此為了讓使用者快速瞭解一個視訊物件的內容，我們需要視訊摘要(video summarization)的技術，將視訊物件的主要內容自動地擷取與呈現。而視訊摘要資料中又以電影資料最為重要。電影是現代人類最重要的文化資產之一，隨著數位化電影資料(VCD、DVD)日漸成為人們日常生活的一部份，因而電影資料的自動化摘要成為許多視訊/電影應用系統的重要技術。

一部電影在要上映之前的第一步就是為這部電影作宣傳，而最直覺宣傳的工具當然就是這部電影的海報、劇照、短片等電影摘要。這些電影摘要的設計，目的是為了讓所有使用者能對此部電影的大致內容能一目了然。尤其是電影海報，在宣傳時期可以說到隨處可見，就連是公車站牌都能看見正當宣傳期的電影海報。除此之外當然還有最動人心弦的電影預告片，可以說是一部電影最精華部分的直接呈現。從預告片中可以看到主要演員在重要場景中的演出，以及最具代表性的鏡頭(shots)。

不論是在電影海報、電影劇照或電影預告片中，都可以清楚的看到其中使用了大量的特寫鏡頭(close-up shot)。因此我們可以藉由特寫鏡頭的偵測及擷取，來自動合成電影摘要(movie summary)。我們也可將自動合成的電影摘要作為索引，用於電影資料庫內涵式查詢。如此一來使用者可在大量的電影資料中，很快的搜尋到有興趣欣賞的電影。此項技術為網際網路電影資料庫，VCD/DVD 線上租售系統，甚至行動式隨選視訊(mobile VOD)等應用系統所必需之技術。

本計劃所提出之電影摘要系統整體架構如圖 2 所示。我們假設要建立電影摘要的電影為 MPEG-4 格式，因為其中的物件已經標明，若為其他格式的電影資料，則須經過物件辨識的程序取出電影中的物件。經由鏡頭偵測模組(shot change detection)，以鏡頭為單位對一部電影進行切割。再將電影鏡頭做鏡頭前製處理(pre-processing)，取出關鍵視訊物件平面後，對視訊物件平面做特寫鏡頭偵測(close-up shot detection)。自動判斷出特寫鏡頭之後，將特寫鏡頭輸出到鏡頭叢集分析器(shot cluster analysis)做鏡頭叢集分析，進而判定演員的戲份。最後將鏡頭叢集輸出到電影摘要合成模組(movie summary synthesis)中，在電影摘要合成模組中，我們可以從演員的戲份決定各個演員的重要性，依照重要性調整其在合成電影摘要時所佔之比重。

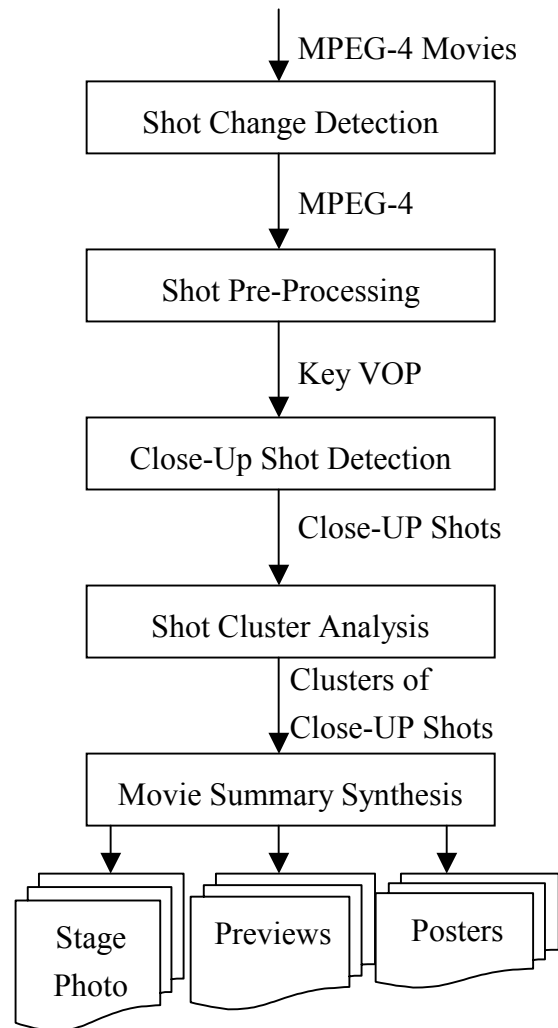


圖 2、電影摘要系統架構圖

(三) MP3 音樂自動化摘要系統

我們所提出的 MP3 音樂物件自動化摘要系統架構大致上可以分成四個模組，分別是前置處理(preprocessing)、發現音樂物件的重複樂句(discovering the repeating phrases)、叢集分類(clustering)和組合重複出現的音樂片段(assembling repeat fragments)四個模組。

前置處理模組主要是先將 MP3 音樂物件做自動斷句，將一首 MP3 歌曲切割成為一連串的樂句。接著透過重複樂句尋找模組來做 MP3 樂句特徵值擷取，進而分析整個音樂物件並發現重複樂句出現的位置。而叢集分類模組則是將前一步驟所得的結果作叢集分析，由此我們可以找到重複出現的 MP3 樂曲以及整個音樂物件的結構。最後組合重複出現的音樂片段模組，主要是要利用一個重複樂句合併演算法來選擇出現頻率較高並且是連續的音樂片段，將這些連續的音樂片段視為最後音樂摘要的結果。整個 MP3 音樂摘要系統架

構，如圖 3 所示。



圖 3 MP3 音樂物件自動化摘要系統架構

(四) MP3 Sniffer: MP3 音樂非法傳輸偵測系統

近來將版權音樂製成 MP3(MPEG 1 Audio Layer 3)的格式，透過網路傳輸、散佈的非法行為越來越猖獗，使得版權所有的唱片公司、歌手蒙受重大損失，也造成了社會大眾對音樂智慧財產權的重視及關心。目前在取締非法 MP3 的散佈與下載，都是靠警方以人力的方式直接進行，不但費時費力，甚至可能造成民怨，而且 MP3 這種以網路進行非法傳輸的方式，也不是以人力可以有效取締的，我們需要的是個更有效率的方式來改善此問題。

目前用來保護音樂智慧財產權的方式大體可以分為兩種，一是使用浮水印式的保護方式，另一是使用內涵式查詢的方法。浮水印式的保護方式是利用特殊的演算法在數位音樂資料中加入特別形式的「記號」，而此記號就算經各種不同的方式破壞仍可以被檢查出來，以達到保護音樂智慧財產權的目的，但令人遺憾的是，目前在網路上氾濫的 MP3 音樂資料絕大部分都是由 CD 直接轉檔成 MP3 的格式，而除非唱片發行公司能預先在 CD 壓製時加入浮水印，否則何來的浮水印可供檢查？因此我們採取的是另一種方式，也就是使用內涵式查詢的技術。

以往在多媒體內涵式查詢的技術之相關研究領域上，許多學者都將焦點擺在影像或視訊方面的研究，較少研究學者針對音訊的內涵做分析與探討。但是近年來，音訊內涵式的分析(audio content analysis)與研究已經逐漸成為非常重要的課題。

如圖 4 所示，本文所提出之“線上 MP3

音樂非法傳輸偵測系統”是由網路封包的攔截與分析器(MP3 Sniffer)、MP3 音樂資料庫(MP3 Database)、MP3 音樂識別器(MP3 Identifier)及 MP3 音樂認證伺服器(MP3 Authenticating Server)等四部分所組成。而 Web 服務的部分則是由 MP3 音樂認證伺服器所提供，後文將詳細說明各個部分的功能。



圖 4. 線上 MP3 音樂非法傳輸偵測之系統架構

- MP3 音樂封包捕捉器：負責將流經該網域區段的封包全數攔截，並利用 MPEG 組織所制定的 MP3 標頭格式，過濾出 MP3 的封包後，再傳送至 MP3 音樂認證伺服器作後續的特徵值取出，以及辨識所傳送的 MP3 是否為版權所有的歌曲。
- MP3 音樂認證伺服器：提供版權 MP3 音樂認證辨識等服務。並配合後端 MP3 音樂特徵值資料庫，提供一套完整的 Web 服務以解決當前 MP3 非法傳輸氾濫的問題。
- MP3 音樂特徵值資料庫：資料庫中存有目前各大唱片公司在國家中央圖書館註冊有案版權所有國語歌曲的 MP3 特徵值，以及其歌曲基本資料。
- MP3 音樂識別器：配合 MP3 音樂特徵值資料庫，MP3 音樂識別器可將 MP3 音樂認證識別器所傳來的之 MP3 特徵值和資料庫中所存之 MP3 特徵值加以比對，並回傳 MP3 相似度及歌曲基本資料。
- Web 服務用戶端程式：提供使用者一個方便使用 MP3 音樂認證伺服器所提

供之 Web 服務的使用介面，唱片公司可隨時利用此程式查詢所出版的音樂到非法下載的統計以及明細資料。

四、計畫成果自評

本計畫成果至 93 年 9 月止共計發表國內期刊論文一篇[1]，國外研討會論文兩篇[2][3]，國內研討會論文八篇[4][5][6][7][8][9][10][11]。另有兩篇國外期刊論文與兩篇國外研討會論文目前正在審稿中。研究成果亦十分豐碩。

在實作方面，我們亦完成了”MP3 音樂非法傳輸偵測系統”、”MP3 音樂自動摘要系統”、”MPEG4 電影特寫鏡頭偵測系統”、”MPEG4 失蹤人口自動協尋系統”等多項系統的開發。其中”MPEG4 失蹤人口自動協尋系統”獲得教育部指導、台灣微軟與雲林科技大學合辦之”2004 潛能創意盃程式設計競賽”佳作。”MP3 音樂非法傳輸偵測系統”實作成果亦十分豐碩。

五、參考文獻

- [1] 劉志俊、吳智偉、鄭煒平、許肇凌, “MP3 Sniffer: MP3 音樂非法傳輸偵測系統的設計與實作,” *Communications of IICM*, Vol. 7, No. 1, 2004.
- [2] Chih-Chin Liu and Hung-Min Yu, “Effective Retrieval of MP3 Music Objects WITH Background Reduction Technique,” in *Proc. of Third International Workshop on Content-Based Multimedia Indexing (CBMI 2003)*, 2003, IRISA, Rennes, France.
- [3] Chih-Chin Liu and Pang-Chia Yao, “Automatic Summarization of MP3 Music Objects,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*, 2004, Montreal, Quebec, Canada.
- [4] 姚邦佳、劉志俊, “MP3 音樂物件的自動化摘要,” 數位生活與網際網路科技研討會, 2003.
- [5] 陳毅修、盧偉智、楊志強、黃鈞蔚、林怡伸、歐家銘、陳竹晟、劉志俊, “基於 MPEG-4 臉部動畫的失蹤人口協尋系統之設計,” 數位生活與網際網路科技研討會, 2003.
- [6] 范世鎮、劉志俊, “利用特寫鏡頭偵測與主角辨識技術來自動建立電影摘要,” 第二屆數位典藏技術研討會, 2003.
- [7] 葉億真、陳建志、古鎬華、賴明楷、劉韋志、劉志俊, “音效資料的內涵式分類及其在電影資料庫的應用,” 第二屆數位典藏技術研討會, 2003.
- [8] 陳竹晟、歐家銘、林怡伸、劉志俊, “網際網路失蹤人口自動協尋系統之實作,” 第五屆網際網路應用與發展學術研討會, 2004.
- [9] 王卓毓、李政憲、謝傳祿、邱緯廷、姚邦佳、劉志俊, “數位音樂自動化摘要技術及其在網際網路 MP3 音樂資料庫之應用,” 第五屆網際網路應用與發展學術研討會, 2004.
- [10] 范世鎮、劉志俊, “MPEG-4 電影資料之內涵式摘要擷取與角色分析,” 數位生活與網際網路科技研討會, 2004.
- [11] 鄭煒平、劉志俊, “MPEG 電影音效自動分段系統,” 數位生活與網際網路科技研討會, 2004.