

行政院國家科學委員會專題研究計畫 成果報告

AC-3 音訊內涵式分析與 MPEG-4 電影摘要系統之設計與實作

計畫類別：個別型計畫

計畫編號：NSC93-2213-E-216-038-

執行期間：93年08月01日至94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

報告類型：精簡報告

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 94 年 10 月 31 日

行政院國家科學委員會專題研究計畫成果報告

AC-3 音訊內涵式分析與 MPEG-4 電影摘要系統之設計與實作

The Design and Implementation of A Content-Based AC-3 Audio Analysis and MPEG-4 Video Summarization System

計畫編號：NSC 93-2213-E-216-038

執行期限：93 年 08 月 01 日 至 94 年 07 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：吳智偉、鄭煒平、陳竹晟、歐家銘、陳毅修、蘇永隆、賴嘉琪、張俊堂、范世鎮 中華大學資訊工程學系

一、中文摘要

多媒體資料的內涵式查詢 (content-based retrieval of multimedia data) 與分析，是近年來學界與業界研究的重點之一。在各種多媒體資料中，以電影的資料的內涵最為豐富，且對我們日常生活的影響最大。因此，本計畫擬以電影作為內涵式分析的主要對象，對其摘要擷取相關技術進行深入探討。一般電影資料包含視訊 (video) 以及音訊 (audio) 兩種媒體。以往相關研究大多由視訊媒體著手，對電影的分段 (segmentation) 與摘要 (summarization) 的技術進行探討。而我們希望同時由視訊與音效的角度進行分析，針對 AC-3 格式的環場音效以及 MPEG-4 格式的視訊擷取其特徵值，建立表達其內涵的資料模型，並收集電影預告片分析其剪輯的規則與樣型，作為設計電影摘要合成時選取重要鏡頭的依據。

關鍵詞：AC-3, MPEG-4, 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. As one of the most content-rich media types, movies play a major role in our culture and everyday life. Therefore, in this project the issues about

content-based analysis, summarization and identification of movies will be investigated. Most of the previous works on the retrieval of movie data were done based on the analysis of video signals. However, both video signals and audio signals will be analyzed in our project. The data model for representing of the events associated with AC-3 surrounding sound effects will be presented. Techniques about how to extract MPEG-4 video features will also be investigated in depth. Based on the analysis of the patterns behinds movie previews, an automatic movie summarization system will also be developed.

Keywords: AC-3, MPEG-4, 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。我們提出一種方法，可以對 MPEG 音訊資料，電影中音效進行分段。藉由分析各種電影音效所具有的不同特性，可以找出特徵值規則，我們並設計出一套斷點偵測方法來搭配電影音效特徵值組來決定音效斷音位置，以完成音效自動分段系統。

三、研究成果

(一) AC-3 電影音效內涵式分析系統

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。

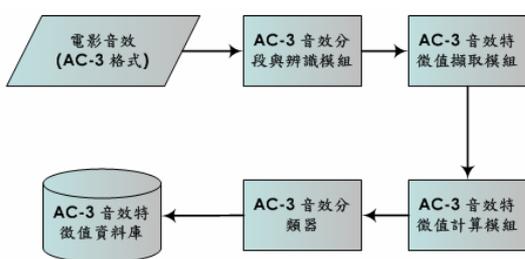


圖 1 AC-3 電影音效內涵式分析之整體架構

如圖 1 所示，我們將電影音效內涵式分析系統分成了許多的模組。電影中的音訊資料是相當龐大且複雜的資料，其包含了音樂、音訊及語音的訊息，直接分析是相當困難的。因此我們先利用 AC-3 音效分段模組將龐大的音訊進行分段，再利用 AC-3 音訊辨識模組進一步判斷音訊分段的身分，最後將身為音效的音訊分段收

集並進行 AC-3 特徵值的擷取程序。

要取得 AC-3 環場音效的資訊，我們必須先取得各個聲道的音訊特徵值。本文使用的特徵值來自於 AC-3 的解碼程序中的 6 聲道解耦合(de-coupling)模組之輸出。

我們由許多音訊分類的相關研究中，收集了各種不同的特徵值，並結合了國際規範 MPEG-7 中的音效描述子，提出一組 AC-3 音效特徵值。

(二) 電影音效自動分段系統

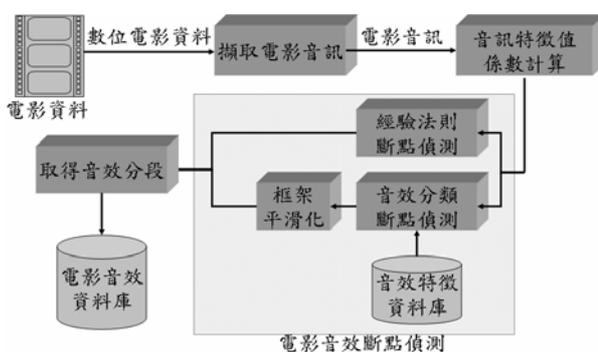


圖 2 電影音效分段之系統架構圖

圖 2 為本計劃所提出之電影音效分段架構圖，此系統可自動化將電影中所有音訊分段出來，並將其斷點位置及其音效片段存入電影音效資料庫。系統主要流程可分為以下六大步驟：

● 取得數位電影資料

今日生活中數位電影相當普及，不論是影音光碟或是網際網路，這類型的多媒體資料幾乎是垂手可得的。

● 電影音訊之擷取

MPEG 影片檔中包含了視訊及音訊兩種媒體。在取得原始 MPEG 影片檔後，首先要做的前置處理，就是將影音檔案中的音訊部份分離出來，以便做之後的音訊分析。

● 音訊特徵值係數計算

多媒體資料是由一組相當龐大且關聯性不大的數值所組成。在進行多媒體分析前我們必須先在這群資料中找出一套規則做為特徵值。在[40]文中提出了以能量、頻率、頻譜、能量頻率四大類 21 個特徵值，能描述電影中各種音效的特徵。我們利用這 21 個特徵值再配合 4ME 特徵值，將所

有的電影音訊擷取出其特徵值組，以便進行之後的音效段點偵測參考。

然而聲音訊號是一個連續性的資料，而且各個音效長度的差異度大，斷點出現位置也無規則性，因此我們先將電影的音訊以預先設定的固定長度進行音訊切片，擷取特徵值也就是以音訊切片為單位。

● 電影音效斷點偵測

影片中各音效段落的長度不同，要將音效片段切割出來，必須先找出各個音效片段的斷點位置，也就是相鄰兩段音效的交界處。由於之前送入的資料是連續的音訊切片，音效片段的斷點位置就會座落在當中的幾個音訊切片上。詳細的音效斷點偵測方法會於之後章節詳述。

● 電影音效分段

系統完成所有音效斷點偵測後，就可進行音效的分段工作。任意兩個斷點位置，分別表示某一段音效的開始位置和結束位置，只要將此兩點間所有的音效切片進行組合，就可獲得一段完整的音效分段。

● 建立電影音效資料庫

整部影片的音訊都被分斷後，就可將這些資訊建立資料庫，包含電影中所有段落音效、斷點位置及音效長度，將這些資料建立電影音效資料庫，可提供日後作電影音效摘要以及電影音效分析等工作。

(三) 系統實作

我們整合了先前的研究成果，擴充 MPEG-7 音效描述子，使得 AC-3 音效描述子更能夠反映 AC-3 的環場音效特性。

我們使用 Microsoft Visual Studio .NET 開發環境來實作 AC-3 音效特徵值計算工具，有興趣者可聯絡作者取得。此工具之使用方式說明如下：首先，先開啟預備分析的 AC-3 檔案，如圖 3 所示。接下來針對開啟的 AC-3 音效檔案進行特徵值擷取與計算的工作，並將擷取的特徵值存入檔案與資料庫中，以便重複利用。如圖 4 所示，擷取特徵值後，我們會將特徵值還原成 MDCT 係數，以便進一步計算音效特徵描述子。在工具畫面中，我們繪出 AC-3 音效在時間軸的能量分布概況，畫面右下角則會簡單地描述 AC-3 檔案特性。



圖 3 AC-3 音效特徵描述工具

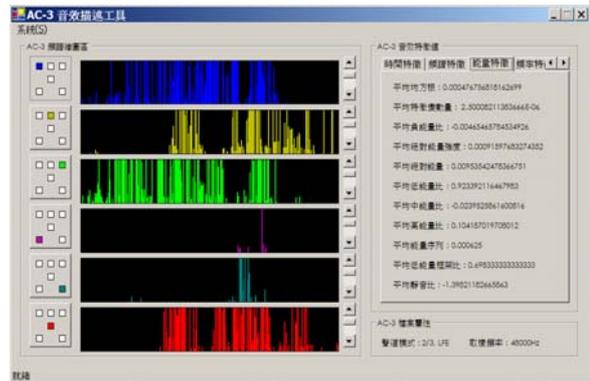


圖 4 AC-3 頻譜與音效特徵值

畫面左方的六個按鈕分別表示 AC-3 的六個聲道，根據使用者按下的聲道按鈕，畫面右方的音效特徵值列表將隨之改變。在特徵值列表中，我們進一步將音效特徵值區分為「時間特徵」、「頻譜特徵」、「能量特徵」、「頻率特徵」以及「頻率能量特徵」五大類型。

由於 XML 可適用於不同平台，具有便於資料交換的優點，以及具備自我擴張與定義標籤的特性。因此 MPEG-7 的核心描述定義語言的基本架構是利用 XML 來呈現。針對 AC-3 音效特徵值，我們也可將其輸出為 XML，以便與其他 MPEG-7 工具互相支援。

在完成 AC-3 音效特徵值計算後，我們使用 Microsoft SQL Server 2000 來儲存特徵值，建立了 AC-3 音效特徵值資料庫，以備後續的音效分析使用。

```

<?xml version="1.0" encoding="utf-8" standalone="no" ?>
<!--
AC-3 音場特徵描述文件
Filename = Audio_01.AC3
-->
<!--features-->
<!--FRONT_LEFT-->
<FrameNumber>600</FrameNumber>
<AveSpectralCentroid>0.760238751765964</AveSpectralCentroid>
<AveSpectralRollOff>2.41</AveSpectralRollOff>
<AveRMS>0.000476756818162699</AveRMS>
<AveFeatureVariance>2.50008211383666</AveFeatureVariance>
<AveHegPower>0.0045465784534926</AveHegPower>
<AveIntensity>0.00091597683274352</AveIntensity>
<AvePower>0.00953542478366751</AvePower>
<AveLowEnergy>0.923392116467983</AveLowEnergy>
<AveMidEnergy>0.0239525861600816</AveMidEnergy>
<AveHegEnergy>0.104187019708012</AveHegEnergy>
<AveEnergySequences>0.000625</AveEnergySequences>
<AvePercentage_of_Low_Energy_Frames>0.698333333333333</AvePercentage_of_Low_Energy_Frames>
<AveSR>1.39821187605063</AveSR>
<AveBandwidth>30.9375</AveBandwidth>
<AveFeg>6896.08137639201</AveFeg>
<AveMaxFeg>647.03125</AveMaxFeg>
<AveLowFegPower>0.77971603221227</AveLowFegPower>
<AveMidLowFegPower>0.128332781796961</AveMidLowFegPower>
<AveMidFegPower>0.173160702823757</AveMidFegPower>
<AveMidHegFegPower>0.156038827578746</AveMidHegFegPower>
<AveHegFegPower>0.00293851176076456</AveHegFegPower>
</FRONT_LEFT-->
<!--CENTER-->
<FrameNumber>600</FrameNumber>
<AveSpectralCentroid>12.8497793874437</AveSpectralCentroid>
<AveSpectralRollOff>10.605</AveSpectralRollOff>
<AveRMS>0.00012100558031783</AveRMS>
<AveFeatureVariance>1.202863275708E-08</AveFeatureVariance>
<AveHegPower>0.0012495751159809</AveHegPower>
<AveIntensity>1.01906716688581E-05</AveIntensity>
<AvePower>0.00290439248109406</AvePower>
<AveLowEnergy>0.89207724914594</AveLowEnergy>
<AveMidEnergy>0</AveMidEnergy>
<AveHegEnergy>0.0918790023950193</AveHegEnergy>
<AveEnergySequences>0.000703125</AveEnergySequences>
<AvePercentage_of_Low_Energy_Frames>0.713333333333333</AvePercentage_of_Low_Energy_Frames>
<AveSR>0.776038718252311</AveSR>

```

圖 5 AC-3 音效描述子之 XML 範例檔

- 位生活與網際網路科技研討會, 2005.
- [8] 吳智偉、劉志俊, “AC-3 環場音效與電影劇情關聯之資料探勘模型,” 第三屆數位典藏技術研討會, 2004.
 - [9] 劉謙易、歐陽儒、江彥志、林秋旺、劉志俊, “點對點 MP3 音樂分享系統歌曲傳輸之偵測技術,” 第三屆數位典藏技術研討會, 2004.
 - [10] 鄭煒平、劉志俊, “網際網路電影資料庫之音效自動分段索引系統,” 第六屆網際網路應用與發展學術研討會, 2005.
 - [11] 陳竹晟、劉志俊, “MP3 指紋與網際網路 MP3 音樂傳輸即時偵測技術,” 第六屆網際網路應用與發展學術研討會, 2005.

四、計畫成果自評

本計劃成果至 94 年 9 月止共計發表國內期刊論文一篇[1]，國內研討會論文十篇[2][3][4][5][6][7][8][9][10][11]。

在實作方面，我們亦完成了”AC-3 電影音效內涵式分析系統”、”MP3 音樂自動摘要系統”、”電影音效自動分段系統”等多項系統的開發。

五、參考文獻

- [1] 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,” 中華理工學刊, Vol. 3, No. 1, pp. 169-176, 2005.
- [2] Chih-Chin Liu and Wei-Yi Kuo, “Content-Based Segmentation of MP3 Music Objects,” in Proc. of Workshop on Computer Music and Audio Technologies (WOCMAT), 2005, Taipei, Taiwan.
- [3] Chih-Chin Liu and Pang-Chia Yao, “Content-Based Summarization of MP3 Music Objects,” in Proc. of Workshop on Computer Music and Audio Technologies (WOCMAT), 2005, Taipei, Taiwan.
- [4] 吳智偉、劉志俊、蘇永隆、賴嘉琪、張俊堂, “支援 MPEG-7 之電影 AC-3 環場音效內涵描述工具,” 二〇〇五數位生活與網際網路科技研討會, 2005.
- [5] 歐家銘、劉志俊, “使用 AC-3 音效特徵來進行電影資料的自動場景偵測,” 二〇〇五數位生活與網際網路科技研討會, 2005.
- [6] 陳毅修、劉志俊, “MP3 音樂哼唱式查詢在嵌入式系統上的實作,” 二〇〇五數位生活與網際網路科技研討會, 2005.
- [7] 陳信光、劉志俊, “在無線網路環境下使用不均等可靠性來傳輸 JPEG2000 影像,” 二〇〇五數