

MPEG-7 音效內涵描述自動擷取工具之實作

The Implementation of an MPEG-7 Tool for Extracting Audio Content Descriptions

賴明楷 陳建志 古鎬華 劉韋志 劉志俊
中華大學資訊工程學系
ccliu@chu.edu.tw

摘要

隨著大量地多媒體資料的產生，多媒體資料的內涵式查詢成為目前極重要的研究主題。為了統一多媒體內涵特徵值的表示方式，建立多媒體內涵描述的標準，因而產生了 MPEG-7 此一項新的多媒體國際標準。MPEG-7 雖然對影像(images)、視訊(video)、與音訊(audio)制定了大量的內涵描述子，但對聽覺性媒體如音樂與音效的內涵描述子便有所欠缺。因此，本篇論文針對此項缺失，整合目前 MPEG-7 的音訊描述子，以及以往在音效與音訊的內涵式查詢與分類相關文獻所提出之音訊特徵值，按照 MPEG-7 標準所制定的資料定義語言以及內涵描述的資料綱目的語法格式，定義了 26 種音效描述子，補充 MPEG-7 在音效特徵描述不足的部分。我們實作了一套 MPEG-7 音效特徵描述自動擷取工具，可以對每一個 MP3 音效檔進行特徵值的擷取動作，並產生各個音效檔的特徵描述內容的 XML 檔案。我們也將電影音效的 26 種描述子及其類別，儲存在關聯式資料庫中，建立一個電影音效特徵資料庫，作為電影與音效資料探勘技術研究發展的基礎。

關鍵詞：MPEG-7、音效描述子(sound effect descriptors)、內涵描述子擷取工具(content descriptor extracting tools)、音效內涵式查詢(content-based sound effect retrieval)、音效資料庫(sound effect databases)

一、序論

在科技發達的現代化社會裡，由於以往的傳統多媒體資料儲存方式，不僅不容易保存，

且需要佔用大量的實體空間來存放。因此隨著科技的進步，數位化多媒體設備越來越普及，諸如 CD、VCD、DVD Player、掃描器、數位相機，等數位影音資料的存取設備等，皆以融入人們的日常生活中，可見我們今日已經完全處於數位化多媒體的時代。

在網路頻寬急速成長的今日，網際網路上的多媒體資料與日遽增。根據學者的估計，一年全世界所生產的多媒體資料量高達 1.5 EB = 15 億個 GB。換算成地球上每人每年可分到 250 MB 的多媒體資料。一般使用者如今可以透過網際網路非常方便地下載到數位化多媒體資料，但我們要如何在大且不斷成長的多媒體資料中，有效率的找到我們所想要的資料呢？一般直覺的想法是採用現在相當普遍的搜尋引擎，如 Google、Yahoo 等來幫助我們做多媒體資料搜尋。但是我們發現這樣做的成效並不太理想，這主要是因為現今的搜尋引擎絕大部分所採用的都是以傳統的關鍵字的方式來做搜尋，但使用者所想要的多媒體資料往往是很難用文字去表達。因此多媒體資料內涵式查詢(content-based multimedia data retrieval)成為目前學術界與多媒體內涵產業界的研究焦點。所謂的內涵式查詢，簡單地說，就是對所想要尋找的多媒體資料做內容的分析並擷取其特徵值，作為相似性查詢比對搜尋的依據的技術。

因為休閒娛樂事業的發達，電影成為重要的人類文化資產。因此，電影資料的內涵式分析技術變得非常重要。我們知道電影的組成可區分成視訊資料與音訊資料這兩大部分。以往對於視訊資料內涵式分析的研究成果非常眾多，但電影音效資料內涵式分析的研究十分稀少。所以本篇論文我們針對 MPEG 格式的數位電影音效資料進行分析，計算其各種特徵值來建立音效內容描述，如此便可以利用電腦做音效資料的內涵式搜尋。而為了能有效地描述音效內容並使其具有互通性，我們採用

*本論文研究為國科會補助之研究成果，計劃編號 NSC 92-2213-E-216-015

MPEG-7 標準中的描述定義語言(Description Definition Language, DDL)來制定電影音效檔的綱要定義(Schema Definition)。我們亦實作了一套 MPEG-7 音效特徵描述自動擷取工具,可以對每一個 MP3 音效檔進行特徵值的擷取動作,並產生各個音效檔的特徵描述內容的 XML 檔案;最後我們利用此工具建立一個包含 20 部電影中所有一千餘個 MP3 音效資料的音效特徵資料庫,提供給有興趣的人員作為從事相關研究的比較參考。

本篇論文結構如下:在第 2 節我們將簡介 MPEG-7 與音效特徵描述子的涵義;第 3 節介紹音效描述子的詳細定義及由 MPEG 音效檔案取出特徵值的計算方式;第 4 節說明音效描述子的資料綱目;第 5 節為電影音效內涵描述工具實作說明;第 6 節說明 MPEG-7 音效資料庫的建立方式;第 7 節為結論及未來的工作。

二、MPEG-7 與音效特徵描述子的涵義

2.1 MPEG-7 標準簡介

MPEG-7 為 MPEG 所制定的一項國際標準(ISO/IEC 15938)[7][8][9][10]。MPEG-7 與以往 MPEG 所制定的標準如 MPEG-1, MPEG-2 與 MPEG-4 等多媒體壓縮標準相比其制定的目的有很大的差異。MPEG-7 不是一個多媒體壓縮標準,而是多媒體內涵描述介面(Multimedia Content Description Interface)。MPEG-7 規範了描述各種多媒體資料如影像(images)、視訊(video)、與音訊(audio)等資料的內涵,所使用的資料定義語言(DDL, data definition language)的語法、內涵描述的資料綱目(Description Schemes)[10]與各種描述子(Descriptors)[9]的涵義,有關 MPEG-7 介紹性的論文可參考[5][6][14][15][17][19]。

由於多媒體內涵描述的種類繁多且格式複雜, MPEG 組織在 2000 年決定採用 XML[23][24][25][26]為多媒體內涵描述的資料定義語言。XML 文件為半結構化的資料型態,其可擴充性十分適合表達涵義豐富且格式複雜的多媒體內涵,且容易閱讀與適合程式進行自動處理。MPEG-7 標準對影像(images)、視訊(video)、與音訊(audio)制定了大量的內涵描述子,並希望成為所有多媒體內涵相關研究所曾經提出的特徵值(features)的總集合(superset)。但以目前的標準文件來看,內涵描述子以視覺性媒體如影像與視訊定義的較為完整,而聽覺性媒體如音樂(music)與音效(sound effects)的內涵描述子便顯得不夠完整而有所欠缺。因此,本篇論文針對此項缺失,整合目前 MPEG-7 的音訊描述子[9],以及在

音效與音訊的內涵式查詢與分類相關文獻[1][2][4][11][12][13][16][18][21][22]所提出之音訊特徵值,按照 MPEG-7 標準所制定的資料定義語言以及內涵描述的資料綱目的語法格式,定義了 26 種音效描述子,以補充 MPEG-7 在音效特徵描述不足的部分。

2.2 系統架構

如同其他 MPEG 標準的做法, MPEG-7 標準定義了多媒體內涵描述子的語法格式與涵義,但如何由多媒體原始資料中擷取這些內涵描述子的方法,並不在標準的範圍之中。因此我們亦實作了一個音效內涵描述自動擷取工具,可以由 MPEG 音效檔案中,擷取出 26 種音效描述子,並以 XML 檔案儲存這些音效描述子,如圖 1 所示。

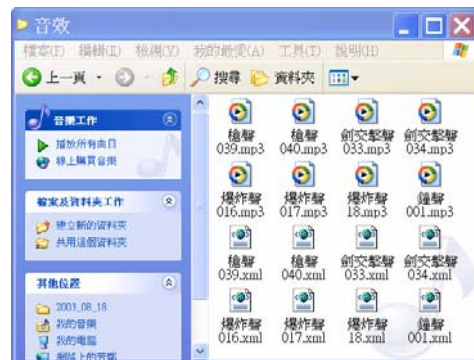


圖 1 MP3 音效檔與音效內涵描述子範例

音效描述子的 XML 檔案可以應用在多媒體資料搜尋引擎(multimedia search engines)、智慧型代理人(intelligent agents)、多媒體資料庫(multimedia databases)等各種多媒體音效的應用系統之中,系統架構如圖 2 所示。我們亦針對電影資料的內涵分析此一問題,採用其音效組成輪廓(sound effect profiles)的做法,提出一種將音效自動分類的方法,並可自動進行電影分類[2][3]。

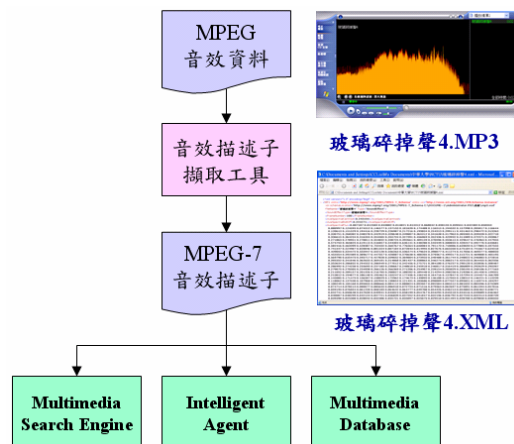


圖 2 音效內涵描述自動擷取系統架構

三、MPEG-7 音效特徵描述子

3.1 MPEG 音效特徵值的擷取來源

對於一般音訊資料可大致分為語音(Speech)、音樂(Music)、音效(Sound Effect)三大類別來分析[22]。對電影資料而言，語音是指角色對話的聲音、音樂是指電影中的背景音樂、音效即是如槍擊聲、爆破聲、尖叫聲等特殊音效。在本文中我們取音效作為討論的對象，其原因除了 MPEG-7 對音效的描述子的定義有所欠缺之外，對於電影資料而言，音效往往較能表現出各種電影類型的區別。

在本文中我們所討論的音效為 MP3(MPEG 1 Audio Layer3)檔案格式的電影音效。由於 MP3 音效格式的壓縮原理是利用聽覺遮罩(Masking)，濾掉一些人們所聽不到的頻率(聲音)，所以可把音效壓縮成原始的十分之一大甚至是十二分之一，如此即可減少資料的儲存空間。雖然說壓縮造成音效失真，但是對於一般人人耳來說聽起來差別不大。所以我們將原始的電影音效，以人工方式逐一取自 VCD 檔案格式電影的音軌之後，儲存為 MP3 檔案格式。

MP3 是經過赫夫曼編碼(Huffman Encoding)的數位音樂格式，無法從 MP3 原始資料(raw data)中找出具內涵代表意義的特徵值，故我們必須在 MP3 解碼的過程中來擷取內涵特徵值，再進一步進行 MPEG-7 音效描述子的計算。MP3 原始資料每一框架(frame)在經過赫夫曼解碼後為 576 個修正離散餘弦轉換係數(MDCT, Modified Discrete Cosine Transform)係數，這些 MDCT 係數接著經過反離散餘弦轉換後產生 32 個次頻帶各 36 個多相位濾波器組係數(polyphase filter bank coefficients)。我們 MDCT[i]表示第 i 條頻率線的修正離散餘弦轉換係數，據以計算頻率域(frequency domain)的描述子；而以 S[j]表示第 j 個次頻帶的多相位濾波器組係數，據以計算時間域(time domain)的描述子。

3.2 MPEG-7 音效描述子的擴充

作為一種標準多媒體內容描述介面，MPEG-7 主要制定的目的是為了要規範描述視訊及音訊的內涵的涵義與語法。因為每個人對於多媒體的內容描述不盡相同，所以 MPEG-7 希望能規範的描述各種多媒體資料型態內涵的標準。多媒體內容提供者(content provider)與多媒體搜尋引擎等應用系統若能遵循 MPEG-7 標準，使用者即可在不同平台上交換多媒體資訊。

過去在音效特徵值分析的部分已有許多的學者發表出許多的計算公式來描述各種音效的

特徵涵義[1][2][4][11][12][13][16][18][21][22]，我們整理這些音效的特徵並匯整 MPEG-7 適用於音效的音訊描述子(audio descriptors)，擴充定義了一組音效描述子(sound effect descriptors)，其涵義如表 1 所示。

表 1 音效描述子涵義與計算方式

Sound Effect Descriptors	涵義
FrameNumber	框架數，即音效檔的長度
LogAttackTime	上升時間對數值，即 ADSR 特性的 A 上升時間(attack)之對數值
AveSpectralCentroid	平均頻譜質量中心
AveSpectralRollOff	平均頻譜偏斜
AveSpectralFlux	平均頻譜通量，即 576 條頻率線的個別通量
AverageFlux	總平均頻譜通量
AveNZFlux	平均非零頻譜通量
AveRMS	平均均方根
AveFeatureVariance	平均特徵變動量
AveNegPower	平均負能量比
AveIntensity	平均絕對能量強度
AvePower	平均絕對能量，即絕對音量
AveLowEnergy	平均低能量比
AveMidEnergy	平均中能量比
AveHigEnergy	平均高能量比
AveEnergySequences	平均能量序列：能量差大於 0.7 倍的平均能量的次數比
AvePercentage of Low-Energy Frames	平均低能量框架比
AveSR	平均靜音比：小於等於 0.05 倍的最大能量所佔之能量比
AveBandwidth	平均頻寬
AveFeq	平均頻率，即平均音高
AveMaxPfeq	平均最大能量之頻率
AveLowFeqPower	平均低頻能量比：低於 200Hz 頻率所佔之能量比
AveMidLowFeqPower	平均中低頻能量比：200Hz 到 500Hz 頻率所佔之能量比
AveMidFeqPower	平均中頻能量比：500Hz 到 1kHz 頻率所佔之能量比
AveMidHigFeqPower	平均中高频能量比：1kHz 到 2kHz 頻率所佔之能量比
AveHigFeqPower	平均高频能量比：高於 2kHz 頻率所佔之能量比

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。我們根據表 1 所定義之音效描述子，以一個槍聲音效範例，其音效描述子之 XML 檔如圖 3 所示。

```

<?xml version="1.0" encoding="Big5"?>
<MP3 xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.mpeg7.org/2001/MPEG-7_Schema C:\mp3.xsd"
FileName="槍聲11" Type="SoundEffect">
<SoundEffectType>武器：槍聲</SoundEffectType>
<FrameNumber>12</FrameNumber>
<AveSpectralCentroid>0.158775</AveSpectralCentroid>
<AveSpectralRollOff>0.160446</AveSpectralRollOff>
<AveSpectralFlux>0.02566 0.01762 0.058257 0.083456 0.058260 0.112132 0.199819
... 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
</AveSpectralFlux>
<AverageFlux>0.017060</AverageFlux>
<AveMF1Flux>0.025999</AveMF1Flux>
<AveMF2Flux>0.000144</AveMF2Flux>
<AveFeatureVariance>0.000002</AveFeatureVariance>
<AveLegPower>0.029563</AveLegPower>
<AveIntensity>0.000028</AveIntensity>
<AveEnergySequence>0.092174</AveEnergySequence>
<AveLowEnergy>0.015131</AveLowEnergy>
<AveMidEnergy>0.060944</AveMidEnergy>
<AveHighEnergy>0.776708</AveHighEnergy>
<AveEnergyPercentage>0.092174</AveEnergyPercentage>
<AvePercentageOfLowEnergyFrames>0.083333</AvePercentageOfLowEnergyFrames>
<AveBandwidth>0.012008</AveBandwidth>
<AveF0p>0.482199</AveF0p>
<AveMF0p>0.013600</AveMF0p>
<AveLowF0pPower>0.114308</AveLowF0pPower>
<AveMidF0pPower>0.118671</AveMidF0pPower>
<AveHighF0pPower>0.163366</AveHighF0pPower>
<AveMidHighF0pPower>0.190311</AveMidHighF0pPower>
<AveHighF0pPower>0.246677</AveHighF0pPower>
<AveSR>0.145076</AveSR>
</MP3>

```

圖 3 MPEG 音效描述子之 XML 範例檔

四、音效特徵描述綱目

根據 MPEG-7 標準，描述子的資料型態與結構以及語義是定義在描述綱目 (description schemes) 之中 [10]。根據 XML 語言的規範，描述綱目是儲存在 XSD 檔案之中。我們考慮表 1 中所定義之音效描述子所制定的描述綱目如圖 4 所示。

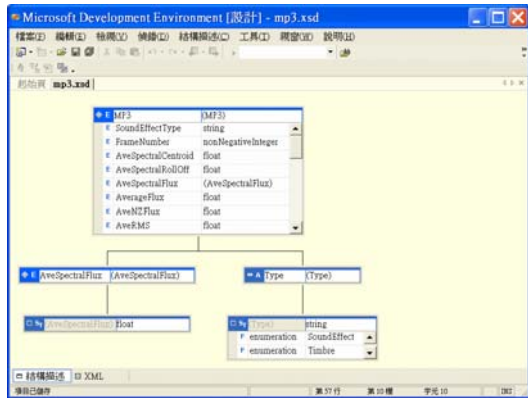
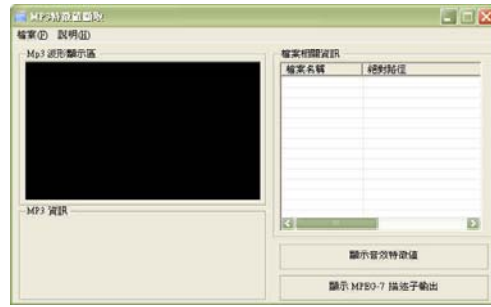


圖 4 音效描述綱目圖示

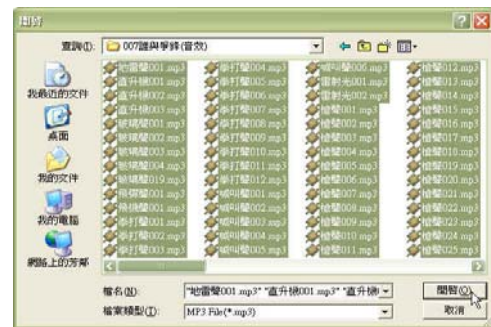
五、音效特徵描述子擷取系統之實作

我們在 Window XP 的環境下，以 C++ 與 C# 語言完成了一套 MPEG-7 音效特徵描述自動擷取工具，對此工具有興趣者可與作者聯絡取得。其中修正離散餘弦轉換係數與多相位濾波器組係數的擷取功能修改自 amp11 此一 MP3 解碼器，作者為 Niklas Beisert

(<http://www.ph.tum.de/~nbeisert/amp11.html>)。音效特徵描述子擷取流程如下：



(a)



(b)



(c)



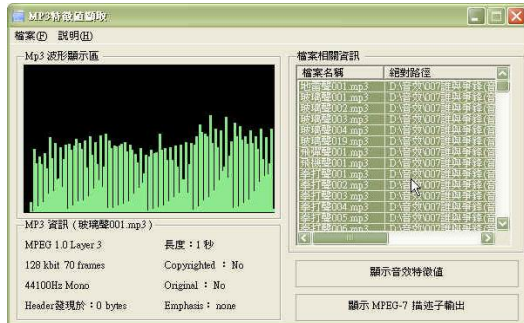
(d)

圖 5 音效特徵描述自動擷取工具操作畫面

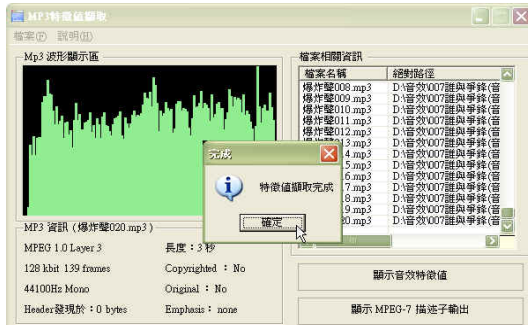
- (1) 當執行此程式時，會出現如下圖 5(a) 的畫面。
- (2) 而後接下來的動作就是要讀入 MP3 檔案，來進行後續的特徵值擷取的動作，點選左上角 Menu 中的“檔案(F)”，然後點選開啟檔案。
- (3) 接下來如圖 5(b)，到儲存音效檔的資料夾

中選取要擷取特徵值的檔案(可複選),選取完成後按下右下角的開啟按鈕。

- (4) 如下圖 5(c), 選取的音效檔會出現在右方的檔案相關資訊中, 點選其中的檔案會出現其相關資訊並且讀取其 MP3 檔案的 Header 並將其資訊顯示於左下角。



(a)



(b)

圖 6 音效特徵描述自動擷取工具操作畫面

- (5) 接下來的動作就是要擷取 MP3 檔案中的特徵值, 如圖 5(d)所示在右方的檔案相關資訊中選取要擷取其特徵值的檔案, 選取完成後按下滑鼠右鍵, 而後會出現有三個選項的選單, 第一個“清除所有的項目(A)”是將所有的 List 中的檔案全都清除, 第二個“清除選定的項目(S)”為清除 List 中被選取到的檔案, 而最後一個“執行選定的項目(E)”就是執行擷取 MP3 檔案特徵值的動作, 並將其結果輸出於 XML 檔案中。擷取特徵值後產生的 XML 檔將會與來源音效存放於同一個目錄下, 主檔名與來源音效檔主檔名相同, 但副檔名由 MP3 改為 XML, 如由“槍聲 037.mp3”音效檔中擷取特徵值後產生的 XML 檔案, 其檔名將會是“槍聲 037.xml”。

- (6) 圖 6(a)為特徵值擷取執行中的畫面, 擷取進行時會同時撥放其音效, 並將各框架能量顯示於左方的 MP3 波形顯示區中。當全部選取的音效檔案的特徵值都擷取完畢後, 將如下圖 6(b)中所示, 出現特徵值擷取完成的訊息。

- (7) 特徵值擷取完成後, 可以點選檔案相關資

訊中的其中一個檔案, 然後可如圖 6(b)按右下方的“顯示音效特徵值”按鈕來讀取 XML 檔案並視窗中依照特徵的類型來顯示其特徵值, 其畫面如圖 7 所示, 或按下“顯示 MPEG-7 特徵值描述子輸出”按鈕直接顯示 XML 檔案, 其畫面如圖 8 所示。

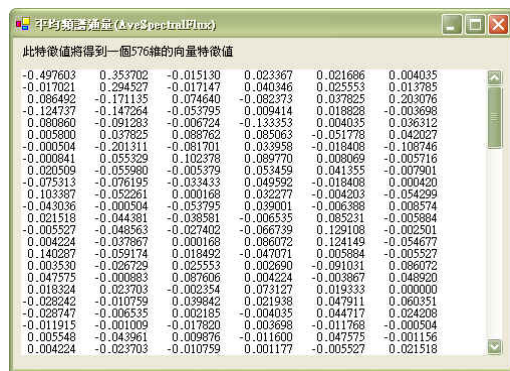


圖 7 螢幕顯示各種音效特徵描述子之數值

Movie 類別記錄關於電影的中文名稱、英文名稱、導演、出版年份與片長等資料；SoundEffect 類別記錄關於音效的 MP3 音效檔名、在影片出現的開始時間、音效長度等資料；SoundType 類別記錄音效的分類名稱、類別的輪廓特徵 (sound profile) 等資料；SoundFeatures 類別記錄電影音效的 26 種描述子之特徵值。我們使用 MS SQL Server 來實作此資料庫。資料庫部分內容如表 2 與表 3 所示。

七、結論與未來工作

在本篇論文中，我們整合目前 MPEG-7 的音訊描述子，以及以往在音效與音訊的內涵式查詢與分類相關文獻所提出之音訊特徵值，按照 MPEG-7 標準所制定的資料定義語言以及內涵描述的資料綱目的語法格式，定義了 26 種音效描述子，補充 MPEG-7 在音效特徵描述不足的部分。我們亦實作了一套 MPEG-7 音效特徵描述自動擷取工具，可以對每一個 MP3 音效檔進行特徵值的擷取動作，並產生各個音效檔的特徵描述內容的 XML 檔案。此外我們也建立了一個 MP3 電影音效資料庫，作為電影資料典藏與音效資料探勘等有關研究的基礎。

由於我們目前在電影音效的分段上還是採取人工的方式，每一個音效的尋找及其前後邊界的決定需要反覆視聽，因此在電影音效特徵值資料庫的建立上需要花上很長的人工時間。在未來的研究工作方面，我們有以下兩個目標：第一是電影音效自動化切割 (sound effect segmentation)。由於目前是以人工的方式擷取音效，所以目前音效的樣本數量上還不夠多，無法應付大量的電影資料之建置要求。因此，我們的首要目標在於能建立一個自動擷取音效的程式，可以讓我們在收集音效上的速度能加快，以節省資料收集的時間。第二是建立良好的索引結構來進行查詢。當資料庫中的資料個數越來越龐大的時候，我們就必須使用良好的多維度特徵空間索引結構來增加資料庫搜尋的速度，提高查詢處理的反應時間。

八、參考文獻

- [1] 游弘明、劉志俊，”以哼唱方式查詢 MP3 音樂資料庫，” 中華民國九十年全國計算機會議，2001。
- [2] 葉億真，”音效資料的內涵式分類及其在電影資料庫之應用”，中華大學資訊工程碩士論文，2003。
- [3] 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 Pfeffer, S., S. Fischer, and W. Effelsberg, “Automatic Audio content Analysis,” ACM Multimedia 96, Boston MA USA. 術研討會，2003。
- [4] Agostini, G., Longari, M. and Pollastri, E., “Musical instrument timbres classification with spectral features,” IEEE Fourth Workshop on Multimedia Signal Processing, pp. 97-102, 2001.
- [5] Shih-Fu Chang, Thomas Sikora, and Atul Puri, “Overview of the MPEG-7 Standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 688-695, June 2001.
- [6] Hunter, J., “An Overview of the MPEG-7 Description Definition Language (DDL),” IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 765-772, June 2001.
- [7] ISO/IEC JTC1/SC29/WG11 N4980, “MPEG-7 Overview,” Version 8, Klagenfurt, July 2002.
- [8] ISO/IEC FDIS 15938-1/PDAM1:2001(E), “Information Technology — Multimedia Content Description Interface — Part 1: Systems,” Awaji, Japan, December 2002.
- [9] ISO/IEC 15938-4:FPDAM, “Information Technology — Multimedia Content Description Interface — Part 4: Audio,” Shanghai, China, Oct 2002.
- [10] ISO/IEC 15938-5, “Information Technology — Multimedia Content Description Interface — Part 5: Multimedia Description Schemes,” Version 7.0, Singapore, March 2001.
- [11] Chih-Chih Liu and Chuan-Sung Huang, “A Singer Identification Technique for Content-Based Classification of MP3 Music Objects” 11th International Conference on Information and Knowledge Management, 2002.
- [12] Guojun Lu and Hankinson, T., “A technique towards automatic audio classification and retrieval,” ICSP '98., vol. 2, pp. 1142 -1145, 1998 ”.
- [13] Martin, K. D., Scheirer, E. D., and Vercoe, B. L. “Music content analysis through models of audition,” ACM Multimedia Workshop on Content Processing of Music for Multimedia Applications, Bristol, England, September, 1998.
- [14] Nack, F. and Lindsay, A., “Everything You Wanted to Know About MPEG -7 : Part 1,” IEEE Multimedia , vol. 6 issue 3 , July-Sept.

- 1999.
- [15] Nack, F. and Lindsay, A., "Everything You Wanted to Know About MPEG -7 : Part 2," *IEEE Multimedia* , vol. 6 issue 3 , July-Sept. 1999.
- [16] Pfeiffer, S., S. Fischer, and W. Effelsberg, "Automatic Audio content Analysis," *ACM Multimedia 96*, Boston MA USA.
- [17] Quackenbush, S. and Lindsay, A., "Overview of MPEG-7 audio," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 725-729, June 2001.
- [18] Rossignol, S. et al., "Feature extraction and temporal segmentation of acoustic signals," *International Computer Music Conference 1998*, Ann Arbor, MI, USA, Oct. 1998.
- [19] Philippe Salembier and John R. Smith, "MPEG-7 multimedia description schemes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 748-759, June 2001.
- [20] De Santo, M. et al. , "Classifying audio of movies by a multi-expert system," *Image Analysis and Processing*, pp 386-391, 2001 .
- [21] Tzanetakis, G., G. Essl, and P. Cook, "Automatic Musical Genre Classification of Audio Signals," In *Proc. Int. Symposium on Music Information Retrieval (ISMIR)*, Bloomington, Indiana, 2001.
- [22] Wold, E. et al., "Content-based classification, search, and retrieval of audio," *IEEE Multimedia*, vol. 3, no. 2, pp. 27-36, 1996 .
- [23] Extensible Markup Language (XML) 1.0, W3C Recommendation, 2000. <http://www.w3.org/TR/REC-xml>
- [24] XML Schema Part 0: Primer, W3C Candidate Recommendation, 2001. <http://www.w3.org/TR/xmlschema-0/>
- [25] XML Schema Part 1: Structures, W3C Candidate Recommendation, 2001. <http://www.w3.org/TR/xmlschema-1/>
- [26] XML Schema Part 2: Datatypes, W3C Candidate Recommendation, 2001. <http://www.w3.org/TR/xmlschema-2/>