

音效資料的內涵式分類及其在電影資料庫的應用

葉億真 陳建志 古鎬華 賴明楷 劉韋志 劉志俊

中華大學資訊工程學系

ccliu@chu.edu.tw

摘要

在本篇論文中，我們提出了一種 MPEG 電影音效的分類方法。透過對 MPEG 電影音效作內涵分析，可以擷取出代表其內容的音效特徵值。這些音效特徵值可以用來做兩個音效物件間的相似度比對，再利用音效分類器便能對一部電影中的音效進行自動分類的工作。最後，透過電影的音效組成分佈情形可以建立各種電影類型，如動作片、警匪片、驚悚片等的音效輪廓。對於一部未知的電影，我們取出其音效部分來進行音效自動分類後，再將結果進一步與各種電影類型的音效輪廓比對其相似性，我們即可判別出此電影所屬的影片類型。

關鍵字：音效分類 (sound effect classification)、內涵式分析 (content-based analysis)、電影資料庫 (movie databases)、電影音效輪廓 (sound effect profiles of movies)

1. 簡介與相關研究

由於數位時代潮流的衝擊下，世界各國都致力於數位資料典藏的工作，例如：聯合國的世界記憶計畫、美國的美國記憶計畫等。以期能藉此提昇國家的競爭力與永久保存國家的文化資產。因此我國政府成立了數位典藏國家型科技計畫，將國家重要的文物典藏數位化。此外各地方政府也開始注重電影文化資產的保存與呈現，例如：新竹市的影博館、高雄市電影圖書館等。而其中高雄市電影圖書館珍藏了約三千多部的外中經典電影。根據統計在 2000 年香港就上映了 151 齣電影，因此我們可以感受到電影片數的快速成長。而要管理如此龐大的電影資料勢必將要花費許多人力與時間。因此，要如何有效地分類電影成為相當重要的研究主題。

我們知道電影內容是由影像資料與音訊資料所結合而成的。所以可以藉由分析影像或音訊資料兩方面來進行電影的分類。但影像資料量遠大於音訊資料，而在電影的音訊資料中，又以音效資訊為最重要。這是因為電影的音效組成與電影的類型息息相關，例如：動作片會具有較高的槍聲音效比例、喜劇片則會具有較高的笑聲音效比例等等。所以我們可以利用分析電影的音效組成，來進行電影分類的工作。

以往在多媒體的相關研究領域上，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；雖然如此，可是關於音訊內涵式方面的現有研究，大多著重於音訊資料流的分割 (segmentation) 及其音樂分析，卻很少研究學者將焦點擺在電影音效的分析上。但以往在音訊內涵相關研究所提出的特徵值，亦可作為電影音效分析的參考。

關於音訊物件內涵探討可以粗分為語音 (speech)、音樂 (music)、音效 (sound effect) 三大類研究。其中以語音與音樂類的相關研究較多。在 Saunders[9] 中，利用音訊的音調、頻寬、音高、音調的持續時間、能量序列等特徵值，從無線電廣播 FM 之音樂中分辨出語音。經由對所有的特徵值進行正規化的動作來提昇分辨率至 90%，然後又利用統計能量輪廓及越零率使其分辨正確率可達到 98%。而 Scheirer [10] 由語音或音樂訊號中計算出 13 種不同的特徵值，然後利用這些聲音特徵來分析聲音內涵。另外，在 Lu[3] 一文中，利用靜音比、頻譜質量中心、和諧度、音高此 4 個特徵值來分辨音樂與語音，其平均正確率約為 80%。

*本論文研究為國科會補助之研究成果，計劃編號 NSC 91-2213-E-216-003

在 Tzanetakis [11][12][13][14]的系列研究中，提出利用結合多項音訊分析工具：頻譜質量中心、頻譜通量、音高、和諧度、低能量等特徵，便可以藉由音訊分析模組來進行音樂風格分類、音訊切割、音訊查詢、瀏覽等應用。此外，作者還利用 MFCC、高斯分類器和 kNN 分類器透過音訊分析之統計結果來對音訊加以分類。

在 Chih-Chin Liu [2]中提出一個根據歌手自動分類 MP3 物件的方法。作者先從 MP3 的解碼過程中，擷取出多相位濾波器係數，來作為 MP3 的特徵值。再利用此特徵值來切割 MP3 物件成為一音素的序列。然後，依據每個音素的框架之 576 個 MDCT 係數，計算得其音素特徵值。再利用音素特徵值作為 MP3 歌手分類的依據，其分類正確率約為 80%。

在音效方面的研究敘述如下。在 Wold[15]一文中，首先提出利用音訊的知覺特徵來代表音訊，再結合這些知覺特徵為一組特徵向量。然後便可以計算音訊查詢樣本和資料庫中的音訊樣本之間的歐幾里德距離，對兩個音訊樣本作相似度比對。但是這樣的作法可能會讓許多不懂那些特徵含意的人在查詢時感到困惑，因為他們可能不懂那些特徵的含意。如同在 Martin[6]中，提到傳統依據音樂理論及音樂訊號處理的方法尚無法建構出成功的音樂多媒體系統。這是因為一般的使用者無法透過這些專業音樂理論來理解音樂的內涵性。所以，作者希望能建立一個類似人類的聽覺系統，有著與人類處理、判斷音樂的相同行為，而這也正是研究音訊處理之終極目標。若要達到此一目標，則必須朝向“結合許多有用的聲音感知特徵，然後利用數學統計的方法來分析這些特徵值，最後使用圖形識別技術來訓練系統，使系統可以達到最佳化”此一方向研究。

而在 Foote[4]一文中，提出一個以資料導向的方法來擷取音訊資料。此方法其使用語音處理常用的 MFCC 特徵值來做音樂與音訊的分類。而分類的技術則是使用量化樹。其對音效分類可達 60%~75%的正確率，而在音樂分類的正確率則為 30%~40%。Foote

音效分類的結果與 Wold 等人的分類結果近似。在 Pfeiffer[8]音訊的研究中，提到音訊內涵可以從兩個方面來觀測：一個是從音訊本質的角度，例如：音訊的振幅、波形等；另一方面是可由人類認知的角度來探究。在此文中，作者主要是針對暴力影片的相關音訊（包括哭聲、槍聲、爆炸）來作分析，並且測試每個特徵屬性其所應該採用屬性比重應是如何設定。

本篇論文結構如下：在第 2 節我們將說明電影音效分類系統之整體架構，第 3 節介紹電影音效的分類方法，第 4 節為實驗與結果分析，第 5 節為結論及未來的的工作。

2. 電影音效分類系統之整體架構

在本節中，我們將介紹整個電影音效分類系統架構的執行流程，以及簡述每個音效分類系統模組的功能。從圖 1 MPEG 電影音效分類系統整體架構圖中，可以大致了解我們對於電影音效分類的處理程序。以下我們將會對每一個過程做一簡單地描述。

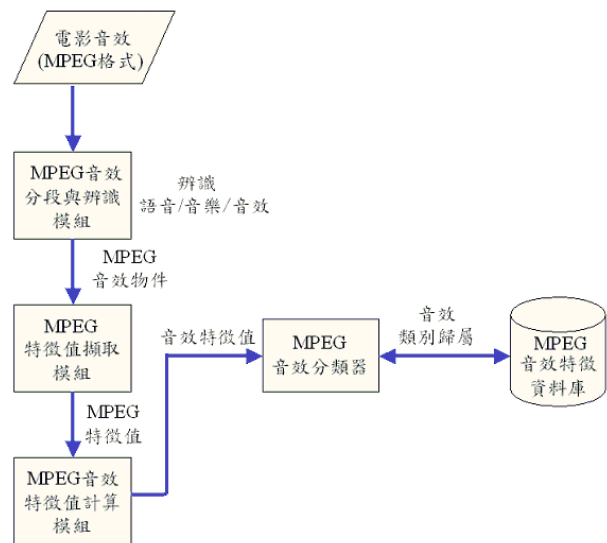


圖 1. MPEG 電影音效分類系統之整體架構

2.1 電影音效(MPEG 格式)

由於 VCD 格式為目前最普遍的電影儲存格式，而且使用 MPEG-1 的壓縮技術。此外，由 MPEG 組織所制定的數位音樂檔案的壓縮格式 MP3(MPEG-1 Audio Layer3)，其壓縮比約為 CD 音樂的十分之一至十二分之一。基於 MP3 檔案格式具有高壓縮比與近乎 CD 音樂高品質的此兩項優勢，所以我們

將原始的电影音效儲存為 MP3 的檔案格式。

2.2 MPEG 音效分段與辨識

由於整個電影 MPEG 音效物件是非常龐大並且混雜各種音效，因此很難直接從龐大的聲音資料去做分析與辨識。所以，我們先利用 MPEG 音效分段模組將 MPEG 音效檔進行分段，以使用來作音訊內涵分析。然後我們使用 MPEG 音效辨識模組來識別每個 MPEG 音訊片段是屬於語音、音樂或是音效類別。最後將所有屬於音效類別之 MPEG 音效檔案進行特徵值擷取動作。

2.3 MPEG 特徵值擷取模組

由於 MP3 檔案是一種經過赫夫曼編碼的數位音樂格式，因此，我們無法直接從 MP3 的原始資料中，找出有意義的資訊。所以我們必須在 MP3 的解碼過程中擷取 MP3 特徵值。因為我們是從壓縮領域中抓取特徵值，不需要將 MP3 檔案完全解碼，所以將會大幅縮短擷取特徵值所花費的時間。

依據 ISO 國際標準組織對於數位音樂的壓縮標準，MP3 的位元流會一個框架接著一個框架地轉入成 MDCT(Modified Discrete Cosine Transform)係數，然後藉由將頻率樣本轉換為時間樣本的動作，將 MDCT 係數(576 條頻率線)對應到 32 個子頻帶，而此 32 個子頻帶又稱為多相位濾波器係數(Polyphase Filter Coefficients)。最後，將合成每個子頻帶成為原始的聲音訊號(PCM Audio)。

我們所擷取的 MPEG 音效檔之特徵值，即是從 MP3 解碼過程中擷取出 576 個 MDCT 係數當為 MPEG 音效之原始特徵值。然後，經由對擷取出的原始特徵值作正規化後，再將正規化的音效特徵值傳送到音效特徵值計算模組，來作各種有意義的音效特徵值計算及統計分析。

2.4 MPEG 音效特徵值計算模組

由音訊研究 Saunders [9] 中，可以發現特徵值若經過正規化動作後，可以提昇音訊分類的準確率。這是因為透過將特徵值的能量/此音訊的總能量的正規化動作，如此一來就可以消除聲音大小的影響。因此我們將經

由 MPEG 特徵值擷取模組所取出的 576 個 MDCT 係數正規化。然後透過 MPEG 音效特徵值計算模組在時間領域與頻率領域上作各種 MPEG 音效特徵值的計算。最後利用這些音效特徵值來作音效的分類工作。

我們在此所使用的 MPEG 音效特徵是從音訊分類相關研究中所使用的特徵與 MPEG-7 中音效描述子[7]，經過匯集整理出來的。我們對每項音效特徵值，深入研究其如何由 MPEG 音效特徵(576MDCT 係數)計算而得。表 1 整理我們所使用的音效特徵值。

音效特徵值	特徵值涵義說明
框架數	記錄 MPEG 音效檔的檔案長度
平均頻率	表 MP3 音效檔的平均音訊頻率
平均頻寬	表 MP3 音效檔的平均音訊頻寬
平均頻譜質量中心	表音訊頻普能量分佈的平衡點
平均頻譜偏斜	量測頻普形狀的不對稱性
平均頻譜通量	觀測兩兩框架之間的平均變化量
總平均頻譜通量	計算 MP3 音效檔中 576 個平均頻普通量之總平均值
平均非零頻譜通量	求所有大於零的頻普通量的平均值
平均均方根	表框架的大小聲
平均特徵變動量	求整個 MP3 音效檔的平均特徵變動量
平均負能量比	計算 MP3 音效檔中，能量小於零之 MDCT 係數的次數所佔的比例
平均能量強度	計算 MP3 音效檔的平均能量強度
平均能量	求平均總能量值
平均低能量比	計算低於 0.3 倍平均能量之能量和佔總能量的比例
平均中能量比	計算介於 4.5 倍平均能量與 5.5 倍平均能量之能量和佔總能量的比例
平均高能量比	計算高於 0.7 倍平均能量之能量和佔總能量的比例
平均能量序列	計算 MDCT 係數中，能量差大於 0.7 倍的平均能量的次數比例
平均低能量框架比	計算 MP3 音效檔中，框架的平均均方根小於 0.5 倍總平均均方根所佔的框架比例
平均靜音比	計算小於等於 0.05 倍的最大能量 MDCT 係數所佔之能量比
平均最大能量之頻率	計算 MPEG 音效檔擁有最大能量的平均頻率
平均低頻能量比	計算低於 200Hz 頻率所佔之能量比
平均中低頻能量比	計算介於 200Hz 到 500Hz 頻率所佔之能量比
平均中頻能量比	計算介於 500Hz 到 1kHz 頻率所佔之能量比
平均中高頻能量比	計算介於 1kHz 到 2kHz 頻率所佔之能量比
平均高頻能量比	計算高於 2kHz 頻率所佔之能量比

表 1 音效特徵值表

3. 電影音效的分類

3.1 MPEG 音效特徵資料庫的建立

我們將每個 MPEG 音效物件所擷取的特徵經過正規化動作後，再作各種音效特徵值的計算，然後將計算所得的特徵值儲存於 MPEG 音效特徵資料庫中，如圖 2，在此我們利用計算特徵模組所得的 25 類特徵，每個電影音效樣本總共有 600 個音效特徵值，對 MPEG 音效特徵資料庫中每筆音效特徵紀錄做相似度比對。然後透過 kNN 分類器[5]進行音效分類並且儲存到 MPEG 音效特徵資料庫中，以作為電影分類的參考。

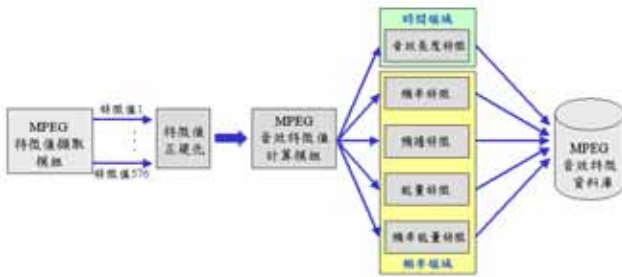


圖 2 MPEG 音效特徵資料庫之建立架構圖

3.2 音效分類

在我們建立了 MPEG 音效特徵資料庫之後，我們就可以利用資料庫中的音效特徵資料來做電影音效分類的工作。如圖 3 即是我們的音效分類的示意圖。在此所謂的“欲分類的 MPEG 音效特徵向量”是指經由 MPEG 特徵值擷取模組與 MPEG 音效特徵值計算模組後所得之音效特徵向量。然後把此特徵向量與資料庫中的所有特徵向量進行相似度比對，再利用 kNN 分類器分辨出所屬之音效種類。



圖 3 MPEG 音效分類的示意圖

在計算電影音效特徵的相似度上，我們使用了歐基理德距離來評估兩個 MPEG 音效

特徵之間的相似度。在比較兩個 MPEG 音效特徵的相似度時，由於每個特徵值的辨識能力與重要性是不一樣的，所以我們會分別計算兩個 MPEG 音效特徵向量中各個的歐基理德距離後，再乘上此特徵的權重值，然後將所有音效特徵計算所得的值加總起來。此外為了方便比較相似度，我們將最後加總所得之值進行正規化，使其介於 0 到 1 之間。其公式如下：

$$Similarity = 1 - \sqrt{\frac{\sum (Wi(Feai - Febi))^2}{600}} \quad (\text{公式 1})$$

其中，Wi: 特徵之權重值 Feai, Febi: 兩個音效的特徵向量

3.3 電影音效輪廓的建立

若要我們試圖去辨識電影的種類時，那我們會怎麼做呢?直覺地，我們會想說直接拿電影來看一次，如此就可以知道電影的類別了。但這種方式是相當費時的，不適用於擁有大量電影資料的電影資料庫。之前我們曾經提到，可以透過分析電影的音效組成來進行電影的分類。這是因為各種類型的電影，其音效的組成分佈基本上有一定規則，譬如說：警匪片的槍聲音效會佔有較高的比例等等。

我們統計每種電影類型中的音效分佈，可以建立代表此電影類型的電影音效輪廓(profile of sound effects)。建立方式如下：我們先分析多部電影的音效組成分析，如表 2 所示，藉以取得武俠片、警匪片與驚悚片，此三類型的電影平均音效組成分布概況，如表 3 與圖 4 所示。由圖 4 電影類別音效輪廓圖中，我們可以得知武俠類電影具有高兵器音效的比例(76.9%)特性、警匪類電影中槍聲音效佔有較高的比例(50.2%)特性、驚悚類電影擁有較高的尖叫音效的比例(35.5%)特性。

電影類別	電影名稱	爆炸	槍聲	聲音	兵器	過橋	尖叫	叫聲	哭聲	哭聲	哭聲	哭聲	其他
武俠片	臥虎藏龍	0	0	0	76.4	12.6	0	3.1	0	0	0	1.6	6.3
武俠片	英雄	0	0	0	77.4	0	0	17.2	0.6	0	0	0	4.8
警匪片	教父 I	1.8	44.7	0	0	0	10.7	0	0	12.5	8.9	8.9	12.5
警匪片	教父 II	1.8	56.1	0	0	0	1.7	0	0	5.3	0	3.5	31.6
警匪片	教父 III	0	64	0	0	0	2.6	0	0	0	0	2.6	30.8
警匪片	魔鬼終結者 I	12	36.1	2.3	6.8	0	2.3	0	0	0	0	3.7	1.5
警匪片	魔鬼終結者 II	5.5	72.4	9.7	0	0	0	0	0	0.7	0	0	6.2
警匪片	驚聲尖叫 I	0	2.8	0	0	1.8	39.4	12.8	4.6	0	13.8	2.8	22
警匪片	驚聲尖叫 II	0	13.1	0	0.8	13.9	27.9	7.4	3.3	0	18	1.6	14
警匪片	驚聲尖叫 III	0	14.5	0	0	17.9	0	39.3	8.6	2.6	0	11.1	4.3

表 2 電影音效組成分析表

電影類別	爆炸	槍聲	震彈	警車	兵器	過招	尖叫	叫聲	笑聲	哭聲
武俠片	0	0	0	0	76.9	6.3	0	10.1	0.3	0
警匪片	3.9	50.2	0.6	1.7	0	0	4.3	0	0	4.5
驚悚片	0	10.1	0	0.3	11.2	0	35.5	9.6	3.5	0

表 3 電影類別平均音效組成分布表

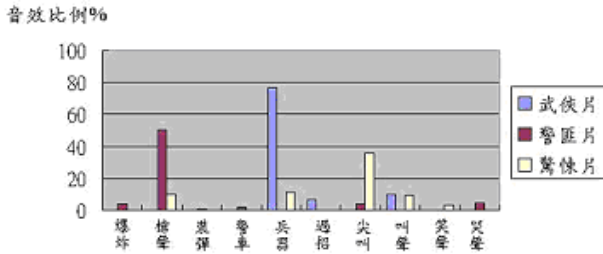


圖 4 電影類別音效輪廓圖

4. 實驗

在實驗的部分，首先我們會先介紹我們的實驗環境，接著我們也會提到影響我們音效分類準確率的因素有哪些，以及最後實驗的結果與探討。

4.1 實驗環境

我們用來進行電影音效分類實驗所採用的資料樣本，皆是擷取自 VCD 格式的電影中。在初步實驗中，我們從多部電影檔中，擷取出各種類別的電影音效。以人工方式建立電影音效分類樹，如圖 5 所示。

而在音效分類系統的訓練階段中，我們將先使用槍聲、刀劍聲、尖叫聲、笑聲、過招聲，與電話鈴聲這 6 種音效與一些其他的音效，如警車聲、玻璃碎掉聲等等，共 200 個音效，如表 4 所列。然後利用這些 MPEG 音效檔來對音效分類系統進行訓練的動作。以求得最佳的 kNN 分類器的各項參數，如：有效半徑、決策者個數，與各種音效特徵的權重值。之後我們將利用此資料樣本來進行電影分類，而有關電影分類部分請參考 4.3.3 節。

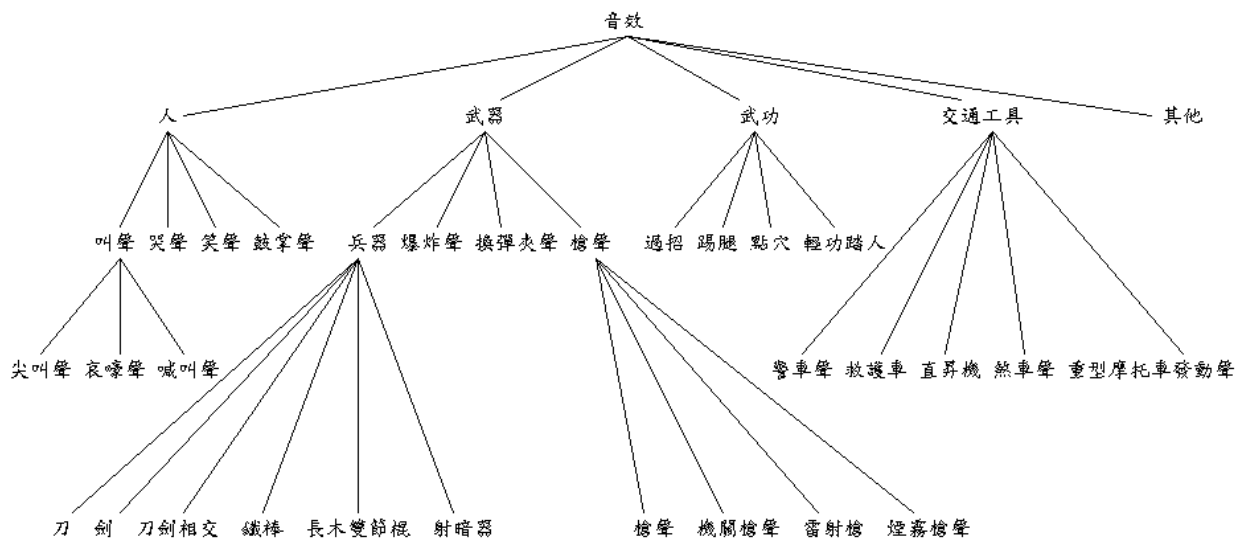


圖 5 電影音效分類樹

音效類別	資料來源	資料數
槍聲	教父全集	40
刀劍聲	臥虎藏龍	40
尖叫聲	驚聲尖叫 I、II	40
笑聲	驚聲尖笑	16
過招聲	臥虎藏龍	11
電話鈴聲	教父全集 & 驚聲尖叫 II	40
其他音效	警車聲、玻璃碎掉聲、鼓掌聲、...等	13

表 4 音效分類系統訓練之資料樣本來源表

4.2 影響音效分類準確率的因素

在我們的實驗中，我們會探討幾個會影響 MPEG 音效分類正確率的因素，其中包含了：kNN 分類器中的決策者個數(k)與音效特徵之權重值(w)、有效半徑(r)等。以下即針對這三個因素做些簡短的說明：

決策者個數 k 值是指：最多可以有多少個具有資格的物件來進行音效分類工作。所以若 k 值太小，則可能造成分類決策不夠公

正，而降低分類準確度；反之，若 k 值太大的話，則會使得分類準確率與分類效率都會降低。

由於每個特徵值所代表的音效涵義是不同的，所以每個特徵對於音效分類工作的影響力也會有所差異的。因此我們不能將所有的特徵一視同仁，必須依照各個特徵對音效分類的影響程度，調整其重要性亦即所謂的音效特徵權重值(w)。適當的音效特徵權重值將可以提升音效分類的準確度。

於 kNN 分類器中，有效半徑是用來設定與欲分類物件的有效判別距離。而唯有在有效判別距離內的物件，才具有成為分類決策者的資格。如果有效半徑設定的太小，則會造成沒有任何物件可以符合資格；反之，若設定過大的話，則等於沒有設定有效半徑，如此不但會降低分類的正確率而且也喪失了設定有效半徑的意義。由此可知有效半徑的大小將會深深地影響分類的準確率。

4.3 實驗結果

在此節中，我們將分別針對三個影響音效分類正確率的因素來加以探討，並利用實際真實的電影 MPEG 音效檔案來進行電影音效分類實驗，以求得最佳的實驗參數，進而提升電影音效分類系統的整體準確率。

4.3.1 實驗影響因素探討

首先我們先針對 k 值的大小，分析其對於電影音效分類準確率的影響程度。我們將控制音效特徵權重值、有效半徑這兩個影響因素。經由圖 6 顯然可以看出當 $k = 5$ 時可以獲得最佳的分類準確率。

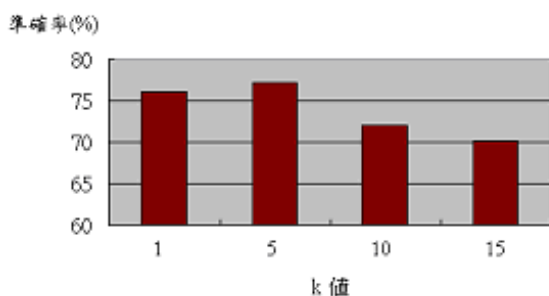


圖 6 k 值對分類準確率之影響

然後我們將分析不同組合的音效特徵權重值，其對於音效分類準確率的影響。為了簡化問題的複雜度，所以我們將不探討 600

個音效特徵權重影響，而是依據 5 類的音效特徵來做音效權重影響分析。再者我們先將音效特徵權重之重要性劃分為低，中，高 3 個等級，且依序的權重值為 0.1、0.5、1。所以共有 243 種音效特徵權重組合。經由分析此 243 種音效權重組合，我們可以得到：音效長度特徵類別之最佳音效權重值介於 0.1 至 0.5 之間、頻譜特徵類別的最佳音效權重值為 1、能量特徵類別的最佳音效權重值為 1、頻率特徵類別的最佳音效權重值則介於 0.5 到 1 之間、頻率能量特徵類別的最佳音效權重值為 1。因此我們針對音效長度特徵類別與頻率特徵類別的最佳特徵權重值範圍內，每 0.1 為一權重單位，然後以更加精準的角度來觀測。最後我們得知長度特徵類別的最佳音效權重值為 0.5，而頻率特徵類別的最佳音效權重值為 0.6。經由以上的探討與分析，我們得知了三個影響 MPEG 音效分類準確率因素最佳的設定值，而每個實驗影響因素的最佳設定值，如表 5 實驗影響參數值之設定表所列。

影響音效分類準確率因素 最佳設定值	
有效半徑	0.01
決策者個數	5
音效長度特徵類別	0.5
頻譜特徵類別	1
能量特徵類別	1
頻率特徵類別	0.6
頻率能量特徵類別	1

表 5 實驗影響參數值之設定表

最後將探討 r 值大小的變動率，對於電影音效分類準確率所造成的影響。在此我們將固定決策者個數與音效特徵權重值這兩個影響實驗的參數，排除會影響分類準確度的干擾。然後藉由觀測 r 值大小的變動對於音效分類準確率所造成的影響，進行分析以求得最佳 r 值。由圖 7 中我們可以明顯的知道當 $r = 0.01$ 時，可達到最佳音效分類準確率。而當 $r = 0.002$ 時，則由於有效半徑過小，所以導致準確率偏低的結果。

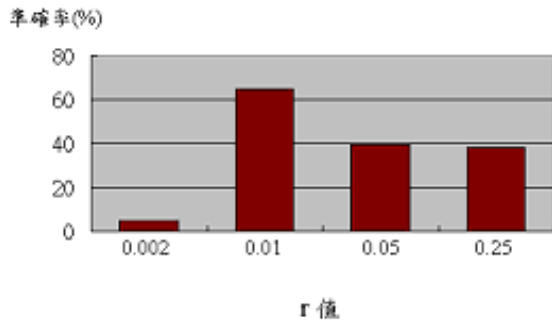


圖 7 有效半徑對分類準確率之影響

4.3.2 MPEG 電影音效分類結果

我們使用 4.3.1 節實驗分析所得的影響實驗因素的最佳設定值，來進行 MPEG 電影音效的分類。表 6 為實驗進行時所使用的測試樣本。電影音效分類實驗結果如圖 8 所示。而 MPEG 電影音效分類的整體平均準確率約為 76.6%。而表 7 為電影音效分類結果的交錯矩陣(confusion matrix)。

音效類別	資料來源	資料數
槍聲	魔鬼終結者 I、II	40
刀劍聲	英雄	40
尖叫声	驚聲尖叫 III、我的左眼看到鬼	40
笑聲	驚聲尖笑、我的左眼看到鬼	16
過招聲	臥虎藏龍	11
電話鈴聲	驚聲尖叫 II、III	40

表 6 音效分類系統測試之資料樣本來源表

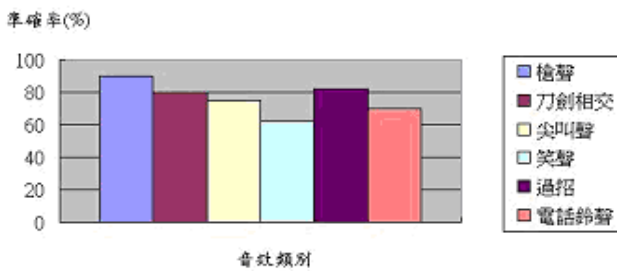


圖 8 MPEG 電影音效分類結果

分類成 原始樣本	槍聲	刀劍相交	尖叫声	笑聲	過招	電話鈴聲
槍聲	18	1	0	1	0	0
刀劍聲	0	16	1	0	1	2
尖叫声	0	1	15	4	0	0
笑聲	2	0	3	10	0	1
過招	0	1	0	1	9	0
電話鈴聲	2	0	1	2	1	14

表 7 電影音效分類結果之交錯矩陣表

4.3.3 利用音效來進行電影的自動分類

表 8 為目前電影音效資料庫所含之音效樣本數。然後我們利用電影音效分類系統分別對英雄、教父III與驚聲尖叫I這三部不同類型的電影作電影音效的分類，以取得此三部電影的音效組成分析。而表 9 與表 10 為英雄的音效組成分析表。最後再將這三部電影各自的電影音效組成分布與表 3 電影類別平均音效組成分布作相似度比較，則結果如表 11 所示。由表 11 中我們可以發現電影自動分類可達相當不錯的結果，而教父III與驚悚片的相似度達 72.2 % 這是因為教父III中叫聲與尖叫声約佔有 15 %。

音效類別	資料來源	資料數
爆炸聲	教父 I、II & 魔鬼終結者 I、II	25
槍聲	教父 I、II & 魔鬼終結者 I、II	236
裝子彈聲	教父 I、II & 魔鬼終結者 I、II	17
警車聲	魔鬼終結者 I	10
兵器聲	臥虎藏龍	135
過招聲	臥虎藏龍	16
尖叫声	臥虎藏龍 & 驚聲尖叫 II-III & 魔鬼終結者 I & 我的左眼看到鬼	148
叫聲	魔鬼終結者 II & 我的左眼看到鬼 & 驚聲尖笑	40
笑聲	教父 I、II	26
哭聲	教父 I、II & 魔鬼終結者 I & 驚聲尖叫 II-III	7
電話聲	教父 I、II & 魔鬼終結者 I、II	81
破碎聲	教父 I、II & 魔鬼終結者 I、II	26
其他	教父 I、II & 魔鬼終結者 I、II	84

表 8 MPEG 電影音效資料庫之音效樣本

音效類別	實際音效分布	電影音效系統分析	音效佔有率(%)
爆炸聲	0	11	2.3
槍聲	0	5	1
裝子彈聲	0	19	4
警車聲	0	7	1.5
兵器聲	369	296	62.1
過招聲	0	1	0.2
尖叫声	0	17	3.6
叫聲	82	44	9.2
笑聲	3	4	0.8
哭聲	0	0	0
電話聲	0	4	0.8
破碎聲	0	0	0
其他	23	69	14.5

表 9 英雄之電影音效分析表

分類成 原始樣本	爆炸	槍	裝	警	兵	過	尖	笑	哭	電	破	其
爆炸	0	0	0	0	0	0	0	0	0	0	0	0
槍聲	0	0	0	0	0	0	0	0	0	0	0	0
裝彈	0	0	0	0	0	0	0	0	0	0	0	0
警車	0	0	0	0	0	0	0	0	0	0	0	0
兵器	2	0	18	1	296	0	11	0	2	0	0	39
過招	0	0	0	0	0	0	0	0	0	0	0	0
尖叫声	0	0	0	0	0	0	0	0	0	0	0	0
叫聲	6	0	0	2	0	0	6	44	1	0	4	19
笑聲	0	2	0	0	0	0	0	0	1	0	0	0
哭聲	0	0	0	0	0	0	0	0	0	0	0	0
電話	0	0	0	0	0	0	0	0	0	0	0	0
破碎	0	0	0	0	0	0	0	0	0	0	0	0
其他	3	3	1	4	0	1	0	0	0	0	0	11

表 10 英雄音效分類結果之交錯矩陣表

測試電影	電影分類		
	武俠片	警匪片	驚悚片
英雄	97.1%	36.0%	62.8%
教父 III	21.1%	98.1%	72.2%
驚聲尖叫 I	23.9%	63.3%	97.7%

表 11 未知電影依音效組成自動分類結果

5. 結論

在本篇論文中，我們提出了一個 MPEG 電影音效分類架構，透過對 MPEG 電影音效作內涵分析，擷取出其本身的內涵特徵值，得到 MPEG 電影音效的特徵向量。然後就可以使用音效的特徵向量來做相似度比對，利用 kNN 分類器對電影音效進行分類的工作。最後，透過對電影的音效組成分佈，我們可以分析出電影所屬的影片類別。

在 MPEG 電影音效分類上，最大的問題就在於從電影中所擷取出的 MPEG 音效包含了背景音樂，而背景音樂不只會降低 MPEG 電影音效分類的準確率，也會使得分類的正確率變得很不穩定。而於 [1] 論文中，提出了一個削減 MP3 背景音樂的方法，來提昇 MP3 音樂內涵式查詢的準確率。而在未來的研究方向，我們大致上有四個目標：一、消除 MPEG 電影音效之背景音樂的影響；二、充實 MPEG 音效特徵資料庫；三、建立良好的索引結構；四、電影音效自動切割。

6. 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] Chih-Chin Liu and Chuan-Sung Huang, “A Singer Identification Technique for Content-Based Classification of MP3 Music Objects” *11th ACM International Conference on Information and Knowledge Management*, 2002.

[3] Guojun Lu and Hankinson, T., “A technique towards automatic audio classification and retrieval,” *Signal Processing Proceedings, 1998. ICSP '98.*, vol. 2, pp. 1142–1145, 1998.

[4] Jonathan T. Foote. et al., “Content-Based Retrieval of Music and Audio,” *Multimedia Storage and Archiving Systems II, Proc. of SPIE, Vol. 3229*, pp. 138-147, 1997.

[5] Fukunaga, K. “Introduction to Statistical Pattern Recognition,” 1972.

[6] 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.

[7] MPEG Requirements Group, “Information technology - Multimedia Content Description Interface - Part 4 : Audio” *ISO/IEC CD 15938-4*, Oct. 2000.

[8] Pfeiffer, S., S. Fischer, and W. Effelsberg, “Automatic Audio content Analysis,” *ACM Multimedia 96*, Boston MA USA.

[9] Saunders, J., “Real-time discrimination of broadcast speech/music,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96)*, vol. 11, pp. 993-996, 1996.

[10] Scheirer, E., and Slaney, M. “Construction and evaluation of a robust multifeature speech/music discriminator,” *IEEE Transactions on Acoustics, Speech, and Signal Processing (ICASSP-97)*, pp. 1331-1334, 1997.

[11] 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

[12] Tzanetakis, G., and P. Cook, “A Framework for Audio Analysis Based on Classification and Temporal Segmentation,” *In Proc. EUROMICRO Conf.*, vol. 2, pp. 61-67, 1999.

[13] Tzanetakis, G., and Cook, P. “Multifeature audio segmentation for browsing and annotation,” *In Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPA99*, New Paltz, NY, 1999.

[14] Tzanetakis, G., and Cook, P. “Sound analysis using MPEG-compressed audio,” *In Proc. Int. Conf on Audio, Speech and Signal Processing, ICASSP*, 2000.

[15] Wold, E. et al., “Content-based classification, search, and retrieval of audio,” *IEEE Multimedia*, vol. 3, no. 2, pp. 27-36, 1996.