

國立臺灣師範大學
資訊工程研究所碩士論文

指導教授：方瓊瑤 博士

以觀眾為拍攝主體之虛擬攝影師系統

An Automatic Virtual Cameraman System for Audience

研究生：蔡侑廷 撰

中華民國一百零二年六月

摘要

本研究的目的乃在於建立以觀眾為拍攝主體之虛擬攝影師系統。由於人力成本的不斷提高，聘請商業攝影團隊為演講或會議拍攝影片對一般預算有限的非營利學校或公司行號是一個沉重的負擔。若為了節省成本而讓缺乏攝影經驗的人來掌鏡，往往會讓影片缺乏美感及流暢度，甚至是降低觀賞的意願。本研究所提出之系統除了能夠節省人力開銷，還提供專業攝影技巧以製作高規格影片。

本研究以兩台 PTZ 攝影機作為實驗設備，架設於拍攝場地的左或右前方，兩台 PTZ 攝影機具有不同的功能，一台為全景 PTZ 攝影機，另一台為運鏡 PTZ 攝影機。全景 PTZ 攝影機代表攝影師的雙眼，將拍攝場地的全景的連續影像輸入到系統協助系統的全景監控與主體偵測。運鏡 PTZ 攝影機則是攝影師手上的攝影機，在系統決定運鏡所需之一切資訊後，運鏡 PTZ 攝影機就會實際執行運鏡動作。

本系統的功能為模仿攝影師的拍攝技巧並自動進行運鏡動作，每次運鏡皆需要運鏡的種類、景別、主體等要素。系統從全景 PTZ 攝影機輸入連續影像，從中擷取四種具描述觀眾行為的 motion 特徵。接著將這些特徵經過模糊化(fuzzifier)處理讓這些數值轉換成自然語言的表達方式，以便於分析攝影師的運鏡習慣。接著將這些特徵輸入自動運鏡模型(automatic camera movement model)，該模型能夠紀錄專業攝影師的運鏡習慣，並依據輸入的特徵輸出適合該情況的運鏡以及景別種類。拍攝主體的挑選則以五種分別代表不同美感的特徵作為判斷標準。在完成運鏡要素的計算後，將資料傳送至運鏡 PTZ 攝影機執行運鏡動作。

實驗結果顯示，本系統可進行即時且流暢的運鏡動作且具專業攝影的運鏡水準，符合演講錄製的需求。

關鍵字： 攝影師、運鏡、美學、自動機

Abstract

This study proposed an automatic virtual cameraman system for audience. Because of increasing of personnel costs, to employ commercial photography team filming speeches or conferences is a heavy burden for the nonprofit school or company line number that budget limited. If In order to save costs and let person who lack of photography to shoot, it often causes the videos which not smooth and aesthetics. Our system not only can save personnel costs, but makes professional videos with photographic.

System equipment includes two PTZ web-cameras which are located right front or left front of the lecture theatre. These two PTZ web-cameras have different functions. One, which named full-shot PTZ camera, grabs auditorium screen and input continuous images to system. Another one , which named camera movement PTZ camera, do camera movement after system computes all necessary camera movement information.

Camera movement information includes camera movement class, shot class, and subject. To get this information, system first input continuous images from full-shot PTZ camera. Then extracting four motion features which represent four kinds of audiences' behavior. For combining motion features and natural language, system fuzzifier motion features. To decision camera movement and shot, we construct automatic camera movement model (ACMM), an automata model. ACMM records photographers' habit of camera movement and shot. It can pick up suitable camera movement and shot by input fuzzy motion features. After that, system will choose

subject by using four aesthetic features, which are continuous, repeated, luminance and composition. Last, system operates camera movement PTZ camera to finish recording.

In the experiment, there are eleven kinds of camera movement and six kinds of shot in the ACMM. And ACMM was trained with six lecture videos. Compared to amateur's recording and point shooting, our system is really performance well than them. Beside, system can real time work, so it won't lose any shot on the auditorium.

Index item: cameraman, camera movement, aesthetics, automata.



誌謝

兩年的碩士生活結束了，雖然時間不是很長卻讓我有不小的成長，和大學時期不同，碩士更需要獨立思考以及表達的能力，這是方老師一再向我強調的。不論是哪個領域、哪種研究主題都不見得都有前人鋪路。因此唯有不斷的思考和吸收新知才能突破以往的思維，發現新的可能，即使結果不如想像中美好，但歷經思考的過程，懂得如何獨立解決問題正是我所學習到的寶貴經驗。

在這兩年的求學之路上我受到許多朋友及老師的幫助，讓我能夠一步步走到這裡，讓我能夠抬頭挺胸的走向人生的下一個路程。感謝方瓊瑤老師在這兩年對我的教導，即使我停留在同一個瓶頸，老師每次都能給我很多想法和建議，沒有老師積極的催促和監督，我的進度很難維持在跑道上。感謝陳世旺老師的高等影像和電腦視覺課程讓我更了解影像處理的技術和理論，同時要感謝梁祐名老師在百忙之中抽空擔任我們的口試委員。另外要感謝系辦每位成員的照顧，還有阿谷的幫忙讓我的實驗能夠順利。

一個契機讓我踏入原本應該額滿的 CVIU LAB，從看著俊麟大師兄和派大星的水深火熱到和欣紘以及瑋鴻一同努力奮戰，都讓我感到慶幸能夠來到這裡。感謝 CVIU 的各位夥伴們這兩年來的共同奮鬥。感謝派大星帶給 LAB 歡樂，感謝俊麟替大家解惑以及好歌喉。感謝欣紘和瑋鴻在這段時間共同的努力和扶持，即便大家最後都一副要崩潰的樣子，但是我們終究挺過來了！感謝巧珊、懿禎和良謙的大力幫忙和支持，沒有你們分享和維持 LAB 的能量可就一切免談了。另外要感謝米奇、胡禎和其他 LAB 的常客們以及 IPCV LAB 的各位與我們互相照顧。

最後，感謝我的家人在在這段時間的鼓勵和陪伴，如今我順利抵達畢業的大門，希望往後每位師長、每位夥伴、每位朋友都能夠繼續互相扶持，在接下來的人生都能一路順遂，謝謝大家！

目錄

摘要.....	I
Abstract.....	II
誌謝.....	IV
目錄.....	V
圖目錄.....	VII
表目錄.....	IX
第一章 緒論.....	1
第一節 研究背景與目的.....	1
第二節 研究困難.....	5
第三節 論文架構.....	7
第二章 文獻探討.....	8
第一節 攝影與錄影技術.....	8
第二節 虛擬攝影師系統之相關研究.....	17
第三章 系統概述.....	21
第一節 架設環境.....	21
第二節 系統流程.....	23
第四章 資料前處理與特徵擷取.....	30
第一節 影像前處理.....	30
第二節 特徵擷取.....	32
第三節 特徵模糊化.....	35
第五章 運鏡選取.....	38
第一節 運鏡切換條件.....	38
第二節 自動運鏡模型(automatic camera movement model, ACMM).....	41
第三節 景別決定(shot decision).....	49
第六章 主體選取與運鏡路線規劃.....	51
第一節 主體候選者產生.....	51
第二節 主體挑選.....	53
第三節 運鏡路線規劃.....	59
第七章 實驗結果與討論.....	64
第一節 系統設定.....	64
第二節 系統運作範例.....	71
第三節 自動拍攝結果與討論.....	72
第八章 結論與未來工作.....	78
第一節 結論.....	78

第二節 未來工作.....	79
參考文獻.....	82



圖目錄

圖 1.1 自動化演講錄製系統架構.....	3
圖 1.2 專業與業餘的攝影實例.....	4
圖 1.3 不同的室內演講場地.....	7
圖 2.1 三種基礎構圖法.....	9
圖 2.2 不同光線運用的攝影範例.....	9
圖 2.3 顏色的對比與色系.....	10
圖 2.4 不同景深下所拍攝的影像，表現出不同的畫面主體.....	11
圖 2.5 各種景別範圍示意圖.....	13
圖 2.6 大遠景與遠景的拍攝實例.....	14
圖 2.7 全景與中景的拍攝實例.....	15
圖 2.8 特寫與大特寫的拍攝實例.....	15
圖 3.1 PTZ 攝影機.....	21
圖 3.2 PTZ 攝影機架設於水平地面演講廳示意圖.....	22
圖 3.3 PTZ 攝影機架設角度示意圖.....	23
圖 3.4 全景 PTZ 攝影機實際拍攝影像.....	23
圖 3.5 系統流程圖.....	24
圖 3.6 Motion feature extraction 流程圖.....	25
圖 3.7 Camera movement decision 流程圖.....	26
圖 3.8 Path planning 流程圖.....	28
圖 4.1 語言變數之類屬函數示意圖.....	37
圖 5.1 FSM 結構圖.....	41
圖 5.2 Moore 和 Mealy 模型結構圖.....	42
圖 5.3 自動運鏡系統結構圖.....	43
圖 5.4 Training data 分別轉換成各 TPM 之範例.....	45
圖 5.5 ACMM 運作流程圖.....	48
圖 6.1 Integral image 與 Harr feature 示意圖.....	52
圖 6.2 Cascade classifier 概念圖.....	53
圖 6.3 井字構圖法與井字構圖遮罩.....	54
圖 6.4 畫面中直線與井字遮罩標準線距離示意圖.....	55
圖 6.5 HSV 色彩模型.....	57
圖 6.6 重複性評比範例.....	58
圖 6.7 第一類運鏡流程示意圖.....	59
圖 6.8 第二類運鏡流程示意圖.....	60
圖 6.9 沿觀眾席方向示意圖.....	61

圖 6.10 景別特徵.....	62
圖 7.1 經過訓練的 ACMM 模型.....	67
圖 7.2 CMS 轉移示意圖.....	72
圖 7.3 實驗結果 1(運鏡 1~4).....	73
圖 7.4 實驗結果 1(運鏡 5~9).....	74
圖 7.5 實驗結果 1(運鏡 10~14).....	75
圖 7.6 實驗結果 2(運鏡 1~5).....	76
圖 7.7 實驗結果 2(運鏡 6~9).....	77



表目錄

表 1.1 工商錄影價目表.....	2
表 2.1 各種景別的介紹.....	12
表 4.1 Fuzzy motion feature 與語意變數對照表	36
表 5.1 Fuzzy motion feature 與其語意值對照表	38
表 5.2 攝影機運動的數據資料.....	40
表 5.3 運鏡持續時間與運鏡種類關係.....	41
表 6.1 Value 值與亮度評比對照表.....	56
表 7.1 Training set 介紹.....	64
表 7.2 TPM 元素值列表	67
表 7.3 TPM 權重值列表	69
表 7.4 SPM 元素值列表.....	69
表 7.5 T_{0j} 元素值列表.....	71
表 7.6 T_{0j} 權重值表.....	71
表 7.7 S_{02} 元素值表.....	72

第一章 緒論

第一節 研究背景與目的

本研究的主要目標乃在於建立一套具專業水準的自動化演講錄製系統。在網路寬頻快速提升以及儲存設備價格日漸下降的影響之下，各種媒體的傳輸及保存已經不是資訊科技發展的瓶頸，反倒是數位內容的缺乏才是目前所面臨的問題。

優質數位內容的製作通常需要高額的預算，以演講錄製為例，若是要獲得專業的錄製影片，現今的做法是雇用專業的攝影團隊錄製演講過程，再經由選鏡、剪輯等後製處理，以得到高品質的影片。攝影團隊拍攝的過程分成三個主要的部分：(1)事前場勘作業，(2)現場拍攝，以及(3)事後剪輯後製。支出成本則分為：(1)以購置器材為主的固定成本以及(2)以聘請人員為主的人事成本。通常固定成本在器材設置妥當之後就僅剩定期維修的費用，但攝影師則是在每次錄影都要雇用，因此人事成本會不斷的重複支出。對於預算有限的公司行號或非營利的學校教學機構來說，這樣長期的成本支出將是難以承受的負擔。

表1.1為一般攝錄影公司所提供的錄影服務價目表，由於每次進行實況錄影就有一定的成本在，所以錄影時間越少反而每單位時間的價格越昂貴。而且大多數公司都是以單一攝影機作業作為前提進行報價，若是想要用兩台以上的攝影機來拍攝，那報價可能超過倍以上。若為了節省成本，選擇單一固定攝影機針對演講者及其演講內容拍攝錄製成影片，內容勢必較為枯燥、單調，對觀看者的觀看意願以及學習成效可能有負面的影響。為了提供專業的錄影品質並節省人事成本，開發並建立一套具專業水準的自動化演講錄製系統是必要的。

表1.1 工商錄影價目表

公司	項目	價格
DVworld專業錄影製作	活動實況錄影	12000/4 hrs 15600/8hrs
OIS 原創映像多媒體工作室	活動實況錄影	6000+/3hrs
絕攝錄影工坊	三小時會議錄影	8000/3hrs
采攝錄影工坊	三小時會議錄影	8000/3hrs
喬崑攝影中心	畢業典禮錄影	12000/6hrs
ALIC專業攝影工坊	工商活動攝錄影	9000/8hrs(單機) 16000/8hr(雙機)
憶彩錄影工作室	三小時會議錄影	8000/hrs
恆彩錄影工坊	工商活動攝錄影	9000/8hrs(單機) 16000/8hr(雙機)
數位錄影公司	工商錄影	4000/2hrs
南廣傳播事業有限公司	婚禮錄影	12000全天

要開發自動化演講錄製系統以達到專業的攝影水準，必須先了解攝影團隊的整個拍攝的過程。首先，拍攝團隊會在演講開始之前勘察場地，並決定適用攝影器材及器材的擺設位置。為了讓拍攝的角度和畫面的內容更多元活潑，通常會架設多台攝影機，同時拍攝演講者、觀眾或是整個會場。在進行拍攝時，每個攝影機都由一個攝影師掌鏡，攝影師除了要確實掌握拍攝的目標，同時拍攝的畫面必須兼具攝影的美感。拍攝完畢後，所有的影片將經由攝影團隊進行影片剪接、配音、字幕搭載、畫面特效等處理，所有步驟完成後才算是一次完整的拍攝工作。

自動化演講錄製系統的架構如圖1.1，包含四個子系統，分別稱為虛擬導播子系統，虛擬攝影師子系統，演說文譯子系統，以及手動操控子系統。虛擬攝影師子系統控制著多台攝影機，分別用來拍攝演講者、聽眾、以及全場，因此虛擬攝影師子系統又可分成演講者攝影師次子系統，聽眾攝影師次子系統，以及全場攝影師次子系統，在自動化的架構下所有這些次子系統均可自行運作，並將拍得的視訊直接傳給虛擬導播子系統。虛擬導播子系統便從虛擬攝影師子系統傳來的視訊中選取最佳者，配合演講者的聲音，演講內容(如Power Point 檔和展示用

videos)，以及由演說文譯子系統所提供的字幕，結合成畫面一起輸出。總結虛擬導播子系統的工作將包括有：視訊剪輯，配音，搭載演說字幕及演講內容成一畫面輸出。

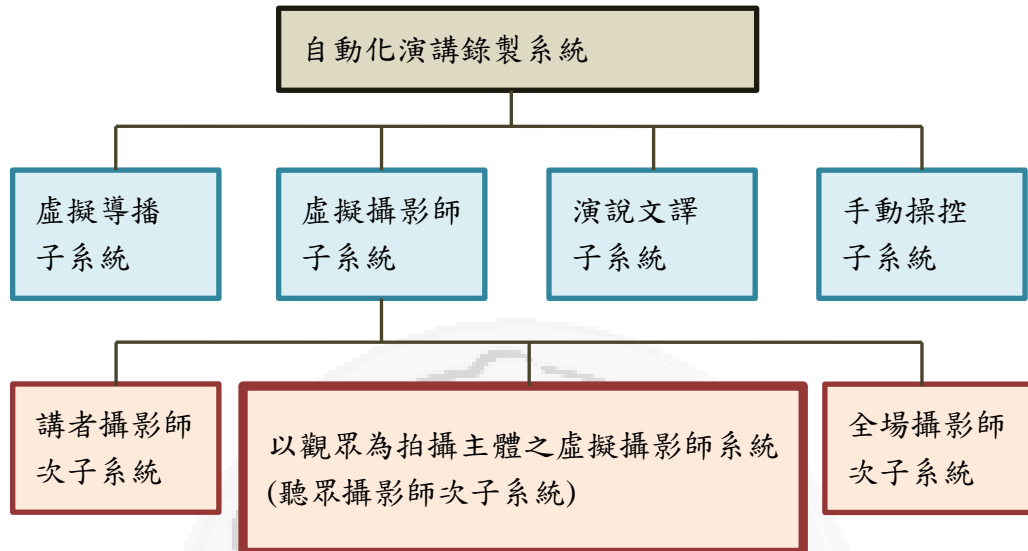


圖 1.1 自動化演講錄製系統架構

本研究的目的為自動化演講錄製系統中之虛擬攝影師子系統--聽眾自動錄影系統的研發，內容為開發以觀賞演講的觀眾作為主體的錄影系統。其功能為系統能自動選取拍攝目標，並操作攝影機完成錄影程序。然而專業的錄影工作不只需要挑選適合的拍攝目標，採取合適的「運鏡」更是決定影片好壞的關鍵，並同時以美學的觀點選取適合的主體位置。而且本系統的目標不局限於演講的聽眾，而是希望能夠透過輸入不同的場地參數的方式讓以適應各種場地，例如球賽、音樂會、戲劇等不同場合都能夠提供正確地而專業的錄影效果。因此本研究以自動運鏡、目標挑選、美學專業為主軸，將聽眾攝影師次子系統重新命名為「以觀眾為拍攝主體之虛擬攝影師系統」。

以觀眾為拍攝主體之虛擬攝影師系統的輸入資料為攝影機所拍攝的畫面，在以往節省成本的思維下大多只使用一台攝影機。然而單一攝影機所能捕捉到的鏡頭非常有限，大多數的鏡頭都集中在演講者身上，而觀眾席上的活動往往都會被

忽視。由觀察網路上的演講影片可發現觀眾在其中出現的時間通常十分短暫。然而觀眾不只是配角，更是整個活動的一部分。觀眾的各種反應能表現當下活動進行的情況，是帶動活動氣氛的重要指標。在賽況激烈時，觀眾席上群情激昂的畫面能夠帶起激情；在表演的高潮，觀眾鼓掌的畫面能夠表達感動。藉由穿插現場觀眾的畫面，能增加轉播畫面的多樣性與可看性，還可以激發影片觀看者的臨場感與參與感，是在各類轉播中不可或缺的畫面。



圖1.2 專業與業餘的攝影實例[Niu12] (a)(b)專業的攝影實例 (c)(d)缺乏專業攝影技巧的攝影實例

因此，雖然觀眾所占的拍攝時間或許不長，但是專業且具規畫的運鏡能使整部影片流暢且不枯燥，然而大多數的非專業攝影師卻無法達到這個標準。舉例而言，從觀察網路上的演講影片可以發現，當鏡頭轉到觀眾席時最常用的手法是以定點攝影的方式，特寫觀眾席的一部分。依照專業攝影的觀點來看，定點攝影屬於偏靜態的運鏡方式，在沒有其他運鏡手法輔助下，過多靜止的鏡頭會使得原本就以靜態為主的演講更為僵硬，甚至會降低觀賞者觀看的意願。Niu 等人在2012年提出影片在專業程度上的評比方法[Niu12]，圖1.1為專業和非專業的攝影實例，除了主要目標的強調、顏色的控制之外，還提到停留在同一個運鏡的攝影時間也是影

片專業度的重要指標，因為這代表攝影師無法在有限的空間進行變化，而且缺乏詳細的運鏡路線規劃。另外，這些定點拍攝的鏡頭通常不具有主題性，違反了此類運鏡手法的使用原則，有時連觀眾起身離開這種不該入境的情況也出現在影片之中。正確的運鏡安排，應該搭配動態的運鏡手法，像是用搖攝的方式呈現整個場面，或是搭配推移或是拉遠鏡頭營造靜態畫面所缺乏的動感。

總之，攝影觀念或是手法上的差異足以讓影片的評價呈現兩極。一個專業的攝影師必須能夠藉由專業的攝影觀念，因地制宜選用最適合的運鏡方式讓觀賞者有足夠的興趣繼續觀賞。而目前的演講拍攝影片大都缺乏專業的攝影技術，尤其在觀眾畫面的拍攝上。本研究的目標即為模擬專業的攝影師的拍攝方式，提供使用者專業的攝影技術並自動化完成攝影工作。不但能夠將數位內容以最佳的品質保存，同時節省長期拍攝團隊錄製影片的大量成本。另外，本系統為了增加泛用性和實用性，系統將會依攝影師的拍攝習慣來建立各種不同的拍攝場合的運鏡習慣模組，其目的是為了讓虛擬攝影師系統能配合不同的場合都能提供完整的功能。

第二節 研究困難

為了順利捕捉聽眾席上各式各樣的精彩畫面，聽眾及全場自動錄影系統開發時需要考慮與克服的環境因素眾多，包括攝影主體如何決定的問題、聽眾席座位設計的多樣性問題、聽眾席座位排列的坡度問題、聽眾入座和離開時的干擾問題、以及配合不同環境場合的運鏡習慣等。以下分別說明之：

1. 攝影主體如何決定的問題：由於本系統是以觀眾為主體，觀眾席上的每個觀眾都能夠當作拍攝目標。亦或是可能有比起觀眾更具有吸引力的目標，像是有強烈 motion 的區域、或是主體不局限於單一個體之上。該如何偵測，以及如何從中挑選出合適的主體來拍攝將是一大難題。
2. 觀眾席座位設計的多樣性問題：聽眾座位依場地的不同有各式各樣的規格(如圖1.2(a)(b)所示)，有些較寬有些較窄，亦有不同的顏色，這些不確定因素的影響在系統開發時均應納入考慮。例如聽眾臉部的偵測技術能否處理座位的顏色貼近膚色的問題。另外，若拍攝背景會隨著時間而改變，那聽眾整體

或個別反應的偵測技術能否過濾這類的背景變化。

3. 聽眾席座位排列的坡度問題：一般聽眾席的分為平面和階梯式二種，且階梯式聽眾席的坡度並不統一(如圖1.2(c)(d)所示)。聽眾席座位排列的坡度會影響到攝影機拍攝到的聽眾角度，例如聽眾臉部與身上的特徵會因為攝影角度而改變。另外，在聽眾席坡度比較緩和的情況下後排聽眾容易被前排聽眾遮蔽而影響到拍攝的畫面。
4. 聽眾入座和離開時的干擾問題：在典禮或是演講進行的時候，常會有聽眾起身離開座位或是中途入座的情況。如此現象導致畫面中常有物體在移動、進入、離開，系統必須自動判斷當下情境的畫面是否必要擷取並輸出。例如若是頒獎典禮時領獎人自聽眾席起身上台領獎，則系統可能需全程追蹤並擷取該領獎人的相關鏡頭。但若只是一般聽眾起身離開座位，則系統最好忽略此類的動作。
5. 配合不同環境場合的運鏡習慣：從觀看電視轉播的經驗中可知不同的性質的節目或是運動賽事中運鏡的方式差異頗大。例如在轉播演講的場合，講者的畫面通常佔大部份的比例，而聽眾的畫面則會分段地穿插在其中，且該類畫面常是聽眾做筆記或是專注於聽講的畫面。通常在聽眾提問時才會包括較多的聽眾特寫畫面。而若是運動類的轉播，像是網球或棒球比賽，運鏡的方式比較有規律，觀眾的畫面除了遠景外，一般會在暫停時間對選手的教練或親人有較多的特寫畫面。因此，自動的配合不同場合進行不同的運鏡習慣是本系統開發時考慮的因素之一。



(a)



(b)



(c)



(d)

圖1.3 不同的室內演講場地 (a)(b)兩種不同造型的觀眾席座椅[20][21] (c)(d)兩種不同坡度的觀眾席[22][23]

第三節 論文架構

本論文的第二章將會介紹專業攝影師所需要具備的專業攝影技巧；第三章會說明系統的架設環境以及系統的流程；第四章介紹如何從輸入影像擷取系統所需之特徵；第五章為系統的運鏡選擇，此章節會詳述如何模仿專業攝影師對運鏡使用的判斷；第六章為拍攝主體的選取，以及各種運鏡的自動路線規劃；第七章為實驗結果，將以實際攝影器材測試在活動現場進行拍攝，測試是否合乎專業水準；最後第八章為結論與未來工作。

第二章 文獻探討

第一節 攝影與錄影技術

想要模仿專業的攝影師的手法拍出優質的作品，必須先了解一個專業的攝影師所應該具備的能力。一位專業的攝影師需要熟稔各種攝影技巧、運鏡手法並具敏銳的美感，才能捕捉到適合拍攝的畫面。

攝影技巧包含構圖、光影以及色彩三個組構成畫面的主要元素。構圖指的是畫面中空間的配置，被拍攝的主體以何種比例、何種位置安排在畫面裡是奠定影片可看性的基礎。適當的構圖能夠突顯主體的存在而不至於讓畫面顯得擁擠或繁亂。一般較常使用的構圖方法有「三分構圖法」、「井字構圖法」以及「對角線構圖法」[1][2]，以「三分構圖法」(圖2.1(a))為例，把欲拍攝的畫面以垂直線或是水平線分成三等份，並保持主體分別存在於被均分的三個區域中。在拍攝風景時，等分線的位置通常是和水平線重合，或是用來區分遠、中、近景。另外「井字構圖法」(圖2.1(b))又被稱為「黃金分割法」是最常被使用、也是入門攝影必須先熟悉的構圖方法。該方法是將畫面以九宮形式切割，並讓主體維持在線上或是交點上。「對角線構圖法」(圖2.1(c))可增加畫面的幾何張力，在處理簡約場景時，可嘗試尋找明顯或隱形的對角線，無論存在一條或兩條對角線，都能讓畫面更戲劇化。其他另有「放射線構圖法」、「延伸性構圖法」[4]等等諸多構圖方法，要靈活運用這些構圖方法，需要累積相當的攝影經驗，也是本研究所要解決的課題之一。

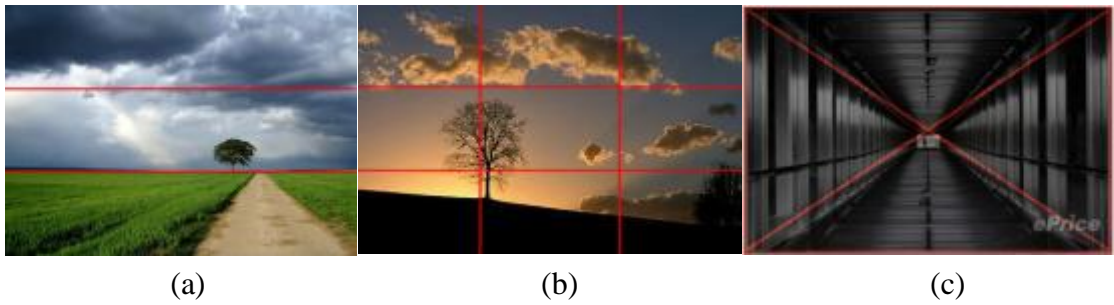


圖2.1 三種基礎構圖法[4] (a)三分構圖法 (b) 井字構圖法(黃金分割法) (c)對角線構圖法

在光線的運用上，不同的光線能夠幫助攝影者讓影像帶出不同的感覺，例如攝影師希望景物清晰呈現時，可使用「正光」(順光)。正光即為讓光線從鏡頭對主體的方向發出，能夠讓主體清晰易見並讓色彩鮮麗而且飽和(如圖2.2(a))，但是由於主體上沒有陰影，較欠缺層次及立體感。想帶出層次感的話，從主體側邊打上的「側光」能令主體更顯凸出，展現景物細節質感(如圖2.2(b))，拍攝人像經常用到側光拍攝。若想表現具戲劇感的氣氛，可將主體放於「背光」(逆光)的位置。若拍攝人像時配合閃光燈補光，則更能加強氣氛，獲得具戲劇性的效果(如圖2.2(c))。但用閃光燈補光的方式易失去前景或背景的細節，而且容易造成背景過度曝光造成畫面模糊。靈活運用不同的光線除了能夠獲得理想的攝影效果，還能營造各種截然不同的意境和氣氛。礙於硬體上的限制，本研究並沒有將操控燈光器具納入研究範圍。但是藉由攝影機獲得的影像資訊，鏡頭能夠從全景畫面中自動判斷並挑選光影效果較佳或是符合現在情況氣氛的局部畫面加以特寫，同時避開不符專業標準的區域。

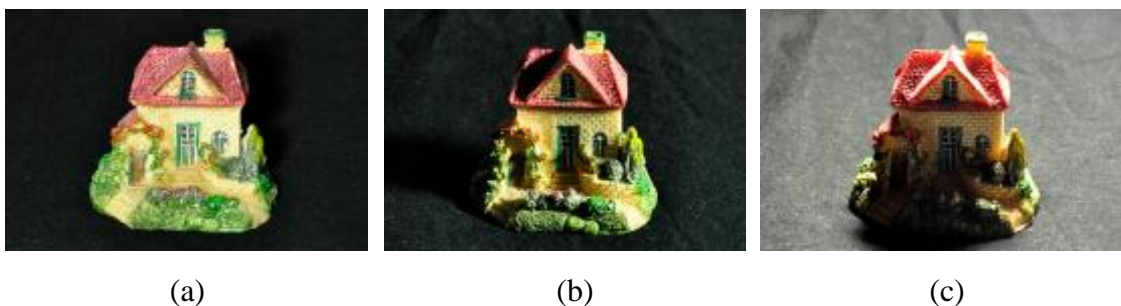


圖 2.2 不同光線運用的攝影範例[5] (a)順光 (b)側光 (c)逆光

在色彩方面，簡單的色調和鮮明的對比能襯托出前景的存在感(如圖2.3(a))。此外，畫面氛圍會隨著的畫面整體顏色的不同而有所改變，暖色系的顏色會讓人覺得放鬆、溫暖(如圖2.3(b))，冷色系則帶給觀眾理性、清楚的感受(如圖2.3(c))。由於在演講場地的顏色並不容易隨時改變，而且色彩的調整通常在影片的後製階段處理。因此本研究從影像的資訊判別每次取景的畫面顏色是否符合標準，若是背景混雜或找不到適合的畫面，可適當地把焦距專注在前景，讓背景模糊以降低其對前景的影響。



圖2.3 顏色的對比與色系 (a)以鮮明的對比突顯主體[9] (b)(c)不同的整體色調帶來不同的視覺效果[10]

除了以上三個畫面的構成元素之外，「景深」的調整對攝影而言也是一門重要的學問。景深就是當對焦完成後，景物在底片上呈現完全清楚的距離範圍，也就是「焦點前後清楚的範圍」。因此景深是「一段距離」，在這段距離內的物體都應該是清楚的。不同的景深下所拍攝的影像會表現出不同的畫面主體，如圖2.4所示。圖2.4 (a)中使用較深的景深，因此並無法明確定義主體；而圖2.4(b)的景深較圖2.4(a)為淺，較遠的景物變得模糊，主體逐漸浮現；(c)則是三張圖片裡使用景深最淺者，唯有較靠近焦距的景物有清楚的輪廓，主體的位置十分明顯。影響景深的要素有三，光圈的大小、鏡頭焦距的長短、攝影距離的遠近。用大光圈、長焦距、近距離拍攝的組合可以霧化背景、保留清晰主體來突顯主題。攝影時景深的遠近並沒有一定的標準，完全因應攝影師的需求來做調整。但通常風景照會使用深景深(因為想整個風景變得清楚銳利)，而人像照便會使用淺景深(因為想人物在

相中變得突出)，不過並不代表使用其他景深的方法是錯誤的，完全要視攝影需求而定。



圖2.4 不同景深下所拍攝的影像，表現出不同的畫面主體[8] (a)光圈 $f/8$ (深景深) (b)光圈 $f/5.6$ (中景深) (c)光圈 $f/2.8$ (淺景深)

此外，攝影機自動拍攝時取景的角度與運鏡的方式也是本研究一個重要的課題。攝影機取景的角度可表達明確的主題或目標物，並交代目標物的動作與表情，呈現有秩序的美感，且可使畫面產生立體結構[6]。以人眼的視平角為水平的標準，攝影機的視角比視平角高稱為高角或仰角，反之為低角或俯角。攝影機視角的改變，會使影像的構圖傳達不同感覺。例如攝影機的視角與人眼的視平角齊高時(水平鏡頭)，是最常被運用且為表現現實生活的最佳角度；攝影機鏡頭由低處往上拍攝(仰角鏡頭)則會使主題或目標物明顯加大，呈現威嚴感令人敬畏，可營造誇張戲劇效果或恐怖氣氛，突顯衝擊力；攝影機鏡頭由高處往下拍攝(俯角鏡頭)，則會使主題或目標物變小，可表現出渺小、哀傷、孤寂與無助感；若攝影機鏡頭由主題或目標物側邊拍攝(側角鏡頭)，則可增加立體感，保持主體與背景和環境間適當的距離，增加構圖的深度；最後，若攝影機鏡頭傾斜拍攝(傾斜鏡頭)，會讓主體或目標物產生失去平衡的感覺，可表達主觀的不安、打鬥與災難的現場氣氛。

主體在畫面中所佔的比例以及呈現的景觀範圍是影響整體平衡的重要因素，這項

畫面的結構要素在攝影技術當中被稱作「景別」。景別的影響要素有二：一為攝影機和景物的距離；二為攝影機鏡頭的焦距。無論是主體和攝影機距離的變化，或是攝影機在固定位置調整鏡頭焦距都能引起主體在畫面上的大小變化。這種畫面上景物大小的變化所構成的不同取景範圍即為景別的變化，而掌握景別的變動是一種重要的表現手段。不同的景別，表現的是不同的視野、不同的空間範圍以及視覺節奏。藉由各種畫面景別的單獨展現、前後排列、段落組合等鮮活運用，將能表現出截然不同的情境和視覺效果。景別依照主體在畫面中的比例可以分為七類，由小而大為大遠景、遠景、中景、近景、特寫、大特寫，表2.1簡單的定義了這七種景別，其詳細定義與適用場合將在以下段落詳述。圖2.5則說明了以人為主體的情況下，主體和畫面的比例關係。

表 2.1 各種景別的介绍[12]

景別	中文名稱	說明
Extra Long Shot	大遠景	指整個大環境，此時主體的大小已無關緊要
Long Shot	遠景	景物主體約占鏡頭的 3/4 到 1/3，其範圍定義較為模糊
Medium Long Shot	全景	景物全身外加上下空間
Medium Shot	中景	景物主體的 1/2 左右
Medium Close-Up	近景	景物主體的 1/4 左右
Close-Up	特寫	景物主體的局部細節
Extra Close-Up	大特寫	景物主體的細部放大

- (1) 大遠景：大遠景是視距最遠、展現空間最遼闊的景別(圖2.6(a))。其目的在於強調多個主體間以及主體和環境之間的關係，並呈現整個背景訊息，同時帶來開闊和整體感。主體在大遠景中所佔的比例極小，不容易描述事件的過程。大遠景大多是肉眼難以觀察到的，因而具有非實體的感受，適合用來表達情緒或意境，通常使用於影片的開頭或結尾，以說明完整的場景內容。

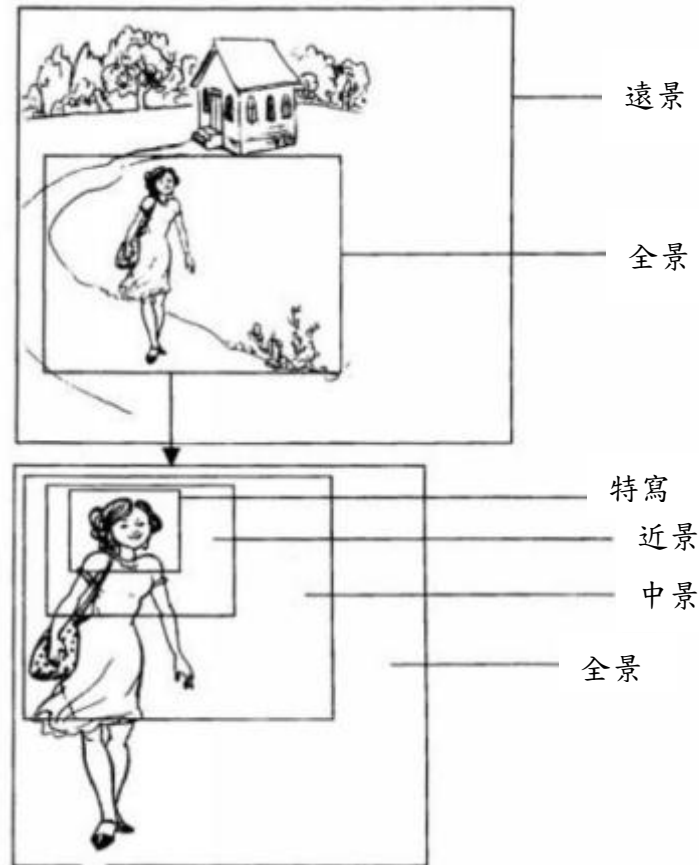


圖 2.5 各種景別範圍示意圖[12]

- (2) 遠景：遠景主要用於表達主體的動向以及和環境間的空間關係。遠景中的主體較大遠景顯著，雖然畫面整體上看來背景的訊息仍佔多數，但多個主體間和主體與環境間的互動更為明顯。比起大遠景所象徵的情境，遠景將連結主體和環境的狀況和氣氛清楚的展現出來(圖2.6(b))。而大遠景和遠景這兩種景別，都常被用於定場鏡頭(establishing shot)的拍攝[註1]，以說明主體間的關係，例如互相的方向與動向。
- (3) 全景：全景一般用在表現人物全身形象或是特定場景全貌的畫面。全景在涵蓋主體全貌的同時保留一定範圍的環境或主體的活動空間，藉此能夠完整地表現人物的形體動作，並透過對人物或人物間形體動作的描述刻劃內心情感。相較於遠景所帶有的氣氛和大遠景的意象，全景著重於主體的特殊結構與其意涵，並藉由特定環境和特定場景表現或是解釋特定人物及場景。例如圖

2.7(a)所呈現的是兩個足球員之前的進攻與防守，藉著觀察畫面中的兩個主體的動作和相互位置，讓觀賞者了解此一畫面的意涵。

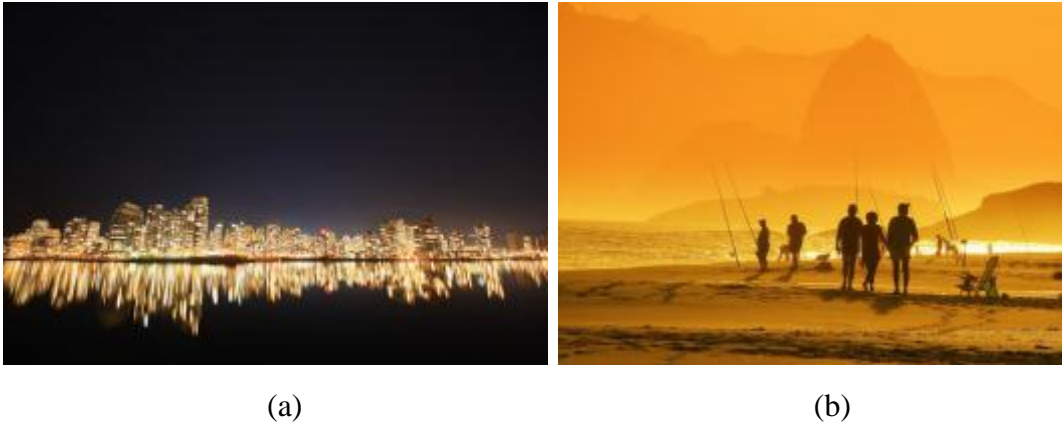


圖 2.6 大遠景與遠景的拍攝實例[14] (a)大遠景 (b)遠景

- (4) 中景：中景多是指人物膝蓋以上或是場景局部的畫面，是人眼所熟悉的畫面裁切方式，就好似乎時所注意到的客觀世界一般。中景能夠忠實呈現人物的線條和細部動作，能夠具體描述人物的神態和姿勢，藉此更進一步透露出心理的狀態。從圖2.7(b)中球員的動作和表情即清楚的展露出其生動的欣喜之情。此外，比起遠景的飄渺不定，中景帶有指向性，意指讓觀眾聚焦於目標之上，指示觀眾畫面的重點。因此中景非常適合用來表述敘事的內容，像是交代故事情節或是描寫事物之間的關係，環境和人物的整體形象則降於次要角色。
- (5) 特寫：特寫是指只拍攝某一景物來對該景物做特別描寫的一種構圖方式，雖然保有主體旁且與主體有關者之背景，但影像中與環境相關的訊息並不多。特寫可對主體的特徵進行充分的描寫，並能產生強烈的視覺印象。對人物的刻畫來說，就如同為了描述人物的表情而產生的景別。使用特寫常會利用縮短景深來模糊背景，藉此突顯主體的存在，如圖2.8(a)所示。

[註1] 定場鏡頭是影片一開始，或一場戲的開頭，用來明確交代地點的鏡頭，通常是一種視野寬闊的遠景。

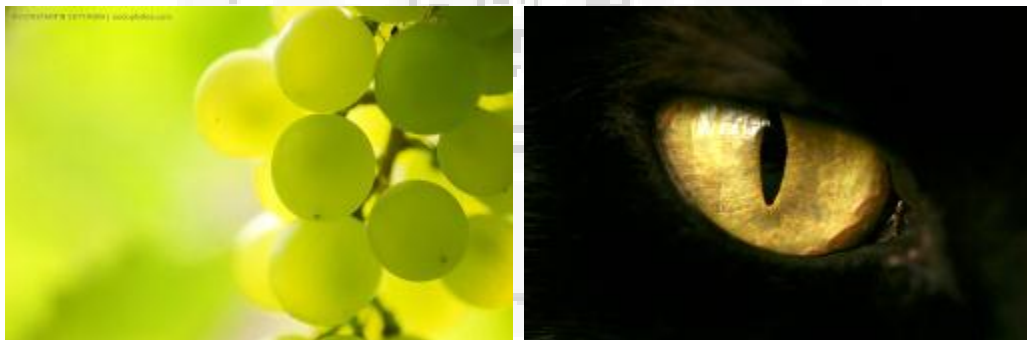


(a)

(b)

圖 2.7 全景與中景的拍攝實例[15] (a)全景 (b)中景

- (6) 大特寫：大特寫又稱作細部特寫，是為單純對景物的某一細部做誇張描寫的構圖方式，影像中並不保留主體周圍的背景。一般而言，大特寫更清楚的描寫景物的特徵或是細節，由於幾乎沒有什麼背景的消息，大特寫本身必須帶有特別強烈的意義才足以推動敘事的情節。通常，大特寫的表演方式帶有誇大情節的效果，圖2.8(b)為貓眼大特寫，帶有被注視、緊盯的強烈意圖。



(a)

(b)

圖 2.8 特寫與大特寫的拍攝實例[14] (a)特寫 (b)大特寫

專業的運鏡處理是提升影片品質另一個重要的因素。運鏡與否可謂靜態和動態攝影最大的不同之處，靜態攝影不須運鏡，其重點在於化瞬間為永恆，能捕捉的唯有一刻的畫面，因此單一畫面的配置和張力顯得十分重要；動態攝影則利用運鏡方式來結合時間的要素，藉由連續的畫面表達單張畫面難以呈現的「故事性」，讓觀眾的焦點能有順序的流動在畫面的各個主體上。運鏡可以藉此突顯拍攝目標物並吸引觀眾的注意力。運鏡的技巧[3]就是利用攝影機的推、拉、晃、移、跟和甩等形式的運動進行拍攝的方式，是一種在鏡頭邊界侷限中，拓展畫面視野

的方法。藉由鏡頭的移動，讓畫面從靜態變成動態，串聯並組織不同的連續畫面。常用的運鏡方式列舉如下：

- (1) 推拉鏡頭：推鏡頭相當於攝影機以直線的方式向該主體逐漸地靠近拍攝，而拉鏡頭則是攝影機以直線的方式逐漸地遠離拍攝主體。推鏡頭的作用是突顯並介紹在影片中出現的主體，而拉鏡頭則能夠展現出主體所在的環境。
- (2) 搖鏡頭：搖鏡頭指攝影機的位置不動，只是鏡頭變動拍攝的方向。其作用可逐一的展示所要描述的場景。在起始點到終點的過程裡所拍攝到的內容都有被展示的需要，而非單單帶過。
- (3) 移鏡頭：移鏡頭即是把攝影機放置在移動車上沿著軌道一邊移動一邊拍攝運鏡方式。這種鏡頭的作用是為了表現場景中的主體之間的空間關係，或者把一些事物連貫起來加以表現。
- (4) 跟鏡頭：跟鏡頭指攝影機跟隨著被拍攝物體運動的方式運鏡拍攝，使處於動態中的主體在畫面中的位置保持不變，而其他的前後景則可能不斷的變換。這種運鏡方式可以突出運動中的主體。
- (5) 升降鏡頭：升降鏡頭則是上下移動攝影機的運鏡方式，是一種從多視點表現場景的方法，其變化的技巧有垂直方向，斜向升降和不規則升降。在拍攝的過程中不斷改變攝影機的高度和仰俯角度，會給聽眾造成豐富的視覺感受。
- (6) 甩鏡頭：甩鏡頭是指一個場景拍攝結束後不停機，鏡頭急速「搖轉」向另一個方向，使得鏡頭搖轉過程中所拍攝下來的內容變得模糊不清楚。此方式主要用來強調空間的轉換和同一時間內在不同場景中所發生的並列情景。
- (7) 旋轉鏡頭：旋轉鏡頭則是使得被拍攝主體或背景呈旋轉效果的運鏡方式。
- (8) 晃動鏡頭：晃動鏡頭則是用來產生強烈的震撼力和描述主觀情緒，像是精神恍惚、乘車搖晃等情況。此外運鏡時亦應注意鏡頭運動應該保持均速、平穩、果斷，另外還要考慮到的就是前後鏡頭節奏和速度的一致性。不應無故停頓或者上下左右前後晃動，這樣不但影響內容的表達，而且使得觀眾眼花繚亂，毫無頭緒。

另外，轉場的處理，特別是在不同攝影機畫面相互轉換時的轉場處理，亦是

自動化演講錄製系統開發時需要考慮的因素[1]。理論上一部影片由多個片段所組成，而每個片段又由許多鏡頭所組成，每個鏡頭的切換或是片段的結尾都是一個場景的轉換，稱之為轉場。在轉場時要考量影片內容的流暢度。轉場的方式可分為「技巧轉場」與「非技巧轉場」。其中技巧轉場是利用特效後製來處理場景轉換，例如淡出、淡入、翻頁、定格等；「非技巧轉場」則是仰賴畫面間直接切換的手段來實現轉場，用鏡頭的移動自然地完成場景的轉換。自動演講錄製系統同時提供不同的轉場處理方式自動插入或讓導播手動選擇。

第二節 虛擬攝影師系統之相關研究

以觀眾為主體之虛擬攝影師系統的目的為研發一套以演講為場合、觀眾為主體、並自動操作攝影機進行運鏡之系統。該系統的主要功能之一在於挑選與拍攝現場狀況最為適合的運鏡手法。目前尚未發現和該功能相關之論文，因此以下將以技術為導向介紹相關技術之文獻。

2.1 景別(shots)

景別代表的是被拍攝的主體所占整個畫面的空間比例，其類別在前一個小節已稍微介紹。儘管在各種景別的區分上仍有模糊空間，不過大抵有個標準能夠參考。Cherif 等人[Che07]曾探討利用何種特徵可判斷幾種電影中常見的景別，其中主體皆為人物。在這份研究報告中，人物的臉部高度與臉的底部和畫面底部的距離被發現為一個分類景別的良好特徵。

Lang 等人[Lan09]則是對運動場上的選手們做主體的辨識與景別的界定，首先以 saliency map 方法找出畫面中的專注區域(attention field)，也就是被認定為觀賞者比較容易集中注意力的區域。然後以該區域的長寬為基礎，和畫面的長寬做比較所得到的比例做為特徵對系統進行 training，所得到的分類結果即為景別的分類。

另外，Chakraborty 等人[Cha12]同樣對運動場上的主體進行鎖定並判斷景別。Chakraborty 等人把重心放在排球場上的攝影需求，以辨識球場上正在被擊飛的排球軌跡作為主體。另外，該研究以事前架設好的四個不同位置的攝影機來代表不同的景別。每個攝影機代表一種景別，每種景別代表能夠拍攝到的不同賽事重

點。景別是利用事先定義的攝影機角度和鏡頭的縮放程度來判斷，主要原因為鏡頭所拍攝的主體(排球軌跡)和周圍的環境(球員的位置和動作)都有其限制和方便球評解說的攝影角度。

從景別判斷的相關文獻可以得知，大多數研究對景別的定義多是以主體的長寬與輸入影像的長寬的比例做為景別的判斷標準，也有少數研究和[Cha12]一樣事先定義幾個不同的拍攝位置做為不同景別。以前者的方法來說，系統必須要先偵測並決定什麼是主體才能利用比例做判斷，若是限定以人物為主體，那人臉偵測的技術將扮演重要的角色。後者則需要多台固定位置的攝影機作為前提，因此拍攝的內容內容可能缺少了運鏡能所具有的運動效果。

2.2 主體選取(target selection)

從景別判斷的相關研究可以得知，欲確認景別必須先決定主體的位置或是大小，因此選擇主體的方法將會大大影響景別判斷的成效。目前有多種選取主體的方法應用於不同類型的研究，例如追蹤物體、辨識身分等應用都需要先偵測到主體的存在才進行後續的流程。Xing 等人[Xin11]提出同時追蹤運動場上的多名運動選手的技術，其偵測主體的方法是採用 templates matching 的技術。Xing 等人先把選手在場上的完整影像以及被部分遮蔽的影像當作 templates，然後當影像輸入系統，就用 templates matching 的技術找出場上的主體。Kelly 等人[Kel11]則是提出從多個攝影機拍攝同一主體的畫面中挑選出最好的鏡頭，由於該研究的有對畫面中的人數進行限制(一人)，所以只要偵測畫面中前景的部份就能夠確認主體的位置與涵蓋的範圍。

Saliency map detection(專注分布偵測)也是被廣為研究的主體選取技術之一[Xu12][Ber06][Dor12][Lin11]，Lang 等人的研究[Lan09]就是使用 saliency map detection 的技術作為景別主體區域的選取。Saliency map detection 多以顏色作為特徵[Xu12][Dor12][Lin11]，並以高斯模糊函數將相近的顏色分作同一個類別，或是以分水嶺演算法將不同顏色的區塊區分開來，然後再擷取出顏色相較於周圍特別突出的區塊。Berengolt 等人[Ber06]則是以 edge(邊緣)的長度作為主要特徵，他們認為主體上的邊緣應該有特定的長度。他們所提的偵測方式也可以被看作一種紋理的偵測，在他們的實驗結果中，多數細小破碎的邊緣會被去除，留下完整而

長度較長的邊緣作為專注區域。以上兩類 saliency map 的偵測方式雖然有所不同，其意都在偵測畫面中集中而和周圍特徵相異的區塊，所以以這類方式取得的主體必須和背景在特徵上有明顯的差異。

以上兩類偵測目標的方法在主體的確定性上有所不同，templates matching 事先擬定的 templates 作為目標的模型，因此主體勢必與 templates 有一定程度的相仿；而 saliency map detection 則是以設定規則的方式求得符合規則，但並不確定實際模樣為何的主體。

2.3 美學(aesthetic)相關研究

Aesthetic 在影像的領域中係指畫面的美感，由於「美感」這個主觀程度較高的個人感受難以被量化，因此與美感相關的研究大多以專業攝影的觀點作為基礎對一個畫面做出審美評比。例如已經有許多研究者提出以構圖(composition)判斷評比美感的方法[Ban07][Lia12][Su12]，其利用攝影常用的構圖法，如三分構圖法、井字構圖法、對角線構圖法當作標準。接下來對原影像擷取邊緣特徵以及用 saliency map 挑選出主體位置，然後檢視這些邊緣特徵及主體的位置是否符合這些構圖法的結構。Liang 等人[Lia12]利用這個方法重新調整畫面的窗格大小及位置，等同於以攝影師的觀點重新調整畫面的構圖。在構圖法的選擇上，Banerjee 等人[Ban07]以三分構圖法，Su 等人[Su12]以井字構圖法，Liang 等人[Lia12]則使用了多種構圖法來對畫面作構圖美感評比。

另外，Li 等人[Li09]提出以其他特徵作為美感評比標準的方法，其做法為從影像中擷取兩大類的特徵。第一種是整體特徵，其內容包括畫面中的顏色在色彩空間的分布、模糊程度、邊緣密度等等，該類特徵帶給觀賞者強烈的第一印象。第二種是局部特徵，先把原畫面依 graph cut 技術[Bov01]分割成多個區塊，再以每個區塊內部的模糊程度、邊緣密度，以及區塊間的相似程度和構圖作為特徵。最後透過事先定義的規則。

以上的研究皆為探究單張畫面美感程度為主，Niu 等人[Niu12]除了單張畫面的美感分析之外，另外提出了包含時間要素的影片美感評比系統，該研究從影片中擷取多個靜態與動態特徵作為美感評比的標準。靜態的特徵包含雜訊、對焦、曝光、色彩，其中雜訊指的是干擾畫面的任何外在因素，例如光源不佳的環境或

是儀器問題造成畫面模糊等等。對焦指的是焦點的位置、景身的深淺，主要會影響主體的清晰程度。曝光指的是曝光時間是否適當，其功能為調節畫面的亮度。色彩代表畫面的色彩是不是足夠亮麗、飽和。動態的特徵則有 motion 雜訊、拍攝時間、畫面的連續性。motion 雜訊在該研究中的定義為攝影機晃動所產生的 motion，這類雜訊不只多餘、還會干擾觀賞影片的舒適度。拍攝時間指的是鏡頭停留在同一個畫面的持續時間，由於畫面長時間的停留缺乏變化性，因此應試時的改變運鏡或是切換場景以保持影片的活力。畫面的連續性代表鏡頭移動的平滑程度，快速的移動鏡頭而使得畫面在短時間內完全改變多為轉場時使用，若有過多這類移動鏡頭的方式出現會使得影片的節奏過於雜亂而難以觀賞。

不論是靜態或是動態的美感特徵都是一個專業的攝影師應該要考慮的，靜態的特徵多為畫面的邊緣結構、色彩、主體位置和模糊程度。動態的特徵則從 motion 的擷取作為基礎，衍伸出 motion 的變化活潑度、有沒有多餘的 motion 產生等等特徵。

第三章 系統概述

第一節 架設環境

此節將介紹本研究之系統所使用的攝影器材以及攝影機架設的環境，用以說明實驗的實際情況。

為了能夠讓電腦遠端操縱攝影機，並且同時具備運動攝影機的功能，本研究採用 Pan Tilt Zoom Camera (PTZ 攝影機，如圖3.1(a))作為本系統的輸入及輸出窗口。所謂的 PTZ 攝影機，和坊間一般的監視用攝影機相較之下，具有較靈活的角度變化能力，意指其鏡頭可以同時進行左右轉動(Pan)、上下傾斜(Tilt)和拉近與拉遠(Zoom)等鏡頭運動。而透過 PTZ 攝影機的靈活性，可以在拍攝過程中隨時改變攝影的角度所涵蓋的區域以及清晰程度。和傳統僅能單一方向運動的攝影機相比，PTZ 攝影機擁有更好的監控效果。而且藉由架設多台 PTZ 攝影機並利用其空間上的相對關係，就能夠以較少的攝影機數量達到智慧型監控的目的。



(a)



(b)

圖 3.1 PTZ 攝影機 (a)PTZ 攝影機 (b)全景 PTZ 攝影機與運鏡 PTZ 攝影機

本系統同時使用兩台 PTZ 攝影機，其中一台負責拍攝會場的全景，另一台負責模仿攝影機鏡頭的運作。圖3.1(a)為 PTZ 攝影機的實際外貌，圖3.1(b)則為運鏡

PTZ 攝影機及全景 PTZ 攝影機的相對位置示意圖。全景攝影機的主要任務在監控全景，取代攝影師雙眼的功能，而運鏡攝影機的主要任務則是為了取代攝影師手中的攝影機。運鏡攝影機會因為操作運鏡的關係使得觀察區域變得較為狹窄。很可能會在運鏡進行當中錯失全景範圍才能偵測到的資訊，接連影響到後續運鏡的決策；除了視野受到限制，運鏡操作時同時會增加 motion 上的雜訊，同樣會干擾判斷後續的運鏡方法。像是左右或上下移動鏡頭都會造成偵測畫面中主體 motion 的雜訊產生。因此，本研究中全景 PTZ 攝影機負責監控並擷取拍攝場地中用於判斷運鏡方法的資訊，而運鏡 PTZ 攝影機則是在系統分析完全景 PTZ 攝影機所擷取的資訊並做出下一個運鏡指令後，把指定的運鏡手法實際展現出來。

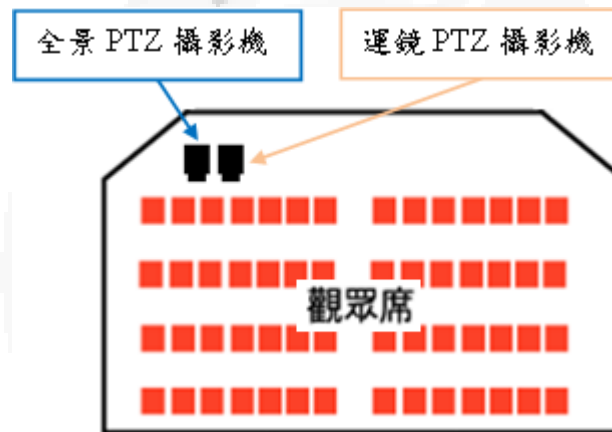


圖 3.2 PTZ 攝影機架設於水平地面演講廳示意圖

圖3.2為全景和運鏡 PTZ 攝影機架設示意平面圖，圖中紅色的方塊出現的區域為觀眾席，黑色的攝影機圖示即為 PTZ 攝影機的架設位置。為了減少校正上的誤差，兩台 PTZ 攝影機被設置在相鄰的位置，水平距離10到15公分、垂直距離15到20公分之內為容許範圍。另外，為了準確地偵測到主體位置以避免運鏡失誤，全景 PTZ 攝影機和觀眾正面的夾角必須控制在一定範圍之內。圖3.3為全景 PTZ 攝影機架設角度示意圖，圖3.3(a)為俯角 θ_d ，其範圍在 0° 到 30° 之間；圖3.3(b)為水平偏差角 θ_h ，其範圍為 0° 到 40° 。兩者的角度大小皆於影響主體辨識的重要因素。因此在系統能夠容忍的範圍內定義最大化適用角度以方便全景 PTZ 攝影機的架設。



圖3.3 PTZ 攝影機架設角度示意圖 (a)垂直俯角 (b)水平偏差角

本系統藉由調整攝影機架設高度和位置即可適用於不同類型的觀眾席場地，包括不同顏色、不同形式的座椅以及坡度。圖3.4為本系統實際運作時全景 PTZ 所拍攝的影像，圖3.4(a)為階梯式演講廳，座位較為擁擠且觀眾間有互相遮蔽的情形；圖3.4(b)為一般會議室，地面為水平且座位間的間隔較寬，兩者在場地的坡度、座椅顏色和造型上都不同。另外，本系統要求室內的光線以能夠清楚辨識觀眾正面為原則，因此在架設系統時會先排除觀眾背光的情形。



圖3.4全景 PTZ 攝影機實際拍攝影像 (a)階梯式演講廳 (b)會議室(水平地面)

第二節 系統流程

以觀眾為拍攝主體之虛擬攝影師系統之流程圖如圖3.5所示，其主要架構分為運動特徵擷取(motion feature extraction)、運鏡判斷(camera movement decision)以及路線規劃(path planning) 三個部分。Motion feature extraction 包含連續畫面的

motion 擷取、motion feature 的擷取以及 motion feature 的模糊化(fuzzification)。Camera movement decision 包含判斷是否需要切換運鏡、運鏡的選擇以及景別的選擇。Path planning 包括拍攝主體候選者的偵測、拍攝主體的選擇以及運鏡路線的規劃，最後操作運鏡 PTZ 攝影機進行運鏡的動作。以下會對這三個主要流程做進一步的說明。

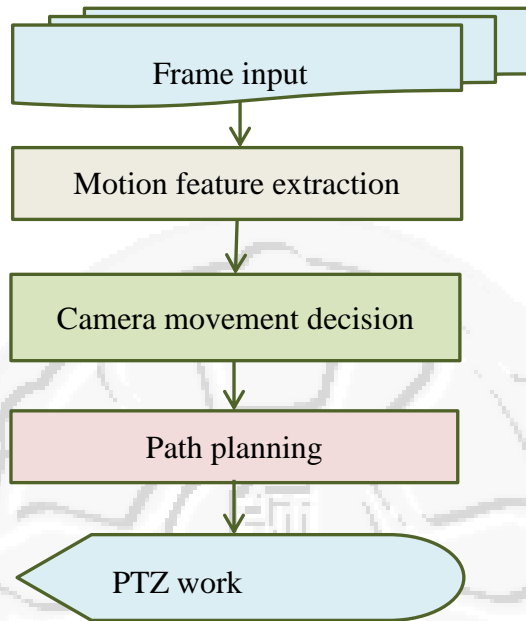


圖 3.5 系統流程圖

圖3.6為 motion feature extraction 之流程圖，分為 MEHI calculation、feature extraction 以及 feature fuzzification 三個部分。首先從全景 PTZ 輸入連續影像，然後對每張影像偵測 motion，為求運算的便利及速度，偵測 motion 的方法使用差值影像法(intensity difference image, IDI)。以此法所得到的 motion 為單一時點的資訊，只能知道 motion 產生，而不能得到 motion 的方向和持續性。因此使用 motion history image (MHI) 技術來擷取連續的 motion 資訊，MHI 以重疊 IDI 和隨時間衰減的方式記錄 motion 的時序性。但是由於 MHI 具覆蓋先前 motion 資訊的性質，此性質會造成被覆蓋 motion 能量的佚失而不利於某些 motion feature 的擷取。因此本研究利用改變 MHI 重疊的方式另創 motion energy history image (MEHI)讓 motion 的能量不會被覆蓋以便於 feature 的擷取。

在完成 motion 資訊的擷取後，接著進行 feature extraction 的步驟，該步驟會從 MHI 及 MEHI 共擷取四種 motion features。此四種特徵分別為運動強度(motion strength)、運動集中度(motion concentration)、運動強度亂度值(motion strength entropy)、運動方向亂度值(motion orientation entropy)。這四種 features 能夠代表觀眾的熱烈程度和動作的一致性，可以來判斷適合的運鏡。

由於這五種 motion features 都是明確的數值，難以和人類的感覺做連結，例如由上個步驟擷取的 motion strength 是一個數值，但是以人的觀點來度量這項特徵通常會以「強」、「弱」來區分。Fuzzification 是個連結這兩種關係的方法，可將純數值和人類的語言中的語義結合對應，讓使用者能夠用語言表達的方式定義規則。如此不但可將 motion features 和攝影師判斷運鏡的標準作結合，同時也利於人工擷取 motion features。關於 motion feature extraction 之內容將在第四章詳細介紹。

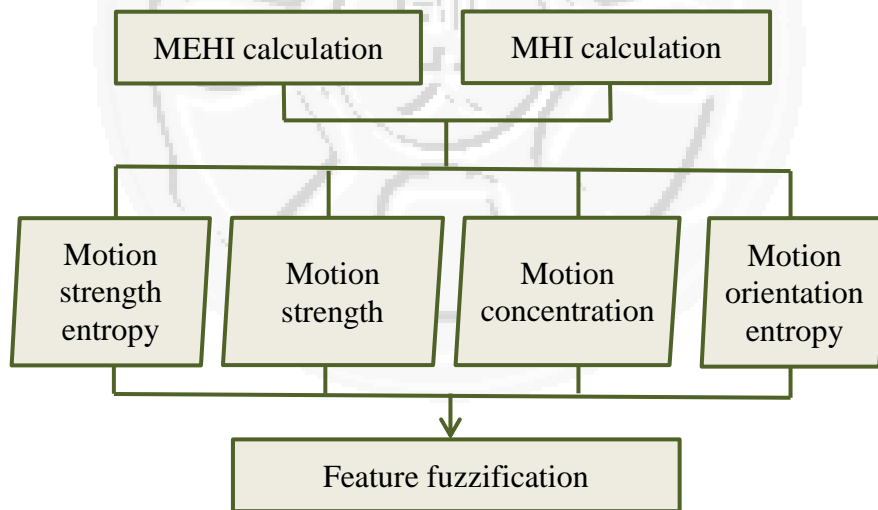


圖 3.6 Motion feature extraction 流程圖

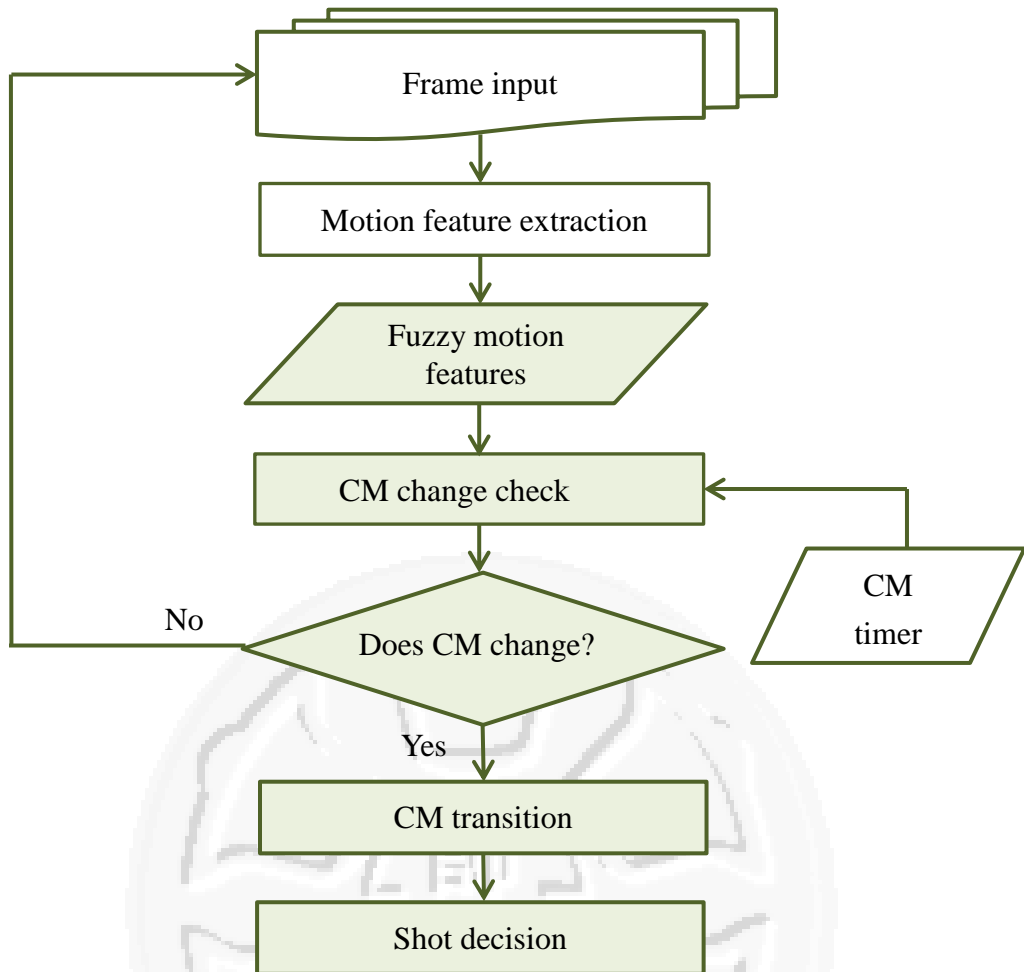


圖 3.7 Camera movement decision 流程圖

圖3.7為 camera movement decision (CM decision)的流程圖，分為 CM change checking、CM transition、shot decision 三個主要步驟。Change check 的步驟主要在確認切換運鏡的必要性，影響運鏡是否切換的要素有二，第一是 fuzzy motion feature 值的改變，第二是前一個運鏡的持續時間。Fuzzy motion features 是上一個步驟所得到的模糊化 motion features，若是這些特徵值發生變化表示畫面中的 motion features 已經有明顯足以辨認的強弱差異，系統在其發生大幅變化時應該要採取改變運鏡的動作。運鏡的持續時間是指運鏡開始執行到切換至下一次運鏡的時間間隔，運鏡持續時間太短會造成短時間內連續的切換而使得影片混亂而且瑣碎。若是運鏡持續時間太長又會讓人感覺缺乏變化，透過相關研究[Chu09]得知，一般影片的運鏡持續時間約在5.3秒左右。因此本系統設定運鏡持續時間的最短

為3秒(3秒內不得切換運鏡),最長不超過6秒(6秒內必須切換運鏡)來避免以上問題。若運鏡持續時間不到最短時間或是 fuzzy motion features 沒有變化就被視為不需要切換運鏡(回到輸入連續影像),反之則進入下一個步驟 CM transition。

CM transition 的目的在於選擇合適的運鏡種類,這個步驟除了必須考慮如何模仿攝影師的經驗判斷外,另外還需考慮鏡頭間的關聯性。系統要模仿攝影師必須要有能夠學習並記錄攝影師習慣的能力,並且要能夠將攝影師的經驗轉化成系統能夠接受的型式。鏡頭間的關聯性指的是連續運鏡之間的關係,前幾個運鏡可能間接或直接的影響到下一個運鏡的選擇,例如連續的定點攝影、定點攝影接著拉近或拉遠以及拉近後橫移等等。本研究提出 automatic camera movement model (ACMM)以解決上述的問題,ACMM 以 finite state machine (FSM)為基礎,主要結構分為 camera movement state (CMS)以及 transition probability model (TPM)。CMS 如同 FSM 的 states,先前的每個輸入都會影響到 CMS 的轉移,也可以說是具有紀錄先前輸入的性質,這個性質有利於運鏡間關聯性的建立。任意兩個 CMS 間皆有 TPM 連接,TPM 具有指向性,其功能為記錄攝影師的運鏡習慣。ACMM 的輸入為 fuzzy motion features,並將這些特徵值放入與當前 CMS 連接且方向向外的 TPM 中。每個接收到輸入的 TPM 將會產生一轉移機率,然後挑選具最高機率的 TPM 做為轉移方向。而該 TPM 所指方向之 CMS 的輸出即為下一個運鏡。

在決定合適的運鏡之後,shot decision 步驟將決定與該運鏡配合的景別。ACMM 的每個 TPM 上都具有景別機率模型 (shot probability model),該模型統計了通過此 TPM 的景別機率分布。藉由上一步驟所挑選出的 TPM 其上的景別機率模型可以得到每種景別的使用率,為避免每次都只用同一種景別,在挑選景別時以使用率的高低作為選取的機率以增加選擇的多元性。以上關於 CM decision 的部分將在第五章作詳細介紹。

圖3.8為 path planning 之流程圖,主要步驟分別為 face target detection、motion target detection、target decision 以及 CM path planning。當 ACMM 決定運鏡及景

別後，path planning 將會決定拍攝主體並依據所決定的運鏡規劃運鏡 PTZ 攝影機鏡頭的移動路線。拍攝的候選主體有兩種產生方式，一種是以觀眾的臉部作為目標，另一種是以產生 motion 的區域中心做為目標。由於本系統以觀眾為拍攝主體，因此以觀眾臉部的拍攝較為優先。Motion targets 則做為備用以及排序 face targets 的條件之一。

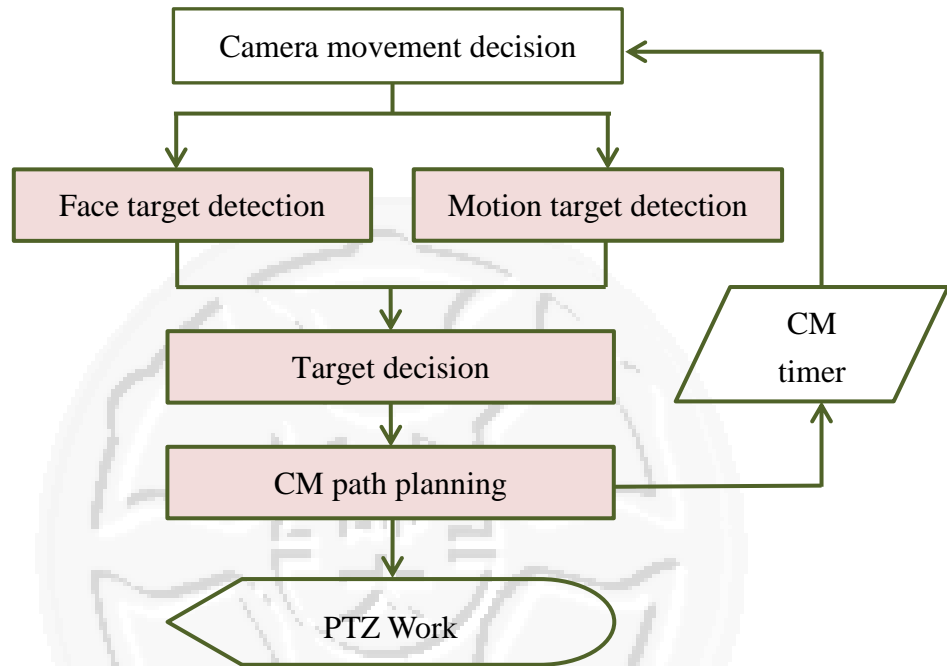


圖 3.8 Path planning 流程圖

利用前一步驟挑選出拍攝候選主體之後，target decision 將從中決定拍攝主體。其決策標準包含構圖(composition)、專注分布(saliency map)、光線明暗、主體與畫面中心的距離以及主體和 motion targets 的距離。經由以上條件的評比將會選出此次運鏡的拍攝主體。Composition 代表畫面的構圖，在構圖法中以井字構圖法最為普遍。因此本研究以此構圖法作為標準，將攝影主體候選者放置在井字構圖法中的理想位置來計算構圖評分[Ban07]。Saliency map 表示畫面中被觀賞者注意區域的分布，該值越高表示越容易受到注意。由於光線的明暗會影響到拍攝的效果，畫面中過於黑暗或是明亮的部分將被評為不適合拍攝的區域。主體與畫面中心的距離會影響運鏡 PTZ 攝影機移動的幅度，移動幅度越大越容易影響到與前一

張畫面的連續性。另外，主體和 motion target 的距離越近表示該觀眾可能正在動作，相較於其他沒有 motion 相鄰的觀眾較為突出，因此兩者之間的距離越近越適合當作拍攝主體。

最後，CM path planning 將會依運鏡種類不同替拍攝主體規劃運鏡 PTZ 的移動路線及產生運鏡 PTZ 的控制參數。運鏡 PTZ 的控制參數包含起始位置(start point)、結束位置(end point)、運鏡速度(move speed)及推拉速度(zoom speed)。其中 start point 及 end point 與運鏡的種類相關；move speed 會依 motion strength 強弱作改變；zoom speed 的快慢則和景別與拍攝主體的面積比有關。最後這些參數將會轉成 PTZ 的指令傳送給運鏡 PTZ，執行運鏡的動作。Path planning 的詳細作法將會在第六章說明。



第四章 資料前處理與特徵擷取

第一節 影像前處理

4.1.1 差值影像(intensity difference image)

為了擷取 motion 的時間特徵，首先必須先得到影像中的 motion 資訊。差值影像法是一個快速且有效的 motion 偵測方法。當物體移動時，在物體當前位置和前一個位置有顏色差異的情況下則差值影像上會呈現出 intensity 值的差異。利用這個性質可以知道物體是否移動，並能得到畫面中何處產生 motion。令 I_t 為時間 t 之灰階影像， $I_t(x, y)$ 為 I_t 中座標為 (x, y) 的 pixel 的 intensity 值； I_{t-1} 為時間 $t-1$ 之灰階影像， $I_{t-1}(x, y)$ 為 I_{t-1} 中座標為 (x, y) 的 pixel 的 intensity 值，其中 $1 \leq x \leq W$ ， $1 \leq y \leq H$ ， W 為 I_t 的寬度， H 為 I_t 的高度。則 t 時間的差值影像 D_t 計算方式如下：

$$D_t(x, y) = |I_t(x, y) - I_{t-1}(x, y)|$$

當 D_t 中某個 pixel 的 intensity 值越大代表該處的 motion 越明顯，反之則越微弱，微弱的 motion 可視為雜訊，由於雜訊可能干擾 motion feature 的擷取構成影響，所以系統對 D_t 進行二值化處理。令 T_{the} 為二元化之門檻值，則時間 t 之二值化差值影像 B_t 公式如下：

$$B_t(x, y) = \begin{cases} 1 & \text{if } D_t(x, y) > T_{the} \\ 0 & \text{otherwise} \end{cases}$$

上式中若 $D_t(x, y)$ 大於門檻值 T_{the} ，將 $B_t(x, y)$ 值設為 1，反之，則設為 0。

4.1.2 Motion history image (MHI)

利用上述差值影像擷取 motion 的方法雖然能夠擷取當前 motion 的位置，卻無法得知 motion 的移動方向和不同 motion 發生的先後關係，若利用 motion history image(MHI)則可保留更多 motion 資訊[Bob01]。令 τ 為 intensity 值的上限(在灰階

影像中為255)， B_t 為時間 t 之二值化差值影像， $0 < k < 1$ ， k 為遞減倍率，則 MHI 的計算方式如下：

$$MHI(x, y, t) = \begin{cases} \tau & \text{if } B_t(x, y) = 1 \\ \max(0, k \times MHI(x, y, t-1)) & \text{otherwise} \end{cases}$$

當 $B_t(x, y)$ 值為1代表該 pixel 的位置 (x, y) 產生 motion，則 $MHI(x, y, t)$ 值設為 τ ，其餘沒有在時間 t 偵測到 motion 的 pixel 的 intensity 值皆乘上遞減倍率 k 。其上述公式的目的在疊加連續的差值影像保存連續的 motion 資訊，並隨時間降低沒有 motion 產生的 pixel 位置的 intensity 值。因此 MHI 上 intensity 值較高的 pixel 代表其所在位置 motion 發生的時間較晚，反之則代表 motion 發生時間較早。藉此能夠由 MHI 影像中 intensity 值的高低判斷 motion 產生的先後，進一步能夠分析物體移動的方向。

4.1.3 Motion energy history image (MEHI)

MHI 會把 motion 發生所在位置的 pixel 之 intensity 值更新設為最大灰階值，此舉的用意在於區別 motion 發生的次序。然而這個作法同時也產生了一個缺點，當物體沿著相同的軌跡來回移動時，MHI 影像中在軌跡上的 pixels 的 intensity 值會不斷地更新到最大值。因此即使物體在同樣的軌跡上重複移動數次，在 motion 發生的位置上只會存留最近一次的 motion 資訊。雖然 MHI 作法能保留 motion 移動方向的資訊，但卻無法提供 motion 是否重複發生的資訊。

為保留 motion 是否重複發生的資訊，本研究改良 MHI 計算公式，建構 motion energy history image (MEHI)。令 k 為遞減倍率， $0 < k < 1$ ， a 為增加倍率， $1 < a < 2$ ， B_t 為時間 t 之二值化差值影像，MEHI 初始值為一全零影像，則 MEHI 如下所示：

$$MEHI(x, y, t) = \begin{cases} \min(\tau, a \times MEHI(x, y, t-1)) & \text{if } B_t(x, y) = 1 \\ \max(0, k \times MEHI(x, y, t-1)) & \text{otherwise} \end{cases}$$

每當 $B_t(x, y)$ 值為1時將 $MEHI(x, y, t-1)$ 值乘上 a 以得到 $MEHI(x, y, t)$ ，即使同一個 pixel 的位置不斷有 motion 產生，也能和單次移動所造成之 motion 有所區隔，

讓每個 motion 的資訊都能保存下來而不被覆蓋。雖然 MEHI 能夠留下能量的資訊，但因為重複發生 motion 的位置上的 pixel 有較高的 intensity 值會影響紀錄 motion 產生的先後的資訊。所以本研究會依照不同 motion features 的需求分別在 MHI 或 MEHI 上進行擷取。

第二節 特徵擷取

本節將介紹五種 motion features，包含 motion strength (m_s)、motion concentration (m_c)、motion strength entropy (m_{se})、motion orientation entropy (m_{oe})和 action (act)等五種特徵。

4.2.1 Motion strength

Motion strength 的數值高低代表畫面中觀眾的動作多寡。一段影片若具有較高的 motion strength 即為畫面中出現較多的 motion，若以拍攝觀眾席的角度來看，這象徵著觀眾的氣氛熱烈。例如賽場上的球員展現美技時，觀眾揮手狂歡的情形，在這種觀眾情緒高昂情況下，由於興奮使然，動作會變多幅度會變大，motion strength 自然會較高。若一段影片的 motion strength 較小，則反映出觀眾平靜的態度，也可能是賽況緊張的令人屏息，亦或是專注於講台上的演講者。因此 motion strength 可以說是代表現場氛圍的一項重要資訊。

Motion strength 的擷取方法為累計 MEHI 中每個 pixel 的 intensity 值。令 I_{ME} 為一 MEHI 影像， W 為 I_{ME} 之寬度， H 為 I_{ME} 之高度， τ 為 intensity 值的上限(在灰階影像中為255)，則 I_{ME} 的 motion strength m_s^u 為：

$$m_s^u = \sum_{x=1}^W \sum_{y=1}^H I_{ME}(x, y) / \tau$$

為了讓不同大小的影像都能有等質量的 motion strength，本研究將 m_s^u 除以 I_{ME} 的面積進行正規化，正規化結果 m_s 為：

$$m_s = m_s^u / (W \times H)$$

4.2.2 Motion concentration

Motion concentration 代表 motion 發生區域的集中程度，在同樣 motion strength 值的情況下，motion 可能集中發生在畫面中的一個區域，也可能平均散佈在整個畫面中。而當 motion 發生的區域較為集中，攝影師較可能將目標放在 motion 發生的位置，並且調整景別和使用的運鏡以符合 motion 集中區域的大小。若是同樣的 motion 量平均分布在畫面當中，代表幾乎每位觀眾都有動作發生，此時的場景可能是群起揮手、起立拍手或是散場。在這種狀況下所使用的運鏡多為緩慢的左右移動或是定點拍攝，景別也以能夠包含大多數觀眾的遠景或大遠景為主。

Motion concentration 的擷取方式如下：首先對 MHI 影像作 connected component 處理，將 MHI 中 intensity 值不為0且相連的 pixel 當作一個區塊，因為面積較大的區塊代表活動的區域較大，比起小幅度的動作，大範圍的運動較容易引起攝影師的注意，因此每個區塊以面積做為權重。令共 MHI 中有 b 個區塊，其中一個區塊 p_i 的重心座標為 (C_i^x, C_i^y) ， w_i 為 p_i 區塊的權重， $1 \leq i \leq b$ ， $i \in N$ 。

(C_m^x, C_m^y) 為區塊 p_i 的加權中心，則 C_m^x 、 C_m^y 分別為：

$$C_m^x = \frac{\sum_{i=1}^b w_i \times C_i^x}{\sum_{i=1}^b w_i}, \quad C_m^y = \frac{\sum_{i=1}^b w_i \times C_i^y}{\sum_{i=1}^b w_i}$$

令 $dis(\cdot)$ 為幾何距離函數， d 為 MHI 影像的對角線長度，則 motion concentration m_c 如下式所示：

$$m_c = \frac{\sum_{i=1}^b w_i dis((C_i^x, C_i^y), (C_m^x, C_m^y))}{d \sum_{j=1}^b w_j}$$

當 motion 的區塊面積差異小且中心座標分散時， $dis(C_i, C_m)$ 值會上升，進而造成 m_c 值的下降。若有面積特別大的區塊存在，加權中心會靠近面積較大的區塊，因此會取得較大的權重值，所以 m_c 在區塊集中的情況下會上升。

4.2.3 Motion strength entropy

在4.2.1節所介紹的 motion strength 雖然能夠記錄整體的 motion strength，卻不能從中分別得知每個 motion 區塊的 motion strength 差異。即使一張 MEHI 中有強烈的 m_s ，卻無法提供每個 motion 區塊的相對強弱關係。任何 motion 的表現方式，不論是整體或是相互關係都可能是影響運鏡和的景別的因素，例如各 motion 區塊的 motion strength 都差不多，極可能是一種整體觀眾的動作，而不再單是個別觀眾的動作。整體觀眾的動作的意思是指由多位觀眾共同完成的動作，像是大家一起拍手或是波浪舞一類的共同動作。如果各 motion 區塊間的 motion strength 差異大，則真正在發生的有效 motion 可能只有少數幾個，其餘強度較弱的 motion 多為雜訊。

在分析 motion strength 歧異度的部分，本研究選擇了 entropy 作為訂定歧異程度的方法。Entropy 的概念源自於物理學，本用來表達熱力學系統的無序程度，現今則廣泛應用於資訊科學上。Entropy 在資訊領域代表的意涵為對資料類別不確定性的測量，當資料的類別越一致，entropy 將會越低，也代表資料大多屬於同一類。

為求得 motion strength entropy，首先將 m_s 值依照大小分為 k 類($m_s^1 \sim m_s^k$)，令 b 為 MEHI 做 connected component 處理後產生的 motion 區塊數量， p 為機率質量函數，則 m_{se} 如下式所示：

$$m_{se} = -\sum_{i=1}^k p(m_s^i) \log p(m_s^i)$$

當 m_{se} 越大代表各個區塊的 m_s 值分布於不同的強度類別，表示觀眾間動作強度差異大，反之則代表觀眾間動作強度差異小。

4.2.4 Motion orientation entropy

不同的動作發生在觀眾席上具有不同的意義。當初不只是動作上的差異，有

多少觀眾做同樣的動作也同樣具有不同的意義。許多觀眾同時做不同的動作所代表的是互相獨立的行為，沒有整體氣氛的连接，每個動作都適合做為運鏡目標。而許多觀眾做同樣的動作則可能是配合講者的演出，或是具有強烈氣氛的加油動作，讓多數觀眾入鏡是比較好的選擇。若是每位觀眾的動作都不同，那 motion 方向的分布就不會全部屬於相同的方向。由於個體間的差異較大，若用大範圍的景別和運動幅度較大的運鏡反而會讓畫面更為混亂。如果觀眾的動作大多集中於同一個方向，即使放大景別或是用運動較為激烈的運鏡也不至於讓畫面顯得混亂。

Motion orientation entropy 的計算方式如下：首先計算 MHI 影像中每個 pixel 的運動方向，由於 MHI 影像紀錄了 motion 發生的時序關係，因此利用計算二維梯度影像就能夠得到每個 pixel 上運動方向，梯度影像 ∇MHI 的計算方法為：

$$\nabla MHI = \left(\frac{\partial MHI}{\partial x}, \frac{\partial MHI}{\partial y} \right)$$

接著將 ∇MHI 中的 pixel 值依代表的角度方向分為 k 類 ($o_1 \sim o_k$)，然後使用 entropy 公式來計算代表全部 motion 區塊的 motion orientation entropy m_{oe} 。令 p 為機率質量函數，則 m_{oe} 為：

$$m_{oe} = -\sum_{i=1}^k p(o_i) \log_2 p(o_i)$$

第三節 特徵模糊化

被擷取出來的五種 motion features 除了 *act* 以外皆為連續實數，而這些 motion feature 的數值其實和人類的直覺是難以連結的。當攝影師看到一個拍攝觀眾的畫面，會以語言的方式來形容這個畫面中的情況。例如「觀眾的動作多不多」、「大部分的觀眾正在做同樣的動作」、「那邊有個觀眾正在揮手」等等描述觀眾席上情況語句。每一個 motion feature 都對應到攝影師在決定運鏡時所考慮的因素，但是由於表示法的差異而難以將攝影師的運鏡經驗納入本系統，造成模仿攝影師運鏡

的困難。因此本研究採用模糊化(fuzzy)技術將單純的數值轉為描述特徵強弱的語句，讓系統可以將攝影師對特徵的描述和所偵測到的特徵數值做結合。

Fuzzification 這項技術起源於1965年由美國加州大學查德(L. A. Zadeh)教授所提出的模糊集合(fuzzy set)的觀念，被用於描述現實中各種事物的「程度」。在傳統的二元邏輯中，一個元素和一個集合只有屬於該集合和不屬於該集合兩種關係。模糊集合是以隸屬函數(membership function)值來描述元素和集合的關係，意即以0到1的數值來表示一個元素屬於一個集合的隸屬程度，該值被稱作隸屬程度值(membership grade)。另外，傳統集合以數值作為變數，模糊集合則是以語意變數(linguistic variables)作為變數。語意變數以自然語言中的字或句子而不是以數字為值的變數，例如定義一個語言變數「氣溫」，其語意值(linguistic value)可以是「很冷」、「有些冷」、「不冷不熱」、「有點熱」、「很熱」。

表4.1 Fuzzy motion feature 與語意變數對照表

特徵	模糊集數量	語意變數	語意值
m_s	3	觀眾活潑度	高，中，低
m_c	3	運動集中度	集中，稍微集中，分散
m_{se}	3	運動強度一致性	強度一致、強度稍微一致、強度不一致
m_{oe}	3	運動方向一致性	方向一致、方向稍微一致、方向不一致

本研究所擷取的 motion features 都需要進行模糊化來和攝影師的感覺做連結，因此定義四個語言變數以對應四個 motion features (表4.1)，分別為觀眾活潑度、動作集中度、動作強度一致性、動作方向一致性。首先必須先替語言變數決定輸入值的範圍(論域)，這四個 motion features 都是0到1的實數，所以四個語言變數的論域皆為0到1的實數。模糊集 (fuzzy partition)代表以多少數量的語意值來描述該模糊變數，就人的感覺來較難細分這些特徵。為了讓每個模糊集合之間皆有能明顯判斷的區隔，每個語言變數的模糊集數量皆為3個，其各自的語意值如表4.2所示。

隸屬函數的圖形常見的有三角形、高斯函數、梯形為三種。本研究採用三角

形的隸屬函數，因為三角形隸屬函數的計算速度較快且與其他兩種函數的效果相差並不大。圖4.1為語言變數的隸屬函數示意圖，其中 x 為運動特徵值， $\mu(x)$ 為隸屬函數值，三個三角形分別為該語意變數的三個模糊集合的隸屬函數圖形。

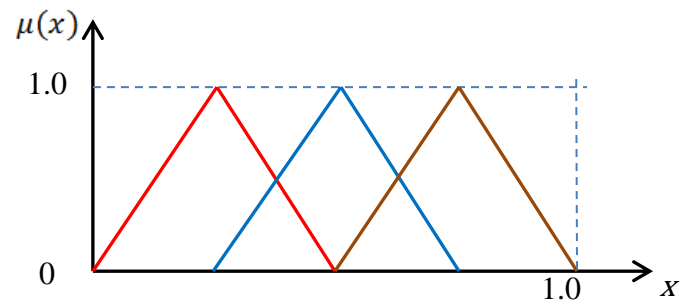


圖 4.1 語言變數之類屬函數示意圖

經過隸屬函數的轉換後，輸入的 motion feature 會產生對每個隸屬函數的隸屬程度值，接著選擇最大隸屬程度值的模糊集其語意值作為模糊化後的特徵值並命名為 fuzzy motion features。

第五章 運鏡選取

第一節 運鏡切換條件

影響是否切換運鏡的要素有二，其一是 motion feature 值的改變，其二是前一個運鏡的持續時間。由於 fuzzy motion features 是由 motion feature 模糊化而來，每個 fuzzy motion feature 都具有三個模糊集。只分割成三個模糊集的目的為的是讓每個模糊集之間的差異較為明顯，所以每當 fuzzy motion feature 值產生變化，代表該特徵有明顯的變動，而且是足夠被攝影師觀察到的明顯差異。不過並非每次 fuzzy motion feature 值發生變化都一定會造成運鏡切換，每個 fuzzy motion feature 值的改變都具有不同的意義。因此本研究對於每個 fuzzy motion feature 個別設定觸發運鏡切換的條件以進行適當的運鏡切換，表5.1為 fuzzy motion feature 與其語意值對照表，本研究將依照此表設定並說明切換運鏡的觸發條件。

表5.1 Fuzzy motion feature 與其語意值對照表

特徵	變數名稱	語意值
\tilde{m}_s	觀眾活潑度	高，中，低
\tilde{m}_c	動作集中度	集中，稍微集中，分散
\tilde{m}_{se}	動作強度一致性	強度一致、強度稍微一致、強度不一致
\tilde{m}_{oe}	動作方向一致性	方向一致、方向稍微一致、方向不一致

\tilde{m}_s 為 m_s (motion strength) 模糊化過後的模糊特徵，可用來描述觀眾活潑度。當 \tilde{m}_s 對於「低」模糊集的隸屬程度值降低，對「高」模糊集的隸屬程度值提升時，代表畫面中的 motion 正在增加、觀眾的動作開始變多。此時因為 motion 增加而可能有新的拍攝主體產生，或是讓某些畫面的某些區域更容易受到注目。為求畫面呈現的多變性，若有新的拍攝主體能夠選擇就是一個適合切換運鏡的時機。如果 \tilde{m}_s 對「高」的隸屬程度值降低，而對「低」的隸屬程度值升高時，多數的情況是由動作產生的 motion 正在消退，觀眾的動作趨近和緩但 motion 的資訊仍殘留

在 MEHI 之中。此時並非適合切換運鏡方式的時機。

\tilde{m}_c 為 m_c (motion concentration) 模糊化後的模糊特徵，可用來形容觀眾在畫面中的動作集中或是分散。當 \tilde{m}_c 對「集中」模糊集的隸屬程度值降低，「分散」模糊集的隸屬程度值升高，代表大多數的觀眾都開始做出動作，主體則由少數觀眾轉換成多數觀眾。如果 \tilde{m}_c 值對「分散」模糊集的隸屬程度值降低，「集中」模糊集的隸屬程度值提高，則代表主體從大部分的觀眾轉移到少部分的觀眾。這兩種情況都會造成拍攝主體的轉移，因此皆為適合運鏡方式切換的時機。

\tilde{m}_{se} 為 m_{se} (motion strength entropy) 模糊化後的模糊特徵，可用來形容觀眾的動作的劇烈程度。當 \tilde{m}_{se} 對「強度一致」模糊集的隸屬程度值降低，「強度不一致」模糊集的隸屬程度值增加時，代表部分觀眾做出特別劇烈的動作，則該觀眾可能被當作拍攝主體，此時適合轉換運鏡方式。若 \tilde{m}_{se} 值對「強度不一致」模糊集的隸屬程度轉弱，「強度一致」模糊集的隸屬程度轉強，代表沒有任何觀眾的動作特別突出，此時並沒有特別明顯的新主體需要拍攝，所以沒有轉換運鏡的必要。

\tilde{m}_{oe} 為 m_{oe} (motion orientation entropy) 模糊化後的模糊特徵，可用來形容觀眾的動作方向是否一致。當 \tilde{m}_{oe} 對「方向不一致」模糊集的隸屬程度值降低，而對「方向一致」模糊集的隸屬程度值提高時，代表動作觀眾做出相同的方向的動作，像是波浪舞、一起起立或是舉手等情形，此時應該選擇大部分觀眾作為拍攝主體。若動作方向由一致轉為不一致時，可能代表動作並無特殊趨向，沒有整體性的動作產生，或是動作結束後的剩餘部分，沒有必要轉換運鏡方式。

另外一個影響是否切換運鏡的因素是運鏡的持續時間，其定義為從運鏡 PTZ 攝影機開始運鏡動作到下一個運鏡動作開始的這段時間。運鏡持續時間太短會發生上一個運鏡動作才剛開始進行，下一個運鏡動作就開始的情況。由於每次的運鏡動作都會產生新的主體以及新的運鏡路線，運鏡 PTZ 攝影機會被移動到新的運鏡路線的起始位置。在運鏡的途中轉換運鏡會使得影片無預期中斷而使得畫面不連續，甚至造成畫面切換的混亂，即使有新的拍攝主體產生，這類的情況仍極需

避免。

另外，若是運鏡持續時間過長則會產生兩種情況，第一種是上一個運鏡已經結束，但下一個運鏡卻尚未開始，甚至是待機一段長時間。另一種是因為 fuzzy motion features 的變化而發現新主體產生，但是系統卻因為運鏡持續時間的限制而拒絕切換運鏡。前者會造成畫面的停滯而讓影片失去活力，而且明顯讓人感覺到運鏡的中斷，後者容易錯失更合適的拍攝主體，像是發生鼓掌、起立這類整體性的動作時，系統都應該立刻改變運鏡以捕捉最即時的畫面。

表5.2 攝影機運動的數據資料[Chu09]

		Pan	Tilt	Roll	Dolly Truck	Crane	Zoom	Hand held	綜合	None
玩具 總動員	個數	48	32	1	53	28	18	70	11	1288
	%	3.1%	2.1%	0.1%	3.4%	1.8%	1.2%	4.5%	0.7%	83.2%
	秒數	285.85	155.67	2.57	215.48	294.27	102.22	205.38	88.6	3255.56
	%	6.20%	3.40%	0.10%	4.70%	6.40%	2.20%	4.50%	1.90%	70.70%
平均秒	5.96	4.86	2.57	4.07	10.51	5.68	2.93	8.05	2.53	
史瑞克	個數	93	62	4	69	78	56	52	58	811
	%	7.2%	4.8%	0.3%	5.4%	6.1%	4.4%	4.1%	4.5%	63.2%
	秒數	496.82	261.47	9.34	316.53	502.58	253.77	157.11	437.59	2492.7
	%	10.10%	5.30%	0.20%	6.40%	10.20%	5.10%	3.20%	8.90%	50.60%
平均秒	5.34	4.22	2.34	4.59	6.44	4.53	3.02	7.54	3.07	
冰原 歷險 記	個數	22	31	1	131	23	27	60	17	1072
	%	1.6%	2.2%	0.1%	9.5%	1.7%	2.0%	4.3%	1.2%	77.5%
	秒數	101.9	117.54	2.23	402.75	165.78	180.9	239.64	145.25	3115.52
	%	2.30%	2.60%	0.00%	9.00%	3.70%	4.00%	5.40%	3.20%	69.70%
平均秒	4.63	3.79	2.23	3.07	7.21	6.70	3.99	8.54	2.91	
總個數%	4.0%	3.0%	0.1%	6.1%	3.2%	2.5%	4.3%	2.2%	74.6%	
總平均秒	5.31	4.29	2.38	3.91	8.05	5.64	3.32	8.05	2.84	

因此運鏡持續時間不能夠太短促但也不能太冗長，需要一個適度的時間範圍來限制。Chung 等人[Chu09]提出好萊塢3D 動畫長片的分鏡技術研究，分鏡技術包含景別、運鏡、攝影機角度以及人稱視點的規劃。表5.2為該研究對運鏡持續時間的分析表，表中列出數個影片中出現的運鏡手法以及平均運鏡持續時間。本研究只探討本系統提供的運鏡手法的部分，包括橫向移動(pan)、垂直移動(tilt)、伸縮鏡頭(zoom)以及無運鏡(none，指定點攝影)，其平均運鏡持續時間分別為5.31、4.29、5.64以及2.84秒。本系統將對不同的運鏡手法設定不同的運鏡持續時間以符

合該研究的運鏡持續時間平均值。

表5.3運鏡持續時間與運鏡種類關係

運鏡種類	運鏡持續時間(s)	
	下限	上限
Pan	4.5	6
Tilt	3.5	5
Zoom	4.5	6
Cut	2	3.5

表5.3為運鏡持續時間與運鏡種類關係表，本系統依據[Chu09]之研究結果設定每個不同的運鏡手法的運鏡持續時間，不採用統一的運鏡持續時間是因為每種運鏡手法適合的平均運鏡時間並不相同。例如定點攝影容易造成畫面的活力降低，所以的運鏡持續時間較其他運鏡種類為短，因此本系統針對每種運鏡手法設定各別的運鏡持續時間上下限。

第二節 自動運鏡模型(automatic camera movement model, ACMM)

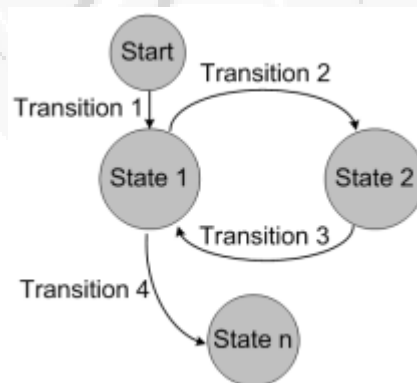


圖5.1 FSM 結構圖[25]

自動運鏡模型(automatic camera movement model, ACMM)是以有限狀態機(finite-state machine, FSM)作為樣板所構築的模型，其功能為決定使用何種運鏡，同時保持運鏡間的關聯性。FSM 如圖5.1所示[25]，其架構包含了輸入(input)、狀態(state)、轉移(transition)。State 的功能為紀錄系統的狀態，而 transition 則代

表著狀態之間轉移的方向。例如圖5.1中若輸入的資料經由 transition 1 進入 state 1，接著可能經由 transition 2 前往 state 2 或是經由 transition 4 進入 state n。若是前者的情況第二次的轉移只可能經由 transition 3 回到 state 2，而後者則會一直停留在 state n，兩者的發展趨勢將會有所不同。

圖中 state 之間的箭頭符號則代表 transition，transition 代表由一個 state 前往下一個 state 的動作，每個 transition 的發生都必須滿足該 transition 所帶有的條件。FSM 依照輸出方式的不同可分為兩種不同的模型。一種被稱為 Moore 模型，另一種為 Mealy 模型。Moore 模型中其輸出值依附於 state 之下，依 state 不同而不同。圖5.2(a)為一 Moore 模型的範例，其中 x 、 y 、 z 為輸入值； q_0 、 q_1 、 q_2 、 q_3 為 state； a 、 b 、 c 為輸出值。依照 Moore 模型的輸出方式，當 state 為 q_0 時輸出 b ，state 為 q_1 時輸出 a ，以此類推。Mealy 模型的輸出則同時建立在 state 和 input 上，輸入值和當時的 state 會共同決定輸出值。圖5.2(b)為一 Mealy 模型的範例， x 、 y 、 z 為輸入值； q_0 、 q_1 、 q_2 、 q_3 為 state； a 、 b 、 c 為輸出值。當 state 為 q_0 、輸入為 z 時，輸出為 a 。當 state 為 q_1 且輸入為 y 時輸出為 a ，輸入為 x 或 z 時則輸出為 c 。

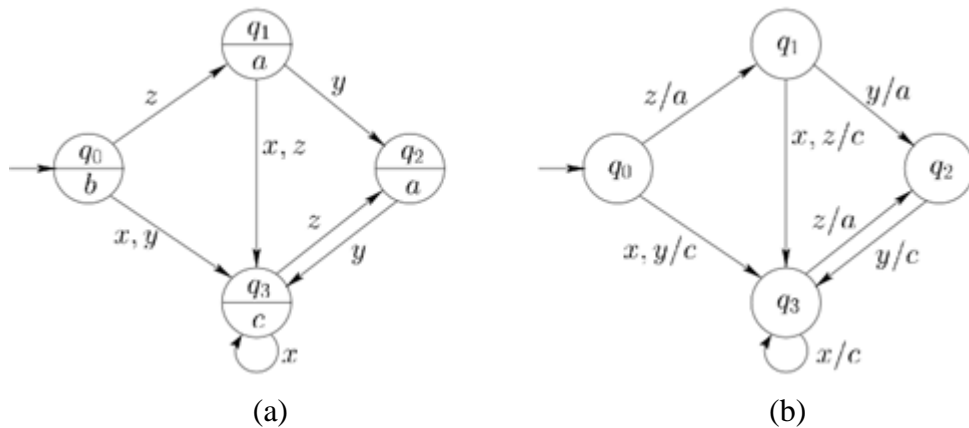


圖5.2 Moore 和 Mealy 模型結構圖[26] (a)Moore 模型 (b)Mealy 模型

而本研究使用的是 Moore 模型和 Mealy 模型的綜合體，亦即一部份的輸出依附於 state 之下，另一部份的輸出則同時考量 state 和 input 的情況。因為本系統以

運鏡種類為 state，所以運鏡種類的輸出可以交由 state 自行決定。而景別會受到不同運鏡變化的影響，在兩種不同的運鏡間轉換會使用不同的景別，因此景別輸出需要由 input 和 state 共同決定。

5.2.1 ACMM 結構

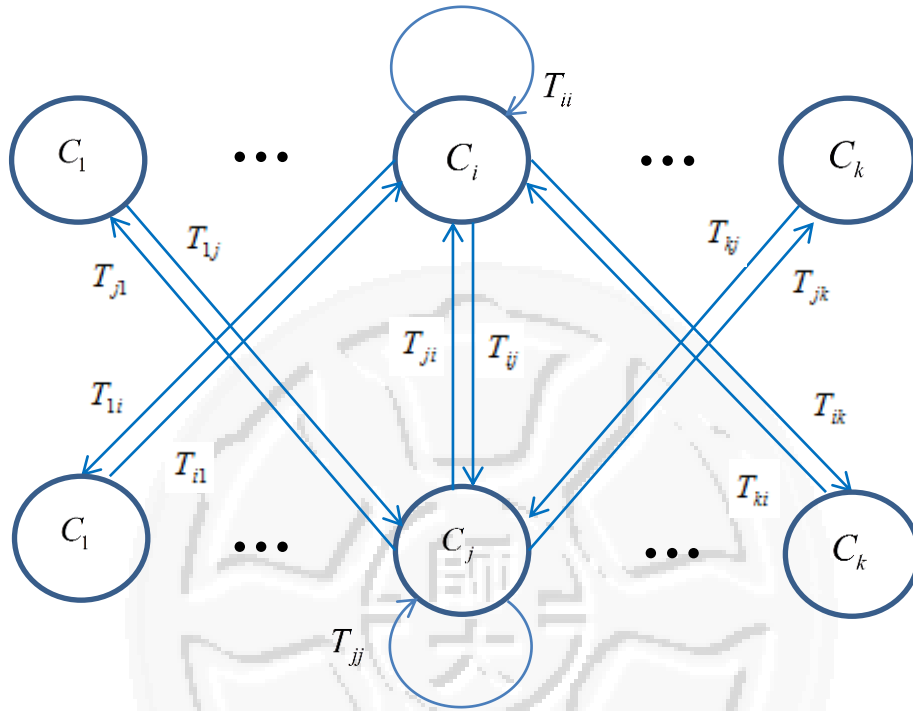


圖 5.3 自動運鏡系統結構圖

本研究提出的 ACCM 主要結構分為輸入(input)、運鏡狀態(camera movement states, CMS)以及轉移機率模型(transition probability model, TPM)，如圖5.3所示。圖中每一個圓形圖示代表一個 CMS，每一個箭頭符號皆代表一個 TPM。ACMM 中共有 k 個 CMS， $k \in N$ ，則每個 CMS 擁有 $k-1$ 個方向向外的 TPM 和 k 個方向向內的 TPM。ACMM 的 input 為前章所述的四種 fuzzy motion features，分別為 \tilde{m}_s 、 \tilde{m}_c 、 \tilde{m}_{se} 、 \tilde{m}_{oe} ，而 ACMM 的輸出為景別以及運鏡種類。前面曾經提到其輸出方式為 Moore 模型和 Mealy 模型的綜合體，系統將根據 CMS 輸出運鏡種類並由 CMS 和 input 同時決定景別。

CMS 為類似 FSM 中的 state，令 C_i 為 label 為 i 之 CMS， $1 \leq i \leq k$ ， $i \in N$ ， k 為運鏡的種類數量，每一個 C_i 都對應到不同的運鏡狀態。每次的運鏡特徵輸入都

會影響 CMS 的轉移，例如在某些特徵出現的情況下造成系統由 C_i 轉移到 C_j ，而後 C_j 又可能因為下一個輸入轉移到和 C_j 較為相關的 CMS，其中 $1 \leq i, j \leq k$ ， $i, j \in N$ 。藉由 CMS 之間的相互關係，當一段連續的現場特徵向量輸入 ACMM，系統將會依照該特徵序列挑選出與其最合適的連續運鏡。

TPM 是一個具有指向性，並存在於兩個 CMS 之間的機率模型。TPM 的實際結構為一個五維矩陣，四個維度分別對應四個 fuzzy motion features，每個維度大小取決於其對應的 fuzzy motion feature 的模糊集數量。矩陣中所記錄的值代表在不同輸入的情況下該 TPM 被選為 CMS 轉移路線的機率。令 $m_1 \sim m_4$ 分別為四個 fuzzy motion features 之模糊集數量， T_{ij} 為方向由 i 往 j 之 TPM， $1 \leq i, j \leq k$ ， $i, j \in N$ ， k 為 CMS 之數量，則 T_{ij} 矩陣中所對應 $m_1 \sim m_4$ 之機率值可表示成 $T_{ij}(m_1, m_2, m_3, m_4) = p_{ij}$ ， $p_{ij} \in [0, 1]$ 。

TPM 可用來記錄專業攝影師就觀察畫面中的 fuzzy motion feature 所作出對運鏡選擇的判斷經驗，每當系統需要切換運鏡，ACMM 即利用 TPM 得知在該 input 的情況下專業攝影師會如何選擇運鏡。本研究將攝影師以 fuzzy motion features 作為判斷運鏡依據的經驗經過分析和統計後放入 TPM 之中，其處理方式將在後面做說明。

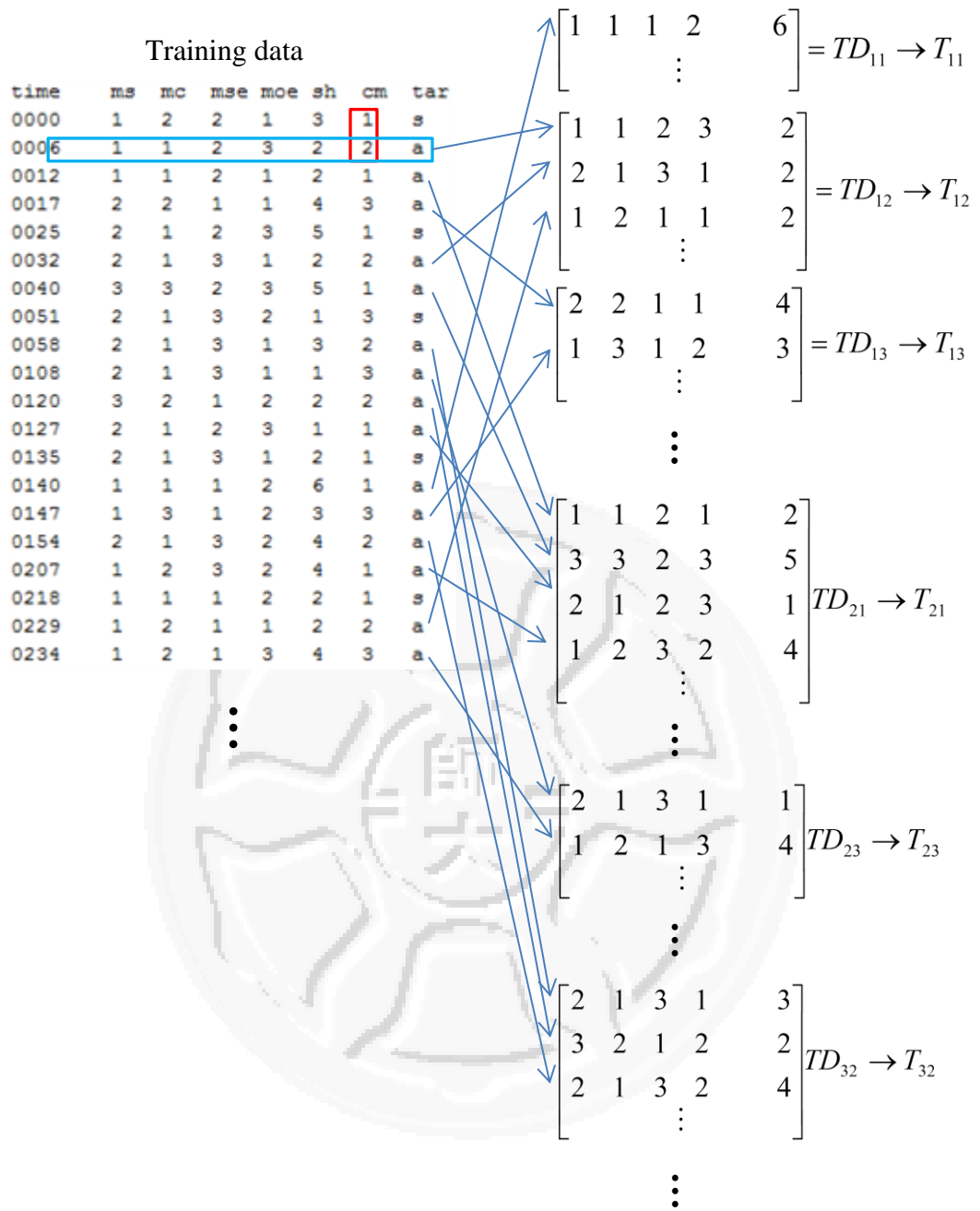


圖 5.4 Training data 分別轉換成各 TPM 之範例

TPM 的 training data 的形式如圖5.4所示，每一筆 training data 所記錄的資訊包含運鏡切換的時間(time)、四個 fuzzy motion features 的模糊值、使用的景別、運鏡以及拍攝主體屬於觀眾或講者。為方便記錄，除了時間以外的所有資料項都以數字或單一英文字母來表示，fuzzy motion features 以小於其模糊集數量的自然數表示其變數值，景別和運鏡則是不同數字代表運鏡或景別的類別，主體的部分

則以“a”代表觀眾、“s”代表講者。圖5.4的左方為原始 training data，每一筆資料由左至右包含運鏡發生時間(time)、 \tilde{m}_s 、 \tilde{m}_c 、 \tilde{m}_{se} 、 \tilde{m}_{oe} 、景別、運鏡以及主體類別，圖5.4的右方則是經過分析後並拆解後的結果。圖5.4左方的藍色框是為同一筆資料，紅色框的內容則為判斷該筆資料屬於哪個 TPM 的依據。紅色框所指的資料項為運鏡的種類，以紅色框內上半部的數字1作為起始點，下半部的數字1做為終點，所以藍色框中的資料屬於 T_{11} 的 training data。因此原始的 training data 被分類到 k^2 個資料矩陣 TD_{ij} 再分別整合至對應的 T_{ij}'' ， T_{ij}'' 為方向由 i 往 j 且未經正規化之 TPM，其中 $1 \leq i, j \leq k$ ， $i, j \in N$ ， k 為 CMS 之數量。

要注意的是，本研究因為以觀眾為拍攝主體，所以只會留下主體為觀眾的資料，每筆資料含有五個 fuzzy motion features 以及景別共六個資料項。在完成所有 training data 的分類後，接下來要將資料放入對應的 TPM 中進行統計。令 TD_{ij} 的四個 fuzzy motion features 資料項之值為 m_1 、 m_2 、 m_3 、 m_4 ， T_{ij}'' 為 TD_{ij} 所屬之 T_{ij}'' 在 t 時間的狀態，則每一筆輸入資料所對應到的 T_{ij}'' 之值的調整方式如下式所示：

$$T_{ij}''(m_1, m_2, m_3, m_4) = T_{ij}''(m_1, m_2, m_3, m_4) + 1$$

根據 TD_{ij} 中的 fuzzy motion feature 值作為一個四維索引，在 TPM 中找到對應矩陣元素並增加該元素之值。在 TD_{ij} 所有資料都輸入 T_{ij}'' 後， T_{ij}'' 中的每個元素值代表在每個元素的索引對應之 fuzzy motion feature 的情況下， T_{ij}'' 中被選為轉移方向的次數，亦可代表其被選中的可能性，但是在不同 TPM 之間這樣的作法其實並不公平。舉例來說，令兩個 TPM 分別為 T_{ij}'' 和 T_{ij}'' ，當 T_{ij}'' 中值的總和較小，可是卻集中在某個索引的元素上，這代表該元素對 T_{ij}'' 的重要性較高，而 T_{ij}'' 雖然在該索引的元素值較大，但佔 T_{ij}'' 值總和的比例卻較少，這代表該元素並不是選擇 T_{ij}'' 的

主要原因。為解決這個問題，接下來將對 T_{ij}^u 做正規化的處理，令 $1 \leq i, j \leq k$ ， $i, j \in N$ ， k 為 CMS 數量， m_1^n 、 m_2^n 、 m_3^n 、 m_4^n 分別為 m_1 、 m_2 、 m_3 、 m_4 的模糊集數量，則正規化後的 T_{ij} 計算方法如下：

$$T_{ij}(m_1, m_2, m_3, m_4) = \frac{T_{ij}^u(m_1, m_2, m_3, m_4)}{\sum_{m_1=1}^{m_1^n} \sum_{m_2=1}^{m_2^n} \sum_{m_3=1}^{m_3^n} \sum_{m_4=1}^{m_4^n} T_{ij}^u(m_1, m_2, m_3, m_4)}$$

總而言之， T_{ij} 即為 T_{ji}^u 中每個元素所存的值除以該 T_{ji}^u 中所有值之總和，因此 T_{ij} 中的每個元素的值代表該元素在整個 T_{ij} 中的重要性。

由於 TSM 經過正規化之後，其矩陣內部所記錄之機率值皆為 0 到 1 的連續實數。但經由統計輸入每個 TPM 的數量可以發現不同 TPM 被選為轉移方向的機率並不相同，當 TPM 之值總和較大代表該 TPM 被選中的機率較高，反之則較低，因此每個 TPM 的重要性有差異存在。為解決此問題，本研究以 weight map 來記錄每個 TPM 的權重，令 W_{ij} 為其對應 T_{ij} 的權重， $1 \leq i, j \leq k$ ， $i, j \in N$ ， k 為 CMS 數量，則

$$W_{map} = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1k} \\ W_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ W_{k1} & \cdots & \cdots & W_{kk} \end{bmatrix}$$

上式中 W_{map} 為所有權重值之集合， W_{ij} 之值為屬於 T_{ij} 的 training data TD_{ij} 的資料筆數。

5.2.2 ACMM 運作流程：

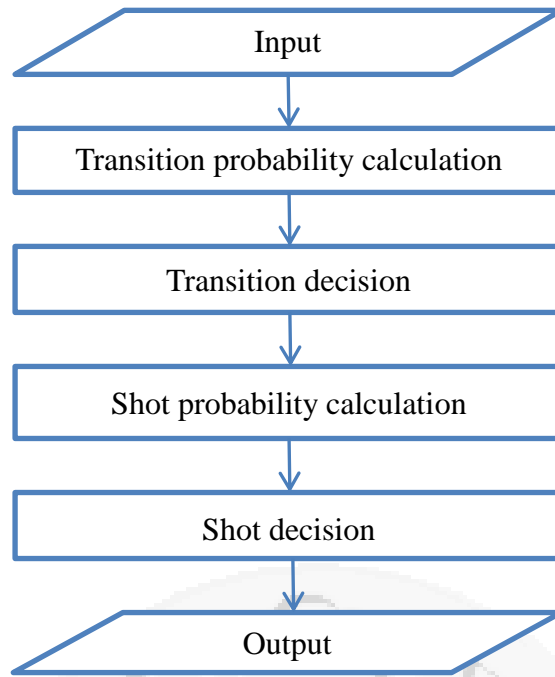


圖 5.5 ACMM 運作流程圖

圖5.5為 ACMM 運作流程圖，系統判定需要運鏡轉換後，會將當前所擷取到的 fuzzy motion features 輸入自動運鏡模組。令 C_{now} 為當前 CMS， TPM_{now} 為以 C_{now} 為起始、方向為離開 C_{now} 的 TPM 集合。這些輸入值會代入 TPM_{now} 的每個 TPM 矩陣之中，並在對應 input 的座標找出前往每一個 CMS 的機率。若 T_{ij} 是方向由 C_i 往 C_j 的 TPM ($1 \leq j \leq k$ ， $j \in N$ ， k 為 CMS 的數量)， i 為 C_{now} 的 label，label 代表 CMS 所代表之運鏡種類的編號， $m_1 \sim m_5$ 為五個作為輸入的 fuzzy motion features，則機率 P_{ij} 產生公式如下所示：

$$P_{ij}(m_1, m_2, m_3, m_4) = T_{ij}(m_1, m_2, m_3, m_4)$$

P_{ij} 為由 C_i 前往 C_j 的機率函數，所得結果為由 C_i 前往 C_j 的機率。另外，由於每個 TPM 具有不同的權重，因此將每個 P_{ij} 乘上 T_{ij} 的權重值 W_{ij} ，如下式所示：

$$P_{ij}^w(m_1, m_2, m_3, m_4) = W_{ij} \times P_{ij}(m_1, m_2, m_3, m_4)$$

P_{ij}^w 為加權過後的轉移機率函數，所得結果為加權過後的轉移機率。

在得到全部的經過加權的轉移機率後，系統會從中挑選機率最高者作為轉移路線並得到下一個運鏡狀態 C_{next} 的 label。令 i 為 C_{now} 的 label， C_{next} 的 label 之計算公式如下：

$$\text{label of } C_{next} = \arg \max_j (P_{ij}^w)$$

其中 $1 \leq j \leq k$ ， $j \in N$ ， k 為 CMS 的數量。決定 C_{next} 之後代表系統的當前 CMS C_{now} 將轉移到 C_{next} ，並決定最後的運鏡種類輸出值。運鏡輸出值的決定方式如同 Moore 模型是由 CMS 所決定，輸出值為 C_{next} 的 label 所代表之運鏡種類，同時系統之 state 由 C_{now} 轉移至 C_{next} 。

第三節 景別決定(shot decision)

ACMM 的另一個輸出是景別(shot)，景別指的是主體在畫面中所佔的比例大小，不同的運鏡會搭配不同的景別以呈現該運鏡的最佳效果。景別機率模型(shot probability model, SPM)以 fuzzy motion features 為輸入，並輸出每個景別被挑選的機率，該模型依附於每個 TPM 之下。一旦 state C_{next} 被決定，被選為轉移路線的 TPM 之下的 SPM 將會決定並輸出景別の種類。SPM 的結構為一維陣列，其陣列大小為景別種類の數量，令 $s_1 \sim s_n$ 為第 1~n 種景別發生的機率， n 為景別種類之數量， k 為 CMS 數量，則依附於 T_{ij} 之下的 SPM S_{ij} 結構為：

$$S_{ij} = [s_1, s_2, s_3, \dots, s_n], \quad 1 \leq i, j \leq k, \quad i, j \in N。$$

SPM 之 training data 與所依附的 TPM 相同，並使用其中的景別(shot)資料項 s ，令 t 為當前時間，則未經正規化的 SPM S_{ij}^u 的調整過程為：

$${}^t S_{ij}^u[s] = {}^{t-1} S_{ij}^u[s] + 1$$

上式中顯示當 training data 的 shot 資料項為 s ，則 SPM 的第 s 項元素加 1。接著對 SPM 進行正規化以便於觀察，若 z 為景別種類の數量，則正規化過程如下：

$${}^t S_{ij}[s] = \frac{S_{ij}^u[s]}{\sum_{z=1}^n S_{ij}^u[z]}$$

其中 $1 \leq s \leq n$, $s \in N$ 。上式中顯示所謂的正規化是指 SPM 中每個元素都除以陣列中所有元素值的和。由於 SPM 附屬於 TPM 之下，當所屬的 TPM 被選為轉移路線時才會啟用，並沒有和其他 SPM 相比較，因此不需要被賦予不同的權重。

SPM 決定景別的方法為以矩陣內每個元素值的大小作為該元素被選取的機率，這是為了讓景別更富變化性，令 R 為一個下限為 0，上限為 $S_{ij}[s]$ 之隨機變數產生函數，則下一個景別 S_{next} 的產生方式如下式所示：

$$S_{next} = \arg \max_s R(S_{ij}[s])$$

由這個方法產生景別的不會永遠是有最高機率值的景別種類，因此讓其他景別依照其使用頻率而仍有被選取的機會。至 S_{next} 被決定並輸出至下一個步驟，ACMM 的流程才算完成。

第六章 主體選取與運鏡路線規劃

第一節 主體候選者產生

主體是攝影師所要呈現的目標物，同時也是畫面的結構重心。本系統為以觀眾為拍攝主體之虛體攝影師系統，拍攝主體以觀眾為主。由於臉部是大多是攝影師拍攝人像時最為關注的部位，因此若以觀眾作為拍攝的主題，臉部的資訊是相當重要。同時臉部也是對觀眾席上的觀眾較容易完整擷取的部分，在觀眾席上觀眾多坐在座位上，大部分的座位排列都會使得後排座位被前排座位遮蔽，因此不易偵測到全身或是肩膀以下的部位特徵。另外，由於輸入影像不會有上下翻轉的情形發生，觀眾臉部的畫面一定在身體的畫面之下，根據 Chiref 等人的研究 [Che07] 顯示，人臉部的長度與頭頂至身體軀幹的各個位置有固定比例關係存在。綜合以上兩點，本研究先擷取觀眾臉部的區域，再以該區域的高度作為臉部的長度就可以換算出該主體觀眾的身體各部位的相對位置。

人臉的偵測方法已經被研究多時，Viola 等人 [Vio04] 提出了穩定且能夠 real-time 執行的人臉偵測技術，本系統採用該技術來偵測畫面中的人臉。Viola 等人提出的技術分為三個主要的部份，第一個是 integral image，第二個是 adaptive boost，第三個是 cascade classifier。

Integral image 是一種改良自 Harr feature 的特徵擷取方式，Harr feature 使用如圖 6.1 (a) 的檢測窗格中指定位置的相鄰矩形，計算每一個矩形的 intensity 值之和，以這些值的差值作為特徵。Harr feature 的缺點就是計算時間過長，integral image 即為解決此問題的方法。Integral image 為一個二維矩陣的查找表，該矩陣的每一個元素值為左上方矩形內 intensity 值之和。如圖 6.1(b) 中座標點 (x,y) 之值左灰色部分區域內 intensity 值之和。想得到矩形間 intensity 值之和的差值只需要從

integral image 中查詢端點之值進行運算，藉此可以大幅度地降低運算時間。

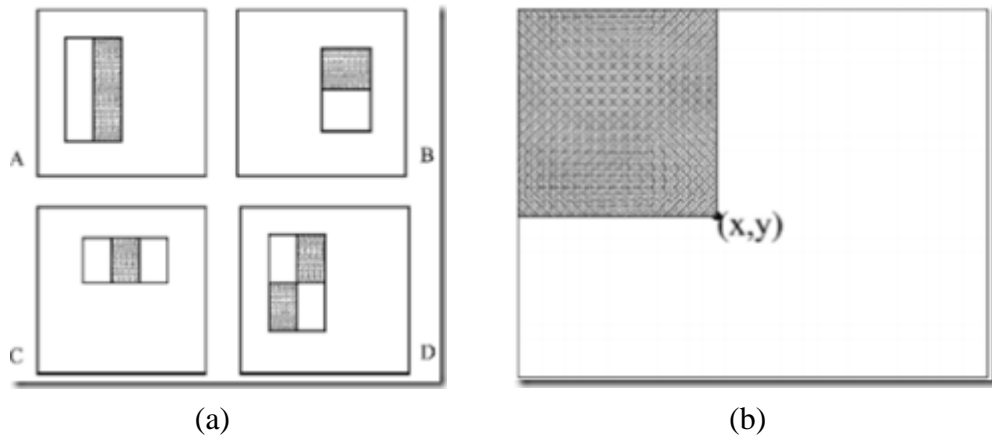


圖6.1 Integral image 與 Harr feature 示意圖[Vio04] (a)Harr features (b) integral image

Adaptive boost 演算法的構想為對同一組 training set 訓練出多個不同的 classifiers，而單一個這類的 classifier 為弱分類器，意即辨識正確率較低的分類器，然後以這些弱分類器構築成一個辨識能力較高的強分類器。首先對每筆測資給予一個權重值，接著對以 Harr feature 為特徵之弱分類器進行訓練之後得到每筆測資是否辨識正確以及總體辨識正確率。根據這些資訊進行每筆測茲的權重更改，若辨識正確則該筆測資的權重不變，若錯誤則提高該筆測資權重，以增加該筆資料的重要性。更改權重後再次將測資放入原弱分類器進行訓練，重複數次後即完成該弱分類器調整。最後將不同 Harr feature 訓練得到的弱分類器串聯起來成為強分類器，並以內部弱分類器投票表決的方式決定輸入圖像是否為人臉。

Cascade classifier 的概念如圖6.2所示，首先將 Harr features 分為數類，並對每一類的 Harr features 訓練各自的 classifier。最前方的 classifier 的辨識能力最低，但仍能刪去大部分非臉的圖像，未被刪去的圖像則交給後面的 classifier。越往後的 classifier 辨識能力越強，即使要付出較高的時間或運算成本，也因為在前面的 classifier 已經刪去大部分的圖像，總成本仍是較低的。經過所有的 classifier 篩選後，最後留下的圖像就是臉部的圖像。圖6.2中的圓圈1、2、3個代表一個 adaptive boost 分類器，其使用的 Harr features 複雜程度以及分類器的辨識率由1、2、3依

序提升,藉此先以前方的分類器濾掉多數能以簡單 Harr feartures 為特徵之 adaptive boost 分類器辨識的影像,後面的分類器則以較複雜 Harr feartures 過濾較難辨識的影像。

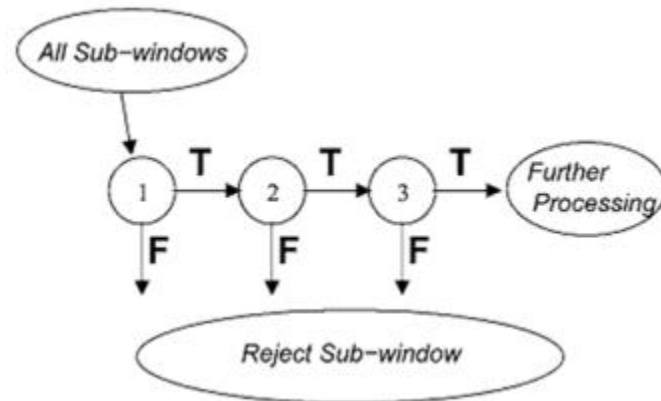


圖6.2 Cascade classifier 概念圖[Vio04]

在完成人臉的偵測後,系統記錄每一個人臉的中心座標、高度以及寬度以便作為主體挑選之用,每一個人臉都是一個主體候選者。本系統採用 Intel(英特爾)公司開發的影像函式庫 opencv 中所提供的,根據 Viola 等人的研究實作之 face detection 函式做為本系統偵測臉部的之用。

第二節 主體挑選

當主體候選者產生後,接下來要從中挑選最合適拍攝主體。本研究以五個標準對每個主體候選者做評比,這五個標準分別為運動鄰近程度(motion field neighborhood)、構圖(composition)、亮度(luminance)、畫面連續性(continuous)以及畫面重複性(repetition)。運動鄰近程度代表被作為主體候選者的觀眾做出動作的可能性;構圖指的是主體候選者是否符合攝影學中常用的畫面結構。畫面的亮度會影響觀賞者對主體的辨識程度,因此利用該評比項目來降低主體亮度過低的主體候選者。另外,不斷重複的畫面會讓影片變的單調,所以影片的重複性亦為評比的標準之一。以下將就這五個評比標準作介紹:

6.2.1 運動區域鄰近程度(motion field neighborhood)

由於拍攝主體候選者多為觀眾的臉部，且在觀眾席上產生 motion 的物體多為觀眾所造成，所以當主體候選者附近有 motion 產生代表該觀眾可能正在做出某種動作。動作中的觀眾相較於其他觀眾更容易被注意，一般在演講廳的觀眾的動作並不多，若有觀眾做出動作很可能是舉手或是需要被其他人注意的情況。運動區域鄰近程度即為用來計算該主體候選者做出動作可能性，首先對運動能量歷史影像(MEHI)做二元化處理，以凸顯有 motion 產生的 pixels。接著對其進行 connected component labeling，此步驟可以把影像中相連且 intensity 值不為0且相連的 pixel 標記為同一個區塊。當每一個產生 motion 的區塊都被標記為不同的 label，接著對每個不同 label 的區塊計算其重心座標，並計算主體候選者的中心座標和每一個 motion 區塊的中心座標的距離。當距離越小代表該觀眾做出動作的可能性越高，應要在此項評比中獲得較高的分數。令 C_s 為主體候選者的中心座標， C_m^i 為 label 為 i 的 motion 區塊中心座標， $0 \leq i \leq n$ ， $i \in N$ ， n 為 motion 區塊的數量， dis 為兩點間的幾何距離函數，則主體候選者的運動區域鄰近程度分數為 A_n ：

$$A_n = \min(dis(C_s, C_m^i))$$

6.2.2 構圖(composition)

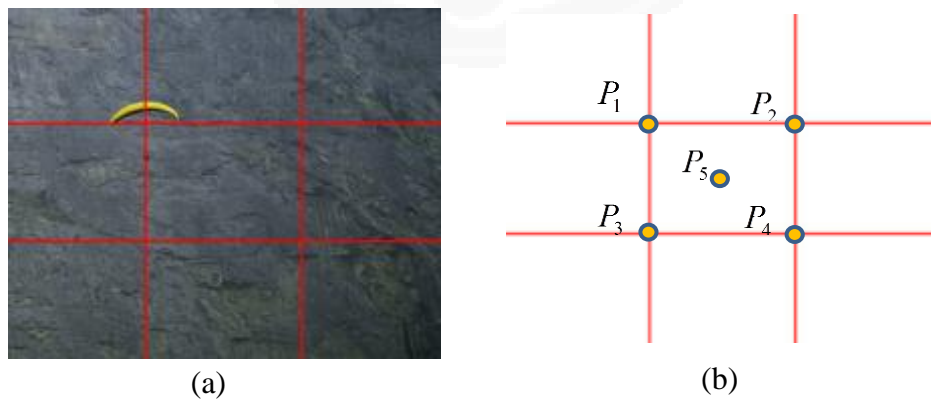


圖6.3 井字構圖法與井字構圖遮罩 (a)井字構圖法 (b)井字構圖遮罩

畫面的構圖是一個較為抽象的攝影詞彙，凡以畫面的結構、光影、顏色等構

成畫面的要素達到畫面的平衡，都能夠稱作構圖。本研究所探討的構圖是以達到畫面中結構的配置平衡為訴求，結構的配置意指畫面中物體的位置分配，或是背景層次的呈現。該種類的構圖所常用的構圖方法有「井字構圖法」、「三分構圖法」以及「對角線構圖法」，雖然構圖並沒有實際的標準或規則，這些構圖方法為學理上的基礎構圖法，若符合該構圖法的規則仍具有一定的公信力。

本系統以最為常用的「井字構圖法」(rule of thirds, 圖6.3(a))作為構圖評比的標準，首先建立井字格線遮罩 M_{rot} (如圖6.3(b))，該遮罩的大小以 ACMM 輸出的景別做調整。接著將遮罩中的點 P_1 和主體候選者中心點 C_s 對齊，然後對遮罩內的影像作 line detection [Ham99] 處理。將得到的直線依照斜率分為水平、垂直左45度角和右45度角四類，留下垂直和水平兩類的直線，並計算遮罩的井字標準線到畫面中最近的同類斜率直線之距離。例如令距離標準線 L_1 的水平類直線為 L_2 ，如圖6.4所示，則 L_1 和 L_2 的距離定義為到點 U_0 的到 L_2 的距離 $d(L_2, U_0)$ ，點 U_0 為 L_1 標準線和其中垂線的交點。將四條遮罩標準線使以同樣方法計算該距離，並相加得到構圖評比 A_{cp} 。

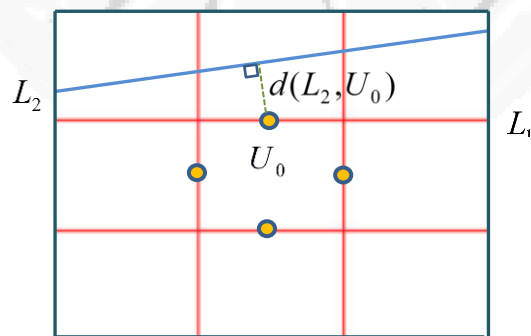


圖 6.4 畫面中直線與井字遮罩標準線距離示意圖

6.2.3 亮度(luminance)

畫面的亮度會影響畫面中景物的清晰度以及觀看的舒適度，如果畫面亮度太高會使得景物變得模糊，而且給人刺眼的感覺。若畫面亮度太低會讓景物難以辨識，並會使畫面的顏色變得暗沉而不鮮麗。為避免主體在畫面中所在的位置有亮

度過高或過低，本研究在 HSV 色彩空間中做出亮度的限制，在主體候選者的區域中不符該限制的 pixels 越多則亮度評比越低。

HSV 色彩空間的三個維度分別是 hue(色調)、saturation(色度)和 value(亮度)，如圖6.5(a)所示。令 R、G、B 分別為 RGB 色彩空間中的紅色、綠色、藍色的 intensity 值，則其經由 RGB 色彩空間轉換成 HSV 色彩空間的轉換公式如下：

$$H = \begin{cases} \frac{G - B}{\max(R, G, B) - \min(R, G, B)} \times 60 & \text{if } \max(R, G, B) = R \\ \frac{B - R}{\max(R, G, B) - \min(R, G, B)} \times 60 + 120 & \text{if } \max(R, G, B) = G \\ \frac{R - G}{\max(R, G, B) - \min(R, G, B)} \times 60 + 240 & \text{if } \max(R, G, B) = B \end{cases}$$

$$S = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)}$$

$$V = \max(R, G, B)$$

相較於傳統 RGB 色彩空間，HSV 將亮度的維度獨立出來較符合本研究對亮度做出限制的需求，圖6.5(b)為一張彩色影像及其 V 值影像。在 HSV 色彩空間中，V 值約在50~70之間是辨識物體最為清楚的明亮程度[李04][周12]，因此將 V 空間分成數個區間(如表6.1所示)，一個 pixel 的 V 值所處的區間越靠近50~70區間的亮度評比越高，接著統計主體候選者區域內每個 pixel 的亮度評比總和作為該主體候選的亮度評比 A_l 。

表6.1 Value 值與亮度評比對照表

Value 值	0~20	20~50	50~70	70~120	120~200	200~255
亮度評比	0	0.5	1	0.5	0.25	0

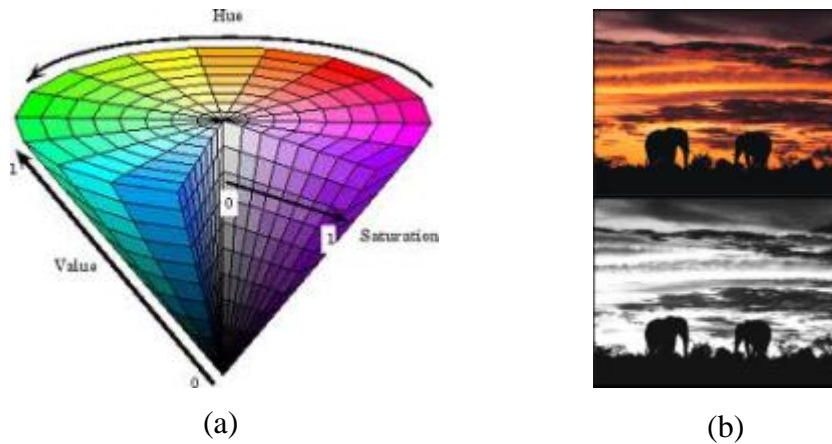


圖6.5 HSV 色彩模型 (a)模型示意圖[27] (b)上方為原始彩色影像，下方為 V 值亮度影像[28]

6.2.4 畫面連續性(continuous)

在運鏡切換的過程中，若鏡頭的移動幅度較大會造成前後兩個運鏡的畫面差異較大，造成畫面間不連續的情況，該情況會使得畫面突兀且會失去和運鏡間的關聯性。解決此問題的方法為減少攝影機在兩個運鏡之間過大幅度的移動，因此當一個運鏡結束後，運鏡 PTZ 攝影機的鏡頭所停留的位置和下一個運鏡的主體候選者之間的距離將作為畫面連續性的評比標準。令 (x, y) 為主體候選者在全景 PTZ 攝影機的畫面中的中心座標，全景 PTZ 攝影機的畫面中心座標為 $(0,0)$ ， ε 為避免分母為 0 的常數，則主體候選者的畫面連續性評比 A_{cn} 為：

$$A_{cn} = \frac{1}{\sqrt{x^2 + y^2 + \varepsilon}}$$

當該主體候選者的中心距離畫面中心越近，代表運鏡 PTZ 攝影機在連續兩個運鏡間的移動幅度越小，因此在此項評比中的分數越高。

重複性(repetition)

當一個主體在短時間內重複被選取到會讓觀看者有時間錯亂的感覺，尤其演講廳的環境並不容易有大幅度的改變，若選到同樣的主體極可能使得目前的畫面和先前的畫面極為相像。為避免畫面重複的狀況，本研究以主體歷史分布圖

(subject history map, SHM)降低此情況發生的可能性。主體歷史分布圖為大小和原輸入影像大小相同的灰階影像，該影像中每個 pixel 的 intensity 初始值為0。令 I 為原始輸入影像， t 為當前時間， τ 為灰階影像的 intensity 值之上限， a_s 為主體在 I 中所涵蓋的區域， $1 \leq x \leq W$ ， $1 \leq y \leq H$ ， $x, y \in N$ ，則 SHM 更新方法如下：

$$SHM_t(x, y) = \begin{cases} \tau & \text{if } (x, y) \in a_s \\ d \times SHM_{t-1}(x, y) & \text{otherwise} \end{cases}$$

每當主體產生，系統會將主體在原影像所涵蓋區域的所有座標點在 SHM 中同樣座標的 pixel 值提升到 τ ，並將 SHM 上非主體區域的 pixel 值乘上衰減常數 d 。圖6.6為一組連續時間的 SHM 的執行結果，其中圖6.6(a)中的主體挑選的結果，圖6.6(b)為 SHM 的執行結果，SHM 的 intensity 值越高的部分代表主體剛出現在這個區域。因此在下一次進行主體挑選時，若該主體候選者位於 SHM 的 intensity 值較高的區域，則在重複性評比的得分較低，藉此降低前一個主體再度出現的可能性。令主體候選區域的左上與右下座標分別為 (w_1, h_1) 與 (w_2, h_2) ， (x, y) 為主體候選者區域內的座標點， $w_1 \leq x \leq w_2$ ， $h_1 \leq y \leq h_2$ ，則該主體候選區域的重複性評比 A_r 計算方式為：

$$A_{r_i} = - \sum_{x=w_1}^{w_2} \sum_{y=h_1}^{h_2} SHM(x, y)$$

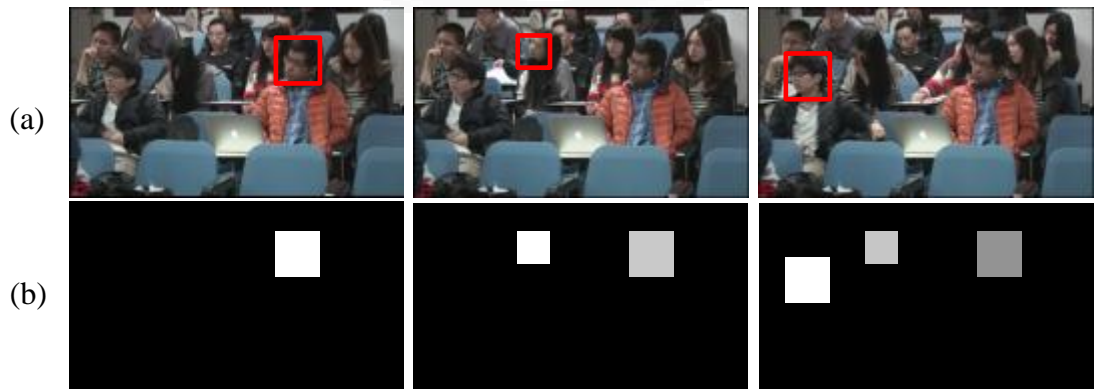


圖6.6 重複性評比範例 (a)主體挑選結果 (b)對應的 SHM

第三節 運鏡路線規劃

運鏡路線規劃提供運鏡 PTZ 攝影機執行運鏡動作時從開始到結束所需的一切資訊。不同的運鏡方式以及景別種類皆會改變運鏡路線的內容，這個步驟正是將系統在先前步驟得到的運鏡種類、景別種類、畫面中心位置和主體面積大小等資訊整合為實際執行運鏡所需要的資訊，並讓運鏡 PTZ 攝影機執行運鏡動作。運鏡路線包含運鏡起點、運鏡終點、伸縮鏡頭以及景別設置，運鏡起點意指運鏡 PTZ 攝影機開始運鏡動作的位置；終點為運鏡動作結束的位置；伸縮鏡頭代表利用焦距的變化讓畫面的放大或縮小；景別的设置則是利用鏡頭的伸縮將主體和畫面的比例調整到系統所決定的景別。

為確保影片的豐富性，本系統共使用11種運鏡手法以及7種景別呈現運鏡動作。其中運鏡手法可以分為兩大類，第一類是 pan、tilt 等轉動式的運鏡，這類運鏡需要鏡頭的轉動來呈現其效果。第二類是 zoom in、zoom out 等利用伸縮鏡頭強調主體來強調主體的運鏡手法。兩類運鏡在執行運鏡的步驟上以及所需的運鏡路線資訊有所差異，第一類運鏡的運鏡步驟如圖6.7所示，首先將攝影機的鏡頭轉動到運鏡起點，接著將畫面的景別調整為系統所決定的景別，然後將鏡頭逐漸轉動到終點的位置。

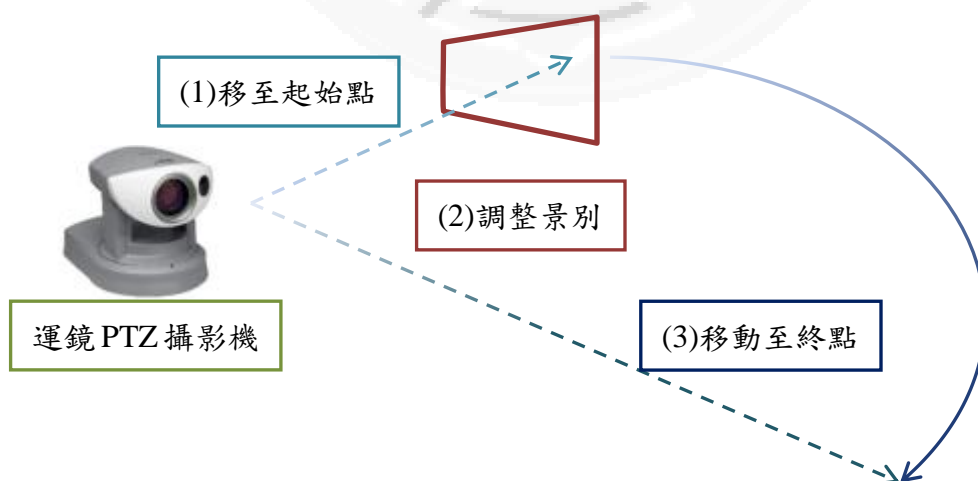


圖 6.7 第一類運鏡流程示意圖

第二類運鏡的運鏡步驟如圖6.8所示，首先將鏡頭移動到運鏡起點，接下來將

景別調整至系統決定的景別，然後進行鏡頭的推或拉的動作。第一類運鏡需要運鏡起點、運鏡終點和景別來完成運鏡動作，第二類運鏡則需要運鏡起點以及鏡頭推拉指令以及景別。以下將介紹如何以運鏡種類、景別種類、畫面中心座標及主體面積計算這兩類運鏡路線。

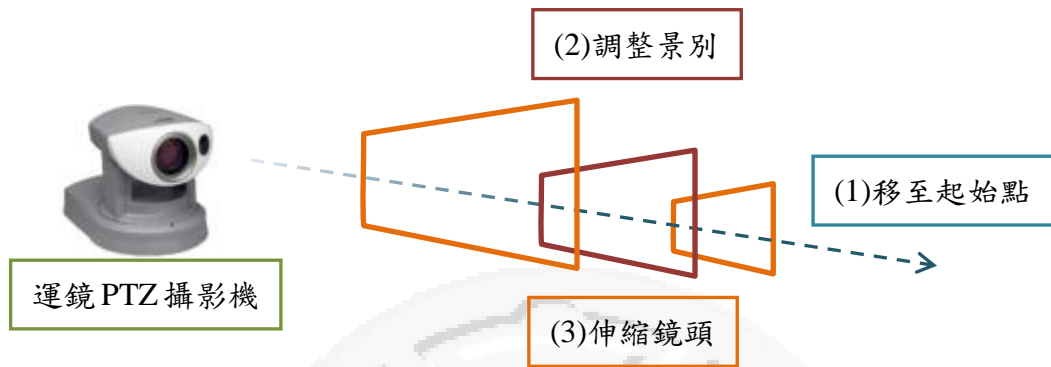


圖 6.8 第二類運鏡流程示意圖

運鏡起點與運鏡終點的座標是由上一節得到的畫面中心座標、主體大小以及 ACMM 輸出的運鏡種類來決定。為確保主體在運鏡的過程中能完整展現，起點與終點間的連線必須通過畫面中心座標，以避免偏離系統決定的拍攝位置。

另外，不同的運鏡種類會使得兩點的相對位置不同，例如由左向右橫移的運鏡，運鏡起點會在畫面中心的左邊，運鏡終點在右邊；若是由下而上的運鏡，運鏡起點會在畫面中心的下方，而運鏡終點則在上方。接下來將以運鏡種類做分類來介紹運鏡起點和終點的計算方式，令畫面中心為 C ，座標為 (C_x, C_y) ， H_s 為主體高度、 W_s 為主體寬度，則運鏡方式說明如下：

1. 第一類運鏡(cut、zoom in、zoom out)：第一類運鏡只需要把運鏡 PTZ 攝影機移動到畫面中心的位置，因為在運鏡過程中不需要移動鏡頭，不需要運鏡終點的資訊，該類運鏡的運鏡起點 (Ps_x, Ps_y) 與終點 (Pe_x, Pe_y) 皆為畫面的中心位置 (C_x, C_y) 。
2. 第二類運鏡(pan、tilt)：第二類運鏡會在運鏡過程中移動運鏡 PTZ 攝影的鏡頭，

因此運鏡起點與終點之計算方式會因為運鏡種類而不同。

- (a) 橫向運鏡(pan)：橫向運鏡可分為從左到右和從右到左兩種移動方式，兩種移動方式的運鏡起點運鏡終點互為相反，令 α 為系統設定之常數(設定為2)，橫向運鏡起點與終點計算方式如下：

由左到右運鏡之運鏡起點為 $(Ps_x, Ps_y) = (C_x - \alpha W_s, C_y)$ ，運鏡終點為 $(Pe_x, Pe_y) = (C_x + \alpha W_s, C_y)$ 。

由右到左運鏡之運鏡起點為 $(Ps_x, Ps_y) = (C_x + \alpha W_s, C_y)$ ，運鏡終點為 $(Pe_x, Pe_y) = (C_x - \alpha W_s, C_y)$ 。

- (b) 垂直運鏡(tilt)：垂直運鏡分為由下往上以及由上往下兩種移動方式，兩種移動方式的運鏡起點運鏡終點互為相反，令 α 為系統設定之常數(設定為2)，其運鏡起點與鐘點計算方式如下：

由下向上運鏡之運鏡起點為 $(Ps_x, Ps_y) = (C_x, C_y + \alpha H_s)$ ，運鏡終點為 $(Pe_x, Pe_y) = (C_x, C_y - \alpha H_s)$ 。

由上向下運鏡之運鏡起點為 $(Ps_x, Ps_y) = (C_x, C_y - \alpha H_s)$ ，運鏡終點為 $(Pe_x, Pe_y) = (C_x, C_y + \alpha H_s)$ 。

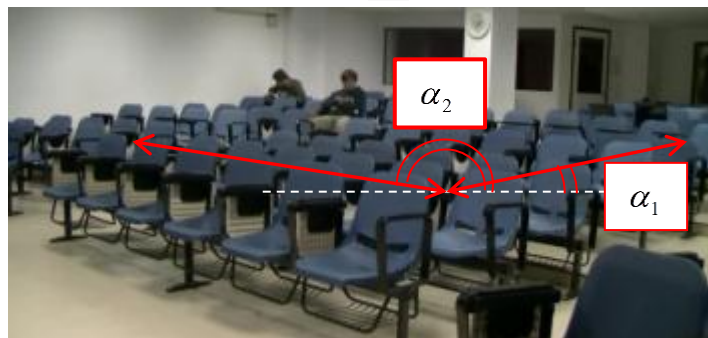


圖 6.9 沿觀眾席方向示意圖

- (c) 沿觀眾席方向運鏡(auditorium direction)：觀眾席方向指的是觀眾座位間走道的

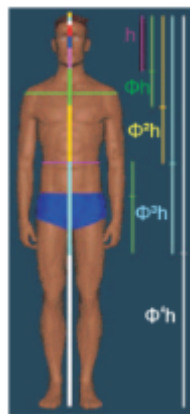
方向，如圖6.9中紅色箭頭所示。令圖中兩條直線和 0° 方向的夾角 a_1 、 a_2 ，首先對輸入影像的灰階影像 I 做 line detection 處理。接著將360度分為8個角度類別，統計每個類別中包含的直線數量。由於座椅的椅背和通道都會有明顯的 edges，偵測出該角度直線的可能性較高，因此挑選包含直線數量最多和次多的角度作為觀眾席方向和 0° 的兩個夾角，其中夾角較小者為 a_1 ，較大者為 a_2 ，則運鏡起點和終點計算方式如下：

沿觀眾席方向 a_1 (由左至右)的運鏡起點為 $(Ps_x, Ps_y) = (C_x, C_y - \alpha_1 H_s)$ ，運鏡終點為 $(Pe_x, Pe_y) = (C_x, C_y + \alpha_1 H_s)$ 。

沿觀眾席方向 a_1 (由右至左)的運鏡起點為 $(Ps_x, Ps_y) = (C_x, C_y + \alpha_1 H_s)$ ，運鏡終點為 $(Pe_x, Pe_y) = (C_x, C_y - \alpha_1 H_s)$ 。

沿觀眾席方向 a_2 (由左至右)的運鏡起點為 $(Ps_x, Ps_y) = (C_x, C_y - \alpha_2 H_s)$ ，運鏡終點為 $(Pe_x, Pe_y) = (C_x, C_y + \alpha_2 H_s)$ 。

沿觀眾席方向 a_2 (由右至左)的運鏡起點為 $(Ps_x, Ps_y) = (C_x, C_y + \alpha_2 H_s)$ ，運鏡終點為 $(Pe_x, Pe_y) = (C_x, C_y - \alpha_2 H_s)$ 。



(a)



(b)

圖6.10 景別特徵[Che07] (a)臉部高度與身體各部位長度關係圖 (b)臉部高度與臉部下緣至畫面下緣示意圖

景別是由主體與畫面的比例所決定的，Cherif 等人[Che07]以主體高度與主體下緣至畫面下緣的距離比例做為用來判斷景別種類的特徵，該研究的主體皆為人的臉部，臉部的高度 h 和臉部上緣至身體各部位的距離 d 有一定的比例關係，如圖6.10(a)所示，圖中 $\Phi^1 \sim \Phi^4$ 為代表 h 倍率的常數。因此主體下緣至畫面下緣的距離代表人臉以下的部位長度 (圖6.10(b))，由臉部的高度和該距離加上臉部高度的比例可以知道畫面的底端在人的什麼部位。若是畫面中底部截到影像中主體的肩膀部位，則該景別應該為肩上景(特寫)，如果截到的部分是在影像中的膝蓋部位，則應該為膝上景(中景)。本研究沿用 Cherif 等人的景別分類方式，並依照觀眾在座位上的坐姿設定倍率常數。

鏡頭的伸縮是一個讓運鏡 PTZ 攝影機做出鏡頭推拉的動作，當運鏡種類為推鏡頭(zoom in)或拉鏡頭(zoom out)時該變數才會對運鏡動作有影響。而鏡頭伸縮的速度取決於景別の種類，當主體在畫面中所佔的比例較大則伸縮速度較慢，可避免畫面變化過快。反之則鏡頭伸縮的速度較快，可讓觀賞者能感覺到主體在畫面比例的變動。

第七章 實驗結果與討論

第一節 系統設定

本系統使用11種運鏡以及7種景別來構築 ACMM。7種景別分別為大遠景、遠景、全景、中景、近景、特寫以及大特寫。11種運鏡方式分別為定點攝影、拉近、拉遠、向左橫移、向右橫移、垂直向上、垂直向下以及沿觀眾席角度1向左、沿觀眾席角度1向右、沿觀眾席角度2向左、沿觀眾席角度2向右。

系統的training data是以表7.1所示之六部演講影片所擷取出來的fuzzy motion features 以及所使用的景別與運鏡。擷取方式是以人工擷取，因此 fuzzy motion strength entropy 以及 fuzzy motion orientation entropy 由於人工判斷上的困難因此只擷取 fuzzy motion strength (\tilde{m}_s)和 fuzzy motion concentration (\tilde{m}_c)兩種特徵，所以系統目前只使用這兩種特徵進行訓練。

表7.1 Training set 介紹

編號	影片名稱	影片長度 (小時：分：秒)	切換運鏡次數	平均運鏡持續時間(秒)
1	丹佐·華盛頓為2011年賓州大學畢業生演講	00:22:44	96	14.2
	攝影機數量：多台		環境：室外	
				
2	J.K.Rowling 給哈佛畢業生的致辭(中文字幕)	00:23:44	67	21.3
	攝影機數量：多台		環境：室外	

				
3	美國總統歐巴馬為2012年巴納德學院畢業生演講	00:30:46	21	87.9
	攝影機數量：多台		環境：室外	
				
4	Steve Jobs 賈伯斯史丹佛大學演講（完整版）	00:15:05	39	23.2
	攝影機數量：多台		環境：室外	
				
5	比爾·蓋茲：在哈佛大學畢業典禮上的演講	00:26:04	65	24.1
	攝影機數量：多台		環境：室外	
				
6	歐普拉為2008年史丹佛大學畢業生演講	00:29:55	85	21.1
	攝影機數量：多台		環境：室外	
				

從表7.1可以得知每部影片的運鏡持續時間分別為14.2秒、21.3秒、87.9秒、23.2秒、24.1秒、21.1秒，比起一般運鏡的持續時間4到6秒還要高出許多。其原因有兩種可能，第一是因為這些影片的運鏡都含有拍攝講者的部分，而在演講的場合畫面停留在講者身上的情況較為普遍。在拍攝講者期間為了讓畫面注意力集中在講者，所以切換運鏡的頻率較低導致運鏡持續時間的增加，其中編號3的影片此現象特別明顯。第二是掌鏡者並非專業攝影師或是人力的不足，未考慮或無法兼顧到運鏡持續時間的問題。

以上六部影片的拍攝環境有皆為室外環境，室內與室外的差異主要在光影的變化，而光影的改變對於運鏡的選取並無顯著差異，也不會影響運鏡方式的擷取，所以即使本系統實驗環境架設於室內仍然能夠以拍攝環境為室外的影片作為測資。另外，每個影片在一個畫面中所能拍攝的最大觀眾人數皆有差異，幾乎都有百人以上的規模。但是每次運鏡幾乎都是以10人以下的觀眾為主體做拍攝，也較少直接以大遠景拍攝全部觀眾的畫面。本系統因為主體偵測的限制，觀眾人數約在5到10人之間，仍然能夠保持多數影片挑選主體的習慣。攝影機的數量代表在影片中除了拍攝講者的攝影機之外，拍攝觀眾的攝影機的數量。本系統使用的六部影片所使用的攝影機數量皆在兩台以上，因為可以從影片中察覺某些運鏡的攝影機位置明顯不同，例如在觀眾席左側及右側，可能是同時在多個位置架設攝影機或是以人工現場移動位置。本系統的功能為多台攝影機的其中一台攝影機，若要表現多台攝影機的效果，則需要再多個位置擺設本系統，並搭配導播子系統進行剪輯。

藉由以上6組測資所建構之 ACMM 的內部結構如圖7.1所示，其中共有7個 CMS(最大11個)、27個 TPM(最大121個)以及27個 SPM(最大121個)。未被標示的 CMS 表示其從未成為當前 CMS_{now} ，未被標示 TPM 及 SPM 的所有元素值皆為 0，代表該 TPM 及所附屬之 SPM 從未被測資選取。

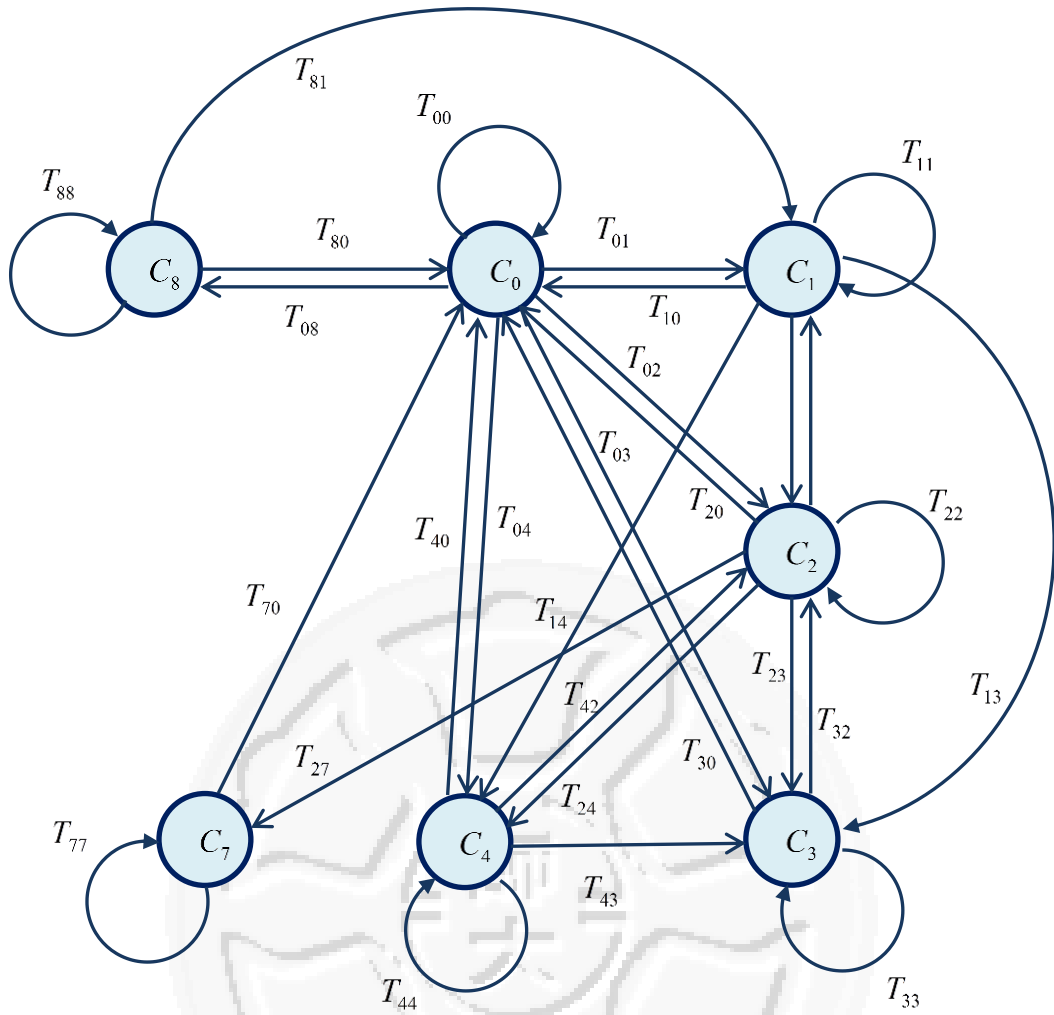


圖 7.1 經過訓練的 ACMM 模型

表 7.2 TPM 元素值列表

T_{00}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.224	0.107	0.341
	2	0.019	0.065	0.234
	3	0.005	0.005	0.000

T_{01}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.500
	2	0.000	0.227	0.227
	3	0.045	0.005	0.000

T_{02}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.115	0.077	0.154
	2	0.154	0.346	0.115
	3	0.000	0.038	0.000

T_{03}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.364	0.000	0.000
	2	0.182	0.273	0.000
	3	0.182	0.000	0.000

T_{04}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.333	0.000	0.000
	2	0.333	0.333	0.000
	3	0.000	0.000	0.000

T_{08}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	0.000	0.750	0.000
	3	0.250	0.000	0.000

T_{10}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.048	0.048	0.857
	2	0.000	0.000	0.048
	3	0.000	0.000	0.000

T_{11}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	0.000	1.000	0.000
	3	0.000	0.000	0.000

T_{12}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	0.000	1.000	0.000
	3	0.000	0.000	0.000

T_{13}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	0.000	1.000	0.000
	3	0.000	0.000	0.000

T_{14}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	0.000	1.000	0.000
	3	0.000	0.000	0.000

T_{20}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.080	0.120	0.560
	2	0.000	0.000	0.240
	3	0.000	0.000	0.000

T_{21}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	1.000
	2	0.000	0.000	0.000
	3	0.000	0.000	0.000

T_{22}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.500
	2	0.000	0.000	0.000
	3	0.500	0.000	0.000

T_{23}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	0.000	1.000	0.000
	3	0.000	0.000	0.000

T_{24}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	1.000	0.000	0.000
	3	0.000	0.000	0.000

T_{27}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	0.000	1.000	0.000
	3	0.000	0.000	0.000

T_{30}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.692
	2	0.077	0.000	0.231
	3	0.000	0.000	0.000

T_{32}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	0.000	1.000	0.000
	3	0.000	0.000	0.000

T_{33}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	1.000	0.000	0.000
	2	0.000	0.000	0.000
	3	0.000	0.000	0.000

T_{40}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.250
	2	0.250	0.250	0.000
	3	0.000	0.250	0.000

T_{42}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	1.000	0.000	0.000
	2	0.000	0.000	0.000
	3	0.000	0.000	0.000

T_{43}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	0.000	1.000	0.000
	3	0.000	0.000	0.000

T_{44}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	0.000	0.500	0.500
	3	0.000	0.000	0.000

T_{70}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	1.000
	2	0.000	0.000	0.000
	3	0.000	0.000	0.000

T_{80}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	1.000
	2	0.000	0.000	0.000
	3	0.000	0.000	0.000

T_{81}	\tilde{m}_c			
	1	2	3	
\tilde{m}_s	1	0.000	0.000	0.000
	2	1.000	0.000	0.000
	3	0.000	0.000	0.000

表7.2為 ACMM 的27個 TPM 中每個元素的數據，未列出的 TPM 拜表其所有元素值皆為0。表7.3為每個 TPM 的權重值列表，紅色代表權種植不為0的部分。

表7.3 TPM 權重值列表

W_{ij}		j										
		0	1	2	3	4	5	6	7	8	9	10
i	0	214	22	26	11	3	0	0	0	4	0	0
	1	21	1	2	1	1	0	0	0	0	0	0
	2	25	2	2	1	1	0	0	2	0	0	0
	3	13	0	1	1	0	0	0	0	0	0	0
	4	4	0	1	1	2	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0
	7	2	0	0	0	0	0	0	0	0	0	0
	8	3	1	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0	0

表7.4為表7.2所列出的 TPM 所附屬的 SPM 每個元素的數據，未列出的 SPM 其所有元素皆為0。

表7.4 SPM 元素值列表

選取機率	景別編號					
	0	1	2	3	4	5
S_{00}	0.042	0.103	0.112	0.089	0.640	0.014
S_{01}	0.000	0.182	0.091	0.045	0.682	0.000
S_{02}	0.077	0.077	0.077	0.192	0.538	0.038

S_{03}	0.091	0.182	0.000	0.182	0.545	0.000
S_{04}	0.667	0.000	0.000	0.000	0.333	0.000
S_{08}	0.000	0.250	0.000	0.000	0.750	0.000
S_{10}	0.048	0.000	0.286	0.000	0.667	0.000
S_{11}	0.000	0.000	0.000	0.000	1.000	0.000
S_{12}	0.500	0.000	0.000	0.500	0.000	0.000
S_{13}	0.000	0.000	0.000	0.000	1.000	0.000
S_{14}	0.000	1.000	0.000	0.000	0.000	0.000
S_{20}	0.040	0.040	0.080	0.240	0.600	0.000
S_{21}	0.000	0.000	0.000	0.500	0.500	0.000
S_{22}	0.000	0.500	0.000	0.000	0.500	0.000
S_{23}	0.000	1.000	0.000	0.000	0.000	0.000
S_{24}	1.000	0.000	0.000	0.000	0.000	0.000
S_{27}	0.000	0.500	0.000	0.000	0.500	0.000
S_{30}	0.077	0.000	0.077	0.077	0.769	0.000
S_{32}	0.000	1.000	0.000	0.000	0.000	0.000
S_{33}	1.000	0.000	0.000	0.000	0.000	0.000
S_{40}	0.000	0.000	0.500	0.000	0.500	0.000
S_{42}	0.000	0.000	0.000	1.000	0.000	0.000
S_{43}	0.000	0.000	0.000	1.000	0.000	0.000
S_{44}	0.000	0.000	0.000	1.000	0.000	0.000
S_{70}	0.000	0.000	0.000	0.000	1.000	0.000
S_{80}	0.000	0.000	0.667	0.000	0.333	0.000
S_{81}	0.000	0.000	1.000	0.000	0.000	0.000

第二節 系統運作範例

ACMM 的起始狀態設定為 C_0 ，假設系統在時間 t 偵測到的 fuzzy motion features 為 \tilde{m}'_s 、 \tilde{m}'_c 。令 $\tilde{m}'_s = 2$ 、 $\tilde{m}'_c = 2$ 、 $0 \leq j \leq k, k = 10, j \in N$ ，則 C_0 的轉移至其他 CMS 的轉移機率 P_{0j} 為表 7.5 紅色框部分為 T_{0j} 索引為 $[\tilde{m}'_s, \tilde{m}'_c]$ 的元素值。則 $P_{00} = 0.065$ 、 $P_{01} = 0.227$ 、 $P_{02} = 0.346$ 、 $P_{03} = 0.273$ 、 $P_{04} = 0.333$ 、 $P_{08} = 0.750$ ，其餘未列出之轉移機率皆為 0。

表 7.5 T_{0j} 元素值列表

T_{00}		\tilde{m}'_c			T_{01}		\tilde{m}'_c			T_{02}		\tilde{m}'_c		
		1	2	3			1	2	3			1	2	3
\tilde{m}'_s	1	0.224	0.107	0.341	\tilde{m}'_s	1	0.000	0.000	0.500	\tilde{m}'_s	1	0.115	0.077	0.154
	2	0.019	0.065	0.234		2	0.000	0.227	0.227		2	0.154	0.346	0.115
	3	0.005	0.005	0.000		3	0.045	0.005	0.000		3	0.000	0.038	0.000

T_{03}		\tilde{m}'_c			T_{04}		\tilde{m}'_c			T_{08}		\tilde{m}'_c		
		1	2	3			1	2	3			1	2	3
\tilde{m}'_s	1	0.364	0.000	0.000	\tilde{m}'_s	1	0.333	0.000	0.000	\tilde{m}'_s	1	0.000	0.000	0.000
	2	0.182	0.273	0.000		2	0.333	0.333	0.000		2	0.000	0.750	0.000
	3	0.182	0.000	0.000		3	0.000	0.000	0.000		3	0.250	0.000	0.000

表 7.6 T_{0j} 權重值表

W_{ij}		j										
		0	1	2	3	4	5	6	7	8	9	10
I	0	214	22	26	11	3	0	0	0	4	0	0

接下來將 P_{0j} 乘上 T_{0j} 的權重值 W_{0j} ，表 7.6 為 T_{0j} 之權重值表，則 P_{0j} 加權後的結果

P'_{0j} 計算的結果為 $P'_{00} = 13.91$ 、 $P'_{01} = 5.00$ 、 $P'_{02} = 9.00$ 、 $P'_{03} = 3.00$ 、 $P'_{04} = 1.00$ 、

$P'_{08} = 3.00$ 。然後對每個 P'_{0j} 計算最後的隨機轉移機率 P''_{0j} ，方法為從 $1 \sim P'_{0j}$ 隨機選

取一個正整數作為 P_{0j}'' 的值。令結果為 $P_{00}'' = 3$ 、 $P_{01}'' = 4$ 、 $P_{02}'' = 7$ 、 $P_{03}'' = 2$ 、 $P_{04}'' = 1$ 、 $P_{08}'' = 2$ ，其中最大值為 $P_{02}'' = 7$ ，則 T_{02} 被選為 C_0 在時間 t 轉移方向。

表 7.7 S_{02} 元素值表

選取機率	景別編號					
	0	1	2	3	4	5
S_{02}	0.077	0.077	0.077	0.192	0.538	0.038

決定轉移方向 T_{02} 後，接著以附屬於 T_{02} 之下的景別機率模型 S_{02} 輸出景別。表 7.7 為 S_{02} 的元素值表， S_{02} 的每個元素值代表該索引值代表的景別被選中的機率，例如定點攝影(景別編號 0)在 S_{02} 為被選中的機率為 $S_{02}[0] = 0.077$ 。假設景別編號 4 被選中，最後將時間 t 的 CMS 從 C_0 轉移至 C_2 (如圖 7.2 所示)，並由 C_2 輸出運鏡方法編號 2， S_{02} 輸出景別編號 4。

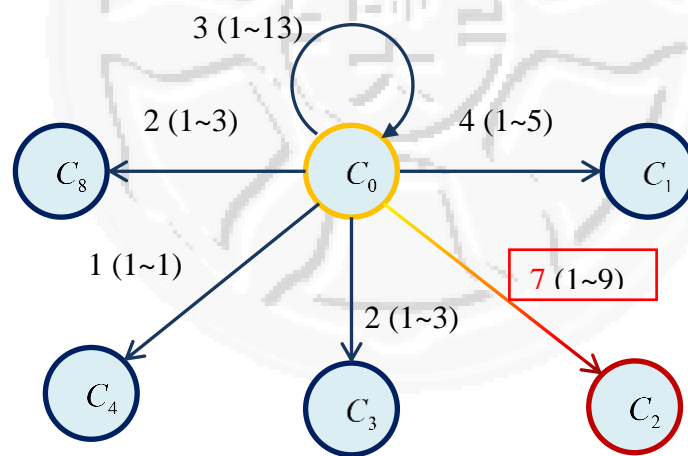


圖 7.2 CMS 轉移示意圖

第三節 自動拍攝結果與討論

以下實驗結果圖片皆分為(a)和(b)兩部分，(a)為全景 PTZ 所拍攝的畫面，時間順序為由上到下，紅色井字框為模擬運鏡 PTZ 攝影機的運鏡畫面，在井字框中

的紅色方框代表系統選出的主體，綠色的線條代表模擬的運鏡路線。(b)為運鏡 PTZ 攝影機實際拍攝結果，時間順序為由左到右。



圖7.3 實驗結果1(運鏡1~4) (a)模擬結果 (b)運鏡結果

圖7.3為實驗1的第1到4的運鏡結果，其中運鏡1為定點攝影，運鏡2同樣為定點攝影，運鏡3為向右水平移動，運鏡4為推鏡頭。運鏡2與運鏡3為同一個主體，由於重複性評比所佔的比重其實較其他美學評比為重，會發生重複選取同一個主體的原因可能是在主體候選者偵測的部分。當臉部偵測只偵測到一個候選者，同時畫面上沒有 motion 可以作為運動主體候選者時就可能造成重複選取主體的情況。



圖7.4 實驗結果1(運鏡5~9) (a)模擬結果 (b)運鏡結果

圖7.4為實驗1的第5到9的運鏡結果，其中運鏡5為水平向左移動，運鏡6為水平向右移動，運鏡7為定點攝影，運鏡8為定點攝影，運鏡9為定點攝影。運鏡5與運鏡6主體相同的問題原因如上一段所述，運鏡7由於最右邊的觀眾將身體前傾而產生 motion，所以被選為主體。由於中間和最右邊的觀眾在近期內被選作主體，所以造成重複性評比的降低，因此運鏡8和運鏡9在主體的選擇上避開者兩位觀眾。



(a)

(b)

圖7.5 實驗結果1(運鏡10~14) (a)模擬結果 (b)運鏡結果

圖7.5為實驗1的第10到14的運鏡結果，其中運鏡10為水平向右移動，運鏡11為定點攝影，運鏡12為定點攝影，運鏡13為水平向右移動，運鏡14為定點攝影。運鏡10當中雖然有觀眾揮手產生 motion，可是該觀眾及旁邊的觀眾的臉部未被偵測到，加上中間的觀眾才剛被選為主體，所以是最左邊的觀眾被選作主體。運鏡14中中間觀眾的衣服被錯誤偵測成臉部，加上附近有揮手所產生的 motion，因此被選為主體。



(a)

(b)

圖 7.6 實驗結果2(運鏡1~5) (a)模擬結果 (b)運鏡結果

圖 7.6 為實驗 2 的第 1 到 5 的運鏡結果，其中運鏡 1 到運鏡 5 皆為定點攝影。在運鏡 1 第三張圖可以看到當前主體揮手試圖成為下一個主體，而運鏡 2 的主體則成功避開選取重複的主體。在運鏡 4 中，畫面左方的主體候選者揮手產生 motion，此時該主體候選者的重複性評比已經恢復，因此被選為主體。在運鏡 5 中，右方的主體候選者揮手產生 motion，因此被選為主體。



(a)

(b)

圖7.7 實驗結果2(運鏡6~9) (a)模擬結果 (b)運鏡結果

圖 7.7 為實驗 2 的第 6 到 9 的運鏡結果，運鏡 6 為定點攝影，運鏡 7 為向右水平移動，運鏡 8 為定點攝影，運鏡 9 為定點攝影。在運鏡 6 與 7 中，由於畫面沒有面積足夠的 motion 區域影響主體的選取，因此藉由美學評比作為選取的依據。在運鏡 4 中，畫面右方的主體候選者產生 motion，但是該候選者的重複性評比較低，因此左方的主體候選者被選為主體。

第八章 結論與未來工作

第一節 結論

本研究的目的是為開發具專業攝影師能力之虛擬攝影師系統，系統能夠自動挑選合適的運鏡、景別與拍攝主體，並以 PTZ 攝影機執行運鏡動作。對專業的攝影師而言，一個完整的拍攝活動首先需要觀察拍攝環境，從環境中的資訊以及運鏡搭配的习惯判斷現在適合使用的運鏡，然後以美學觀點決定拍攝主體以及主體在畫面中擺放的位置。接著以主體為核心規劃運鏡路線，最後以手上的攝影機進行運鏡。本系統以三個步驟模擬上述攝影師的拍攝活動，分別為運動特徵擷取 (motion feature extraction)、運鏡判斷(camera movement decision)以及運鏡路線規劃 (path planning)。

在運動特徵擷取步驟中，本研究提出的四種代表觀眾行為的 motion features，為了讓 motion features 值能對應到攝影師以語言描述與表達的經驗法則，本研究對這四種特徵作模糊化處理得到 fuzzy motion features 後，利用這些 fuzzy motion features 作為攝影師從拍攝環境中作為判斷使用運鏡的特徵。

在運鏡判斷的步驟中，本研究提出 automatic camera movement model (ACMM)進行運鏡的自動選取。ACMM 會依照輸入的 fuzzy motion features 和目前 ACMM 的狀態輸出運鏡以及景別的資訊。ACMM 分為三個部分，分別為 camera movement state (CMS)、transition probability model (TPM)以及 shot probability model (SPM)。CMS 提供不同運鏡方法使用順序的习惯資訊；TPM 提供不同環境特徵下美種運鏡方法的使用機率；而 SPM 則提供運鏡切換時適合使用的景別。

在運鏡路線規劃(path planning)中，本研究依照以及攝影經驗的提出運動鄰近程度(motion field neighborhood)、畫面連續性(continuous)以及畫面重複性(repetition)三種評比，並另外參考攝影美學的相關研究，使用構圖(composition)、

亮度(luminance)等兩種美學評比作為主體挑選的標準。最後根據觀察各種運鏡方式的實際做法設計每個運鏡方法在 PTZ 攝影機上的實作方法，並以挑選出來的主體作為運鏡主軸規劃 PTZ 的移動路線。

實驗結果顯示，本系統能夠配合觀眾席當時情況即時切換合適的運鏡方式以及選擇搭配的景別，並自動從畫面中挑選符合美學條件的拍攝主體。系統所決定之拍攝主體能夠盡量避免重複的主體和不連貫的畫面，並藉由井字構圖法將主體放置在畫面中適當的位置。從運鏡 PTZ 攝影機的拍攝結果可以得知，本系統所提供之運鏡流暢而平穩，符合專業錄影的要求。

第二節 未來工作

本研究所使用之 training data 皆為人工手動擷取，擷取的方式為觀看演講的影片，紀錄影片中切換運鏡時所使用的運鏡方式以及四種 fuzzy motion features。其問題有三，第一是擷取的效率非常低落，必須要花費至少與演講影片同樣的時間才能擷取一部影片的 training data，導致 training data 的數量不足。第二為四種 fuzzy motion features 的標準差異，因為是以人工判斷切換運鏡時的 fuzzy motion features，除了每個特徵擷取人員對每個特徵的強弱標準不同之外，畫面的雜亂程度也會影響判斷特徵的強弱。第三點為特徵擷取的畫面範圍，由於每當影片中出現運鏡切換，畫面必定位運鏡中的攝影機畫面，不會有全景攝影機的畫面供觀眾參考。所以特徵擷取人員只能以運鏡攝影機的畫面判斷整個場面的狀況。

由於 training data 擷取上的各種困難使得 ACMM 在訓練上有所欠缺，從實驗結果可以得知 ACMM 系統並未訓練且使用到所有的 CMS 以及 TPM，甚至使用的數量遠少於最大可用數量。主要原因就在於 training data 數量的不足，無法在所有 fuzzy motion features 可能發生的情況下紀錄 CMS 的轉移情形。使得很多場合都被迫使用可能不適合該情況的運鏡方式，造成運鏡方法選擇上的失誤。

為解決以上問題，training data 的自動擷取是未來必定要完成的部分，大量的

training data 才足以讓 ACMM 趨於完善。自動擷取可以避免不同擷取人員在判斷上標準的差異，而且擷取效率勢必遠高於人工擷取。而自動擷取 training data 的困難在於如何去除影片中多餘的 motion 資訊，所謂多餘的 motion 資訊指的是因為運鏡所產生的 motion，這些 motion 並非觀眾產生而是因為攝影機移動或伸縮造成畫面改變而產生的。若能排除該困難並實現自動擷取 training data 的功能，對本系統將是一大幫助。

當自動擷取 training data 得以實現後，本系統將能提供客製化的服務，將特定攝影師的資料輸入系統做訓練，讓系統成為該攝影師風格的虛擬攝影師。像是模仿電影運鏡的習慣、運動場上的運鏡習慣等等不同風格的拍攝習慣，都能夠藉由擷取足夠份量且符合該風格或是特定作者的運鏡習慣資料，並輸入系統做訓練來建構特定風格的虛擬攝影師。

在主體的選擇上本系統使用 opencv 函式庫中的臉部偵測技術進行觀眾臉部偵測，但是該技術對於光影變化的適應力較低，對臉部角度的變化也較為敏感。由於演講活動不只舉辦在室內，室外的演講活動更是不少，而且攝影機較常架設於觀眾作前或右前方。為增加系統的對環境的適應力以及架設位置的多變性，強化臉部擷取的功能也是要務之一。此外，雖然本研究以五種美學相關特徵來挑選主體，仍然存在更多的特徵可以用來決定主體的選擇。例如 saliency map、色彩飽和度等等，增加更多美學標準有助於對主體挑選的周全性。

另外，互動偵測也是個很重要的工作，所謂的互動指的是講者與觀眾之間產生交流的情況。不少演講影片中可以觀察到講者與台下觀眾們的互動，最常見的就是邀請觀眾發問，或是與講者對談。對於這種情況其實並不適用於一般運鏡習慣，因為在這種情況下畫面會不斷停留在講者與對談者之間，若沒有辦法偵測到互動的發生而繼續使用原本判斷運鏡的方式反而造成錯誤的運鏡。若能夠偵測到講者的某些動作，像是手指向某個方位就提高該方向的觀眾被選為主體的機率等等，可以增加本系統對特殊情況的處理能力。

本系統目前雖然能夠自動完整的完成的拍攝動作，但與實際的攝影師相比仍缺乏隨機應變的能力，希望未來解決以上所提的各項問題後，能夠再加入更多的要素在系統當中。例如不再侷限於 motion 的相關特徵，加入物件資料庫判斷畫面中是否有關鍵人物或標誌。增加動作辨識的功能讓觀眾能和系統作互動，甚至能讓系統操作載具模仿跟拍的運鏡方式等等，讓系統更加完整、更像一個真正的專業攝影師。



參考文獻

- [Che07] I. Cherif, V. Solachidis, and I. Pitas, “Shot Type Identification of Movie Content,” *In Proceedings of 9th International Symposium on Signal Processing and Its Applications*, pp. 1-4, Sharjah, United Arab Emirates Feb. 2007.
- [Xu12] L. Xu, H. Li, and Z. Wang, “Saliency Detection from Joint Embedding of Spatial and Color Cues,” *In Proceedings of IEEE International Symposium on Circuits and Systems*, pp. 2673-2676, Seoul, Korea, May 2012.
- [Ber06] A. Berengolts and M. Lindenbaum, “On the Distribution of Saliency,” *In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 543-549, Washington, USA, Jun. 2004.
- [Dor12] M. Dorr, E. Vig, and E. Barth “Colour saliency on video,” *In Proceedings of 5th International Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering (ICST) Conference, BIONETICS 2010*, pp. 601-606, Boston, USA, Dec. 2010.
- [Liu11] Z. Liu, Y. Xue, H. Yan, and Z. Zhang, “Efficient saliency detection based on Gaussian models,” *IET Image Processing*, vol.5, no.2, pp. 122-131, Mar. 2011.
- [Ban07] S. Banerjee and B. L. Evans, “In-Camera Automation of Photographic Composition Rules,” *IEEE Transactions on Image Processing*, vol. 16, no.7, pp. 1807-1820, Jul. 2007.
- [Li09] C. Li and T. Chen, “Aesthetic Visual Quality Assessment of Paintings,”

- Journal of Selected Topics in Signal Processing*, vol.3, no. 2, pp. 236-252, Apr. 2009.
- [Lia12] Y. Liang, Z. Su, C. Wang, D. Wang, and X. Luo, "Optimised image retargeting using aesthetic-based cropping and scaling," *In Proceedings of IET Image Processing*, vol. 7, no. 1, pp. 61-69, Feb. 2013.
- [Su12] T. W. Chen, C. C. Kao, W. J. Hsu, and S. Y. Chien, "Preference-Aware View Recommendation System for Scenic Photos Based on Bag-of-Aesthetics-Preserving Feature," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp.833-843, Jun. 2012.
- [Niu12] Y. Niu and F. Liu, "What Makes a Professional Video A Computational Aesthetics Approach," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1037-1049, Piscataway, USA, Jul. 2012.
- [Bov01] Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1222-1239, Nov. 2001.
- [Chu09] S. K. Chung and Y. C. Chang, "Cinematic Continuity in 3D Computer Animation – the Tendency and Expression of Shot in Hollywood's Blockbuster Animation Films," *27th National Kaohsiung Normal University Journal*, pp. 195-217, Dec. 2009.
- [Vio04] P. Viola and M. Jones "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, pp. 747, May 2004.
- [Kha09] O. O. Khalifa, A. A. M. Assidiq, and A. H. A. Hashim, "Vision-Based Lane Detection for Autonomous Artificial Intelligent Vehicles," *In Proceedings of IEEE International Conference on Semantic Computing*, pp.636-641, Gombak, Malaysia, Oct. 2009.

- [Bob01] A. F. Bobick and J.W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257-267, Mar. 2001.
- [JHO12] Z. R. JHOU and W. R. JIAN, "色彩調和理論與網頁自動配色應用之研究," *In Proceedings of 8th International Conference on Digital Content*, Tainan, Taiwan, Dec. 2012.
- [Lee04] F. R. Lee, "A Study of the Legibility of Texts in Small Scale TFT-LCD for the Elderly," Jun. 2004.
- [1] 【攝影入門教學講義一】淺談構圖原理與主題表現
<http://www.wretch.cc/blog/jhyou/13908453>
- [2] 怎麼拍，才精采？基本構圖法演練實戰
<http://www.eprice.com.tw/dc/talk/356/723/1/>
- [3] 攝影教室—【風景攝影】
http://chang-home.myweb.hinet.net/Classroom/Scenery/page_3.html
- [4] 簡單構圖與基礎攝影
<http://blog.duncan.idv.tw/BasicPhotographicComposition.pdf>
- [5] 拍出不同戲劇效果—善用不同的光向
<http://www.fotobeginner.com/3789/light-direction/>
- [6] 李秀美，鏡頭語言與採訪技巧--如何將文字視覺化(PPT)
- [7] 景深的介紹
<http://cck0217.myweb.hinet.net/TNGNEW-7.html>
- [8] 光圈與景深的關係
<http://dawing888.com/photo01.asp>
- [9] 攝影入門之新手必讀—色彩的奧秘
<http://www.sj33.cn/dphoto/syxt/201211/32597.html>

- [10] 利用色彩輕鬆拍攝漂亮照片
http://shijue.me/show_text/4ffef3e0ac1d840d9001c13c
- [11] 長沙印象寶貝專業兒童攝影
http://dcps.blog.hexun.com.tw/42473850_d.html
- [12] 分鏡的設計(PPT)
- [13] 什麼是畫面的景別？
<http://hdtv.baike.com/article-87679.html>
- [14] 攝影取景
<http://whereyou.pixnet.net/blog/post/19970578-%E6%94%9D%E5%BD%B1%E5%8F%96%E6%99%AF>
- [15] 景別(Doc)
- [16] 運鏡技巧基本解說
<http://tw.myblog.yahoo.com/jianchen0627/article?mid=145>
- [17] 柳在天(2010), *就用構圖取勝吧!*, 台北市: 悅知文化
- [18] 山口高志(2006), *構圖實例事典*, 台北市: 宇琉采伊娛樂經紀有限公司
- [19] 佳影在線(2010), *成功構圖300例*, 台北市: 上奇資訊股份有限公司
- [20] 莎頓國際學院
<http://www.sheltoncollege.edu.sg/cn/aboutus-facilities.html>
- [21] 蘿莎會館
<http://hotel.wingnet.com.tw>
- [22] 浣莎室內樂團
<http://www.classical-music.com.tw>
- [23] 公主郵輪的公主劇院
<http://blog.liontravel.com/ckshsu/post/13365/4408/19906>
- [24] 何謂 Fuzzy

<http://ycc.tsu.edu.tw/9415/page/fuzzy.htm>

[25] oracle technology network

<http://www.oracle.com/technetwork/systems/fsm-156381.html>

[26] Wikimedia-commons

<http://www.commons.wikimedia.org>

[27] MathWorks

<http://www.mathworks.com/index.html>

[28] Motion 4User Manual

<http://documentation.apple.com/en/motion/>

