

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
 陳世旺 博士

虛擬觀眾攝影師系統

Virtual Audience Cameraman System

研究生：廖軒嘉 撰

中華民國 一百零三 年 六 月

摘要

虛擬觀眾攝影師系統

Virtual Audience Cameraman System

廖軒嘉

本研究的主旨在於建立一套模擬專業攝影師並以觀眾為拍攝主體的虛擬觀眾攝影師系統。現今社會中許多資訊的傳播都是透過演講方式，而為了讓觀賞者能隨時觀看演講內容，聘請專業攝影團隊紀錄整場演講是最直接的方式。然而現代生活的人力資源成本不斷提升，聘請一組專業攝影團隊的成本並不低，因此本研究發展一套虛擬觀眾攝影師系統來節省人力資源成本，同時提供專業攝影技巧以製作高規格影片。

本研究以兩台 Pan Tilt Zoom Camera(PTZ 攝影機)為一組作為實驗設備，一台稱為 global-view 攝影機，另一台為稱 local-view 攝影機。Global-view 攝影機是用來代表攝影師的雙眼，主要功能是監控畫面與主體偵測並找出畫面中感興趣的區域(Region Of Interesting, ROI)；Local-view 攝影機則是用來代表攝影師手上的攝影機，在系統決定 ROI 與運鏡所需要的一切資訊後，local-view 攝影機就會實際執行運鏡動作並進行拍攝。

本系統的主要目的是模仿專業攝影師的拍攝技巧並自動進行運鏡拍攝動作，而為了符合專業攝影師的拍攝技巧與手法，每次運鏡前系統皆需要先決定運鏡方式、景別、主體等要素。首先系統從 global-view 攝影機所提供的連續影像中擷取具有描述觀眾行為的 motion 特徵，再將這些特徵經過運算處理並找出畫面中的候選 ROI，接著將這些候選 ROI 輸入 STA(spatio-temporal attention neural model)，STA 能夠紀錄並提供相關資訊來協助系統找出最適合拍攝的 ROI。之後系統計算欲拍攝的 ROI 與鏡頭中心位置的對應關係，並依據輸入的資料輸出最適合該情況的運鏡方式和景別用以啟動 local-view 攝影機運鏡拍攝；而 local-view 攝影機所拍攝的主體挑選與拍攝畫面的質感主要是以美學以及光學的特徵分析來做為判斷標準，本研究透過上述流程來模擬專業攝影學的拍攝技巧。

實驗結果顯示，本系統所運用的方法可以進行即時且流暢的運鏡動作並可準

確模擬專業攝影師的拍攝手法，符合專業攝影團隊來拍攝記錄演講錄製的需求。

關鍵字：虛擬觀眾攝影師、STA(Spatio-Temporal Attention neural model)、運鏡、專業攝影學。

Abstract

Virtual Audience Cameraman System

by

Hsuan-Chia Liao

This thesis proposes a virtual audience cameraman system to capture the audience videos automatically. Nowadays the contents of lectures can be broadcast widely and rapidly by digital videos, thus to capture digital videos of important lectures for the viewers is an essential work. However, the cost to hire a video-recording team, including professional photographers, to capture good-quality digital videos is very high. Thus this study developed a virtual audience cameraman system which can obtain good-quality digital videos automatically and reduce the cost of hiring a professional video-recording team.

In this study, two PTZ cameras are mounted together to be a set, one is the global-view camera and the other is the local-view camera. The global-view camera can be regarded as the photographer's eyes. It can be used to monitor the whole audience and help the region of interesting (ROI) detection. The local-view camera can be regarded as the photographer's camera on hand. It can be used to capture the videos from ROI after the system determines the location of ROI.

Since the purpose of this system is to simulate the camera-control behaviors of professional photographers to capture the audience videos, the proposed system needs to decide the camera steering mode, shot class, and the objects before camera steering. First, the system obtains input videos from the global-view camera and then detects the audience motion features to locate the ROI candidates. The ROI candidates are then input into the spatiotemporal attention (STA) neural model. The STA neural model can record and provide information to help the system to identify the most suitable shooting ROI. Further, the system computes the relative distance between the location of the ROI on the frame and the center of the camera lens, and outputs the appropriate steering mode of the local-view camera. The local-view camera then captures the output videos from the location of ROI by considering the viewpoint of aesthetics and the analysis result of optical characteristics. Through the above process this system can simulate professional photography shooting skills.

The experimental results show that the proposed method can steer the camera immediately, automatically, and smoothly. It can also simulate the style of professional photographers accurately.

Keywords : Virtual Audience Cameraman System, STA (Spatiotemporal Attention) neural model, Camera Steering, Professional Photography.

致謝

在碩士生活的兩年生涯中，承蒙眾多人的幫助，雖然時間不是很長卻讓我有不小的成長。誠摯感謝我的指導教授陳世旺老師、方瓊瑤老師，兩位指導老師所傳授的不僅僅只有學術領域，更不時於會談之餘教導我做人處事的道理，讓我在追求真理的過程中不忘重視人情。在此特別感謝方瓊瑤老師，方老師常常不厭其煩的告知我必須注意的事項，讓我做事能比較從容順利，並且提醒我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。

感謝 IPCV 實驗室的所有成員，梁祐銘老師、王俊銘學長與羅安鈞學長每週於百忙之中撥空一起來參與實驗室團體報告，參與討論研究上面臨的困難，並且提供我解決的方法或技巧，讓我的研究得以順利進行與完成。感謝許之凡學長與呂佳儒學姊在碩士生活中提供給我許多意見與幫助，也與我一同分享歡樂。感謝貞佑和俊宇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！感謝昆賢、冠宇和淳雅的大力幫忙和支持，也感謝所有曾經參與我研究與實驗的所有人。

最後，非常感謝大家一直以來對於我的鼓勵與讚美，以及批評與指教，尤其是對於我的包容，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

目錄.....	i
圖目錄.....	iii
表目錄.....	v
第一章 緒論.....	1
第一節、研究背景與目的.....	1
第二節、研究困難.....	4
第三節、研究範圍與限制.....	6
第四節、論文架構.....	6
第二章 文獻探討.....	8
第一節、自動化攝影師系統相關研究.....	8
第二節、自動化攝影師系統相關技術.....	11
第一項、主體選取.....	11
第二項、主體動作偵測.....	12
第三章 系統概述.....	15
第一節、系統架設環境.....	15
第二節、系統流程.....	18
第四章 ROI 選取.....	22
第一節、觀眾動作分析.....	22
第二節、候選 ROI 偵測.....	23
第三節、ROI 挑選.....	27
第五章 運鏡路線規劃.....	30
第一節、攝影機控制.....	30
第二節、運鏡方式與構圖規則.....	33
第一項、主體大小.....	34
第二項、影像構圖規則.....	39
第六章 實驗結果.....	42
第一節、ROI 選取的正確率.....	42

第一項、第一段影片實驗結果.....	44
第二項、第二段影片實驗結果.....	46
第三項、第三段影片實驗結果.....	48
第二節、本系統運鏡評估.....	49
第三節、與過去研究拍攝結果比較.....	53
第四節、本系統對於拍攝場地的適用性.....	57
第一項、第一段影片實驗結果.....	57
第二項、第二段影片實驗結果.....	58
第七章 結論與未來工作.....	60
第一節、結論.....	60
第二節、未來工作.....	60
參考文獻.....	62

圖目錄

圖 1.1、2011-2016 WORLDWIDE SELF-PACED eLEARNING FIVE-YEAR GROWTH RATED BY REGION[1]	1
圖 1.2、各式演講場合(A)新型產品發表[17](B)論文發表[18](C)記者會[19]	2
圖 1.3、自動化演講錄製系統架設圖	3
圖 1.4、演講廳類型(A)階梯式演講廳[3](B)水平式演講廳[4]	5
圖 2.1、AUTOAUDITORIUM 拍攝模擬圖[BIA98]	8
圖 2.2、多台攝影機合作拍攝架構[RUI04]	9
圖 2.3、不同光線運用的攝影範例[20](A)順光(B)逆光	10
圖 2.4、SALIENCY MAP DETECTION(A)原始影像 (B)主體偵測結果[LAN09]	12
圖 2.5、根據群體運動特徵偵測群體走向[WU10]	12
圖 2.6、KINECT 的外部構造[12]	13
圖 2.7、利用 KINECT 所得到的影像(A)彩色影像(B)深度影像[13]	14
圖 2.8、KINECT 捕捉的骨架資訊[12]	14
圖 3.1、PTZ 攝影機(A)正面 (C)硬體設備[14]	15
圖 3.2、PTZ 攝影機架設方式，上方是 GLOBAL-VIEW 攝影機，下方是 LOCAL-VIEW 攝影機(A)側視圖(B)俯視圖	17
圖 3.3、虛擬攝影師之拍攝觀眾系統流程圖	18
圖 3.4、系統決定候選 ROI(A)MOTION 特徵圖(B)MOTION 特徵密度圖(C)候選 ROI 選取結果	19
圖 3.5、系統採用 STA 決定 ROI(A)ROI 選取結果(B)運鏡後拍攝畫面(C)STA 輸出層激發狀態	20
圖 3.6、系統人臉偵測(A)ROI 選取結果(B)運鏡後人臉偵測結果(C)主要拍攝人臉選取	20
圖 4.1、演講觀眾動作分類圖	22
圖 4.2、LUCAS-KANADE OPTICAL FLOW 計算範圍差異圖	24
圖 4.3、金字塔型架構追蹤示意圖[15]	27
圖 4.4、注意力圖像(A)原始影像(B)根據左圖原始影像建立的注意力圖像	27
圖 4.5、STA 類神經網路架構圖[FAN03]	28

圖 4.6、輸入層與輸出層的權重連結圖.....	29
圖 4.7、墨西哥帽函數圖形.....	29
圖 5.1、GLOBAL-VIEW 與 LOCAL-VIEW 攝影機的水平旋轉移動參數示意圖.....	31
圖 5.2、GLOBAL-VIEW 與 LOCAL-VIEW 攝影機的垂直傾斜移動參數示意圖.....	33
圖 5.3、HAAR MASKS 之範例.....	35
圖 5.4、INTEGRAL IMAGE 計算之圖例(A)INTEGRAL IMAGE 中 $I(x, y)$ 之 INTENSITY 值 所涵蓋的計算範圍(B)INTEGRAL IMAGE 中 D 範圍的 INTENSITY 值總和之計 算示意圖	36
圖 5.5、ADABOOST 演算法[VIO04]	37
圖 5.6、多層分類器架構(正圓形為分類器的篩選階段)[13]	39
圖 5.7、三一法則比較圖	40
圖 5.8、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].....	40
圖 6.1、實驗一第一段影片擷取片段.....	45
圖 6.2、實驗一第二段影片擷取片段.....	47
圖 6.3、實驗一第三段影片擷取片段.....	48
圖 6.4、實驗二全景影片擷取片段.....	50
圖 6.5、實驗二本系統拍攝之運鏡影片擷取片段.....	51
圖 6.6、實驗二評分項目觀賞者評分長條圖(A)清楚地表達主體(B)內容的吸引力(C) 畫面的構圖(D)拍攝技巧的運用(E)整體觀看舒適度(F)評分項目平均分數...53	53
圖 6.7、實驗三實驗影片擷取片段.....	54
圖 6.8、實驗三評分項目觀賞者評分長條圖(A)清楚地表達主體(B)內容的吸引力(C) 畫面的構圖(D)拍攝技巧的運用(E)整體觀看舒適度(F)評分項目平均分數...56	56
圖 6.9、實驗四第一段影片全景影片擷取片段.....	57
圖 6.10、實驗四第一段影片本系統拍攝之運鏡影片擷取片段.....	58
圖 6.11、實驗四第二段影片景影片擷取片段.....	58
圖 6.12、實驗四第二段影片本系統拍攝之運鏡影片擷取片段.....	59

表目錄

表 1-1、工商錄影價目表[2].....	3
表 3-1、AXIS 5512 型 PTZ 攝影機詳細規格表[14]	16
表 5-1、運鏡方式	34
表 6-1、實驗一第一段影片結果	44
表 6-2、實驗一第二段影片結果	46
表 6-3、實驗一第三段影片結果	48
表 6-4、實驗二之觀賞者評分結果	52
表 6-5、實驗三之觀賞者評分結果	55

第一章 緒論

本章第一節將介紹本研究的背景與目的；第二節與第三節會說明研究相關的困難與限制；最後第四節則敘述本論文的基本架構。

第一節、研究背景與目的

近年來，隨著資訊與網路的蓬勃發展，資訊不斷的推陳出新，網路寬頻的傳輸速度也不斷的提升，讓人們可以透過網際網路快速的傳輸影像、聲音等多媒體資訊，使得網路逐漸成為資訊傳遞的主要媒介之一。不斷的學習與接收新的資訊也已經成為生活中一個相當重要的課題，因此以網路傳遞為基礎的數位學習及其應用越來越受企業界、學術界等各界的重視，世界各國也紛紛將數位化的科技教育列為國家未來發展重要計畫項目。根據美國 Ambient Insight 研究機構於 2011 年所進行的「全球市場自定進度的電子學習產品和服務」調查報告中，觀察最近幾年來各國數位學習領域的發展，藉由國家型計畫之經費支持，全球市場在 2011 年達到 356 億美元，估計至 2016 年將可以達到 515 億美元，其中年複合成長率為 7.6%(如圖 1.1 所示)。



圖 1.1、2011-2016 Worldwide Self-paced eLearning Five-year Growth Rated by Region[1]

現今資訊傳遞與接收的方式繁多，而最直接的方式是透過各種演講形式來達到資訊的傳遞與接收，在演講的場合上觀眾能直接與演講者進行面對面的溝通，透過直接的對答是資訊傳播最有效的傳達方式之一。近年來社會中許多資訊的傳

播都是以演講方式進行，例如：新型產品發表、論文發表、記者會等(如圖 1.2 所示)，內容涵蓋了各個領域，兼具學術與應用等各層面。演講內容與主講者表達方式皆是學習與教學之重要資源，也是學生及在職員工們相當重視的學習教材。但安排一場演講需要配合場地與演講者的時間，因而並非所有想聽的觀眾皆能抽空前來參與，所以要如何將演講內容記錄並提供給無法親自前來的觀眾觀看是一個相當重要的問題。

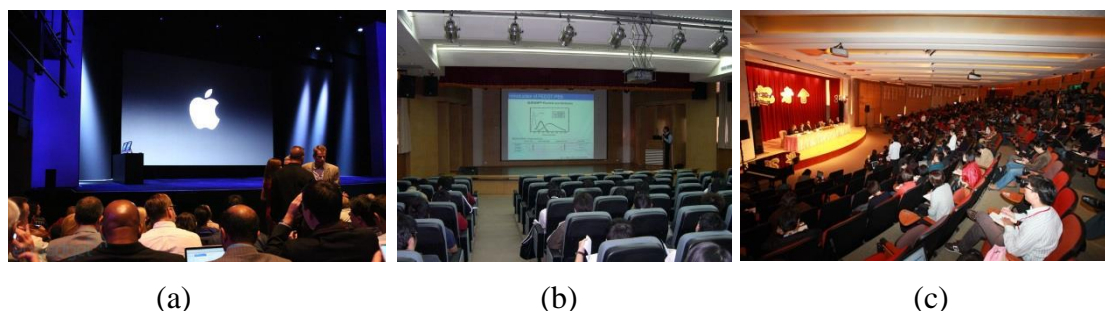


圖 1.2、各式演講場合(a)新型產品發表[12](b)論文發表[13](c)記者會[14]

紀錄演講內容最直接的方式就是聘請一組專業攝影團隊將整場演講拍攝記錄下來，藉由他們專業的攝影技術將整個演講、研討會等活動過程與內容全部收錄，並以數位媒體形式保存這些演講內容，觀賞者可以透過觀看數位媒體的方式來重現該場演講。而要完整錄製一場演講，其攝影團隊主要的工作流程相當繁複，其中必須包含：進行活動地點的場地勘驗、活動當天的拍攝作業、拍攝影片の後製剪輯及輸出等，如此繁複的作業流程，需要相當的成本與人力資源。

根據[Row01]和[Bae03]的研究得知聘請專業團隊拍攝花費的成本，可以歸納成固定成本與重複人事成本。固定成本主要屬於硬體設施部分，主要包含了電腦、麥克風、攝影機等設備，重複人事成本主要屬於拍攝團隊人員薪資(如表 1-1)。然而現代生活的人力資源成本不斷提升，聘請一組專業攝影團隊對一般預算有限但又須經常紀錄活動的非營利學校或公司行號等是相當沉重的負擔。但若為了節省人力資源成本而聘請缺乏專業攝影經驗的人來拍攝，所拍攝出的影片往往會缺乏美感及流暢度等，甚至於降低觀眾觀看的意願，因此開發自動化演講錄製系統有其必要性。

表 1-1、工商錄影價目表[蔡 13]

公司	項目	價格(台幣)
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)
		16000/8hrs(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)
		16000/8hrs(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

利用自動化演講錄製系統(如圖 1.3 所示)拍攝一場完整的演講又可大致分為四個主要部分，第一是以演講者為主體的拍攝系統，主要記錄整場演講的內容與講者的肢體動作和表達方式；第二是以觀眾為主體的拍攝系統，用來記錄觀眾與演講者之間的互動關係，而觀眾的反應與動作也可表現出這場演講活動所散發出的氣氛；第三是以全景畫面拍攝系統，用來拍攝整體演講的場地與所有參與者，可以完整的表達整體演講的畫面；第四是整合前面三者的導播系統，導播系統需要整合前三者的畫面並決定哪一個畫面是目前最佳輸出畫面。

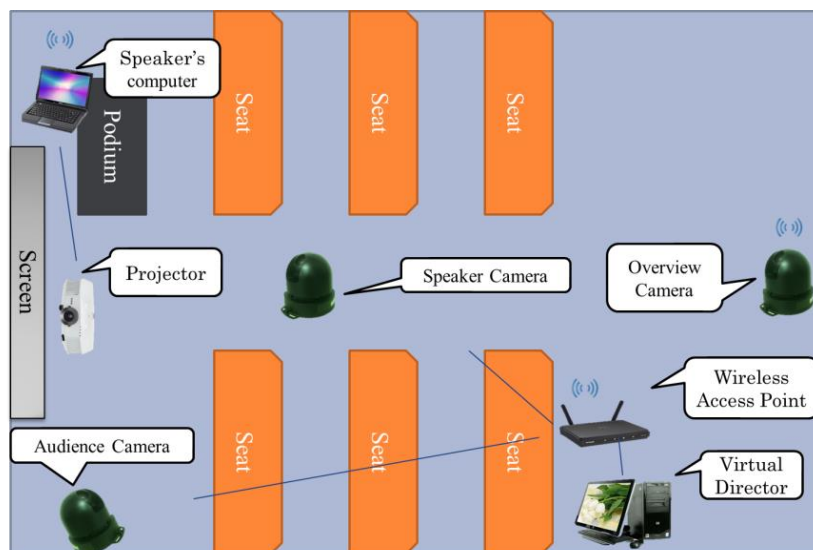


圖 1.3、自動化演講錄製系統架設圖

上述四個部分都是自動拍攝一場完整的演講不可或缺的要素，每部分都有其存在價值，不同的部分所欲表達給觀賞者的訊息也不盡相同，演講者畫面主要給觀賞者完整的演講資訊；觀眾畫面主要給觀賞者本場演講的現場氣氛；全景畫面主要給觀賞者本場演講的整體資訊；而導播需要在適當的時機選擇最合適的畫面給觀賞者。

近年來有許多研究的發展重點都放在以演講者為主體，但觀眾畫面所能帶給觀賞者的訊息也是相當重要的部分，因此本研究的主旨著重於上述自動化演講錄製系統中第二部分，建立一個以觀眾為拍攝主體並能模擬真實攝影師的虛擬觀眾攝影師系統。該系統功能包括自動判定與選取感興趣的區域(Region Of Interesting, ROI)做為拍攝目標，並操作攝影機運鏡動作完成錄影程序。透過合適的運鏡方式可以達到不同的拍攝效果，而且本系統的目標並不局限於觀看演講的觀眾，而是希望能夠透過不同的拍攝角度與運鏡方式來適應各種場合下的觀眾目標，例如球賽、音樂會、戲劇等不同場合都能夠達到專業的錄影效果。

虛擬觀眾攝影師系統的輸入資料為攝影機所拍攝的畫面，在以往拍攝畫面中，大多數的拍攝畫面都是以演講者為主，而觀眾席上的活動往往都被忽視。但觀眾其實不僅僅只是配角，而是整個活動中相當重要的部分，觀眾的各種反應能表現當下活動進行的情況，更是帶動活動氣氛與活動成功與否的重要指標。例如在球賽中，當比賽狀況激烈時，觀眾席上群情激昂的畫面能夠帶起激情；在表演的高潮迭起時，觀眾的表情與反應的畫面能夠表達表演張力，藉由穿插現場觀眾的畫面，不僅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，所以觀眾在各類活動中是不可或缺的重要成員。

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，更讓觀眾從活動中的配角變成活動相當重要的一環。本系統所拍攝出的觀眾畫面皆以觀眾為主要拍攝對象，所拍攝出的畫面不僅具有活動流程中一定的代表性，更能讓導播擁有更多的畫面選擇空間。

第二節、研究困難

為了能夠順利拍攝到觀眾席上各式各樣的精彩畫面，本系統在開發時所需要考慮與克服的因素繁多，在蔡侑廷的論文[蔡13]中已經提出一些與本研究相同的

困難，而本研究將可能遇到的問題作出整理與說明，其中包括演講環境光線的變化問題、拍攝主體如何決定的問題、觀眾席座位設計多樣性的問題、觀眾入座和離席時的干擾問題、以及為了配合不同拍攝環境場合的運鏡手法、觀眾座位背光問題、運鏡與動作偵測同步問題、攝影機與系統間通訊問題等。以下對此分別說明：

1. 演講環境光線的變化問題：自動化演講錄製系統首要考慮的是光線的影響，例如：室內光源的開關、調整與改變，以及其改變時所造成的光影變化，亦會導致觀眾席畫面內容偵測的困難。
2. 拍攝主體如何決定的問題：本系統是以觀眾為拍攝主體，因此觀眾席上的每位觀眾都可能成為拍攝的目標，但拍攝觀眾並不侷限於拍攝單一觀眾，可能也包含群體觀眾。而該如何偵測與如何從中挑選出最合適的主體來拍攝將是本研究一大難題。
3. 觀眾席座位設計多樣性的問題：觀眾座位依場地的不同有各式各樣的規格，一般觀眾席分為平面式和階梯式二種，且階梯式觀眾席的坡度並不統一(如圖1.4(a)(b)所示)，不同觀眾席亦有不同的顏色、擺設與風格。觀眾席的規格對於系統的偵測上會產生許多不確定因素，系統開發時必須要多方納入考慮避免影響系統的運作。



(a)



(b)

圖 1.4、演講廳類型(a)階梯式演講廳[2](b)水平式演講廳[3]

4. 觀眾入座和離席時的干擾問題：在典禮或是演講進行的時候，常會有觀眾起身離開座位或是中途入座的情形，此現象導致畫面中常有物體在移動，系統必須自動判斷當下情況是否必須拍攝擷取並輸出。

5. 配合不同拍攝環境場合的運鏡手法：從專業攝影學的觀點來看，依照拍攝的場景與主體不同運用的運鏡手法也不同，例如拍攝演講的場合與拍攝運動的場合所使用的運鏡手法就有著極大的差異。因此，自動的配合不同場合調整不同的運鏡手法也是本系統開發時需考慮的因素。
6. 觀眾座位背光問題：在許多的演講場合中，演講者需要透過投影片的輔助來陳述演講內容，但由於投影片的放映需要將光線調整成比較暗的環境。因此觀眾座位常常坐落於演講場地中背光的位置，而觀眾背光的問題也會增加系統偵測與運算上的困難，在開發時需要能因應不同光線變化解決觀眾背光的問題。
7. 運鏡與動作偵測同步問題：本系統雖然會自動化偵測 ROI 與自動選擇最適合的運鏡方式，但兩者間存在著一些矛盾。當系統在選取 ROI 與選定運鏡方式時需要一些運算時間，但運算過程都是在動作發生後才去執行，因此可能會導致運鏡反應時間太長的問題，在開發時需要增快運算速度來達到即時拍攝的效果。
8. 攝影機與系統間通訊問題：由於攝影機與系統之間的溝通與傳輸方式是透過網路傳遞，但若牽扯到網路傳輸便需要面對可能的資料遺失、傳輸延遲等問題，這些也是本系統需要克服的重要項目。

第三節、研究範圍與限制

本研究主要以拍攝靜態活動場地中的觀眾為主，針對具有固定位置的觀眾來做偵測與拍攝。因此，本研究假設拍攝期間內每位觀眾的活動範圍不超過一個座位的範圍。其次，由於本研究選用的攝影機是由 AXIS 公司所出產的 P5512 半球型網路攝影機，其攝影機拍攝靜態鏡頭的拍攝範圍為 51.6 度、水平旋轉角度為 360 度、垂直傾斜角度為 180 度、12 倍光學縮放與 4 倍數位縮放，因此整體觀眾的分佈範圍須限制在此範圍內。

第四節、論文架構

本論文的第二章將會介紹相關的文獻探討；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，此章會詳述如何模仿專業攝影師對運鏡使用的判斷；第六章為實驗結果，將實際在活動現場進行拍攝，測試拍攝結果是否合乎專業水準；最後第七

章為結論與未來工作。

第二章 文獻探討

本章將以自動化攝影師系統相關的研究為出發點，探討有關於拍攝演講場合的相關研究與拍攝影像時所需注意到的相關問題。

第一節、自動化攝影師系統相關研究

為了將演講內容記錄並提供給無法親自前來的觀眾觀看，也為了降低拍攝記錄演講的人力資源成本，近年來有許多相關於自動化拍攝演講場合的研究。[Bia98]在1998年發展了一套自動演講轉播系統稱為AutoAuditorium(如圖2.1所示)，AutoAuditorium採用多台可動式攝影機來進行演講拍攝，這些攝影機結合影像分析軟體來偵測並追蹤演講者的動向，增加了影片畫面的活潑性；[Rui01]與[Cho10]提出運用一套智慧型自動攝影系統來降低人力資源成本，對於拍攝演講場合也提出了許多論點，例如：攝影機種類的選用、攝影機架設數量與攝影機擺放位置等；[Oni04]所發展的系統，除了可自動偵測講者的動向外，還能偵測講者的動作，利用這二種資訊來決定攝影機的運作方式，使得系統所提供的錄影畫面更具實用價值。類似以上的自動演講錄製系統還包括[Bia04]、[Zha08]等研究，也提出了在不同的場合，如何紀錄與拍攝整個演講或廣播等。

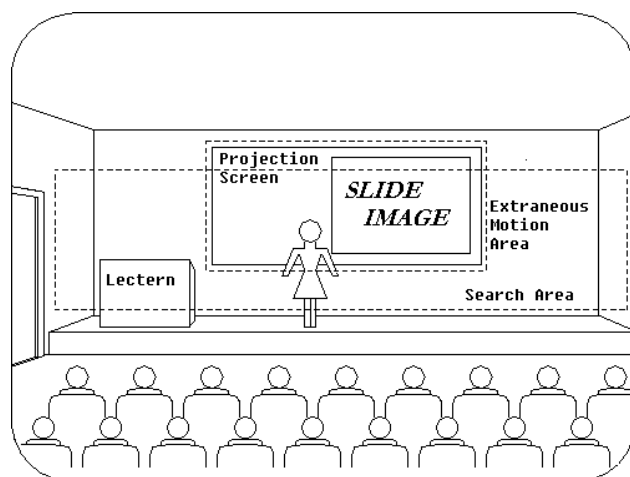


圖 2.1、AutoAuditorium 拍攝模擬圖[Bia98]

事實上，到目前為止，大部份的研究，還是把拍攝重心放在演講者身上，較少考慮到觀眾部份，而觀眾不僅僅只是前來參加本場演講而已，從觀眾的反應還有與演講者間的對應動作也能帶給觀賞者不同的理解。[Rui04]在其研究中，探

討如何用較低的成本，並能同時有效地拍攝演講者與觀眾；[Li05]則探討當演講者與觀眾之間有互動時，應該考慮兩者互動的情況，配合不同的互動情境選擇不同的拍攝手法。

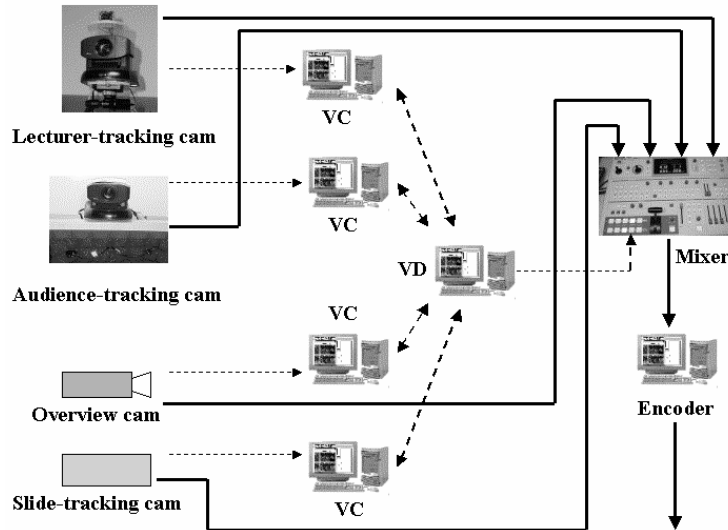


圖 2.2、多台攝影機合作拍攝架構[Rui04]

由上述的相關研究中可知利用多台攝影機同時運作進行演講拍攝是未來的主要趨勢，但在多台攝影機同時運作進行演講拍攝時，需要特別注意各攝影機之間的溝通、協調與合作的機制。[Rui03]提出使用兩台攝影機來模擬真實攝影師(如圖 2.2 所示 F，並將其稱之為虛擬攝影師，其中一台是廣角固定式的攝影機，扮演攝影師的眼睛負責偵測場景中的狀況，另一台是一可動式攝影機，扮演攝影師的攝影機負責追蹤與拍攝。在上述的情境中，攝影機必須要相互溝通與支援才能達成較佳的拍攝品質，而攝影機間相關訊息的傳遞就需要依賴各攝影機之間的溝通、協調與合作的機制，而利用系統自動判斷與控制多台攝影機進行拍攝有下列優點：

1. 攝影機鏡頭的調整將更穩定：鏡頭的轉動或移動由機械控制，不易有不規則的搖動或是震動，畫面拍攝的結果將更平穩。
2. 不易錯失精彩鏡頭：單一攝影師無法在進行局部拍攝的同時觀察會場中其他聽眾的變化，可能因此漏掉了許多精彩鏡頭。若系統規劃兩台以上攝影機同時運作，由一台攝影機監控觀眾席全景，其他攝影機根據全景的畫面內捕捉特寫鏡頭，則能夠同時兼顧較多的場景。

3. 可以有更優質的畫面規劃：上述的多攝影機系統可自動尋找適合拍攝的畫面，通過各攝影機之間的溝通與合作，在全景之下找較具美感的畫面。

此外，拍攝演講場合自動化的同時也不能忘記拍攝畫面所需要的影片品質，為了提升影片品質，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。攝影機取景時需要注意畫面的平衡，在攝影學中畫面組成的平衡感可以透過光影、空間、色彩三樣必要的元素來判斷。在光影方面，「順光」(如圖 2.3(a)所示)時的光線會讓色彩最鮮麗而且飽和，很適合拍攝風景影片，但不適用在人像攝影。「逆光」(如圖 2.3(b)所示)拍攝有時能讓主題或目標物更有立體感，但是強光會造成畫面模糊，易失去前景或背景的細節。在空間方面，攝影機需拍攝到足夠的空間才能把主題或目標物的地位展現，可藉由縮放鏡頭調整畫面裡前景所佔的比例，或是移動鏡頭的動作來完成。另一方面，若畫面前景單薄或結構薄弱，亦可以用同樣的方法找尋支撐畫面豐富性的前景。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。由於色彩比較難調整，若是找不到適合的畫面，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。[Dat06]也提出為了使拍攝畫面更加和諧，拍攝畫面應該依據美學的觀點和主體位置、大小來決定畫面的構圖；[呂 13][蔡 13]提出拍攝畫面結合攝影學中美學分析的部分，此部分包含了拍攝畫面的色彩、光影以及構圖三個元素，將拍攝畫面結合美學分析可以提供觀賞者更舒適的觀看品質。



(a)



(b)

圖 2.3、不同光線運用的攝影範例[20](a)順光(b)逆光

第二節、自動化攝影師系統相關技術

自動化攝影師系統主要目的為拍攝主體，並依據主體的動作來做相對應的反應，所以系統首要功能是選取適當的目標為主體，並偵測該主體動作，且將偵測的資訊轉換成對應的反應動作。以下將分成兩個部分來探討相關的技術，第一部份將探討主體選取的相關技術；第二部份則探討偵測主體動作的相關技術。

第一項、主體選取

目前有多種不同類型的應用需先進行主體選取，例如追蹤物體、辨識身分等。[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 template matching 的技術。首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，然後將影像輸入系統中使用 template matching 的技術找出場上的主體。[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限制為一人，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。

其中 saliency map detection 也是被廣為研究的主體選取技術之一，[Lan09]就是 illuminated 使用 saliency map detection 的技術作為景別主體區域的選取。然而 saliency map detection 多以顏色作為特徵，並以高斯模糊函數去 model 相近的顏色組成同一區塊，並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視為主體(如圖 2.4 所示)。[Ber06]則是認為主體上的邊緣應該有特定的長度，因此以 edge 的長度作為主體偵測的主要特徵。[Ber06]所提的偵測方式也可以被看作一種紋理的偵測，在實驗結果中顯示，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及為主體可能的位置。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，所以用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所偵測到的主體在本質上有所不同，template 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的主體，但並不確定主體之實際模樣。

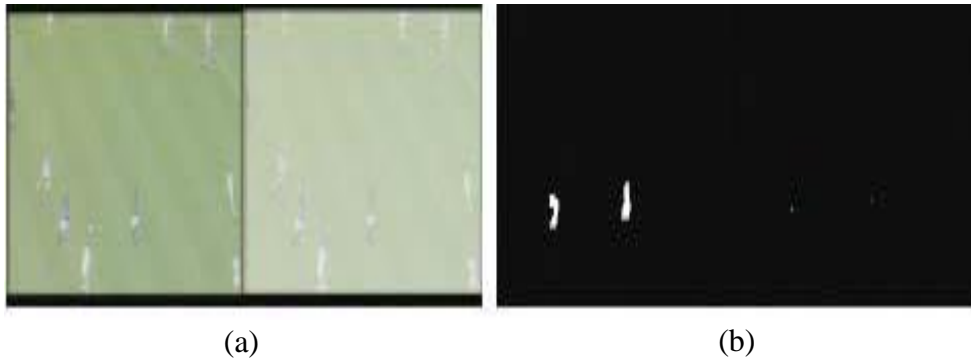


圖 2.4、Saliency map detection(a)原始影像 (b)主體偵測結果[Lan09]

第二項、主體動作偵測

主體動作偵測主要是透過找出畫面中的動作特徵，將這些特徵做適當的分析來偵測主體動作，而因應不同應用需求選擇適當的特徵選取方法可以提升偵測出主體動作的正確率。本研究需要尋找畫面中 motion 性質的特徵，常見的特徵擷取方法是 optical flow，optical flow 是利用影像中的光流變化來找出畫面中具有 motion 性質的像素點位置，可用來描述相對於觀察者的運動所造成的觀測目標、表面或邊緣的運動。[Wu10]與[Sol12]提出針對群體運動來做判定的運用畫面中群體的特徵點，利用 optical flow 將特徵點與時間關係來找出特徵點的軌跡移動，利用群體的軌跡移動來判斷群眾的群體行為走向(如圖 2.5 所示)。圖 2.5 下排是將原始輸入影片經過一段時間後累積出的物體走向軌跡圖，上排是將下排中所選定區域的放大圖，由圖中的軌跡方向可以判斷出群體行為走向。

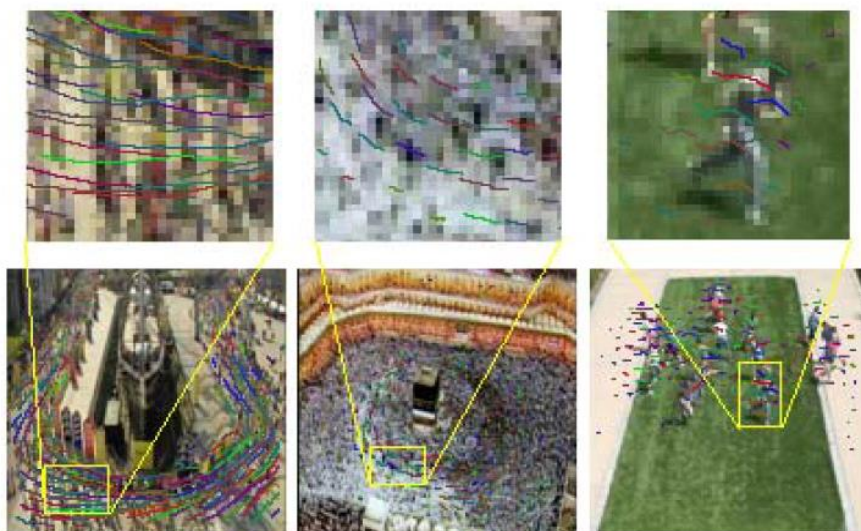


圖 2.5、根據群體運動特徵偵測群體走向[Wu10]

除了利用影像中的 optical flow 特徵來偵測主體動作外，近年來透過各種感應器來偵測主體動作也相當受到注目。微軟(Microsoft)公司所開發的感應器 Kinect(如圖 2.6 所示)，藉由 Kinect 所提供的資源應用，使用者不需使用任何的手持裝置，即可直接藉由身體動作操控系統。Kinect 可接收彩色影像與深度影像兩種訊息來進行主體動作偵測。彩色影像由 Kinect 硬體設備中的 RGB 鏡頭取得；深度影像由左右兩邊的紅外線發射器和紅外線 CMOS 攝影機產生；系統可以透過彩色影像(如圖 2.7(a)所示)與深度影像(如圖 2.7(b)所示)，將兩者所擁有的資訊整合得出使用者骨架資訊(如圖 2.8 所示)，並辨識出該動作。圖 2.7 之左圖是 Kinect 所得到彩色影像，將其與深度影像資訊整合並經過計算後，可以得到右圖的人體骨架資訊，因此使用者只需透過一般的手勢即可直接操作系統介面。



圖 2.6、Kinect 的外部構造[9]

上述兩種運用不同技術偵測主體動作的方法，optical flow 是利用影像上的光流變化，找出畫面中具 motion 性質的像素點位置；而 Kinect 藉由硬體輔助得到彩色影像與深度影像，將兩個資訊整合後得知使用者的骨架資訊，透過骨架分析來偵測主體動作。而本研究所選用的主體偵測方式為 optical flow，因為演講場合中觀眾人數並非僅只兩三人，而 Kinect 所能得知的主體動作雖然較為細微，但其偵測人數僅限於兩人以下，與本研究的目標不符，而 optical flow 可直接偵測整個觀眾席中具 motion 性質的像素點位置，將這些 motion 像素點位置分析計算後可得到觀眾席中擾動現象最大的區域，較適合本研究的應用。



圖 2.7、利用 Kinect 所得到的影像(a)彩色影像(b)深度影像[李 12]

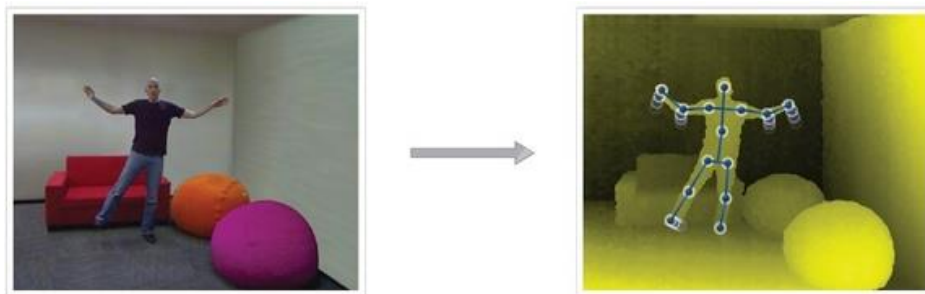


圖 2.8、Kinect 捕捉的骨架資訊[9]

自動化攝影師系統主要目標是選取主體拍攝，因此選取主體的方法將影響系統拍攝上的精彩度，但除了達成選取主體拍攝外，自動化攝影師系統還要能適用於各種拍攝場合，且拍攝畫面需符合專業攝影規則，因此硬體的選用和架設方式與拍攝影像所需應用的技術也是相當重要。下一章將會介紹本研究所選用的硬體與架設環境。

第三章 系統概述

本章將詳細介紹虛擬觀眾攝影師系統的系統架設環境以及流程圖，第一節將描述本系統的硬體結構與架設環境；第二節將介紹本系統的基礎運作流程。

第一節、系統架設環境

為了達成自動化虛擬攝影師拍攝的目的，同時也為了能在各種不同的場地和環境下使用，以及能夠更方便的操控攝影機，所以本研究選用 Pan Tilt Zoom Camera (PTZ 攝影機，如圖3.1所示) 用來作為拍攝觀眾的攝影機。

PTZ 攝影機與傳統攝影機及一般網路攝影機 (Webcam) 不同之處在於，PTZ 攝影機除了具備左右轉動 (Pan)、上下傾斜 (Tilt) 和拉近與拉遠 (Zoom) 等自由地旋轉鏡頭的功能外，同時能讓系統透過網路傳輸直接的遠端操作攝影機各項功能與取得即時影像。



圖 3.1、PTZ 攝影機(a)正面 (c)硬體設備[10]

本研究所選用之 PTZ 攝影機是由 AXIS 公司所製作的 5512 型 PTZ 攝影機 (如圖 3.1 所示)。5512 型 PTZ 攝影機是一種採用半球形鏡頭可大範圍旋轉的網路攝影機，水平旋轉移動範圍 -180 度至 180 度，垂直傾斜移動範圍 -90 度至 90 度，鏡頭縮放倍率可達到 12 倍光學縮放和 4 倍數位縮放，影像設定上可自動化調整顏色、亮度、銳利度、白平衡、曝光控制、背光補償、日夜模式切換等攝影拍攝功能。(AXIS 5512 型 PTZ 攝影機詳細規格見表 3.1 所示)

表 3-1、AXIS 5512 型 PTZ 攝影機詳細規格表[10]

詳細規格-AXIS P5512 / -E PTZ 球型網路攝影機	
攝影機	系統整合
型號	AXIS P5512 60 Hz ; AXIS P5512 50 Hz AXIS P5512-E 60 Hz ; AXIS P5512-E 50 Hz
影像感應元件	1/4吋 CCD
鏡頭	f=3.8 - 46mm、F1.6 - 2.7、自動對焦 自動日夜切換功能 水平視角：51.6°-4.4°
最低照度	彩色：1 lux at 30 IRE F1.6 黑白：0.3 lux at 30 IRE F1.6
快門時間	1/50000秒至4/3秒(60Hz)、 1/50000秒至8/5秒(50Hz)
PTZ	垂直翻轉(E-flip)、自動翻轉、100組預設點 水平：360°旋轉(搭配自動翻轉功能)、100°/秒 垂直：180°、100°/秒 12倍光學縮放和4倍數位縮放、總計48倍縮放
PTZ功能	有限的監守巡代、控制排程、螢幕上指標顯示
影像	系統整合
影像壓縮格式	H.264 (MPEG-4 Part 10/AVC) Motion JPEG
解析度	704 x 480 to 176 x 120 (60Hz) 704 x 576 to 176 x 144 (50Hz)
影像傳送速度	H.264：所有解析度均可達每秒30/25幀 Motion JPEG：所有解析度均可達每秒30/25幀
視訊串流	多重/個別傳送H.264、Motion JPEG影像格式 可控制影像傳送速率(frame rate)與使用的頻寬 可變傳送速率(VBR)/固定傳送速率(CBR) (H.264)
影像設定	手動快門時間、壓縮、亮度、銳利度、白平衡、 曝光控制、背光補強、在低照度下的行為微調、 旋轉、長寬比校正、文字和影像堆疊、在PTZ上的 影像凍結
語音(AXIS P5512 only)	一般
語音串流	雙向語音
語音壓縮格式	AAC-LC 8/16 kHz, G.711 PCM 8 kHz, G.726 ADPCM 8 kHz 可組態位元傳輸率
輸入/輸出	需要多重連接纜線(不包含在其本套件中)，以外接 麥克風/音訊輸入、音訊輸出
網路	一般
安全性	密碼權限保護、IP位址過濾、HTTPS*加密、 IEEE802.1X*網路安全認證機制、摘要驗證(Digest authentication)、使用者存取記錄
支援通訊協定	IPv4/v6, HTTP, HTTPS*, SSL/TLS*, QoS Layer 3 DiffServ, FTP, SMTP, Bonjour, UPnP, SNMPv1/v2c/v3 (MIB-II), DNS, DynDNS, NTP, RTSP, RTP, TCP, UDP, IGMP, RTCP, ICMP, DHCP, ARP, SOCKS
	開放API供軟體整合，包括APIX [®] 及AXIS攝影機應用 平台；ONVIF，規格詳見www.onvif.org 支援AXIS影像代管系統(AVHS)，可搭載一鍵即可連 結攝影機
	智慧型影像 位移偵測、先進守門員功能、AXIS攝影機應用平台 AXIS P5512：語音偵測
	警報觸發 位移偵測、PTZ 預設點、溫度、記憶卡滿載 AXIS攝影機平台 AXIS P5512：外部輸入、語音偵測
	警報事件 透過FTP、HTTP和電子郵件上傳影像 由電子郵件、HTTP 和TCP發出通知 PTZ預設點、自動巡代、本地端儲存 事件前後影像緩衝 AXIS P5512：外部輸入、播放語音片段
	一般
	壓克力(PMMA)透明防護罩
外觀	AXIS P5512：IP51防塵防水等級(鋁製及塑膠外殼) AXIS P5512-E：IP66防塵防水等級和NEMA 4X等級 鋁製及塑膠外殼、遮陽板(PC/ASA)
處理器和記憶體	ARTPEC-3、256 MB記憶體、128 MB快閃記憶體
電源	乙太網路供電(IEEE 802.3af Class 3) AXIS P5512：20-24 V AC，最大23.4 VA；24-34V DC， 最大15.6W RJ-45 for 10BASE-T/100BASE-TX PoE AXIS P5512：針對AC/DC電源，4個可配置的警報輸 入/輸出，內建麥克風，單聲道線路輸出，單聲道 線路輸入置啟動的音箱的多點接器(電纜不包括) AXIS P5512-E：包含IP66等級RJ-45連接器套件
連接器	本地端存取 SD/SDHC記憶卡插槽(不包含記憶卡)
本地端存取	AXIS P5512：0°C至50°C (32°F至122°F)
操作環境	AXIS P5512-E：-20°C至50°C (-4°F至122°F)
安規認證	EN 55022 Class B, EN 61000-3-2, EN 61000-3-3, EN 61000-6-1, EN 61000-6-2, EN 55024, FCC Part 15 Subpart B Class B, ICES-003 Class B, VCCI Class B, C-tick AS/NZS CISPR 22, EN 60950-1, KCC Class B AXIS P5512: IEC 60529 IP51; IEC 60721-4-3 Class 3K3, 3M3, EN/IEC 60068-2 AXIS P5512-E: IEC 60529 IP66, NEMA 250 Type 4X; IEC 60721-4-3 Class 4K1, 4C3, 4M3, EN/IEC 60068-2
重量	AXIS P5512：1.1 kg (2.4 lb.)， 搭配吊掛式支架：1.6 kg (3.5 lb.) AXIS P5512-E：1.2 kg (2.6 lb.)
配件	透明和煙灰色球型防護罩、安裝指南、光碟片(使 用者操作手冊、錄影軟體、安裝和管理工具)、 Windows解碼器(單人版) AXIS P5512：硬式及吊掛式安裝套件 AXIS P5512-E：IP66及RJ-45連接套件

5512型 PTZ 攝影機所具備的功能可以適應各種拍攝場地與環境，包括階梯式場地(如圖1.2(a)所示)與平面式場地(如圖1.2(b)所示)，藉由調整攝影機鏡頭的拍攝角度與攝影機設置於場地的位置來取得較好的拍攝視野與畫面，其中自動調整

的相關功能可以讓系統更輕易的拍攝出較佳的畫面品質。

在本研究中因為觀眾坐落於場地各處，在採用單一攝影機的情況下並無法達到即時拍攝與監控的要求，因而本系統同時使用兩台 PTZ 攝影機來達到較佳的拍攝效果。其架設方式是以一上一下擺放(如圖3.2所示)，該攝影機的架設概念是以攝影學的角度考量後的結果。傳統攝影師在工作時除了操控手中的拍攝鏡頭外，還需要透過攝影師的雙眼，以專業的攝影觀點來觀察全場情境，並選擇適當的目標與運鏡手法來拍攝。[Rui01]的研究中也提到，偵測、辨識與拍攝等工作不應該只單獨由一台攝影機來做處理，應該由兩台以上的攝影機來做分工，以達到較佳的運作效能。

因此本研究的兩台 PTZ 攝影機分別擁有不同的工作內容，擺設於上面的 PTZ 攝影機用來當作攝影師的雙眼，稱為 global-view 攝影機，用以回傳全景畫面提供系統監控整體的會場環境。擺設於下面的 PTZ 攝影機用來當作攝影師操控於手中的拍攝鏡頭，稱為 local-view 攝影機，負責鏡頭移動拍攝與回傳拍攝結果。

本系統透過 global-view 攝影機來收集演講場地中的影像資訊，將所得到的影像資訊經過分析與計算得到影像中的 ROI(詳情請見第四章)與運鏡相關資訊(詳情請見第五章)，然後依所得資訊來操控 local-view 攝影機運鏡拍攝，等到 local-view 攝影機就定位拍攝後，會對所拍攝到的畫面進行專業攝影標準的評估與計算(詳情請見第五章)，最後依照運算結果進行 local-view 攝影機鏡頭微調並將畫面輸出。

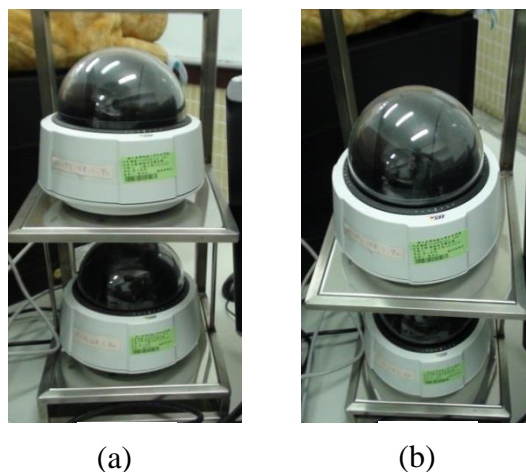


圖 3.2、PTZ 攝影機架設方式，上方是 global-view 攝影機，下方是 local-view 攝影機(a)側視圖(b)俯視圖

第二節、系統流程

本系統運作流程圖(如圖3.3所示)，可分為兩大部份，第一部分是 global-view 的部分，此部分主要處理 global-view 攝影機所提供全景影像中的監控與計算，第二部分是 local-view 的部分，主要處理操作 local-view 攝影機所需要的運鏡與拍攝相關資訊。

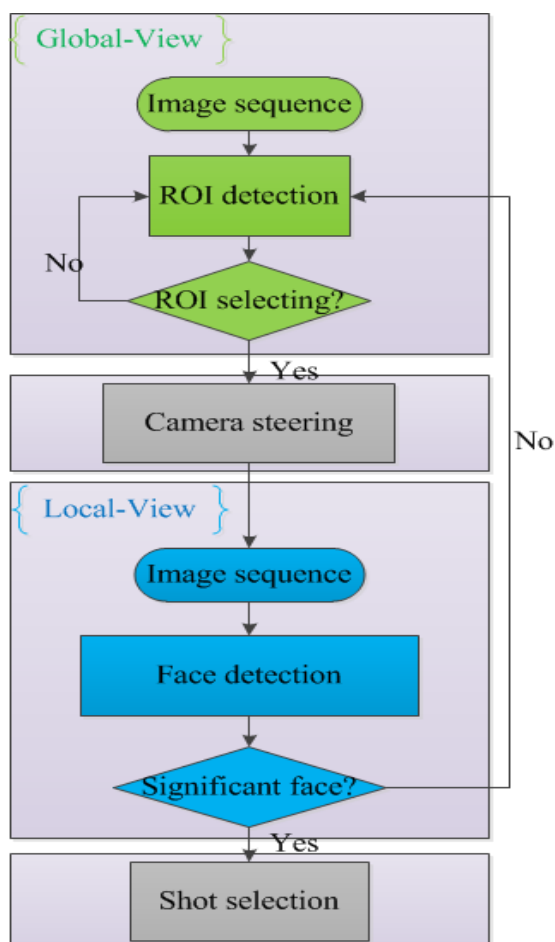


圖 3.3、虛擬攝影師之拍攝觀眾系統流程圖

第一部分系統透過 global-view 攝影機取得演講場合中的連續全景影像，首先需要先找出候選 ROI，候選 ROI 是指畫面中觀眾動作較大與動作較為集中的區域。本研究利用 optical flow 找出畫面中所有 motion 像素點位置與移動距離資訊(詳情請見第四章)，去除雜訊後所得到的結果用以表示影像中觀眾的 motion 資訊。再將所得到的資訊依照密集分散度轉換成 motion 密度圖，motion 密度圖轉換方式是透過 motion 點與其移動距離累計統計而成，依據密度高低給予不同的顏色深淺度，顏色較深的區域 motion 密度較大，顏色較淺的區域 motion 密度

較小，而顏色越深越集中的區域本系統會優先選取為候選 ROI(如圖3.4所示)。圖3.4(a)是利用 optical flow 找出畫面中所有具有 motion 性質的像素點位置；圖3.4(b)是將所找到的 motion 性質的像素點位置與 optical flow 的向量轉換成之密度圖；圖3.4(c)是經由密度圖中找密度最高的區域並選定候選 ROI 之結果。

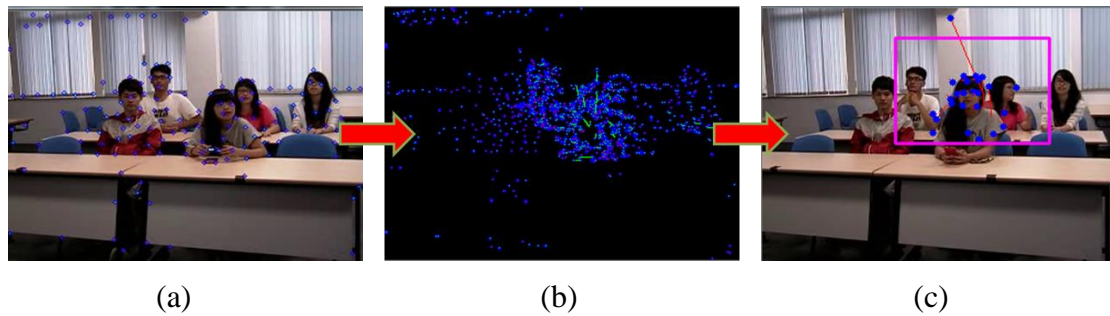


圖 3. 4、系統決定候選 ROI(a)motion 特徵圖(b)motion 特徵密度圖(c)候選 ROI 選取結果

找出候選 ROI 後，需要再決定此候選 ROI 是否適合拍攝，本研究利用 STA 類神經網路(spatio-temporal attention neural model)，STA 能夠紀錄並提供相關資訊來協助系統找出最適合拍攝的 ROI(詳情請見第四章)。將此候選 ROI 輸入 STA 後會先計算候選 ROI 所對應到 STA 輸出層中的激發值是否在安全值以下，如果激發值在安全值以下，則將此候選 ROI 選取為最適合拍攝的 ROI，並將此 ROI 輸入 STA 激發對應區域的激發值，如果激發值在安全值以上，則將此候選 ROI 抑制後並重新找尋新的候選 ROI。

演講場合中會有些持續的動作，這些動作可能造成此區域重複被選取為候選 ROI，而利用 STA 激發值的特性可以解決此問題，STA 可記錄 global-view 攝影機拍攝畫面中曾經被選取為 ROI 的區域，因此可以避免不停的拍攝特定區域，讓其他區域也有被拍攝的可能，而且 STA 會隨著時間降低激發值，因此被抑制過的候選 ROI 再經過一段時間後將有機會再次被選取。圖3.5(a)是系統 ROI 選取結果，ROI 決定後操控 local-view 攝影機運鏡至 ROI 拍攝；圖3.5(b)是 local-view 攝影機運鏡至 ROI 拍攝的結果；圖3.5(c)是將所拍攝的 ROI 輸入 STA 輸出層後之激發狀態圖。

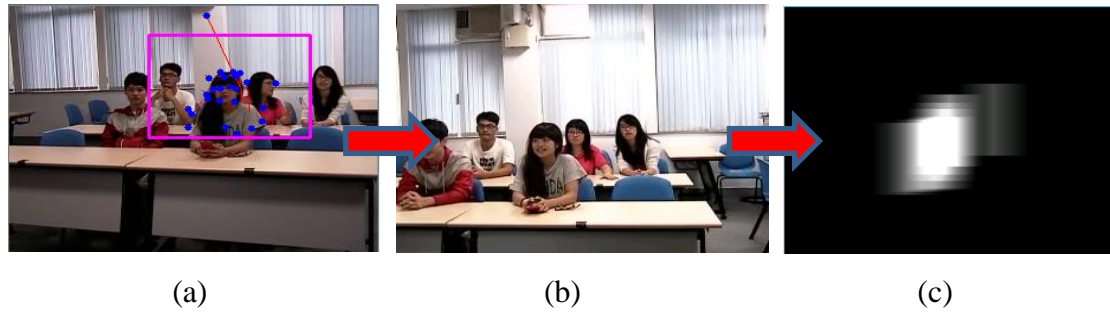


圖 3.5、系統採用 STA 決定 ROI(a)ROI 選取結果(b)運鏡後拍攝畫面(c)STA 輸出層激發狀態

因此第二部分系統會先計算移動至 ROI 所需要的相關運鏡資訊後，操控 local-view 攝影機運鏡拍攝 ROI(詳情請見第五章)，local-view 攝影機到達定位後，進行人臉偵測(詳情請見第五章)。拍攝畫面中可能會有很多的人臉，因此利用所找出人臉中心位置資訊與 ROI 的位置資訊找出最顯著的人臉來做為本次拍攝 ROI 的主體觀眾。由於本系統是以觀眾為拍攝主體，因此在拍攝畫面中觀眾臉部必須要能被完整拍攝到，所以運鏡拍攝時須考量並拍攝到觀眾的臉部位置，否則拍攝畫面將失去主體性(如圖3.6所示)。圖3.6(a)是系統 ROI 選取結果，ROI 決定後操控 local-view 攝影機運鏡至 ROI 拍攝；圖3.6(b)需確定所拍攝畫面中是否有觀眾，因此對畫面做人臉偵測來確定畫面中之觀眾；圖3.6(c)由於所偵測到的觀眾人臉可能有很多，透過計算找出本次主要拍攝人臉。

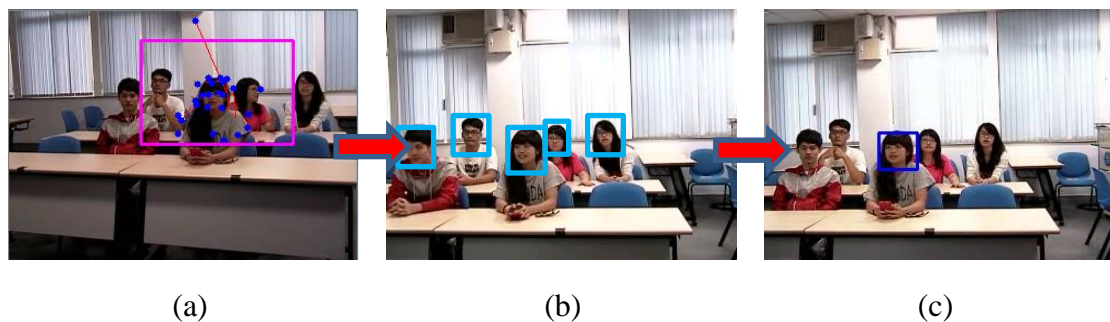


圖 3.6、系統人臉偵測(a)ROI 選取結果(b)運鏡後人臉偵測結果(c)主要拍攝人臉選取

最後本研究的另一個目的是要符合專業攝影學的拍攝標準，將專業攝影學利用美學與光學分析來做評估(詳情請見第五章)，對所拍攝到的畫面做偵測與計算後，操作 local-view 攝影機進行微調，使得拍攝畫面符合專業的攝影標準。

總而言之，本研究利用上述流程來完成拍攝觀眾的目的，對於畫面中觀眾擾

動區域採用 optical flow 特徵做偵測，再根據偵測結果找出候選 ROI，之後將候選 ROI 輸入 STA 中找出適合系統拍攝的 ROI，然後依據系統所計算出的運鏡相關資訊進行運鏡拍攝，攝入的畫面經過人臉偵測再結合畫面的美學分析，最後再微調攝影機取得最佳的拍攝畫面。

第四章 ROI 選取

本系統是以觀眾為拍攝主體，然而觀眾是群體目標並非為個體目標，透過 global-view 攝影機所提供全景影像，並從全體觀眾中找出觀眾動作擾動的區域和找出適合拍攝的 ROI 是本研究的第一要務。首先第一節將介紹觀眾的動作分析；第二節與第三節會說明本系統 ROI 選取方式，從候選 ROI 到決定 ROI 的相關細節。

第一節、觀眾動作分析

本研究的拍攝主體是觀眾，而觀眾的動作分析的結果取決於選取拍攝觀眾區域的優劣。本研究限定觀眾的動作幅度不超過自己所在座位區域。因此依照觀眾的動作幅度大小，可分成靜態動作與動態動作兩個主要類別。而此二類別又可依據做出相同動作的觀眾人數細分成個人動作與群體動作。在此本研究對於個人動作的人數定義是 1 至 3 人，群體動作的人數則為 3 人以上。

透過觀察演講場合中觀眾的動作得知，觀眾產生動作並沒有特定規律與時間，大都與演講內容與演講進度相關，因此本研究擬利用以上區分方式找出觀眾動作中較具代表性也較常出現的幾項動作(如圖 4.1 所示)。圖 4.1 顯示觀眾常見動作包含以下幾項：交談、討論、笑、打瞌睡、點頭表示同意、舉手、起立坐下與鼓掌等動作。

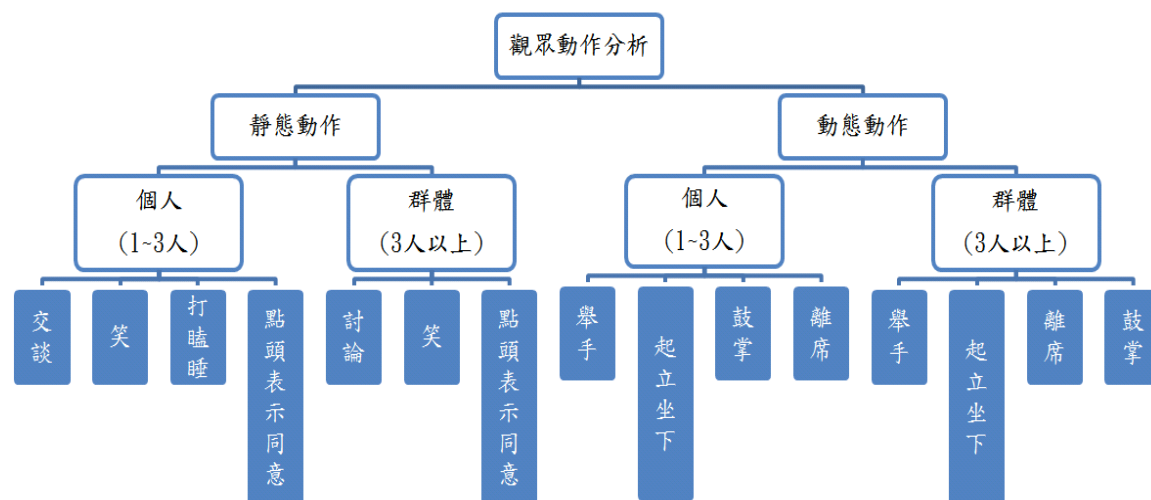


圖 4.1、演講觀眾動作分類圖

根據以上觀眾動作的分類可知，當某些觀眾產生動作時，當下產生動作的觀眾會產生出動作擾動的區域。因此從 global-view 攝影機所提供全景影像中，找出畫面中具有擾動的區域即是本系統所需之候選 ROI。

第二節、候選 ROI 偵測

為了在觀眾全景畫面中找出具有擾動的區域，本研究參考[Luc86]所提出運動向量估計的方法，先從觀眾全景畫面中找出能表示擾動的特徵，再根據這些特徵找出擾動密集區域，將此區域選取為候選 ROI。

擾動的特徵的擷取，本研究選用[Bou00]所提出 Lucas-Kanade optical flow 的方法來進行，因為 optical flow 能找出畫面中具 motion 性質的像素點位置，再依據所找出的 motion 性質的像素點密集度選出候選 ROI。

Lucas-Kanade optical flow 是採用金字塔型的架構，該架構是將影像的解析度分為多個層次，由上至下其解析度則由低至高，先在低解析度的影像進行粗略追蹤，再往底層解析度較高的影像進行較細緻的追蹤。

本節首先介紹一般較常見 optical flow。該方法假設影像上某像素點在極短時間內產生位移時，其 intensity 值並不會改變，此位移的速度向量就是該像素點的 optical flow 方向， $E(x, y, t)$ 表示在時間 t 時，影像中像素點 (x, y) 的 intensity 值，而 Δx 和 Δy 代表經過時間 Δt 的原來落在 (x, y) 位置的像素點位移量，由上述假設可以列出方程式如下：

$$E(x, y, t) = E(x + \Delta x, y + \Delta y, t + \Delta t) \quad (\text{式 4.1})$$

其中，將(式 4.1)右方公式用泰勒展開式展開後可得：

$$E(x + \Delta x, y + \Delta y, t + \Delta t) = E(x, y, t) + E_x \Delta x + E_y \Delta y + E_t \Delta t \quad (\text{式 4.2})$$

接著將(式 4.2)帶回(式 4.1)中並同時除以 Δt 後，化簡可得：

$$E_x \frac{\Delta x}{\Delta t} + E_y \frac{\Delta y}{\Delta t} + E_t = 0 \quad (\text{式 4.3})$$

當取樣時間 Δt 很小時，可將(式 4.3)改寫為：

$$E_x \frac{dx}{dt} + E_y \frac{dy}{dt} + E_t = 0 \quad (\text{式 4.4})$$

設 $u = \frac{dx}{dt}$ 、 $v = \frac{dy}{dt}$ 為影像中像素點 (x, y) 的 x 和 y 的水平與垂直分量的運動速

度，而 (u, v) 即為所求像素點的 optical flow。以下將介紹利用 Lucas-Kanade optical flow 來追蹤特徵點的方法。

此方法考慮一段序列影像 $I, I = \{I_1, I_2, \dots, I_T\}$ ，令 $[P_x, P_y]^t, t \in \{1, 2, \dots, T\}$ 為影像中的一特徵點 P 的位置， $I_t(P)$ 表示影像中特徵點 P 的 intensity 值， τ 為極短的時間變化， Δx 和 Δy 為 P 在 τ 時間內的位移變化量。在假設 P 在間隔相鄰畫面之間的移動夠小且 intensity 值不變的前提下移動到下一個時間的 P' 像素點位置上，可以得知下式：

$$I_t(x, y) = I_{t+\tau}(x + \Delta x, y + \Delta y) \quad (\text{式 4.5})$$

其中 P' 的位置為 $(x + \Delta x, y + \Delta y)$ ， $\vec{v} = (v_x, v_y)$ 是 P 在極短時間內的移動速度，而 Lucas-Kanade optical flow 目的就是要找到最佳的 \vec{v} 值，使(式 4.5)的匹配函數 \mathcal{E} 的誤差值最小。令 w_x, w_y 為 P 周圍的搜尋區域的 $1/2$ 長寬(如圖 4.2 所示)，誤差函數 \mathcal{E} 定義如下：

$$\begin{aligned} \mathcal{E}(\vec{v}) &= \mathcal{E}(v_x, v_y) \\ &= \sum_{x=P_x-w_x}^{P_x+w_x} \sum_{y=P_y-w_y}^{P_y+w_y} (I_{t+\tau}(x + \Delta x, y + \Delta y) - I_t(x + v_x, y + v_y))^2 \end{aligned} \quad (\text{式 4.6})$$

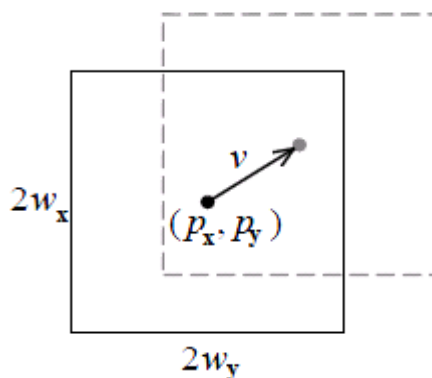


圖 4.2、Lucas-Kanade optical flow 計算範圍差異圖

為了得到 v_x 與 v_y 最佳值 v_{xopt}, v_{yopt} ，將 \mathcal{E} 函數分別對 v_x 與 v_y 偏微並令其為零，可以得到

$$\left. \frac{\partial \mathcal{E}(v_x)}{\partial v_x} \right|_{v_x=v_{xopt}} = 0 \quad (\text{式 4.7})$$

$$\left. \frac{\partial \mathcal{E}(\vec{v})}{\partial v_y} \right|_{v_y = v_{yopt}} = 0 \quad (式 4.8)$$

將(式 4.7)與(式 4.8)經由一次泰勒展開式展開可得

$$\frac{\partial \mathcal{E}(\vec{v})}{\partial v_x} = -2 \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} [(I_{t+\tau}(x+\Delta x, y+\Delta y) - I_t(x, y) - \frac{\partial I_t}{\partial x} v_x - \frac{\partial I_t}{\partial y} v_y)] \cdot \frac{\partial I_t}{\partial x} \quad (式 4.9)$$

$$\frac{\partial \mathcal{E}(\vec{v})}{\partial v_y} = -2 \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} [(I_{t+\tau}(x+\Delta x, y+\Delta y) - I_t(x, y) - \frac{\partial I_t}{\partial x} v_x - \frac{\partial I_t}{\partial y} v_y)] \cdot \frac{\partial I_t}{\partial y} \quad (式 4.10)$$

其中 $I_{t+\tau}(x+\Delta x, y+\Delta y) - I_t(x, y)$ 可解釋為影像對時間的微分

$$\forall (x, y) \in ([p_x - w_x, p_x + w_x], [p_y - w_y, p_y + w_y])$$

$$I_t = \frac{\partial I}{\partial t} = I_{t+\tau}(x+\Delta x, y+\Delta y) - I_t(x, y) \quad (式 4.11)$$

在微分方面，因為影像是數位訊號，其像素座標皆為整數而非連續數值，因此可用有限差分來計算微分值。令影像寬度為 W ，高度為 H 時，則

$$I_x = \frac{\partial I_t}{\partial x} = \begin{cases} I_t(x+1, y) - I_t(x, y) & \text{if } x=1 \\ I_t(x+1, y) - I_t(x, y) & \text{if } 1 < x < W \\ 2 & \text{if } x=W \\ I_t(x, y) - I_t(x-1, y) & \text{if } x=W \end{cases} \quad (式 4.12)$$

$$I_y = \frac{\partial I_t}{\partial y} = \begin{cases} I_t(x, y+1) - I_t(x, y) & \text{if } y=1 \\ I_t(x, y+1) - I_t(x, y) & \text{if } 1 < y < H \\ 2 & \text{if } y=H \\ I_t(x, y) - I_t(x, y-1) & \text{if } y=H \end{cases} \quad (式 4.13)$$

將(式 4.12)與(式 4.13)帶回(式 4.9)與(式 4.10)可得

$$\frac{\partial \mathcal{E}(\vec{v})}{\partial v_x} = -2 \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} [I_t I_x + v_x I_x^2 + v_y I_x I_y] \quad (式 4.14)$$

$$\frac{\partial \mathcal{E}(\vec{v})}{\partial v_y} = -2 \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} [I_t I_x + v_y I_y^2 + v_x I_x I_y] \quad (式 4.15)$$

將所求 (v_x, v_y) 寫成矩陣型式為

$$\begin{bmatrix} \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} I_x^2 & \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} I_x I_y \\ \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} I_x I_y & \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} I_y^2 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} I_x I_t \\ I_y I_t \end{bmatrix} \quad (式 4.16)$$

將(式 4.16)改寫為 $G\vec{v} = -b$ 則 $\vec{v} = -G^{-1}b$ 即為所求結果。

上述式子中 w_x, w_y 的範圍越大表示搜尋的範圍越大，當特徵點的移動較大時，可將搜尋的範圍調大，但也會付出較多的運算時間，本研究中使用的是 $w_x = 10, w_y = 10$ 大小的搜尋範圍。

由於上述是假設相鄰畫面之間特徵點的移動量是相當小的，對於移動距離較大的特徵點無法正確的追蹤，因此 Lucas-Kanade optical flow 的方法會再利用另一種金字塔型架構的影像表示法來克服特徵點移動距離過大的問題。

首先定義 I^L 為影像的第 L 層， $L = \{0, 1, 2, \dots, L_m\}$ ，其中 I^0 為原始影像， I^L 的影像解析度為 I^{L-1} 的一半，以影像解析度 640×480 為例， $L_m = 3$ ，則 I^0 為 640×480 、 I^1 為 320×240 、 I^2 為 160×120 、 I^3 為 80×60 以此類推。利用此金字塔型的影像表示法再以雙線性內插的方式來計算下一層每個像素的強度值，因此可以將計算的精確度提高至半個像素點。以下是此方法的計算公式：

$$\begin{aligned} I^L(x, y) &= \frac{1}{4} I^{L-1}(2x, 2y) \\ &+ \frac{1}{8} [I^{L-1}(2x-1, 2y) + I^{L-1}(2x+1, 2y) + I^{L-1}(2x, 2y+1)] \\ &+ \frac{1}{16} [I^{L-1}(2x+1, 2y) + I^{L-1}(2x-1, 2y+1) + I^{L-1}(2x+1, 2y+1)] \end{aligned} \quad (式 4.17)$$

接著在上述金字塔型的架構中進行追蹤，先計算 L_m 層的特徵點在影像中的移動距離，再依序計算 $L_m - 1$ 層的移動距離，直到第零層為止，圖 4.3 是一個架構為三層的特徵點追蹤示意圖，由高層低解析度至低層高解析度追蹤，圖中不同顏色表示不同特徵，最高層會先找到最初步的特徵，再往下層追蹤，由於解析度提升，因此所上層所追蹤的特徵相對位置可能會改變，並找出原特徵點中附近的新的特徵點，由此方法往下類推追蹤至最低層，利用這樣的方法可以有效的提高整個追蹤器的穩定度。在本論文方法中，所採用的是四層的金字塔架構，也就

是 $L_m = 3$ 。

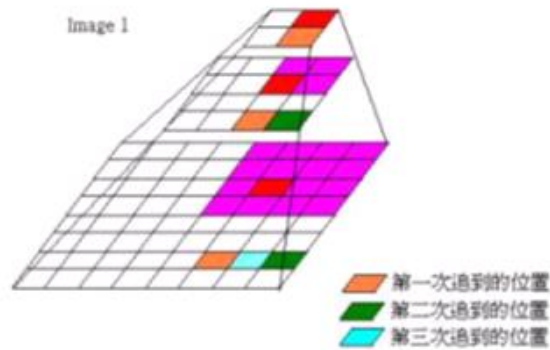


圖 4.3、金字塔型架構追蹤示意圖[謝 95]

利用 Lucas-Kanade optical flow 找出全景觀眾畫面中具 motion 性質的像素點位置，再計算各個像素點間的距離，找出畫面中 motion 點密度最高的區域，將此區域選為候選 ROI。

第三節、ROI 挑選

由於上述方法是將畫面中 motion 點密度最高的區域選取為候選 ROI，而此方法可能會造成畫面中某些持續產生 motion 點的區域重複選取為候選 ROI，因此本研究運用注意力圖像(attention map)來記錄所拍攝過的 ROI，並根據 attention map 中的資訊來決定候選 ROI 是否可成為真正的 ROI，藉此避免 ROI 重複選取拍攝的情形發生。以下將詳述實作方法的內容。

Attention maps 是指在背景是靜態的狀況下，畫面中的動點往往是觀賞者聚集注意力的區域，而此區域與本研究所定義的 ROI 具有相同的性質。因此本研究使用 STA(spatio-temporal attention neural model)[Fan03]來模擬注意焦點集中的過程，並建立注意力圖像(attention maps)(如圖 4.4 所示)。

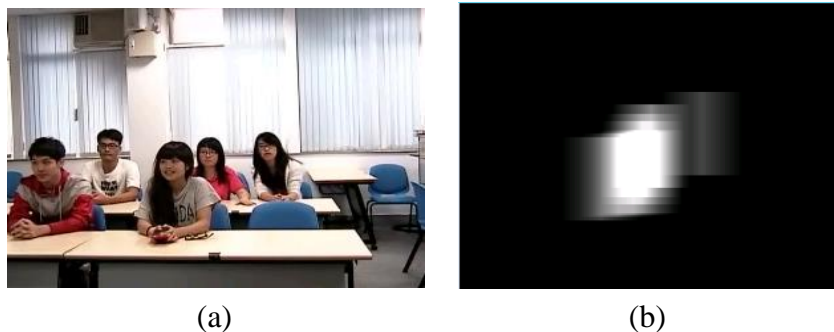


圖 4.4、注意力圖像(a)原始影像(b)根據左圖原始影像建立的注意力圖像

STA 類神經網路分為兩層，一層為輸入層(input layer)，另一為輸出層(output layer)，輸出層亦稱為 attention layer(如圖 4.5 所示)。輸出層中每個神經元都與其鄰近神經元連接，稱之為階層內部連結(within-layer connections)，用來傳遞抑制訊息就是 inhibitory connection。輸入層的每個神經元與輸出層的每個神經元皆有連結稱為 excitatory connection，而這些連結主要用來傳遞刺激訊息，為完全連結(full-connected)。

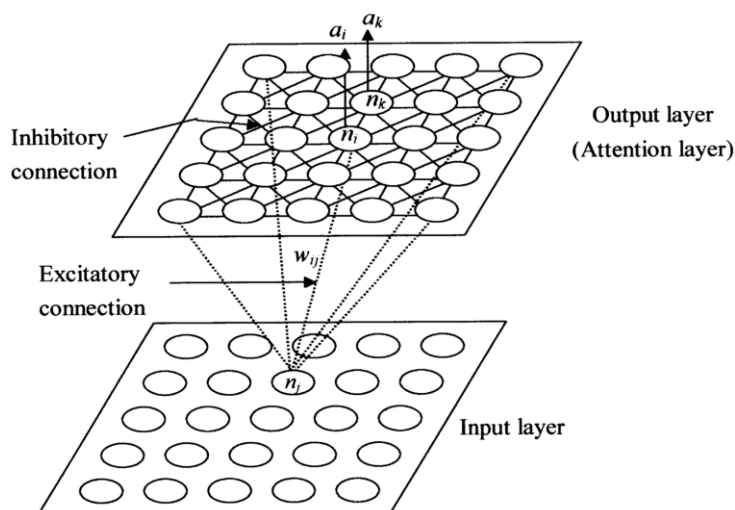


圖 4.5、STA 類神經網路架構圖[Fan03]

輸入層與輸出層的神經元各自組成一個二維陣列，陣列的大小即為輸入影像之大小。令 w_{ij} 表示輸入層神經元 n_j 與輸出層神經元 n_i 連結的權重，而輸入層與輸出層間的連結之權重給定方式，如圖 4.6 所示，假設二維高斯函數 G 的中心位於輸出層神經元 n_i 上，則 n_i 與鄰近的神經元 n_k 形成一位置向量(position vector) r_{ki} ，輸入層神經元 n_j 與輸出層神經元 n_k 的權重 w_{kj} 則可表示為 $G(r_{ki})$ 。二維高斯函數可以使得某個輸入層神經元的刺激在輸入後會刺激到對應的輸出層神經元，進而產生刺激分散的現象。然而輸出層的神經元則會互相抑制，而這種抑制機制可以用「墨西哥帽(Mexican-hat)」函數 $M(r_k)$ 來模擬(如圖 4.7 所示)，其中 r_k 表示神經元 n_k 距離中心神經元 n_i 的距離向量。此種函數的控制方式會造成輸出層的神經元產生激發狀態聚集的現象。

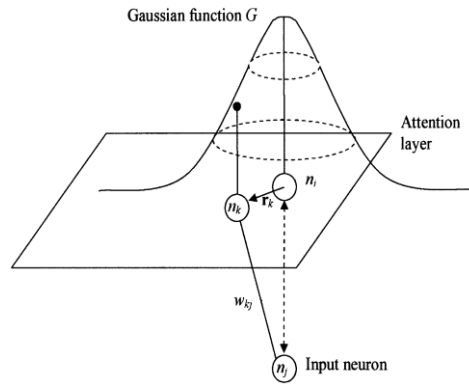


圖 4.6、輸入層與輸出層的權重連結圖

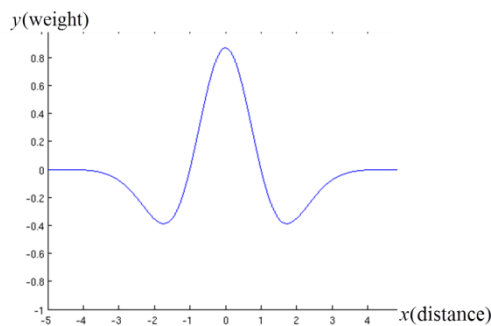


圖 4.7、墨西哥帽函數圖形

運用上述的類神經網路建構出的注意力圖像，如圖 4.4(b)所示。圖 4.4(b)中注意力圖像是一種灰階影像，亮點分佈的區域表示動點出現與系統所拍攝過的位置。STA 連結值向量會經由輸入後激發對應的連結值向量，而其他區域的連結值向量也會被抑制，因而本研究利用此特性，將候選 ROI 與 STA 所對應區域先做相關資訊的比對，確定此候選 ROI 是否重複進行選取拍攝，如果 STA 內此 ROI 對應區域已經被激發，則重新尋找新的 ROI，若 STA 內此 ROI 對應區域未被激發，則將此候選 ROI 選定為正式 ROI 進行運鏡拍攝，並將此候選 ROI 相關資訊作為輸入層輸入 STA 進行對應區域激發。

透過上述方法，先使用 Lucas-Kanade optical flow 可以從 global-view 攝影機所提供全景影像中找出全體觀眾中找出有擾動的 motion 點位置，在找出畫面中 motion 點位置密度最高的區域來當作候選 ROI，再將候選 ROI 透過 STA 相關資訊找出適合拍攝的正式 ROI。

第五章 運鏡路線規劃

系統選定 ROI 後，會啟動 local-view 攝影機進行運鏡拍攝，而其中有關於 local-view 攝影機運鏡到 ROI 進行拍攝的相關參數將在第一節中說明，第二節將介紹運鏡方式與構圖規則。

第一節、攝影機控制

決定 ROI 後，在啟動 local-view 攝影機運鏡前，系統需先決定兩台攝影機的相對運鏡參數，本節將對於兩台攝影機間的運鏡參數計算來做說明。

前述曾提到本系統兩台攝影機是以上下相疊擺放(如圖 3.2 所示)，當 global-view 攝影機輸入之全景影像中之 ROI 被選定後，需要先決定出 local-view 攝影機目前鏡頭的中心位置與 ROI 的中心位置，以及兩者中心位置間運鏡所需要的水平旋轉和垂直傾斜移動角度。假設真實空間中 global-view 攝影機鏡頭中心位置座標為 $G(x_g, y_g, z_g)$ ，local-view 攝影機鏡頭中心位置座標為 $L(x_l, y_l, z_l)$ ，ROI 的中心位置座標為 $R(x_r, y_r, z_r)$ ，global-view 攝影機拍攝 ROI 中心的成像位置座標為 $R'(x_{r'}, y_{r'}, z_{r'})$ 。在計算 local-view 攝影機鏡頭水平旋轉移動角度時(圖 5.1 所示)，由於兩台攝影機是以上下相疊擺放，因此以水平 z 軸視線往 xy 平面垂直觀看，可將兩台攝影機鏡頭中心投影在 xy 平面上，該二投影點會重疊在平面上的同一點，即 $(x_g, y_g) = (x_l, y_l)$ 。另外 ROI 的中心位置 R 與 R' 亦可投影至此平面，分別得到點 T 與 T' ，在此 xy 平面被稱為水平拍攝平面。

經由上述的投影可知 (x_g, y_g) 是攝影機鏡頭中心投影於水平拍攝平面上點 M 之座標， (x_r, y_r) 是 ROI 中心投影於水平拍攝平面時點 T 之座標， $(x_{r'}, y_{r'})$ 是 ROI 中心成像位置投影於水平拍攝平面時點 T' 之座標。圖 5.1 中 $y = y_r$ 直線為真實影像線， (x_{R_L}, y_r) 與 (x_{R_R}, y_r) 是攝影機在水平拍攝平面於真實影像線的拍攝臨界點， $y = y_{r'}$ 直線為影像成像線， $(x_{R'_L}, y_{r'})$ 與 $(x_{R'_R}, y_{r'})$ 是攝影機在水平拍攝平面於影像成像線的拍攝臨界點。此時點 M

對真實影像線 ($y = y_r$) 作垂直線兩者相交於 $M_R = (x_g, y_r)$ ，且點 M 對影像成像線 ($y = y_{r'}$) 作垂直線兩者相交於 $(x_g, y_{r'})$ 。令兩攝影機的焦距皆為 f ， d 是攝影機鏡頭中心到真實影像線的垂直距離， θ_G 是 global-view 攝影機鏡頭於水平拍攝平面拍攝範圍最大的角度， ϕ_L 是欲求得 local-view 攝影機鏡頭中心運鏡至 T 於水平拍攝平面所需的水平旋轉角度。由圖 5.1 中可明顯看出 M_R 與 T 的距離為 $\sqrt{(x_g - x_r)^2}$ ，且 $M_{R'}$ 與 T' 的距離為 $\sqrt{(x_g - x_{r'})^2}$ 。 M_R 與 T 的距離還可利用相似三角形來計算

$$\begin{aligned}\sqrt{(x_g - x_r)^2} &= \frac{\sqrt{(x_g - x_{r'})^2}}{\sqrt{(x_{R_L'} - x_g)^2}} \sqrt{(x_{R_L} - x_g)^2} \\ &= \frac{\sqrt{(x_g - x_{r'})^2}}{\sqrt{(x_{R_L'} - x_g)^2}} (d \tan \theta_G)\end{aligned}\quad (式 5.1)$$

將已知參數帶入(式 5.1)可以求得 d ，然後再將 d 帶入(式 5.2)計算求出 ϕ_L ：

$$\phi_L = \tan^{-1}\left(\frac{\sqrt{(x_g - x_r)^2}}{d}\right)\quad (式 5.2)$$

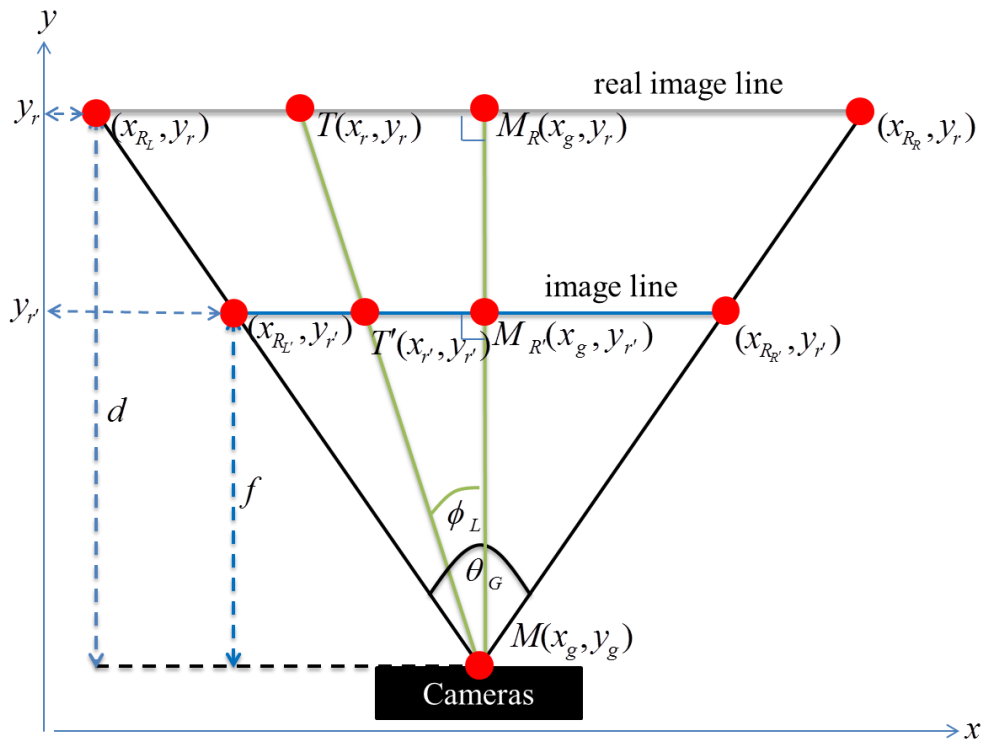


圖 5.1、Global-view 與 local-view 攝影機的水平旋轉移動參數示意圖

在計算 local-view 攝影機鏡頭垂直傾斜移動角度時(如圖 5.2 所示)，由於兩台攝影機是以上下相疊擺放，因此以垂直 z 軸視線往 yz 平面垂直觀看，可將兩台攝影機鏡頭中心投影在 yz 平面上。另外 ROI 的中心位置 R 與 R' 亦可投影至此平面，分別得到點 T 與 T' ，在此 yz 平面被稱為垂直拍攝平面。

經由上述投影可知 (y_g, z_g) 是 global-view 攝影機鏡頭中心投影於垂直拍攝平面上點 G 之座標， (y_l, z_l) 是 local-view 攝影機鏡頭中心投影於垂直拍攝平面上點 L 之座標， (y_r, z_r) 是 ROI 中心投影於垂直拍攝平面上點 T 之座標， $(y_{r'}, z_{r'})$ 是 ROI 中心成像位置投影於垂直拍攝平面上點 T' 之座標。圖 5.2 中 $y = y_r$ 直線為真實影像線， (y_r, z_{T_L}) 與 (y_r, z_{T_R}) 是攝影機在垂直拍攝平面於真實影像線的拍攝臨界點。 $y = y_{r'}$ 直線為影像成像線， $(y_{r'}, z_{T'_L})$ 與 $(y_{r'}, z_{T'_R})$ 是攝影機在垂直拍攝平面於影像成像線的拍攝臨界點。此時點 G 對真實影像線 ($y = y_r$) 作垂直線兩者相交於 $N_R = (y_r, z_g)$ ，且點 G 對影像成像線 ($y = y_{r'}$) 作垂直線兩者相交於 $N_{R'} = (y_{r'}, z_g)$ 。令兩攝影機的焦距皆為 f ， d 是攝影機鏡頭中心到真實影像線的垂直距離， h 是兩台攝影機鏡頭中心的高度差， θ_h 是 global-view 攝影機鏡頭於在垂直拍攝平面拍攝範圍最大的角度， ϕ_h 是欲求得 local-view 攝影機鏡頭中心運鏡至 T 於在垂直拍攝平面所需的垂直傾斜角度。由圖 5.2 中可明顯看出 N_R 與 T 的距離為 $\sqrt{(z_g - z_r)^2}$ ，且 $N_{R'}$ 與 T' 的距離為 $\sqrt{(z_g - z_{r'})^2}$ 。 N_R 與 T 的距離還可利用相似三角形來計算

$$\begin{aligned}\sqrt{(z_g - z_r)^2} &= \frac{\sqrt{(z_g - z_{r'})^2}}{\sqrt{(z_{T'_L} - z_g)^2}} \sqrt{(z_{T_L} - z_g)^2} \\ &= \frac{\sqrt{(z_g - z_{r'})^2}}{\sqrt{(z_{T'_L} - z_g)^2}} (d \tan \theta_h)\end{aligned}\tag{式 5.3}$$

將以知參數帶入(式 5.3)可以求得 d ，然後再將 d 帶入(式 5.4)計算求出 ϕ_h ：

$$\phi_h = \tan^{-1} \left(\frac{\sqrt{(z_g - z_r)^2 + h}}{d} \right)\tag{式 5.4}$$

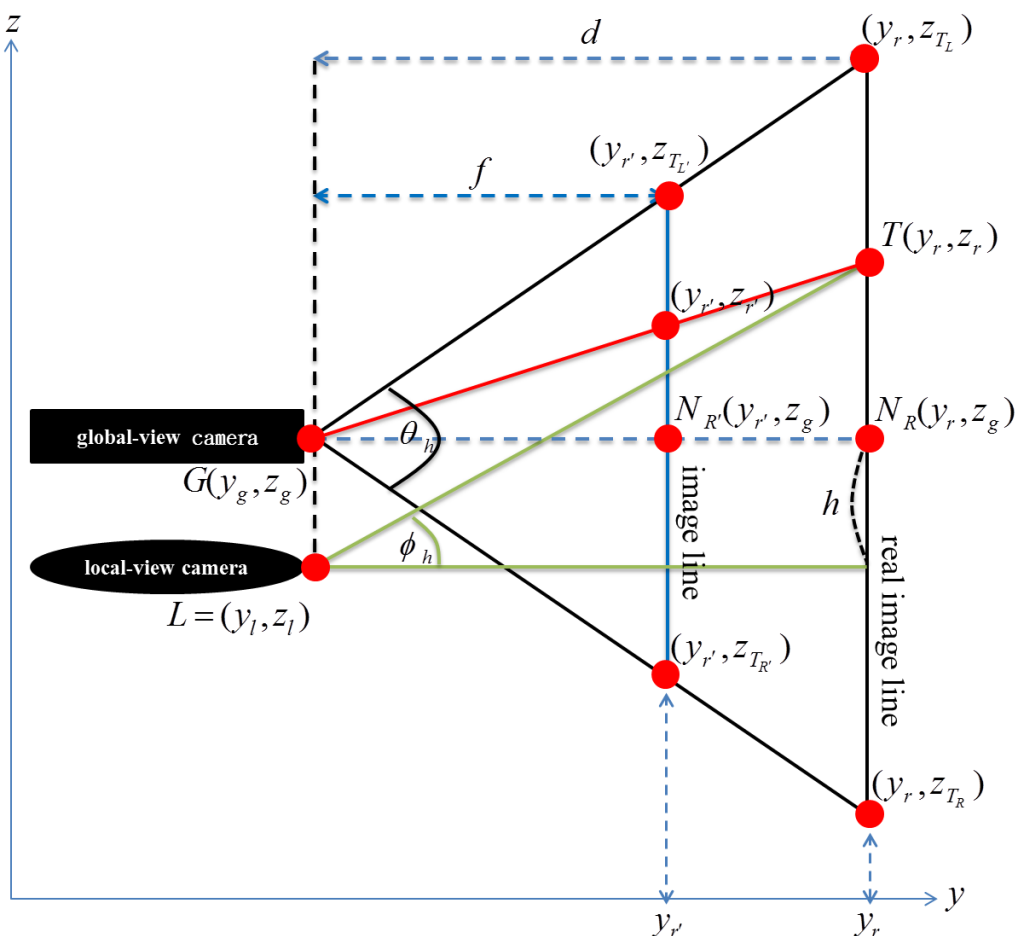


圖 5.2、Global-view 與 local-view 攝影機的垂直傾斜移動參數示意圖

上述式中的參數，除了真實距離 d 和 local-view 攝影機所需求得的水平旋轉角度 ϕ_L 、垂直傾斜角度 ϕ_h 外，其餘皆是已知參數。對於攝影機的移動所需參數只需將已知參數帶入上述式中，所得結果即可得知控制 local-view 攝影機鏡頭運鏡拍攝 ROI 所需要的運鏡參數。

得知 local-view 攝影機鏡頭運鏡拍攝 ROI 所需要的運鏡參數後，需要再決定目前 local-view 攝影機要使用何種運鏡方式前往拍攝 ROI，與 local-view 攝影機拍攝 ROI 畫面後拍攝畫面的構圖規則，這些將在下一節做說明。

第二節、運鏡方式與構圖規則

選定 ROI 後，系統需決定攝影機運鏡前往 ROI 的運鏡方式，選定運鏡方式並操控攝影機拍攝後，本系統會再對所拍攝到的畫面分析構圖與美學標準，依據所分析結果微調攝影機以取得最佳的拍攝畫面與品質。以下將分別對於本系統選定運鏡方式與構圖和美學分析說明。

本研究運鏡方式分為定點拍攝(cut)、推鏡頭(zoom in)、拉鏡頭(zoom out)、水平旋轉(pan)、垂直傾斜(tilt)等運鏡方式(如表 5-1 所示)。其中 zoom in 主要用來掌握運鏡後拍攝畫面中的主體大小；pan 和 tilt 主要配合 ROI 的選定來調整攝影機拍攝 ROI 時的攝影角度，最後根據影像構圖規則來微調影像中主體物位置。

表 5-1、運鏡方式

Camera movement	Chinese name	Definition
Cut	定點攝影	No camera movement.
Zoom in	推鏡頭	A change in the lens focal length with gives the illusion of moving the camera closer.
Zoom out	拉鏡頭	A change in the lens focal length with gives the illusion of moving the camera further away.
Pan right	水平向右	Horizontal movement, right.
Pan left	水平向左	Horizontal movement, left.
Tilt up	垂直向上	Vertical movement of the camera angle, pointing the camera up .
Tilt down	垂直向下	Vertical movement of the camera angle, pointing the camera down.
Pan right (angle A)	沿角度 A 由左向右	Movement along angle A, right.
Pan left (angle A)	沿角度 A 由右向左	Movement along angle A, left.
Pan right (angle B)	沿角度 B 由左向右	Movement along angle B, right.
Pan left (angle B)	沿角度 B 由右向左	Movement along angle B, left.

第一項、主體大小

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。拍攝畫面的主體的大小影響著欲拍攝的目標能否被突現出來，藉由控制主體在畫面中所佔的大小比例，還可以傳達不同含意的主題。例如拍攝主體為觀眾時，為了突顯所選定的觀眾，被選定的觀眾在畫面中的所佔的比例就不能太小，以影像大小的三分之一為佳。

本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。同時臉部也是觀眾席上的觀眾較容易完整擷取的部分。

人臉的偵測方法已經被研究多時，本研究採用的人臉偵測方法是由[Vio04]

所提出的 Adaboost 演算法，其使用的特徵是 Haar features。Haar feature 擷取方式是由兩個或多個黑色和白色的矩形相鄰組成的 Haar masks(如圖 5.3 所示)，計算方式是比較黑色區塊與白色區塊 intensity 值的差異。對於影像中人臉的不同部位使用不同的 Haar mask 會有不同的效果，像是鼻子的 intensity 值總和會大於臉頰的 intensity 值總和等等，Haar features 就是利用這種簡單的概念擷取出的特徵。此外，若搭配積分影像(integral image)技術還能加速 Haar features 擷取時的計算，減少許多執行時間。

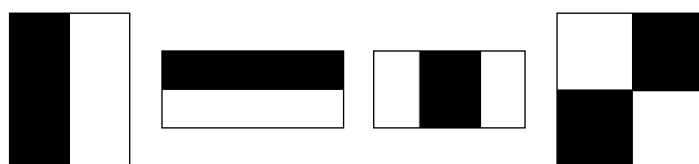


圖 5.3、Haar masks 之範例

利用 integral image 可加速 Harr feature 的擷取，傳統 Harr feature 擷取使用如圖5.3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，以這些和的差值做為特徵。傳統作法的缺點就是計算時間過長，利用 integral image 可解決此問題。Integral image 實際上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5.4(a)中座標點 (x, y) 所在位置之元素值為左上灰色部分區域內 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即可，藉此可以大幅度地降低運算時間。

在 integral image 中，對於一個輸入影像將其表示為一個矩陣 I ，矩陣 I 中之元素 (x, y) 的 integral image 代表向左邊界至上邊界所圍起來的區塊的 intensity 值總和(如圖 5.4(a)所示)，定義如下：

$$I(x, y) = \sum_{1 \leq i \leq x, 1 \leq j \leq y} I'(i, j) \quad (\text{式 5.9})$$

其中 $I(x, y)$ 代表 integral image 中元素 (x, y) 的值， $I'(i, j)$ 代表輸入影像中像素點 (i, j) 的 intensity 值。如此定義，建立好 integral image 後，便可快速求得輸入影像中一矩形內的 intensity 值總和。以圖 5.4(b)為例，欲計算 D 區域內的 intensity 值總和，必須先在 integral image 中取得點 P_1 、 P_2 、 P_3 、 P_4 的值，分別表示為：

$$P_1 = A \quad (式 5.10)$$

$$P_2 = A + B \quad (式 5.11)$$

$$P_3 = A + C \quad (式 5.12)$$

$$P_4 = A + B + C + D \quad (式 5.13)$$

其中 P_1 代表 A 區域的 intensity 值總和、 P_2 代表 A 區域加上 B 區域的 intensity 值總和，以此類推。之後利用簡單的加減，即可算出 D 區域內的 intensity 值總和。

$$P_4 - P_3 - P_2 + P_1 = (A + B + C + D) - (A + C) - (A + B) + A = D \quad (式 5.13)$$

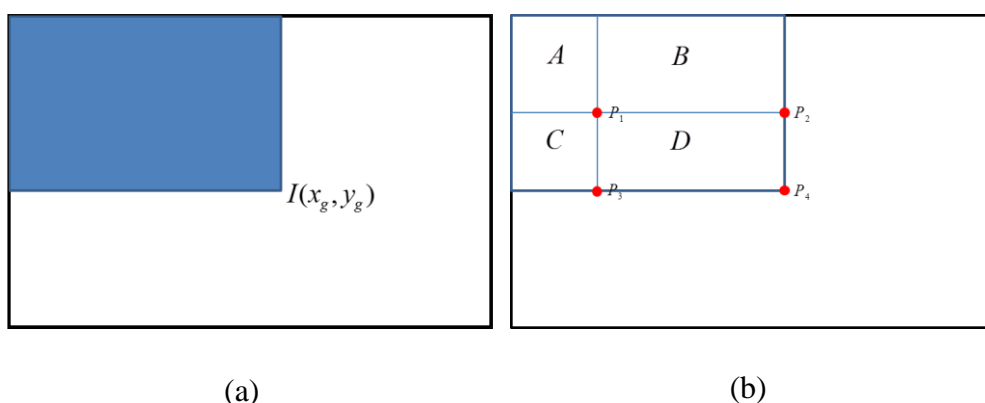


圖 5.4、Integral image 計算之圖例(a)Integral image 中 $I(x, y)$ 之 intensity 值所涵蓋的計算範圍(b)Integral image 中 D 範圍的 intensity 值總和之計算示意圖

藉由 integral image，Haar feature 擷取便可更快速的求出黑色區塊和白色區塊內的 intensity 值總和，再相減求出特徵值。

Adaboost 演算法的基本原理是對整張影像以上述人臉偵 polarity 測的方法，進行地毯式的人臉特徵比對，並從中擷取出不同尺寸、位置的區塊，經由訓練出來的分類器判斷是否為人臉，最後輸出符合分類器檢測的人臉區域。Adaboost 演算法所訓練出來的分類器，是由多個弱分類器(weak classifier)組合而成的一個強分類器(strong classifier)，其原理是在每輪的運算決定一個弱分類器，最後將這些弱分類器組成一個強分類器(如圖 5.6 所示)。

-
- A. Given example images $(x_1, y_1), \dots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
 - B. Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
 - C. For $t = 1, \dots, T$:

1. Normalize the weights, $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$
2. Select the best weak classifier with respect to the weighted error

$$\epsilon_t = \min_{f,p,\theta} \sum_i w_i |h(x_i, f, p, \theta) - y_i|.$$

3. Define $h_t(x) = h(x, f_t, p_t, \theta_t)$ where $f_t, p_t,$ and θ_t are the minimizers of ϵ_t .
4. Update the weights:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

- D. The final strong classifier is:

$$C(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha_t = \log \frac{1}{\beta_t}$

圖 5.5、Adaboost 演算法[Vio04]

根據先前提到每個矩形特徵根據不同大小、位置皆可以當作一個弱分類器，但並非每個弱分類器皆是對於人臉的分類判斷是有益處，所以我們需要一個方法來判斷出這些弱分類器是否能辨別人臉，而 Adaboost 演算法即是拿來進行特徵的挑選，選出對於人臉辨識有用的特徵，如此可大幅降低系統的計算量。

圖 5.5 中 $h(x_i, f, p, \theta)$ 表示一個弱分類器，其中 x 表示所輸入影像， f 表示 feature， p 是表示黑白區塊的方向， θ 表示一個 threshold，而弱分類器分類

函數 $h(x_i, f, p, \theta) = \begin{cases} 1 & pf(x) < p\theta \\ 0 & \text{otherwise} \end{cases}$ 。接下來將解釋在圖 5.5 中演算法的步

驟，在初始的時候先給予一系列的訓練樣本， x_i 為輸入的影像而 y_i 為 0 或是 1 的值，當 $y_i = 0$ 時表示輸入的影像不是人臉，而當 $y_i = 1$ 時輸入的影像則為人臉，而 $y_i = 0$ 與 $y_i = 1$ 的樣本數分別有 m 與 l 個。接下來對每個樣本的權重進行初始

化，初始化方式是以 y_i 為判定，如果 $y_i = 0$ 則 $w_{1,i} = \frac{1}{2m}$ ，如果 $y_i = 1$ 則

$w_{1,i} = \frac{1}{2l}$ ，而在第 C 個步驟時，先對於每個樣本進行正規化的動作

$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$ ，使得所有樣本的權重值加總為 1，然後將每個樣本送入所有弱

分類器中並對每個樣本進行分類的動作，接著去計算所有弱分類器的錯誤率，最後挑選錯誤率小的弱分類器 $\varepsilon_t = \min_{f,p,\theta} \sum_i w_i |h(x_i, f, p, \theta) - y_i|$ ，並定義此較

強的弱分類器中所對應的 f 、 p 與 θ ，最後更所有樣本的的權重

$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$ ，其中 $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$ ，而 e_i 會根據分類結果做最後調整， $e_i = 1$ 表

示本次挑選的較強的弱分類器對於樣本分類結果正確，如果分類結果錯誤則 $e_i = 0$ ，將上述步驟重複 T 次可以找出 T 個較強的弱分類器。而此步驟是相當重要的，因為在此步驟最後會決定 T 個較強的弱分類器，而強分類器即為較強的弱分類器組合而成(如圖 5.6 所示)。

利用上述結果所結合的強分類器來檢測人臉影像時，檢測方式是讓所有較強的弱分類器都對影像去進行偵測，再將每個較強的弱分類器的偵測結果乘上其相對應的權重 w_i ，最後將所得結果全部相加求得一總和，比較所得總和與預設閾值以獲得最終判斷結果。強分類器之判斷公式如式 5.14 所示：

$$C(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases} \quad \text{where } \alpha_t = \log \frac{1}{\beta_t} \quad (\text{式 5.14})$$

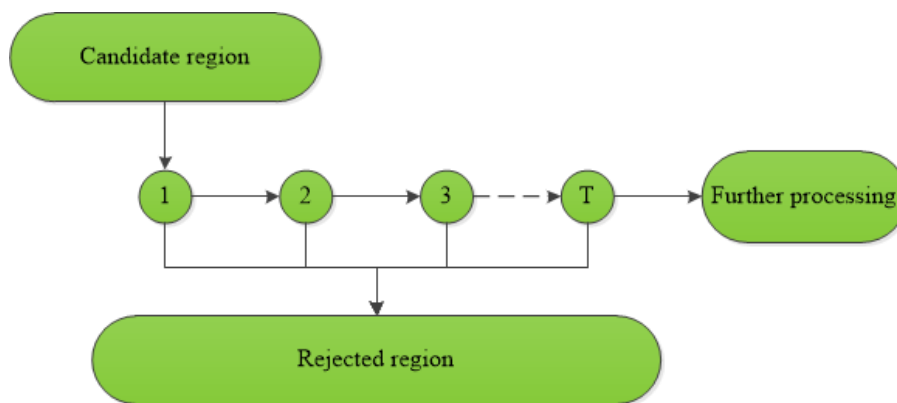


圖 5.6、多層分類器架構(正圓形為分類器的篩選階段)[13]

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。最後找出所記錄的人臉座標中離 ROI 中心最近的人臉，並將其選定為本次 ROI 拍攝的主體目標，再操控攝影機 zoom in 直到拍攝畫面中該人臉的大小佔據畫面的 $\frac{1}{10}$ ，藉此方法來達成專業拍攝標準中主體大小的標準。

第二項、影像構圖規則

構圖指的是畫面中空間的配置，被拍攝的目標佔據畫面的比例與在畫面中的相對位置安排是奠定畫面可看性的基礎。系統完成主體大小調整後，還需要針對影像構圖的規則來做影像的微調，以下將對影像構圖的規則來說明。

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。拍攝的目標佔據畫面的比例則影響著欲拍攝的主題能否被突現出來，藉由控制主體在畫面中所佔據畫面的比例，來傳達不同含意的主題。畫面中的相對位置安排，一般較常用的構圖方法有「三分構圖法」和「井字構圖法」[7][8]又稱為三一法則(rule of third)，或被稱為「黃金分割法」，三一法則的概念是運用兩條水平線及兩條垂直線排成「井」字形，將影像切割成大小相同的九等分，這四條線則稱之為「黃金分割線」，而黃金分割線所交叉形成的四個點，稱作「黃金點」。而畫面中主體位置與黃金點越接近，或是主體延伸的線條(例如地平線)越貼近黃金分割線，皆能夠在視覺上獲得舒適平穩的感覺(如圖 5.7 所示)。圖 5.7 中左圖主體位置與四個黃金分割點的距離較近，給觀賞者較佳的觀賞舒適感；右圖主體位置與四個黃金分割點的距離較遠，給觀賞者主體偏移的感覺。

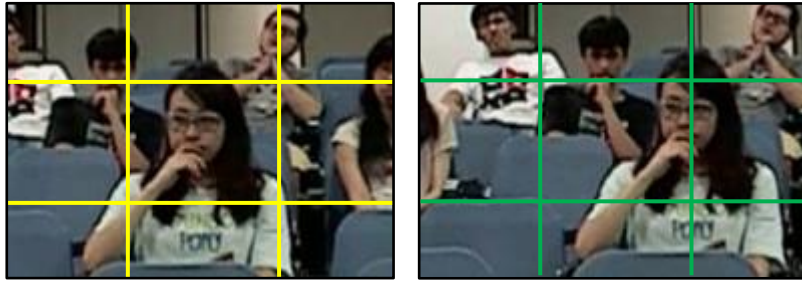


圖 5.7、三一法則比較圖

影像構圖規則還包含在光線的運用上，不同的光線能夠幫助所拍攝出的影像帶給觀賞者不同的感覺。例如攝影師希望景物清晰呈現時，可使順光，順光即為讓光線從鏡頭對主體的方向發出，能夠讓主體更為清晰易見並使得色彩鮮麗且飽和。但是順光會造成主體上沒有陰影，因此較欠缺層次及立體感，而為了增加層次與立體感，可以利用從主體側邊打上的側光，側光能令主體更顯著與展現景物細節質感。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。本研究礙於硬體上的限制，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 5.8、顏色的對比與色系(a)以鮮明的對比突顯主體[7](b)(c)不同的整體色調帶來不同的視覺效果[8]

最後影像構圖規則也包含在色彩方面，簡單的色調和鮮明的對比能襯托出前景的存在感(如圖 5.8(a)所示)。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 5.8(b)所示)，冷色系則帶給觀眾理性、清楚的感受(如圖 5.8(c)所示)。但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。

本研究是以觀眾為主體，其主體判別方式主要是以觀眾的人臉作為主體位置，藉由將觀眾人臉位置調整至符合上述構圖規則，並利用攝影機所具備的相關自動調整功能，讓所拍攝的畫面更具美感與品質，給予觀賞者更佳的觀看效果。

第六章 實驗結果

本章將透過實驗來說明本系統實際拍攝結果與拍攝品質，實驗方式主要是以問卷方式進行，針對常參與演講的觀眾來進行問卷，測試者觀看本系統拍攝結果，並以其參與演講之經驗來回答相關問題，最後透過問卷結果來展現本系統拍攝品質。本研究實驗項目依據研究目標共分成四個項目：(1)ROI 選取之正確率(2)本系統運鏡評估(3)與過去研究拍攝結果比較(4)本系統對於拍攝場地的適用性。以下將針對上述四項實驗方法與結果來做說明。

第一節、ROI 選取的正確率

本實驗目的是測試拍攝演講時的觀眾時，本系統自動選取之 ROI 是否與觀賞者欲觀看之 ROI 相符。本實驗將針對本系統自動選取之 ROI 結果與觀賞者所選定之 ROI 結果進行討論，實驗找了 30 位常參與演講之觀眾來進行實驗(如表 6-1 所示)，由表 6-1 中可以得知本實驗測試者平均每月參與演講次數皆在 2 次以上，本實驗希望藉由測試者每個人參與演講的經驗來選定其欲觀看之 ROI。

實驗方法是分別給予 30 位測試者觀看相同的影片，影片內容主要可以分成三段，三段影片中皆是不同時間與不同場景下所拍攝演講時觀眾的全景影像，由於演講時觀眾動作重複性很高，因此此三段影片皆是由演講所拍攝的觀眾影像中擷取之代表片段。本實驗請測試者在觀看影片的過程中可同時選取其觀看畫面中欲觀賞之 ROI，並將其所選取之 ROI 記錄下來，最後將每位測試者所選取之 ROI 與系統所選取之 ROI 做比較。如果觀賞者所選取之 ROI 與系統自動選取之 ROI 兩者重疊範圍達 70% 以上，則視為系統所選 ROI 與觀賞者所選 ROI 相符，並視系統 ROI 選取正確，反之則視為系統 ROI 選取錯誤，最後將所有測試結果進行統計並計算本系統 ROI 選取之正確率。本實驗將依據第四章第一節所討論觀眾動作整理之結果，先將觀眾動作分為靜態動作與動態動作兩者來分開討論，並針對每位測試者所選取之 ROI 結果來做分析。以下將分別對三段影片內容中，系統自動選取之 ROI 與測試者所選取之 ROI 來進行討論。

表 6-1、實驗測試者資料表

姓名	性別	每月參與演講平均次數
吳 X 仁	男	2~3 次/月
張 X 雨	女	2~3 次/月
吳 X 正	男	2~3 次/月
林 X	男	3~4 次/月
黃 X 威	男	3~4 次/月
林 X 達	男	3~4 次/月
宋 X 逸	男	3~4 次/月
范姜 X	男	3~4 次/月
孫 X 傑	男	3~4 次/月
高 X	男	3~4 次/月
區 X 豪	男	3~4 次/月
鄭 X 謙	男	3~4 次/月
王 X 賢	男	3~4 次/月
廖 X 愷	男	3~4 次/月
劉 X	男	3~4 次/月
黎 X 琦	男	3~4 次/月
劉 X 甫	男	3~4 次/月
周 X 成	男	3~4 次/月
蔡 X 璋	男	3~4 次/月
林 X 立	男	3~4 次/月
陳 X 多	男	3~4 次/月
廖 X 凡	男	3~4 次/月
李 X 漢	男	3~4 次/月
何 X 源	男	3~4 次/月
翁 X 寧	女	3~4 次/月
江 X 男	男	3~4 次/月
李 X 禹	男	3~4 次/月
陳 X 均	男	3~4 次/月
朱 X 琪	女	4~6 次/月
張 X 桂	女	4~6 次/月

第一項、第一段影片實驗結果

第一段影片拍攝場景是位於國立師範大學工館校區應用科技大樓地下室 B1 的演講廳，其觀眾席座位是屬於階梯式觀眾席，拍攝時間是演講進行的實際場景，擷取影片的總 frame 數是 13978 張 frames，而每位測試者觀看第一段影片時所圈選 ROI 之 frame 數是 6989 張 frames，所有測試者觀看第一段影片時所圈選 ROI 之 frame 總數是 209670 張 frames。影片中包含許多觀眾常見的動作，如舉手、討論、鼓掌、離席等動作(如圖 6.1 所示)。圖 6.1 中圖片是依據影片時間序列排序，依據拍攝時間先後由左至右再由上至下排序。

表 6-2、實驗一第一段影片所有測試者與系統圈選 ROI 之比較結果

	圈選 frame 總數	靜態動作正確 率	動態動作正確 率	整段影片正確 率
場景一	209670	69.4%	91.3%	80.35%

第一段影片中由測試者手動圈選 ROI 與系統自動選取 ROI 兩者比較之平均正確率結果如表 6-2 所示。由實驗結果可以得知，第一段影片是拍攝演講時的觀眾，而演講時的觀眾動作大都是靜態動作，因此當觀眾具有動態動作時，能較輕易的吸引測試者視線，而測試者也大都選取該區域為 ROI，因此測試者對於當觀眾具有動態動作時所選取之 ROI 與系統自動選取之 ROI 兩者較相符，而動於觀眾具有動態動作時，系統自動選取 ROI 的正確率達到 91.3%；而測試者對於觀眾具有靜態動作時，測試者所選取之 ROI 與系統自動選取之 ROI 兩間者具有較大的差異，原因是因為當演講者進行演講時，觀眾大都是靜態動作，對於靜態動作時測試者之間所選之 ROI 也較為分散，因為有些測試者並非只注意到觀眾具有 motion 的區域，有時反而著重於專心聽講而毫無動作的區域或是注意特定的人物，而本系統 ROI 選取主要以觀眾具有 motion 的區域，因此當測試者選取觀眾毫無 motion 的區域時或特定人物時，會造成系統自動選取之 ROI 與測試者所選之 ROI 相異，因此降低了對於觀眾具有靜態動作時系統自動選取 ROI 的正確率，系統對於觀眾靜態動作時之選取 ROI 的正確率達到 69.4%，而系統對於第一段影片之 ROI 選取正確率達到 80.35%。

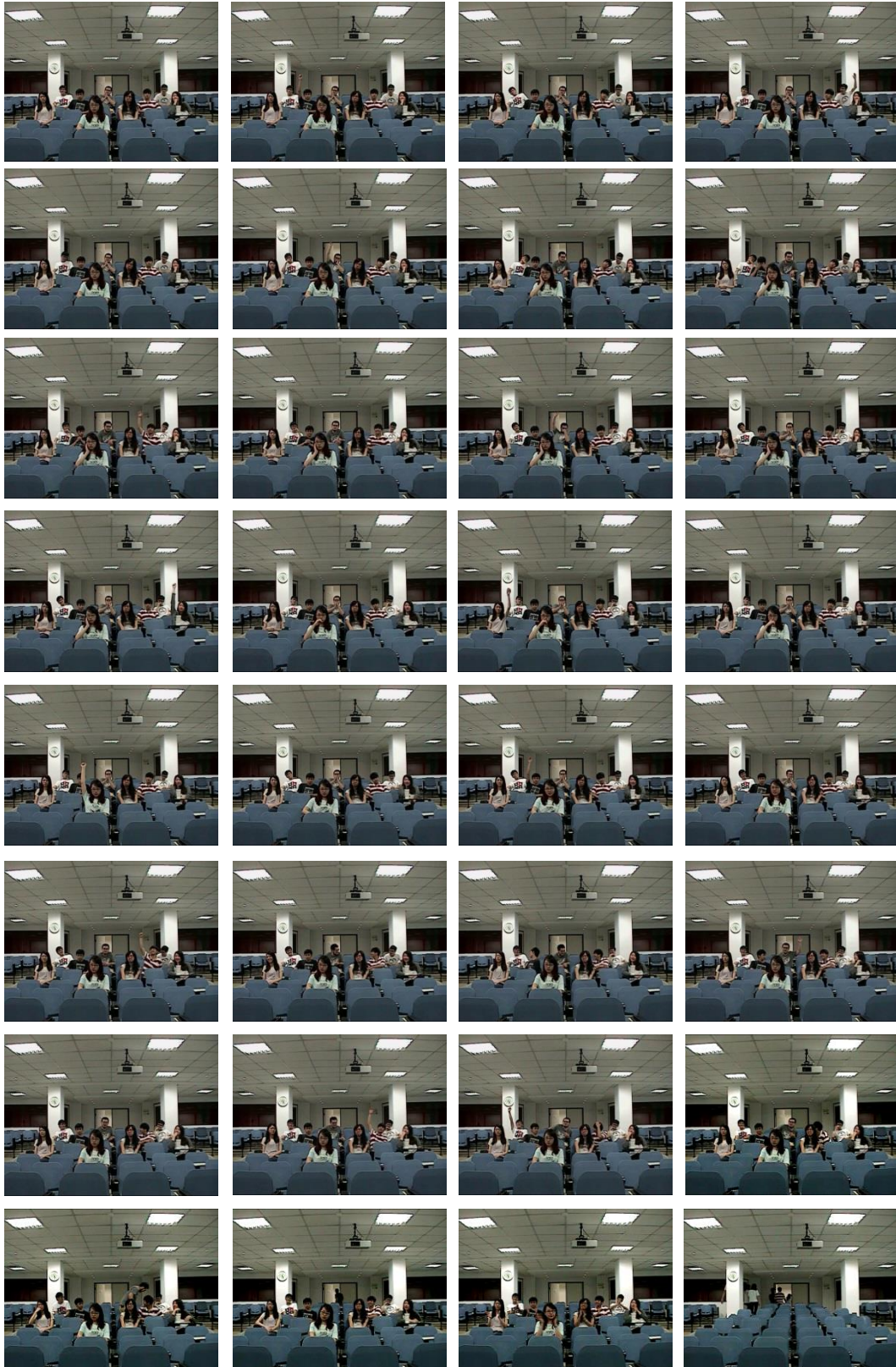


圖 6.1、實驗一第一段影片擷取片段

第二項、第二段影片實驗結果

第二段影片拍攝場景是位於國立師範大學工館校區教學大樓B棟103教室，其觀眾席座位是屬於階梯式觀眾席，拍攝時間是統計課程教師授課進行時的實際場景，影片總 frame 數 11263 張 frames，而每位測試者觀看第一段影片時所圈選 ROI 之 frame 數是 5631 張 frames，所有測試者觀看第二段影片時所圈選 ROI 之 frame 總數是 168930 張 frames。影片中包含許多觀眾常見的動作，如舉手、討論、起立等動作(如圖 6.2 所示)。圖 6.2 中圖片是依據影片時間序列排序，依據拍攝時間先後由左至右再由上至下排序。

表 6-3、實驗一第二段影片結果

	圈選 frame 總數	靜態動作正確 率	動態動作正確 率	整段影片正確 率
場景二	168930	72.6%	84.7%	78.65%

第二段影片中由測試者手動圈選 ROI 與系統自動選取 ROI 兩者比較之平均正確率結果如表 6-3 所示。由實驗結果可以得知，第二段影片是拍攝授課時聽講的觀眾，授課時聽講的觀眾與授課教師之間的互動較為頻繁，且觀眾也可能同時出現動態動作，例如：兩人同時舉手發問等情形，對於此情形發生時，測試者間所選取之 ROI 也差異較大，此情形會降低系統 ROI 選取之正確率，因此對於觀眾動態動作時系統之 ROI 選取正確率達到 84.7%；而對於觀眾靜態動作上，由於此段影片是拍攝授課時的觀眾，授課時的觀眾大都是專心聽講而較無明顯動作，因此當觀眾具有靜態動作時，測試者能較清楚的圈選出具有靜態動作之 ROI，因此對於觀眾靜態動作時系統之 ROI 選取正確率達到 72.6%，而系統對於第二段影片之 ROI 選取正確率達到 78.65%。

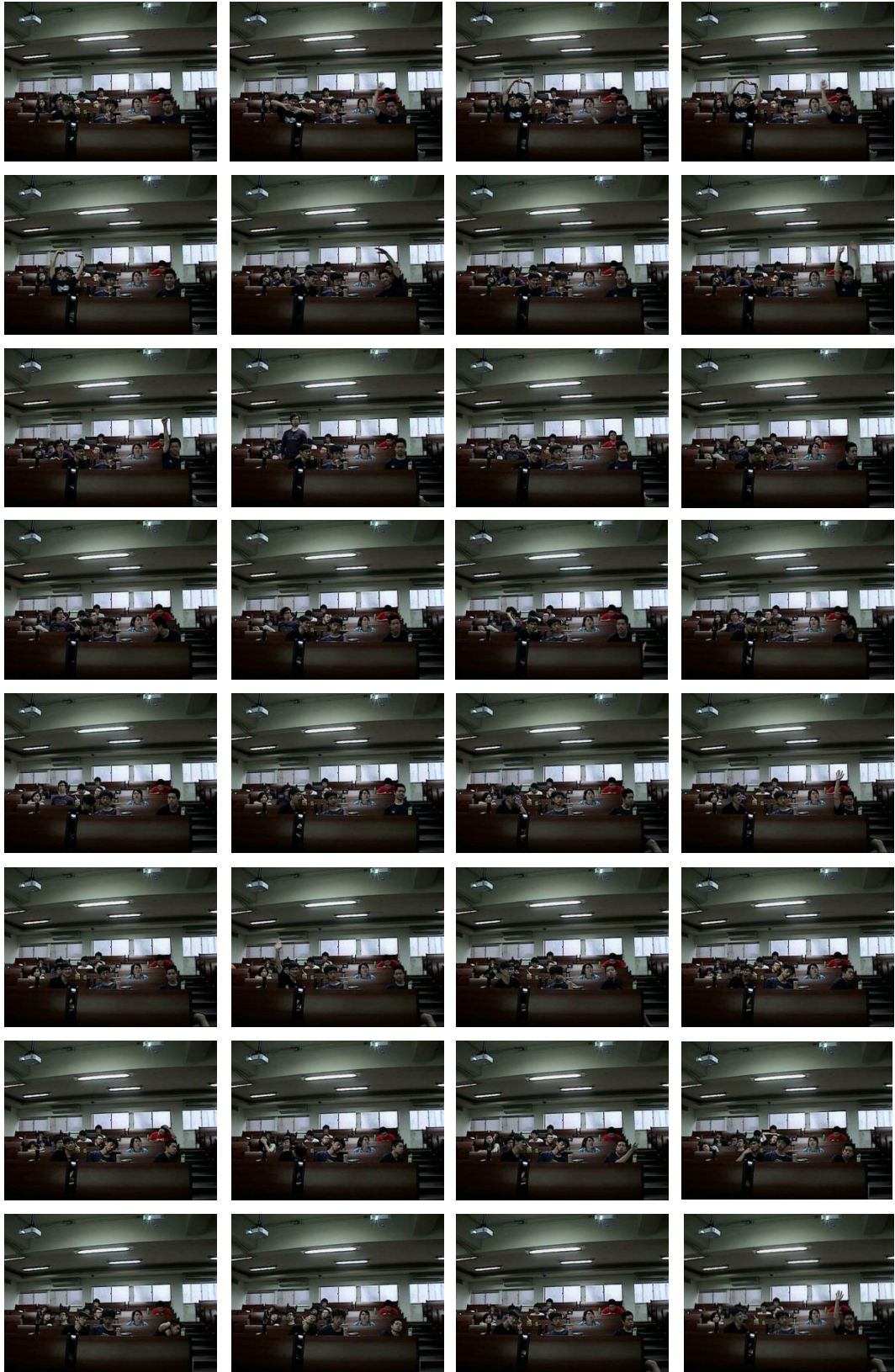


圖 6.2、實驗一第二段影片擷取片段

第三項、第三段影片實驗結果

第三段影片拍攝場景是位於國立師範大學工館校區教學大樓C棟209教室，其觀眾席座位是屬於平面式觀眾席，拍攝時間是演講者進行論文講解時的實際場景，影片總 frame 數 3341 張 frames，而每位測試者觀看第一段影片時所圈選 ROI 之 frame 數是 1670 張 frames，所有測試者觀看第三段影片時所圈選 ROI 之 frame 總數是 50100 張 frames。影片中包含許多觀眾常見的動作，如舉手、討論等動作（如圖 6.3 所示）。圖 6.3 中圖片是依據影片時間序列排序，依據拍攝時間先後由左至右再由上至下排序。

表 6-4、實驗一第三段影片結果

	圈選 frame 總數	靜態動作正確 率	動態動作正確 率	整段影片正確 率
場景三	50100	78.5%	92.6%	85.55%

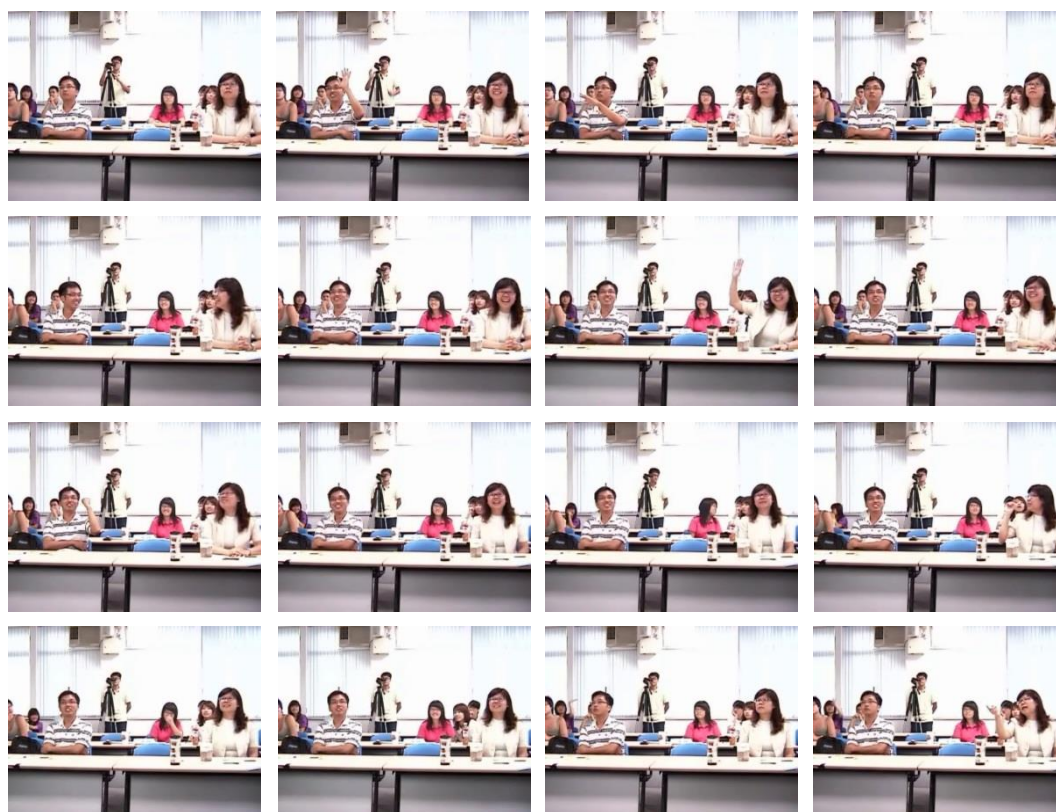


圖 6.3、實驗一第三段影片擷取片段

第三段影片中由測試者手動圈選 ROI 與系統自動選取 ROI 兩者比較之平均正確率結果如表 6-4 所示。由實驗結果可以得知，由於第三段影片是論文講解時聽講的觀眾，由於觀眾的人數相較於前兩段影片是比較少的，且觀眾座落位置較

為分散，因此當觀眾具有動態與靜態動作時觀賞者能清楚的圈出 ROI，對此提升了系統對於 ROI 選取之正確率，因此對於觀眾動態動作時系統之 ROI 選取正確率達到 92.6%，對於觀眾靜態動作時系統之 ROI 選取正確率達到 78.5%，而系統對於第二段影片之 ROI 選取正確率達到 85.55%。

表 6-5、實驗一平均正確率結果

	Frame 數	靜態動作正確率	動態動作正確率	整段影片正確率
實驗一	28582	73.5%	89.53%	81.52%

由實驗一之三段影片中可知，對於觀眾動態動作時，由於第三段影片所拍攝觀眾較少，因此測試者對於此段影片之 ROI 選取之正確率較高，而第二段影片由於觀眾可能同時出現動態動作，因此第一段影片之 ROI 圈選的正確率較第二段影片高；而對於觀眾靜態動作時，由於第二段影片是授課時的觀眾，授課時的觀眾大都專心聽課較無靜態動作，而此時具有靜態動作的觀眾可以很容易的被測試者選取出來，因此而提高了觀眾靜態動作時系統自動選取之 ROI 與測試者手動圈選 ROI 的正確率，所以對於觀眾的靜態動作與系統自動選取之 ROI 正確率與第一段影片正確率有些微提升。

表 6-5 是本系統拍攝各種場合中自動選取的 ROI 之正確率，對於觀眾靜態動作與測試者所選取之 ROI 的平均正確率達到 73.5% 左右，而動態動作的平均正確率達到 89.53% 左右，且總平均正確率有 81.52% 左右，因此本系統自動選取之 ROI 相符觀賞者欲觀看之 ROI。

第二節、本系統運鏡評估

本系統選定 ROI 後會操控攝影機進行運鏡拍攝，而本實驗目的在於確認本系統運鏡拍攝的畫面與無運鏡之全景拍攝的畫面兩者對於觀賞者觀看後之評選結果，因此本實驗測試對象由表 6-1 中找出 20 位測試者來進行實驗。實驗方法是給所有測試者觀看一段影片，而此段影片又分為兩個子畫面，一個子畫面是本系統運鏡拍攝 ROI 的畫面，另一個子畫面是無運鏡之全景拍攝只拍攝觀眾全景影像的畫面，測試者觀看影片完成實驗前並不知兩個子畫面中何者為本系統拍攝影片，測試者觀看此段影片後，針對所觀看兩個子畫面的結果分別給予評分，其評分項目包含：1.清楚的表達主體、2.內容的吸引力、3.畫面的構圖、4.拍攝技巧的運用、5.整體觀看的舒適度五個評分項目，測試者將根據其觀看結果分別針對

兩個子畫面來進行評分。

影片拍攝場景是位於國立師範大學工館校區應用科技大樓地下室 B1 的演講廳，其觀眾席座位是屬於階梯式觀眾席，拍攝時間是演講進行的實際場景，影片總 frame 數 19351 張 frame。影片中包含許多觀眾常見的動作，如舉手、討論等動作。圖 6.4 圖是全景影片擷取片段是依據影片時間序列排序，圖 6.5 圖是本系統拍攝之運鏡影片擷取片段，圖 6.4 與圖 6.5 是依據影片時間序列排序，依據拍攝時間先後由左至右再由上至下排序。

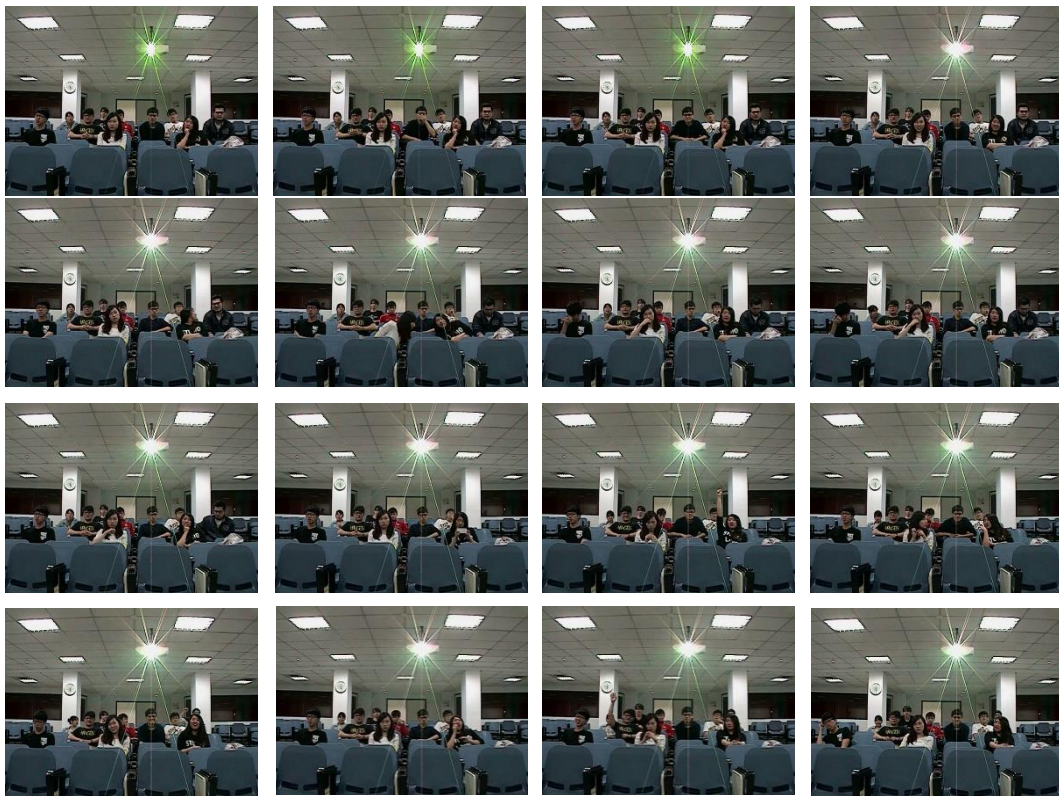


圖 6.4、實驗二全景影片擷取片段

測試者觀看此影片後分別給予兩個子畫面影片做評分，而評分項目分為五個項目，分別是清楚的表達主體、內容的吸引力、畫面的構圖、拍攝技巧的運用、整體觀看的舒適度五個評分項目來讓測試者依據其觀賞結果來做評分，最高 5 分，最低 1 分，表 6-6 是測試者針對所觀看的兩部分影片分別給予評分的結果，最後將表 6-6 的結果，依據不同評分項目分別轉換成長條圖來做討論(如圖 6.6 至圖 6.11 所示)。



圖 6.5、實驗二本系統拍攝之運鏡影片擷取片段

由圖 6.6(a)可知，對於清楚地表達主題這項評分上，本系統拍攝之運鏡影片較能清楚的顯示是場景中的主體，而全景影片由於場景中觀眾主體大小與本系統拍攝之運鏡影片比較結果全景影片的觀眾略小，且全景影片中觀眾靜態動作較不明顯，因此測試者對於顯示主體的評分給予本系統拍攝之運鏡影片的分數較高。由圖 6.6(b)可知，對於內容的吸引力此項評分上，由於全景影片無畫面變化，而本系統拍攝之運鏡影片中拍攝主體變化較多元，因此觀賞者對於內容的吸引力上也給予本系統拍攝之運鏡影片較高的評分，但全景影片中的內容對於大多數的測試者還是有一定的吸引力。由圖 6.6(c)可知，對於畫面構圖的評分上，由於全景影片所顯示畫面較中規中矩，測試者給予其評分較為一致，而本系統拍攝之運鏡影片由於不同年齡層的觀賞者對畫面的要求不同，因而測試者所給予之評分較為分散，但平均結果對於畫面構圖上還是本系統拍攝之運鏡影片略為優異。由圖 6.6(d)可知，對於拍攝技巧的運用評分上，測試者多數認為全景影片雖無運鏡，但其初始拍攝角度也是相當重要的拍攝技巧，因此大都給予全景影片的拍攝技巧中等分數，而本系統拍攝之運鏡影片其所運用的拍攝技巧較全景影片多元，所得到的評分也是相對應的較高。由圖 6.6(e)可知，對於整體觀看的舒適度評分來說，

全景影片較為單調，但對於測試者觀看時較不易造成觀賞時畫面晃動或混亂的情形，而雖然本系統拍攝之運鏡影片畫面因搭配上多元的拍攝技巧讓拍攝畫面較為活潑，但對於不同觀賞需求的觀賞者其拍攝可能會導致畫面較為混亂，因此對於整體觀看的舒適度來，全景影片的觀賞舒適度評分較為集中，而本系統拍攝之運鏡影片的觀賞舒適度較為分散，但其總分結果還是顯示本系統拍攝之運鏡影片的舒適度仍略高於全景影片。

表 6- 6、實驗二之觀賞者評分結果

評分項目			5分	4分	3分	2分	1分	分數 平均
1.清楚地 表達主體	無運鏡	人數			12	7	1	2.55
		百分比			60%	35%	5%	
	有運鏡	人數	16	4				4.8
		百分比	80%	20%				
2.內容的 吸引力	無運鏡	人數		2	10	8		2.7
		百分比		10%	50%	40%		
	有運鏡	人數	5	11	4			4.05
		百分比	25%	55%	20%			
3.畫面的 構圖	無運鏡	人數		1	18	1		3
		百分比		5%	90%	5%		
	有運鏡	人數	1	7	8	4		3.25
		百分比	5%	35%	40%	20%		
4.拍攝技 巧的運用	無運鏡	人數			17	3		2.85
		百分比			85%	15%		
	有運鏡	人數	9	11				4.45
		百分比	45%	55%				
5.整體觀 看舒適度	無運鏡	人數	3	4	13			3.5
		百分比	15%	20%	65%			
	有運鏡	人數	3	8	8	1		3.65
		百分比	15%	40%	40%	5%		

根據由圖 6.6(f)可知，本系統拍攝之運鏡影片在各項評比上皆優於全景影片，可知本系統拍攝之運鏡影片對於觀賞者觀看影片是較適合的選擇，觀賞者對於全景影片與本系統拍攝之運鏡影片兩者的評分，大多觀賞者認為本系統拍攝之運鏡影片有助於增加拍攝畫面的活潑程度，並增加觀眾對於觀賞的意願。

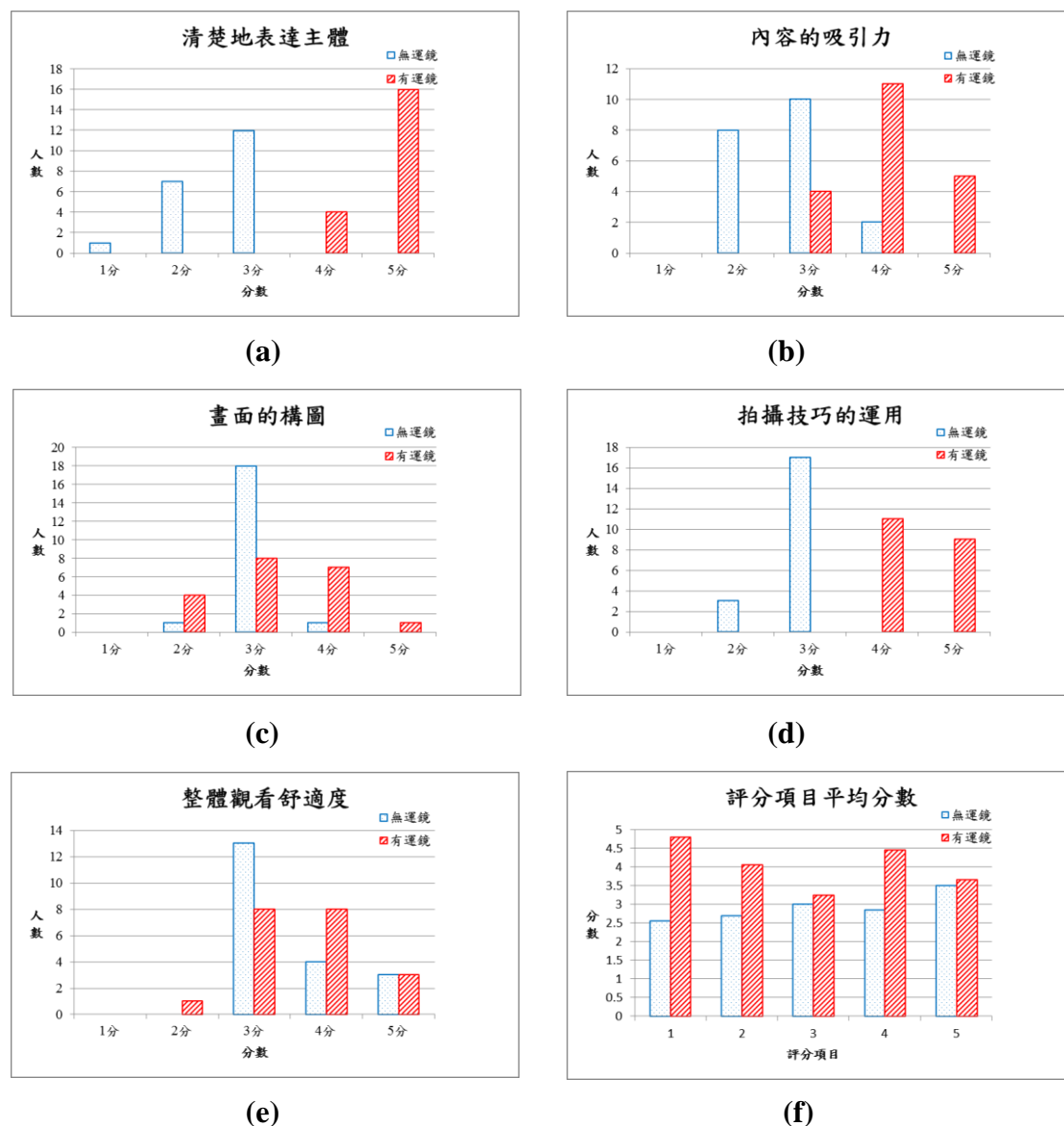


圖 6.6、實驗二評分項目觀賞者評分長條圖(a)清楚地表達主體(b)內容的吸引力(c)畫面的構圖(d)拍攝技巧的運用(e)整體觀看舒適度(f)評分項目平均分數

第三節、與過去研究拍攝結果比較

此實驗將本系統與過去由蔡侑庭所提出之系統[蔡 13]進行相同場地的拍攝，並將拍攝結果給予表 6-1 中 20 位測試者來來進行評選，測試者觀看本系統所拍攝結果與過去研究所提出之系統的拍攝結果，測試者觀看影片完成實驗前並

不知何者為本系統拍攝影片，並在觀看結束後依據其參與演講之經驗選取拍攝與觀看品質較高之系統。

影片拍攝場景是位於國立師範大學工館校區教學大樓 C 棟 209 教室，其觀眾席座位是屬於平面式觀眾席，拍攝時間是演講者進行論文講解時的實際場景，影片總 frame 數 3341 張 frame。影片中包含許多觀眾常見的動作，如舉手、討論等動作(如圖 6.7 所示)。圖 6.7 中圖片是依據影片時間序列排序，依據拍攝時間先後由左至右再由上至下排序，而每張圖片中，最上面較大的圖片是兩系統同時拍攝時的全景影像，左下的圖片是過去所提出之系統拍攝結果，右下的圖片是本系統拍攝結果。



圖 6.7、實驗三實驗影片擷取片段

本實驗將圖 6.7 所表示的影片給予測試者觀看，測試者觀看影片後分別給予本系統與過去所提出之系統拍攝結果做評分，而評分項目分為五個項目，分別是清楚的表達主體、內容的吸引力、畫面的構圖、拍攝技巧的運用、整體觀看的舒適度五個評分項目來讓觀賞者依據其觀賞結果來做評分，最高 5 分，最低 1 分，其評分結果如表 6-7 所示，最後將表 6-7 的結果，依據不同評分項目分別轉換成長條圖來做討論(如圖 6.8)。

表 6-7、實驗三之觀賞者評分結果

評分項目			5分	4分	3分	2分	1分	分數平均
1.清楚地表達主體	前系統	人數	4	8	6	2		3.7
		百分比	20%	40%	30%	10%		
	本系統	人數	9	11				4.45
		百分比	45%	55%				
2.內容的吸引力	前系統	人數		4	9	7		2.85
		百分比		20%	45%	35%		
	本系統	人數	3	5	12			3.55
		百分比	15%	25%	60%			
3.畫面的構圖	前系統	人數		4	12	4		3
		百分比		20%	60%	20%		
	本系統	人數	7	4	9			3.9
		百分比	35%	20%	45%			
4.拍攝技巧的運用	前系統	人數	3	3	14			3.45
		百分比	15%	15%	70%			
	本系統	人數	6	5	9			3.85
		百分比	30%	25%	45%			
5.整體觀看舒適度	前系統	人數		9	9	1	1	3.3
		百分比		45%	45%	5%	5%	
	本系統	人數	8	7	5			4.15
		百分比	40%	35%	25%			

由圖 6.8(a)可知，對於清楚地表達主題這項評分上，本系統較能清楚的顯示是場景中的主體，而過去研究之系統由於主體的選取速度較慢，造成拍攝時主體可以喪失其主題性，因此觀賞者對於顯示主體的評分給予本系統的分數較高。由圖 6.8(b)可知，對於內容的吸引力此項評分上，由於本系統對於 ROI 的選取速度較為即時且較為正確，因此所拍攝出來的影片較能即時的顯示主體的主體性，所以觀賞者對於本系統所拍攝的影片內容較具有觀看的吸引力。由圖 6.8(c)可知，

對於畫面構圖的評分上，由於兩系統的拍攝畫面皆有引用相關的專業美學評估，因此觀賞者給予其評分較為一致，但由於過去研究之系統對於拍攝畫面時並未判定觀眾是否存在畫面中，可能導致所拍攝的畫面中並無拍攝到觀眾，對此觀賞者的評分也稍稍降低，所以平均結果對於本系統的畫面構圖上還是略為優異。由圖 6.8(d)可知，對於拍攝技巧的運用評分上，觀賞者認為兩者的拍攝運用技巧分數一致，但由於本系統運鏡拍攝較為即時，因此觀賞者對於本系統此部分有加分效果。由圖 6.8(e)可知，對於整體觀看的舒適度評分來說，本系統所拍攝的 ROI 較具有主題性且運鏡較為即時，對於觀賞者在觀看的整體舒適度來說，上述的特性給予本系統大大的加分效果。

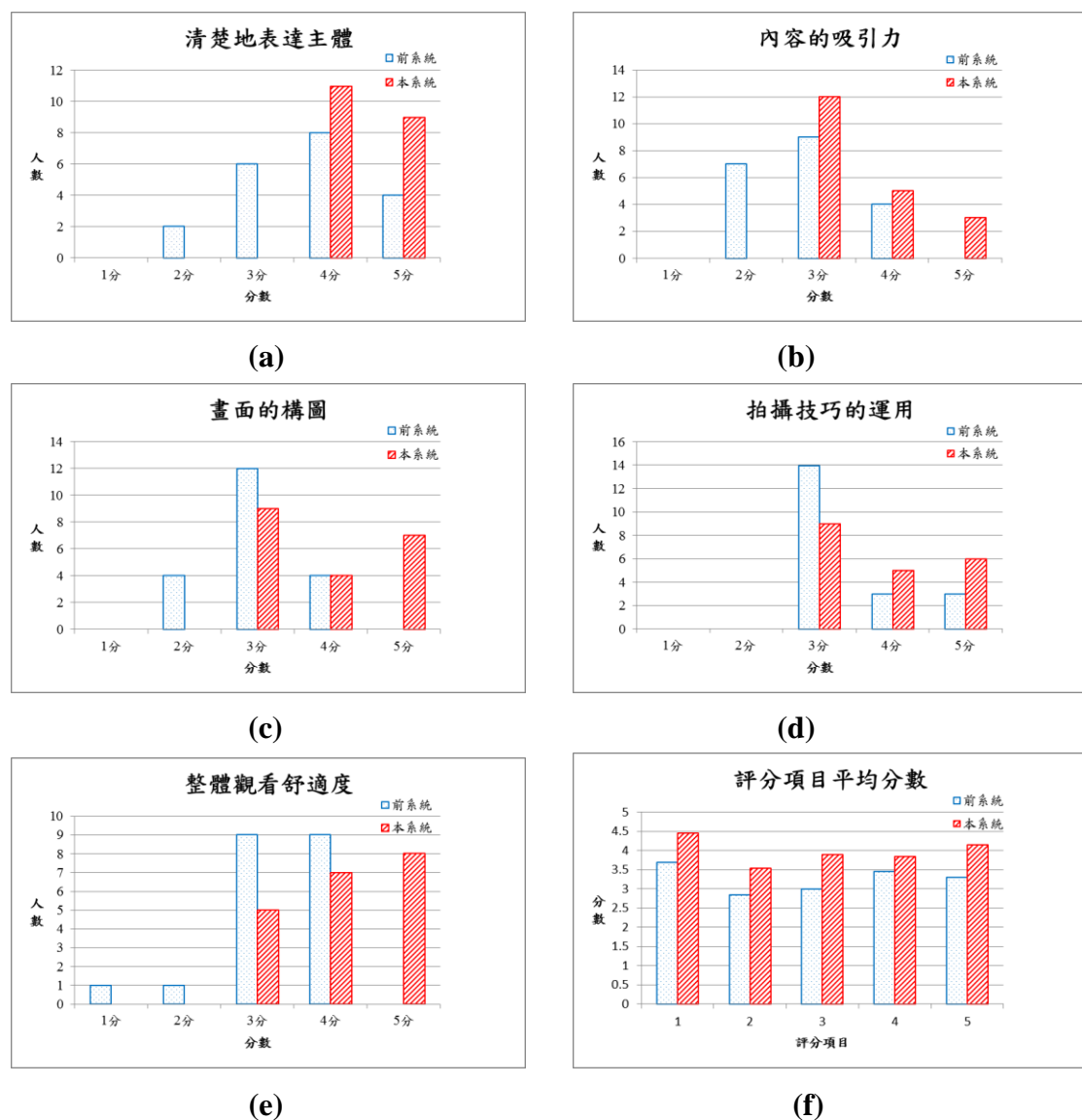


圖 6.8、實驗三評分項目觀賞者評分長條圖(a)清楚地表達主體(b)內容的吸引力(c)畫面的構圖(d)拍攝技巧的運用(e)整體觀看舒適度(f)評分項目平均分數

最後由圖 6.8(f)可知，本系統在各項評比上皆優於過去研究之系統，可知本系統對於拍攝觀眾上較符合觀賞者的觀看需求。

第四節、本系統對於拍攝場地的適用性

本實驗將針對本系統對於拍攝場地來進行討論，本實驗目的在於確認系統能適用於各種拍攝場地來做正確的拍攝。

實驗方法從網路上找到兩段拍攝觀眾的影片，第一段是拍攝課堂中聽講的觀眾，第二段是拍攝幼兒園老師授課時在台下聽講的小朋友，兩段影片中的拍攝場地有著許多相異點，以下將針對兩段影片來做討論。

第一項、第一段影片實驗結果

第一段影片拍攝場景是位於某大學授課時的教室，其觀眾席座位是屬於階梯式觀眾席，拍攝時間是課程進行時的實際場景，影片總 frame 數 1593 張 frame。影片中包含許多觀眾常見的動作，如趴下、討論、寫筆記等動作(如圖 6.9 所示)。6.9 圖是全景影片擷取片段是依據影片時間序列排序，圖 6.10 圖是本系統拍攝之運鏡影片擷取片段是依據影片時間序列排序，依據拍攝時間先後由左至右再由上至下排序。



圖 6.9、實驗四第一段影片全景影片擷取片段

第一段影片中由於拍攝鏡頭是位於拍攝場地的左前方，此種拍攝角度會造成觀眾間的遮蔽問題，且對於離鏡頭較遠的觀眾，其動作較不容易被系統偵測出來，造成系統選取 ROI 時大多選取到離鏡頭較近的前排觀眾，因此透過本實驗

可以得知攝影機拍攝鏡頭的角度與攝影機放置位置會影響到本系統的拍攝。



圖 6.10、實驗四第一段影片本系統拍攝之運鏡影片擷取片段

第二項、第二段影片實驗結果

第二段影片拍攝場景是位於幼兒園上課時的教室，其觀眾席座位是屬於平面開放式，拍攝時間是課程進行時的實際場景，影片總 frame 數 2012 張 frame。影片中包含許多觀眾常見的動作，如舉手、揮手、鼓掌等動作(如圖 6.11 所示)。6.11 圖是全景影片擷取片段是依據影片時間序列排序，圖 6.12 圖是本系統拍攝之運鏡影片擷取片段是依據影片時間序列排序，依據拍攝時間先後由左至右再由上至下排序。



圖 6.11、實驗四第二段影片景影片擷取片段



圖 6.12、實驗四第二段影片本系統拍攝之運鏡影片擷取片段

第二段影片中由於拍攝鏡頭是位於拍攝場地的右前方，且鏡頭離觀眾距離相當近，此種拍攝角度會造成觀眾間的遮蔽問題，而此場景中由於觀眾動作與本研究一開始設定觀眾動作範圍不符，本研究限制觀眾動作不得超過一個人的座位距離，而此場景中觀眾動作屬於比較開放式的動作，因此可能造成系統運鏡拍攝後由於附近觀眾動作超過一個人的座位距離而產生遮蔽問題，因此造成所拍攝畫面失去主體，或是拍攝時無法找到最佳主題的問題，因此造成拍攝畫面效果不佳的狀況，因此可以得知開放式場地與動作範圍較大的觀眾會對本系統造成拍攝上的影響。

由上述第一項實驗可以得知本系統對於一般演講的觀眾場景所選定拍攝的 ROI 與觀賞者欲觀看之 ROI 相符；第二項實驗可以得知本系統所使用之運鏡能提升觀賞者對於觀賞影片的意願，且對於本系統所拍攝的影片也有高度的評價；第三項實驗可以得知本系統相較於過去研究系統有進一步改良，使其所拍攝畫面更符合觀賞者的需求；第四項實驗可以得知本研究對於其他場地的影響，對此是拍攝是需要再改良與注意的方向。根據以上實驗可以得知本系統所拍攝的畫面能符合觀賞者欲觀看畫面，且所拍攝的影片品質也提升了觀賞者的觀看意願。

第七章 結論與未來工作

第一節、結論

本研究的目的為開發具專業攝影師能力且拍攝主體是觀眾之虛擬觀眾攝影師系統。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。本系統能夠自動挑選合適的 ROI 與運鏡方式，並控制 PTZ 攝影機執行運鏡動作前往 ROI 進行拍攝，最後再針對所拍攝的影像進行影像構圖規則的評估，並操控攝影機進行微調讓所拍攝的影像符合拍攝美學的觀點。

在觀察觀眾畫面時，對於演講時的觀眾，攝影師在選取拍攝主體時主要是選取畫面中具有動作的主體來拍攝，因此本研究在 ROI 選取上利用 optical flow 找出畫面中具有 motion 性質的像素點，並依據所有具有 motion 像素點的密度與其每點的 optical flow 向量來找出候選 ROI，之後為了避免相同 ROI 不斷被重複選取的情形，本研究利用 STA 來紀錄每次拍攝的 ROI 資訊並判定候選 ROI 能否成為正式 ROI。

對於操控攝影機前往 ROI 進行拍攝上，本研究運鏡方式會去依據目前攝影機拍攝 ROI 位置與新的 ROI 兩者間的距離進行合適的運鏡方式。而由於本研究拍攝的主體是觀眾，為了避免拍攝畫面主體並非觀眾，因此系統會利用人臉偵測來判斷所拍攝的畫面中是否具有觀眾，最後本研究所輸出的畫面需具有專業攝影技巧，因此將針對所拍攝的畫面進行影像構圖規則的評估並進行攝影機微調，藉此提高拍攝畫面的質感。

最後實驗結果顯示，本系統能自動從畫面中挑選適合 ROI 進行拍攝，系統所選 ROI 與觀賞者欲觀看的 ROI 相符，且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，最後所拍攝畫面皆能符合影像構圖規則與美學條件。

第二節、未來工作

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。

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。

最後，由於本系統選取 ROI 主要是以畫面 motion 點位置來做偵測選取，而不同演講場合中觀賞者所欲觀看的 ROI 並非只有畫面中具有 motion 點位置的 ROI，因此可以再增加更多的選取條件來增加系統選取 ROI 的正確率。或是可以同時使用多套運用不同特徵來選取 ROI 之虛擬觀眾攝影師系統進行拍攝，藉由系統間互相的溝通，讓拍攝內容更多元更能吸引觀賞者。

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。

参考文献

-
- [Bae03] R. Baecker, “A Principled Design for Scalable Internet Visual Communications with Rich Media, Interactivity, and Structures Archives,” *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research*, IBM, pp. 16-19, 2003.
- [Ber06] A. Berengolts and M. Lindenbaum, “On the Distribution of Saliency,” *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, Chicago, pp. 543-549, 2004.
- [Bia98] M. Bianchi, “AutoAuditorium: A Fully Automatic, Multi-camera System to Televisе Auditorium Presentations,” *Proceedings of the Joint DARPA/NIST Workshop on Smart Spaces Technology*, 1998.
- [Bia04] M. Bianchi, “Automatic Video Production of Lectures Using an Intelligent and Aware Environment,” *Proceedings of the 3rd International conference on Mobile and Ubiquitous Multimedia*, New York, pp. 117-123, 2004.
- [Bou00] B. J. Y, “Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm,” *Intel Corporation Microprocessor Research Labs report*, 2000
- [Cho10] H. Chou, J. Wang, C. Fuh, S. Lin, and S. Chen, “Automated lecture recording system,” *Proceedings of the IEEE International conference on System Science and Engineering (ICSSE)*, Taiwan, pp. 167-172, 2010.
- [Dat06] R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Studying Aesthetics in Photographic Images Using a Computational Approach,” *Proceedings of the 9th European conference on Computer Vision*, Vol. 3, Heidelberg, pp. 288-301, 2006.
- [Fan03] C. Y. Fang, S. W. Chen, and C. S. Fuh, “Automatic Change Detection of Driving Environments in a Vision-Based Driver Assistance System,” *IEEE Transactions on Neural Networks*, vol. 14, no. 3, pp. 646-657, 2003.
- [Lan09] C. Lang, D. Xu, and Y. Jiang, “Shot Type Classification in Sports Video Based on Visual Attention,” *Proceedings of the International conference on Computational Intelligence and Natural Computing*, Vol. 1, Wuhan, pp.

336-339, 2009.

- [Li05] T. Y. Li and X. Y. Xiao, “An Interactive Camera Planning System for Automatic Cinematographer,” *Proceedings of the 11th International conference on Multimedia Modeling*, pp. 310-315, 2005.
- [Oni04] M. Onishi and K. Fukunaga, “Shooting the Lecture Scene Using Computer-Controlled Cameras based on Situation Understanding and Evaluation of Video Images” *Proceedings of the 17th International conference on Mobile and Ubiquitous Multimedia*, pp. 781–784, 2004.
- [Row01] L. A. Rowe, D. Harley, and P. Pletcher, “BIBS: A Lecture Webcasting System,” *Berkeley Multimedia Research Center*, 2001.
- [Rui01] Y. Rui, L. He, A. Gupta, and Q. Liu, “Building an Intelligent Camera Management System,” *Proceedings of the 9th ACM International conference on Multimedia*, New York, pp. 2-11, 2001.
- [Rui03] Y. Rui, A. Gupta, and J. Grudin, “Videography for Telepresentations,” *Proceedings of the ACM SIGCHI conference on Human Factors in Computing Systems*, New York, pp. 457-464, 2003.
- [Rui04] Y. Rui, A. Gupta, J. Grudin, and He Liwei, “Automating Lecture Capture and Broadcast: Technology and Videography,” *Multimedia Systems*, Vol. 10, No. 1, pp. 3-15, 2004.
- [Sol12] B. Solmaz, B. E. Moore, and M. Shah, “Identifying Behaviors in Crowd Scenes Using Stability Analysis for Dynamical Systems,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 10, pp. 2064-2070, 2012.
- [Vio04] P. Viola, and M. J. Jones, “Robust Real-time Face Detection”, *International Journal of Computer Vision*, Vol. 57, No. 2, pp. 137-154, 2004.
- [Wu10] S. Wu, B. E. Moore, and M. Shah, “Chaotic Invariants of Lagrangian Particle Trajectories,” *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, San Francisco, pp. 2054-2060, 2010.
- [Zha08] C. Zhang, Y. Rui, J. Crawford, and L. W. He, “An Automated End-to-end Lecture Capture and Broadcasting System,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, New York, Vol. 4, No. 1, pp. 2-11, 2008.
- [郭12] 郭彥佑, “自動化演講錄製系統”, 國立台灣師範大學, 資訊工程研究所

碩士論文，2012/06。

- [李12] 李俊億，“自動化演講錄製系統之虛擬攝影師”，國立台灣師範大學，資訊工程研究所碩士論文，2012/06。
- [蔡13] 蔡侑廷，“以觀眾為拍攝主體之虛擬攝影師系統”，國立台灣師範大學，資訊工程研究所碩士論文，2013/06。
- [呂13] 呂佳儒，“自動化演講錄製系統之虛擬導播子系統”，國立台灣師範大學，資訊工程研究所碩士論文，2013/06。
- [謝95] 謝怡竹，“以光流為基礎之自動化表情辨識系統”，國立中央大學，資訊工程研究所碩士論文，1995/07。
- [1] Ambient Insight Standard Report
<http://www.ambientinsight.com/Resources/Documents/AmbientInsight-2011-2016-Worldwide-Self-paced-eLearning-Market-Premium-Overview.pdf>
- [2] 中臺灣展演藝術平台
<http://perform.culture.taichung.gov.tw/place/Details.aspx?Parser=99,5,16,,,11>
- [3] 莎頓國際學院
<http://www.sheltoncollege.edu.sg/cn/aboutus-facilities.html>
- [4] 攝影入門之新手必讀—色彩的奧秘
<http://www.sj33.cn/dphoto/syxt/201211/32597.html>
- [5] 利用色彩輕鬆拍攝漂亮照片
http://shijue.me/show_text/4ffef3e0ac1d840d9001c13c
- [6] 拍出不同戲劇效果—善用不同的光向
<http://www.fotobeginner.com/3789/light-direction/>
- [7] 【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>
- [8] 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/>
- [9] Kinect for windows
<http://www.microsoft.com/en-us/kinectforwindows/>
- [10] AXIS Communications
http://www.axis.com/zh/products/cam_215/index.htm

- [11] 陳慶瀚，類神經網路
<http://ccy.dd.ncu.edu.tw/~chen/course/Neural/ch3/SOM.pdf>
- [12] 蘋果新一代 iPHONE 發表會
http://blog.chinatimes.com/honey_news/archive/2012/09/13/2984203.html
- [13] 論文發表
<http://www.che.kuas.edu.tw/app/news.php?Sn=241>
- [14] 海基會
<http://www.mac.gov.tw/ct.asp?xItem=72482&ctNode=6643&mp=110>
- [15] 順光、逆光拍攝技巧
http://www.hkzooo.com/backlight_photography/