# 國立臺灣師範大學
# 資訊工程研究所碩士論文

指導教授：　李忠謀　　　博士

基於高斯混合模型之課堂舉手辨識研究
Gaussian Mixture of Model based Arm Gesture Recognition Research

研究生：　蔡承軒　撰

中華民國　101 年 7 月

# 摘要

人體姿勢辨識技術是一項熱門的研究議題，在過去利用影像處理來辨識人體姿勢的辨識系統已經發展一段時間，在學術領域或專業應用上使用這類的辨識系統需要龐大的運算量以及昂貴的設備，使得這類的系統無法普及於一般大眾使用。

因此，在這篇論文中本研究經由偵測與辨識學生舉手的動作設計了一套即時互動或應用的系統。在假設已知上半身範圍的情況下再針對這個範圍採用連續影像差異法(temporal differencing)，利用時間上連續的影像做一對一的像素相減，得到一個移動物件的影像，此影像再透過高斯混合模型(Gaussian Mixture Model)，利用多個高斯函數來描述反覆出現的多種背景值，並透過函數參數值的調適，以適應光線所產生的變化，此目的是為了在複雜的環境中擷取前景(foreground)的影像，並使用尺度不變特徵轉換(Scale-invariant feature transform，SIFT)擷取特徵，將擷取到的特徵套入支持向量機(Support Vector Machine，SVM)對姿勢動作進行辨識。發展此系統的目的在於可以使用方便取得的器材來取代昂貴的設備，使得人體姿勢辨識可以普及於一般大眾所使用。



關鍵字 ： 人體姿勢辨識、連續影像相減、高斯混合模型

# Abstract

Human body gesture recognition is one of the top research topic, and it had been developed for a long time. Due to its massive computational complexity and its expensive equipment, these system can't be used by grassroots. In this paper, we develop a "Gaussian Mixture of Model based Arm Gesture Recognition Research" Real-time system, and using temporal differencing to get a moving object, under the hypothesis of knowing the range of upper body. These image then apply Gaussian Mixture of Model, using multiple Gaussian functions, to describe multiple background status. For adapting illumination effect and extracting foreground in complex environment, we apply parameters adjusting to solve it. We also use SIFT to extract feature and using SVM to classify. We hope this system can let all the people use not expensive and easy-to-get devices to do gesture recognition.


Keyword : Gesture recognition 、Temporal differencing 、Gaussian Mixture of Model

# 誌謝

本人能完成這篇論文，首先必須非常感謝我的指導教授李忠謀老師，在這研究所的兩年來給予我許多細心的指導，畢業的學長姐們也提供我許多研究上的方向，此外博士班的學長也給予我許多的建議，在遇到任何的困難時第一個總是先找學長，學長不會覺得不耐煩，反而細心指導，僅在此表達學生最高的謝意。

在實驗室兩年來的生活過得很充實也很歡樂，這要謝謝實驗室的同學們，在戰場上同甘共苦，”情和義，值千金”這句話總是我和戰友們的默契，在球場上，實驗室的學弟妹們總是熱情邀約打排球，使我身體更佳的健康，讓我的生活更加得多采多姿。

在我的生活中要感謝我的女朋友林圓晴，在我遇到難關的時候，我總是會煩躁，而她總是忍耐並且安撫我，讓我順利地走過每一個難關，謝謝妳的付出

最後要謝謝我的父母，在我的求學路途上，總是給予我最大的鼓勵，並且支持我做的每一個決定,讓我可以全力以赴的完成此論文,我會努力加油，讓我的父母可以快樂的度過每一天。

# 目錄

# 附表目錄

# 附圖目錄

# 第一章　　緒論

## 1.1　　研究動機

人體姿勢辨識技術近年來在日常生活中越來越常見，透過影像處理辨識人體的各種姿勢變化進而提供互動的回饋，例如 2010 年由微軟推出的 Kinect 體感控制系統，藉由 Kinect 控制器上的紅外線感測器(IR-Based)偵測使用者並建立深度資訊，進一步得到使用者骨架資訊，原本 Xbox360 將此技術應用在遊戲主機上，類似這樣的體感控制系統也可以應用在生活中，例如家庭裡直接用手勢來控制電視機，其 Kinect 體感控制系統的控制環境如下圖 1-1 所示。

人體姿勢辨識不僅僅應用在電玩或是家電當中，對於層出不窮的車禍意外也有很大的貢獻，若車上配有行人偵測系統(Pedestrian Detection)，便能即時偵測行人突然竄出的突發狀況，確保駕駛與行人的安全；除了應用在上述的情境中，也可在教室的教學環境中用人體姿勢辨識系統來協助授課老師記錄學生上課反應，並可在課後參考此紀錄做課後檢討與改進，達到教學上的進步。

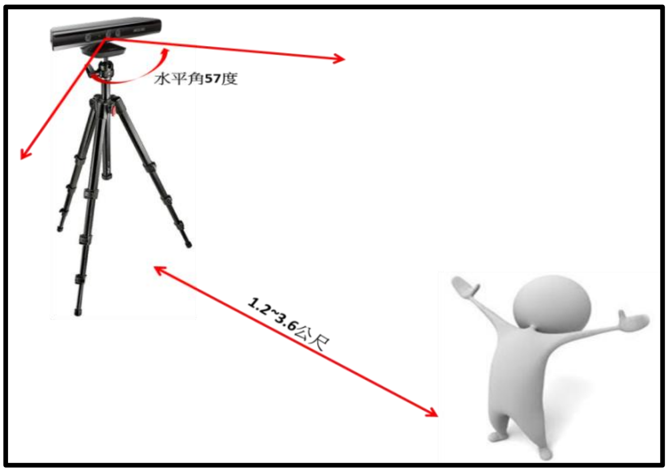



圖 1-1 Kinect 控制環境

現行的人體姿勢辨識技術已發展多年，但大多數是需要額外的輔助器材，例如人體姿勢辨識技術應用在教室中，學生所需要的器材皆昂貴，若使用資訊手套器材將其戴在學生手上，藉由此資訊手套輔助電腦做人體姿勢的判斷，這在使用上也很不方便。此外，不是每所學校都可以負擔這樣昂貴的設備，此因素造成在實際教學上無法普及應用。因此本論文希望藉由一些價格較不昂貴的設備來達到同樣的目標，使得在教學上可以普及應用。

## 1.2 研究目的

過去利用影像辨識技術來辨識人體姿勢的辨識系統已經發展一段時間，但在學術領域或專業應用上使用這種方式的辨識系統需要龐大的運算量以及昂貴的周邊設備，使得這種系統無法普及於一般大眾所能使用到的範疇。因此，本研究的方法將使用網路攝影機(IP camera)得到連續的影像做連續影像差異(temporal differencing)，再經由高斯混合模型(Gaussian Mixture of Background Model)擷取完整的前景，對此前景利用尺規不變特徵轉換(Scale-invariant feature transform，SIFT)，取得特徵點後，將其丟入支持向量機(Support Vector Machine)得到人體姿勢分析的結果。

本論文希望可以藉由方便架設且容易取得的器材來取代昂貴的設備，並且發展出一套能夠普及於教學上人體姿勢的判斷系統，並且藉由這套系統提供授課老師台下學生的反應，進一步可以參考此數據來提升課堂上的教學品質，例如台下學生的專注力或是師生之間的互動。

### 1.3　研究的範圍與限制

本研究方法的環境限制為以室內為主，並且本系統的攝影機必須要為固定式的攝影機。在本研究中總共有三台攝影機，每台攝影機會對準六個座位如圖 1-2 所示，因此可對 18 個座位做人體姿勢判斷，但需要辨識的座位越多本系統就需要更多的處理時間。針對每個座位做人體姿勢的判斷是我們的目標，但是由於人體姿勢有很多種，本系統目前針對三種來判斷，舉左手、正常姿勢以及舉右手，如圖 1-3 所示。

圖 1-2 固定式網路攝影機對準六個座位

圖 1-3 舉左手、正常姿勢、舉右手

# 第二章　文獻探討

對一個成功的人體姿勢行為分析系統而言，人類姿勢辨識是其中最重要的一部分。近年來，研究者對於辨識分析人體的動作或行為更是感興趣，本論文將相關文獻分成下面幾項來做解釋：(1) 物件偵測的方法　(2) 人體姿勢行為分析。

## 2.1 物體偵測的方法

現今，電腦視覺研究已有許多成果，在偵測、追蹤和辨識方面已有長足的發展，這些研究成果可以應用在許多地方，例如在監控系統、汽車駕駛安全、虛擬實境等，我們在此分為靜態攝影機以及動態攝影機兩種情況。

### 2.1.1 動態攝影機

非靜態攝影機在不用建立背景模型的情況下，Wren 等人提出一套針對人的即時追蹤系統 Pfinder [5]，Pfinder 使用色彩和外型的多重統計模型來分離前景物和背景。並能於大範圍的區域持續追蹤人的頭部和軀體。以不考慮行人重疊的問題，目前也有人使用 HOG 加上 SVM 進行行人的偵測[6]，而行人偵測是一項很困難的研究，因此加上機器學習(machine learning)的機制加強準確率。

非靜態攝影機背景模型也是目前監控系統的一個研究方向，把受監控的完整環境建構成一大全景背景模型，全景影像取得方法主要有兩種：1.使用廣角的攝影機，不過由於價錢昂貴，在遠離焦點的部分影像失真情況嚴重，因此通常比較少被使用。2.使用普通攝影機擷取的影像，經過處理建立全景影像，如此一來，

解析度高而且不會發生失真的狀況，這樣的特性可以滿足影像處理的各個應用，S. E. Chen et al. [13]架設固定位置的攝影機，以不同的側擺(panning)角度拍攝，利用軟體的方法比對建立 360 度的全景影像。另外 Y. Ren et al. [14]使用空間上的高斯分佈(spatial distribution of Gaussian)處理攝影機受到風吹的微小震動。

### 2.1.2 靜態攝影機

在使用靜態攝影機情況下，因環境為固定背景，因此大部分都可以建立背景模型來取得完整的前景。我們可分為遞迴方式以及非遞迴方式，若以非遞迴式方法來偵測到物體，以下有幾個常見的方法，時間軸上取中間值法(temporal median)，較常見的建立背景影像的方法，背景影像選取一段時間內視窗中像素值的中間值，R. Cucchiara et al. [7]把此方法推展到彩色影像；計算出像素中間值的複雜度為O(L log L)；B. Shoushtarian et al. [8]比較了三種非遞迴式建立背景影像的方法，分別為非前景像素更新法(selective update using non-foreground pixels)、時間軸上的平均值法(selective update using temporal averaging)與時間軸上的中間值法(selective update using temporal median)，其中非前景像素更新法(selective update using non-foreground pixels)就是前景偵測後得到的背景像素更新為新的背景像素，其餘前景像素部分則用之前的背景像素替代，此法在戶外環境效果會比較不理想，因為戶外光線變化劇烈時，此更新法會使得背景影像無法隨著光線變化更新成符合現狀的背景影像；時間軸上的平均值法(selective update using temporal averaging)取滑動視窗中的像素值的平均值，以此值作為新的背景像素值，

此法會出現錯誤的情形在於存入滑動視窗的像素若包含移動物件，取平均值將會造成新的背景像素失真的現象；時間軸上的中間值法(selective update using temporal median)取滑動視窗中每張影像所有像素之中間值，形成新的背景影像，此法再偵測移動物件有較正確的效果，不過在計算出中間值的計算複雜度頗高，且需要的儲存空間也來得大。

遞迴的方式需要較少的儲存空間，但若是出現誤判的情況，將會導致錯誤持續發生一段時間。目前方法採用指數型的權重來去除過去得到的資訊，並更新為正確的背景像素以更新目前的參數，具代表性的方法為 Kalman Filter，此方法是一個追蹤移動物體的遞迴式方法，目前有許多以 Kalman Filter 為中心建立背景模型的方法，C. Wren et al. [9]就是其中很經典的例子。

現在要說明的高斯混合模型(Mixture of Gaussians Model，GMM)也是遞迴的方法，GMM 不同於 Kalman Filter 地方在於 GMM 可以同時追蹤多個高斯分布，目前 GMM 是在建立背景模型各種方法中普遍被採用的方法，由於 GMM 是參數式模型，所以不需另外的空間儲存一段影像資訊，而可以動態的更新背景模型的各個重要參數，詳細的參數分析可參考 Qi Zang et al. [10]對於 GMM 詳細地分析每個參數的影響。其他相關的，C. Stauffer et al. [11] 以 K 個高斯分佈建立像素點的分佈模型，而 P. KaewTraKulPong et al. [12]針對[11]提出了改進的方法，利用兩種模式的參數估測方法，系統運作初期利用較花時間但是正確性高的估測 (expected sufficient statistics)，可以建立一個穩固的基礎，到達指定的

時間後，使用較不花運算時間的估測方式，如此一來可以有效改善[11]的運算時間與增加背景參數的準確度。

## 2.2 人體姿勢行為分析

姿勢行為分析(behavior analysis)相較於前面章節所提到的領域屬於較高階的技術，也就是整個研究流程的最後部分，主要是將前面章節計算完的資訊加以整理統計做高階的行為判斷，主要的行為分析演算法可分為兩大類：1. 狀態空間法(State-Space Approaches)與 2. 基於模板的方法(Template-Based Methods)。以下就這兩大類分別做說明：

(1)狀態空間分析：

利用狀態空間分析的方法是將姿勢的特徵轉換成一狀態(state)表示，而一個完整的動作姿勢是以連續的狀態轉換機率所組成的，因此包含了時間變化的資訊。其中在判斷人體姿勢方面最常使用隱馬爾可夫模型(Hidden Markov Model，簡 HMM)的分析方法，如圖 2-1 ，此方法是由有限狀態機(Finite State Machine，簡稱 FSM)

衍生出來的數學模型，在 1997 年由 J. Soh et al. [15]提出先利用建立人體姿勢的 3-D 模型來表示每一狀態空間的內容後，再由 HMM 的狀態轉換機率模型判斷動作。由於 HMM 方法是基於機率模型來連接狀態之間的轉換，因此對於一些不能觀察到的資訊也能透過機率模型計算出動作姿勢產生的可能性，而這些機

率模型參數的產生都是在訓練階段時計算得到。HMM 模型一直以來不斷的被提出改良也被廣泛的應用在語音、自然語言、姿勢判斷等應用當中，不過在 2001 年由 J. Lafferty[16] 提出的條件機率域(Conditional Random Field，簡稱 CRF)，改善了 HMM 在空間轉換過程中的資料獨立性(data independent)的缺點，也就是狀態轉換之間只能單方向的進行，因此 CRF 的新想法是讓前後的狀態空間也都能保有互相參考的機率存在，能夠讓長時間的連續動作做前後的關連性，保有動作的特徵。除此之外，在 2006 年由 Quattoni et al. [17]，提出隱藏式條件隨機域(Hidden Conditional Random Field, 簡稱 HCRF)同時結合 HMM 以及 CRF 兩種方法，因此既可以計算出觀察不到的姿勢動作相互關係也能保有讓長時間的連續動作能夠前後的相呼應，如圖 2-2 表示了 HMM、CRF 和 HCRF 三者在狀態轉換的差異。

(2)基於模板的分析：

基於模板分析方法是先將影像資訊轉換到靜態狀態模式或是特殊的動作特

徵，像是利用姿勢移動軌跡的方向、速度、距離等這些靜態的變數，或是利用姿勢的輪廓資訊所計算出的身體重心、投影量、輪廓大小等資訊，再帶入主成分分析(Principal Component Analysis，簡稱 PCA)、線性鑑別分析(Linear Discriminate Analysis，簡稱 LDA)、類神經網路(Neural Network)、支持向量機(Support Vector Machine，簡稱 SVM)等方法來判斷不同動作的差異性。基於模板的分析方法其需要計算的成本很低，實作上也很方便，對於一些簡單動作而言有很不錯的分析

結果，不過相對的，這些輸入給基於模板分析方法的參數很容易受到環境光線、顏色、雜訊等干擾的影響，造成判斷上的失誤。

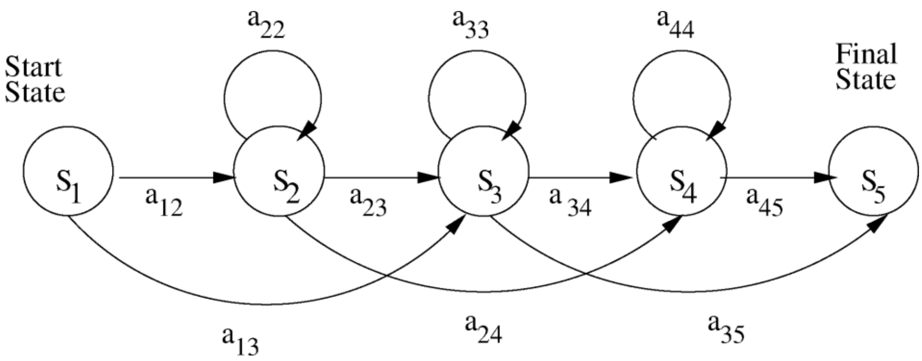
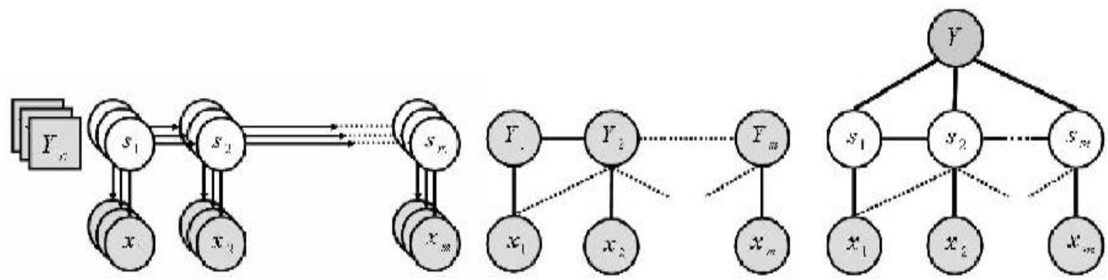



圖 2-1 HMM 是以 FSM(Finite State Machine)為基礎的判斷模型



(a) (b) (c)

圖 2-2　HMM, CRF 和 HCRF 示意圖

(a)表示 HMM 狀態轉換圖, (b)為 CRF 狀態轉換圖, (c)為 HCRF 狀態轉換圖

## 2.3 支持向量機(Support Vector Machine，SVM)分類器

支持向量機(Support Vector Machine,SVM)是一種分類(classification)演算法，由 Vapnik 等根據統計學習理論提出來的一種新的機器學習(Machine Learning)方法。對於一群資料而言，有時候我們會依據資料的特性來將這群資料分為幾群。而就資料分群而言，我們已知有一些效果還不錯的方法。例如:Nearest Neighbor、類神經網路(Neural Networks)、Decision Tree 等等的方式，而如果在正確的使用前提下，這些方法的準確度都差不多，然而，支持向量機(Support Vector Machine，SVM)的優勢在於使用上較為簡易。

支持向量機(Support Vector Machine，SVM)的目的在於尋找一個超平面(hyperplane)以區分此特徵空間中的所有訓練用的資料的不同類別，如圖 2-3 所示，所謂的超平面是因為樣本資料通常是以高維度的方式表現，因此在高維度下的平面就稱之。

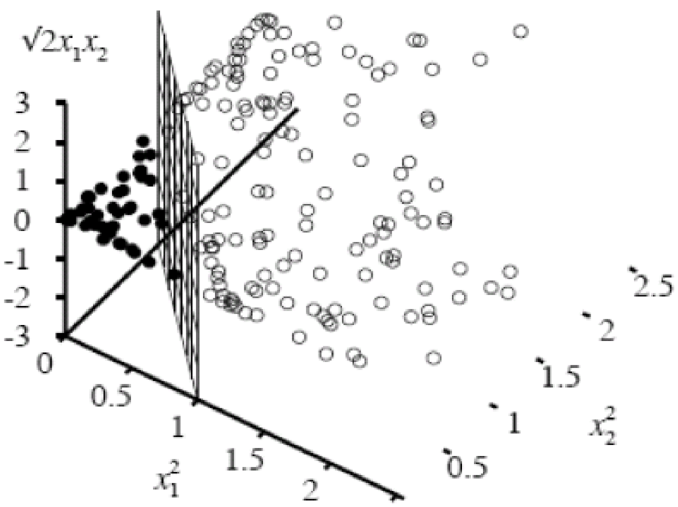



圖 2-3 超平面

最基本的支持向量機(Support Vector Machine，SVM)分類器是將兩個不同類別的集合分開。如上圖 2-3 所示，此資料為二維的矩陣，藍色和紅色分別代表不同類別的資料，在此情況下我們希望找到一個超平面能夠將這兩類別分開，同時這兩個類別與此超平面之間的間隔最大，如此便能更明確的分辨資料位於平面中哪個類別的資料，否則很容易產生判斷誤差，這也是支持向量機(Support Vector Machine，SVM)需要事前透過監督式學習的訓練資料學習出的平面。有此高平面後如下圖 2-4 所示，當有新進的測試資料就能夠依據此超平面來分類。

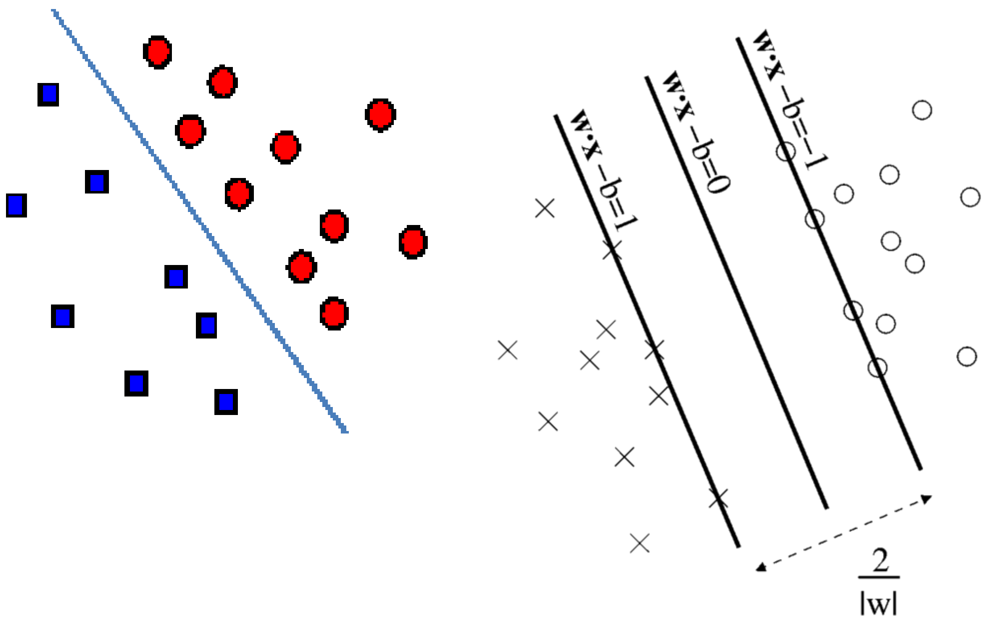



圖 2-4 訓練資料學習出的平面

# 第三章　研究方法與系統架構

一般靜態的高斯混合模型(Gaussian Mixture of Background Model)是描述背景分佈的背景模型，利用多個高斯分佈來描述重複出現的背景值，在背景乾淨的情況下，背景模型方法將提供良好的前景影像給後續的應用，但是假若桌上有水瓶或是其他雜物會被視為前景，進而影響到後續的辨識結果，因此本研究提出連續影像差異(temporal differencing)來解決上述的問題。本研究的系統流程首先使用網路攝影機(IP camera)得到連續的影像做連續影像差異(temporal differencing)，得到差異影像再經由高斯混合模型(Gaussian Mixture of Background Model)擷取完整的前景，對此前景利用尺規不變特徵轉換(Scale-Invariant Feature Transform，SIFT)，取得特徵點後再丟入支持向量機(Support Vector Machine)得到人體姿勢分析的結果，如圖 3-1 為本系統運作的流程圖，大致上可分成五個部分。

第一部分是取得影像，以及取得連續影像差異(temporal differencing)的結果；第二部分是高斯混合模型的背景訓練與更新，本研究使用高斯混合模型(Gaussian Mixture of Background Model)；第三部分是取得前景資訊；第四部份是介紹利用尺度不變特徵轉換(Scale-Invariant Feature Transform，SIFT)演算法來取得特徵點，第五部分是使用 SVM 得到分類的結果。

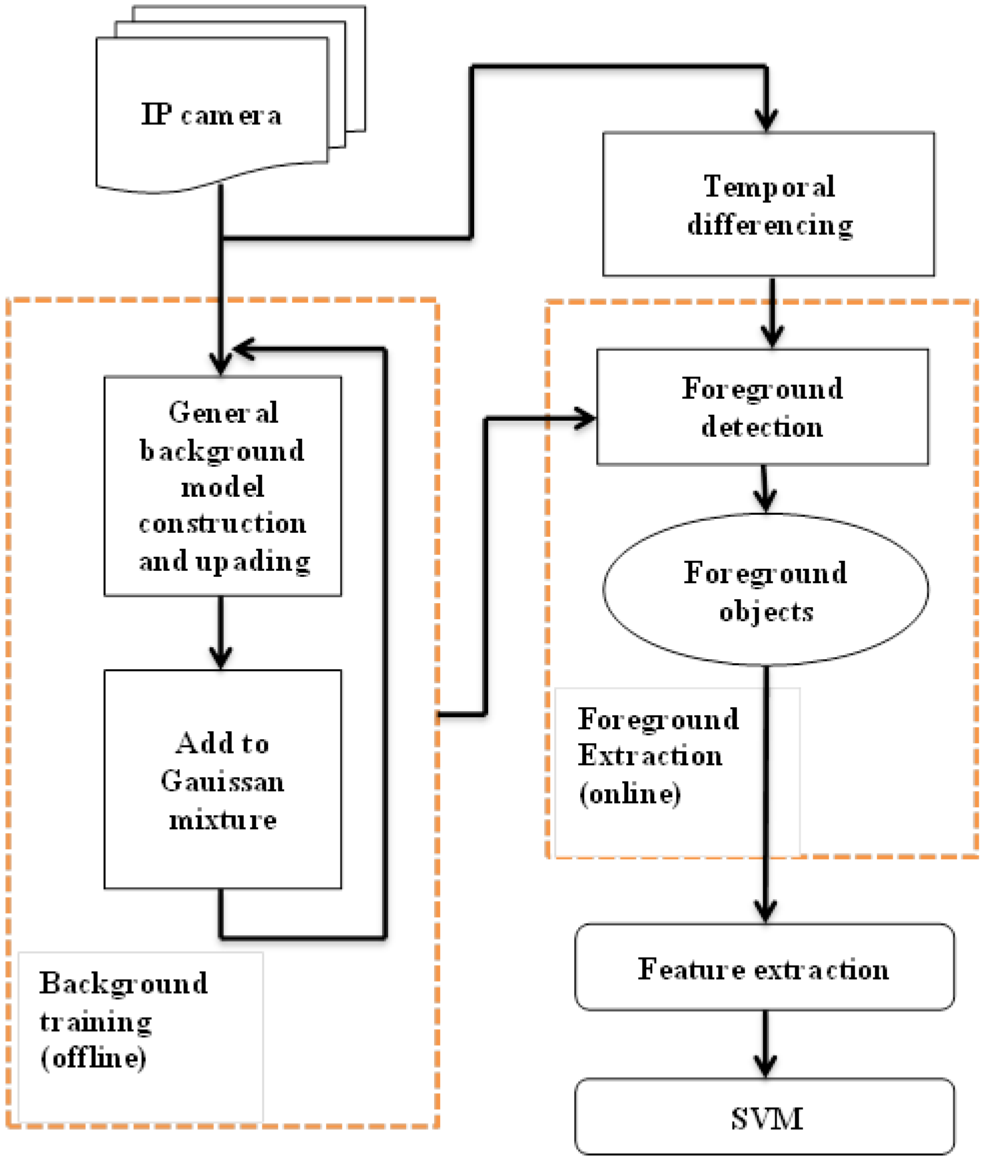



圖 3-1 系統流程圖

### 3.1 連續影像差異(temporal differencing)

此方法針對差異影像中的每個像素點建構一個背景影像(background image)來記錄此像素點灰階值的變化，該背景影像(background image)主要以第一張的灰階影像為基礎，依照此背景影像與目前得到的影像做一個比較，再依照比較的結果對每個像素點以極小的變化作調整。背景影像(background image)會不斷地更新，而調整公式如下列式子(3-1)所表示，$b_i$表示背景影像的灰階像素值，$c_i$表示目前

的影像的灰階像素值。

$$b_i = \begin{cases} b_i = b_i + 1 & c_i > b_i \\ b_i = b_i + 1 & otherwise \end{cases} \tag{3-1}$$

當背景影像(background image) $B(t)$ 於時間 t 建立時，便可用它來決定目前的影像(current image，$c_i$)的像素是否為移動中的物件，其定義以下列來表示 :

$$\delta(b_i, c_i) = \begin{cases} 1 & c_i - b_i > \varepsilon \\ 0 & otherwise \end{cases} \text{，} \varepsilon \text{ 表示一個門檻值。} \tag{3-2}$$

連續影像相減(temporal differencing)的原理是利用時間上連續的影像對此做一對一的像素相減，若差值低於門檻值，就表示此像素可能為靜態物件，反之，則此像素可能為移動中的物件。此方法的優點在於取出的影像都為移動物件，因此可以過濾掉本研究遇到的困難，例如桌面上的雜物，如圖 3-2 所示。

但是假若從正常姿勢到舉手狀態如圖 3-3(a)以及從舉手狀態到正常姿勢如圖 3-3(b)的差異影像結果類似，在做手勢辨識時這兩種情況都可能被判斷成舉手，這類的情況對於後續的移動物件辨識將無法提供正確的結果，因此本研究後續提出高斯混合模型來解決此情況。

圖 3-2 背景桌面有雜物的狀況

(a) (b)

圖 3-3 連續影像差異的結果

### 3.2 高斯混合模型的背景訓練與更新

此處所指之背景通常是指在連續的影像中較常出現的影像，即若考慮同一個位置的像素點，則指較常出現的像素值。於真實環境中，背景像素值會在短時間內因為各種原因而有些許的變化，如光線影響、空氣折射等等。若我們假設其變化的結果呈現常態分布，就可利用高斯模型去描述該情況。

但在很多情況下，如在燈光閃爍情況，同一位置的像素值，其於 RGB 空間中之值會呈現跳動的狀態，而非如常態分佈般於同一值附近變動。若採用單一高斯分布則無法滿足以上的情況，故利用多個高斯分布來解決此問題。基於以上兩種情況，本系統使用高斯混合模型(Gaussian Mixture of Background Model)來建立背景影像模型。

高斯混合模型是利用多個高斯分布，以線性的方式組合而成。假設某一像素位置在時間1~t 觀察到的像素值為 $\mathrm{X} = \{X_1, \dots, X_t\}$，而$\mathrm{P}(X_t)$表示$X_t$出現的機率，若用高斯混合模型來表示，式子如下列所示：

$$\mathrm{P}(X_t) = \sum_{i=1}^{K} \omega_{i,t} * \eta\left(X_t, M_{i,t}, C_{i,t}\right) \tag{3-3}$$

其中 K 代表高斯分布的數目，$\omega_{i,t}$表示第 i 高斯分布在時間 t 的權重值，$M_{i,t}$、$C_{i,t}$分別表示第 i 高斯分布在時間 t 的平均值以及共變異數矩陣，$\eta$ 表示高斯機率密度數值，表示成：

$$\eta\left(\mathrm{X}_t, \mathrm{M}_{i,t}, \mathrm{C}_{i,t}\right) = \frac{1}{(2\pi)^{n/2}|\mathrm{C}_t|^{1/2}} \exp^{-\frac{1}{2}(X_t-\boldsymbol{\mu}_t)^{\boldsymbol{T}}\boldsymbol{C}_t^{-1}(X_t-\boldsymbol{\mu}_t)} \tag{3-4}$$

在 3-2 式子中，n 是特徵向量的維度，在此我們考慮一個像素值的 RGB 三個維度，即$n = 3$。為了降低計算複雜度，假設每一維度在統計上是互相獨立的，所以 C 是一個對角共變異矩陣(diagonal convariance matrix)，以 3-5 式表示:

$$C_{\mathrm{K},t} = \sigma^2_{\mathrm{K},t} I \tag{3-5}$$

### 3.2.1 參數意義

在高斯混合模型中，同個位置的像素值分布可能會由多個高斯函數組成，其各個高斯函數的平均值(mean)表示此像素值在 RGB 空間的中心位置，如 RGB($X_i$, $Y_i$)=(255,255,255)。而共變異矩陣(covariance matrix)則可以描述像素值分布的範圍，當其中的標準差值越小，則代表平均值的變異程度小，即表示該位置點的顏色變化較小。最後，權重值(weight)表示各個高斯函數的重要性，因此較常出現的像素值所建立的高斯分布權重會較高。

在由 C. Stauffer et al. 所提出之“An algorithm to estimate mean traffic speed using uncalibrated cameras”[18]一文中，其以 K 個高斯分佈建立像素點的分佈模型。在針對簡單的室內環境下，K 值可不必設太大就能有效地描述背景，反之若 K 的數目越大，則所需的記憶體容量也必須跟著增加。

### 3.2.2 初始高斯模型

為了減少前處理時間，本系統指派第一張影像的像素值為平均值，且變異數和權重值的初始值亦是事先定義好的常數。這樣的好處是可以減少資料的計算時間，也不用額外的儲存資料的空間；而缺點在於第一張影像必須是要純背景的影像，以本系統的而言則是在無人的教室環境下做初始高斯模型的動作。

變異數和權重值的給定，需考慮系統運作初期的兩項不確定性：1.因短時間內所蒐集的像素值之分布可能是散亂且廣的，其間之變異性過大，故在還不了解像素值分布的情況下，以較大的變異數來描述此未知的分布，能夠避免背景像素被分類成前景。2.剛建立背景時，各高斯分布因資料少，其重要性不明確，故給予較小之權重值來描述新建立的高斯分布，可避免一開始非背景的高斯分布權重過高。

### 3.2.3 更新高斯分布模型參數

為適應背景所在之環境變化，如光線閃爍等情況，使其像素值隨著時間呈現小幅度的變化，故其高斯分布需不斷地隨時間更新以解決上述的問題。在本小節會介紹如何透過更新高斯分布的參數值，來解決光線變化造成的背景影響。

在時間 t 取得像素值$X_t$，便在描述該像素值的混合模型中，計算該值與任一高斯函數的平均值差$\left|X_t - \mu_{i,t-1}\right|, \mathrm{i} = 1 \sim \mathrm{k}$，若其差值小於 N 倍該高斯函數的標準差(standard deviation)， 落在此範圍的值可能會有多個，在此取差異值最小

的，我們稱該$X_t$符合該高斯函數，式子 3-4 為判斷$X_t$是否符合配對到高斯機率密度函數中：

$$\left|X_t - \mu_{i,t-1}\right| \leq N * \sigma_{i,t-1} \tag{3-6}$$

而承上，針對符合的高斯分布有三個參數要更新，分別為 1. 權重值(weight)、2. 平均值(mean)和共變異矩陣(covariance matrix)，以下分別介紹三個參數的更新方式：

**1. 更新權重值**

更新權重值的目的是要調整所有高斯函數的重要性之高低，當高斯分布所符合的資料越多，代表此高斯分布描述的像素值出現頻率高。此權重值在每次有像素值符合時提昇，而較少資料符合的高斯分布其權重值就會慢慢衰退。在時間 t，第 i 個高斯分布根據以下式子 3-8 更新：

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} + \alpha(M_{i,t}) \tag{3-7}$$

$\alpha$是一個常數，可以決定更新的速度。$M_{i,t}$有 0 和 1 兩種值，針對符合的高斯分布$M_{i,t}$為 1，其餘不符合的為 0，因此，只有符合的高斯分布權重值會提升，其餘的會逐漸衰退。當高斯分布的權重值全部更新後，將權重值做正規化，使得加起來的總合等於 1，3-8 式所示：

$$\sum_{i=1}^{K} \omega_{i,t} = 1 \tag{3-8}$$

**2. 更新平均值與變異數**

當要更新平均值與變異數時，必須要考慮到像素點在符合的高斯分布中出現的機率值，若像素值越靠近平均值，對新像素點的信任度就很高，更新時會讓新像素值影響高斯分布越大；若否，則較相信原本高斯分布的平均值與共變異矩陣。

為了能夠在系統運作初期快速的更新成正確的背景模型，所以將更新分為兩部分：

I. 在系統運作初期 η 值很小，若一開始使用 η 去更新速度緩慢，要花很久的時間才能將前景物消去。為了加快更新速度，當 η 小於$1/_T$以前，令$\eta = 1/_T$更新平均值以及變異數加快消去第一張影像的前景物，直到 η 大於$1/_T$時才改以 II 的方式更新。

II. 當 η 大於$1/_t$以後，以 3-9 和 3-10 的式子做更新：

$$\mu_t = (1-\rho)\mu_{t-1} + \rho X_t \tag{3-9}$$

$$\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t) \tag{3-10}$$

$$\rho = \alpha\eta(X_t|\mu_k, \sigma_k) \tag{3-11}$$

ρ 是新像素點$X_t$用來更新參數的權重，其值為 η 值乘上一常數值 α，稱為背景模型的學習比率(learning rate)。參考 Zang et al.” Parameter Analysis of Mixture of Gauissams Model” [19]一文中所提出的高斯混合模型參數分析，對於變化較小的

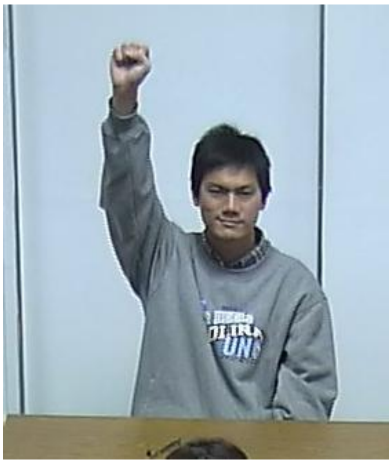
背景需要使用較小的α，而背景變動頻繁的環境下，則需要較大的α去快速適應背景的變化。

### 3.3 前景擷取

本研究目標是從原始影像中擷取出完整的前景。我們所建立的背景模型可以描述背景的高斯分布，在後續的前景偵測時，判斷新進的像素點是否落在背景的高斯分布內，將落在高斯分布內的像素點視為背景像素，權重值比較大的高斯分布是背景的可能性就比較大，所以把所有的高斯分布依照權重值由小到大排列，由最可能為背景的高斯分布權重值開始累加。當累加值超過一個門檻值 T 時，我們就由這 B 個高斯分布所建立的混合高斯模型當作我們背景顏色值的分布。式子 3-12 表示 B 值的計算方式：

$$\mathrm{B} = \arg min_b \left(\sum_{k=1}^{b} \omega_k > T\right) \tag{3-12}$$

如此便可得到完整的前景影像，前景擷取的影像如圖 3-3 所示。



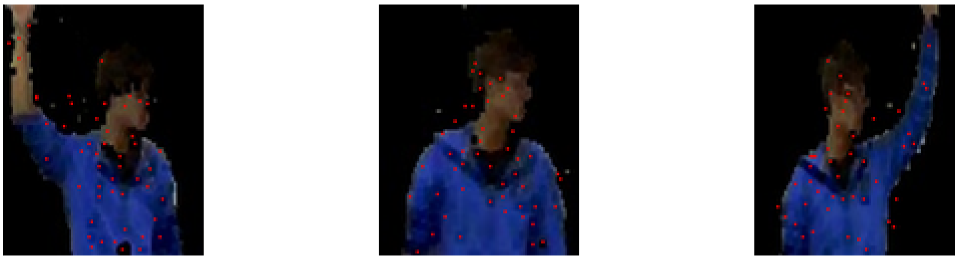
座位上學生舉手行為

去除背景後的結果

圖 3-3 前景擷取的影像

### 3.4 尺度不變特徵轉換(Scale-Invariant Feature Transform ，SIFT)

本系統對每個影像使用 SIFT 演算法得到特徵向量，由 SIFT 演算法主要是偵測興趣點，也就是 SIFT 架構中的關鍵點，影像在不同的尺度下用高斯濾波器(Gaussian filter)進行卷積(convolved)，然後利用連續高斯模糊化影像差異來找出關鍵點。關鍵點是根據不同尺度下的高斯差(Difference of Gaussians,DoG)中的最大值與最小值所導出。利用 SIFT 演算法取得的關鍵點如圖 3-4 所表示，分別表示每個動作訓練影像擷取到的關鍵點(keypoint)分布，關鍵點(keypoint)的數量會根據影像的大小而有所改變當影像大關鍵點(keypoint)數量會增加，影像小則關鍵點(keypoint)數量會減少，因此必須要對得到的前景影像大小調整一致。我們使用支持向量機(Support Vector Machine，SVM)代入的特徵向量長度必須要一致，根據我們的觀察一張前景的影像大小在 70 × 35 可以得到 K 個關鍵點(keypoint)，因此我們利用尺度不變特徵轉換(Scale-Invariant Feature Transform ，SIFT)得到特徵的強度做一個排序的動作取前 K 大的關鍵點(keypoint)並且將每個關鍵點(keypoint)分別取得該關鍵點(keypoint)的 X 座標、Y 座標和強度這三項當該關鍵點的特徵因此總共會有 K*3 維的特徵向量代入支持向量機(SVM)。



(a) (b) (c)

圖 3-4 使用 SIFT 得到的關鍵點

### 3.5 支持向量機(Support Vector Machine，SVM)分類器

本研究辨識結果總共分為三類，分別為舉左手、正常姿勢、舉右手，如圖 3-4 所示，我們利用 3.4 節所提到的 3.4 尺度不變特徵轉換(Scale-Invariant Feature Transform ，SIFT)演算法將得到的關鍵點依照它的強度的排序，並且將排行在前一百的關鍵點的 X-Y 坐標位置記錄下來當作特徵向量的值，因此我們會得到兩百個值在這個特徵向量中，假設關鍵點不足一百個，那麼我們會取所以關鍵點坐標的平均值，將其補足一百個關鍵點，其目的是為了滿足支持向量機(Support Vector Machine，SVM)的特徵向量長度必須相同。

接著將所有影像的特徵向量資料都依支持向量機(Support Vector Machine，SVM)規定的輸入格式排好如下 :

<label> <index1>:<value1> <index2>:<value2>…
<label> <index1>:<value1> <index2>:<value2>…
…

TrainingSet.txt :

| | | | | | |
|---|---|---|---|---|---|
| 1 | 1:17.99 | 2: 32.95 | 3:19.05 | … | 200 : 5.14 |
| 1 | 1:16.99 | 2 :31.74 | 3:19.08 | … | 200: 5.09 |
| ⋮ | | | | | |
| 2 | 1:16.23 | 2: 32.20 | 3:16.29 | … | 200 : 34.21 |
| 2 | 1:11.86 | 2: 31.50 | 3:12.49 | … | 200 : 34.45 |
| ⋮ | | | | | |
| 3 | 1:15.15 | 2: 18.80 | 3:15.69 | … | 200 : 29.97 |
| 3 | 1:16.37 | 2: 7.39 | 3:16.68 | … | 200 : 30.08 |

接著我們將這組資料丟到支持向量機(Support Vector Machine，SVM)之中進行訓練，其結果會產生一個 TrainingSet.model 的檔案

TrainingSet.model :

```
vm_type C_SVC
kernel_type RBF
gamma 0.0001220703125
nr_class 3
total_sv 663
rho -0.0481547762468004 0.131229979604856 0.0458122803962966
label 1 2 3
nr_sv 229 203 231
SV
⋮
```

在測試階段時我們可以將做完訓練的模型(Model)做辨識的動作，來達到本研究的目標。

# 第四章　實驗與結果分析

本研究主要利用固定式攝影機，在一個已知的上半身範圍內，使用基於Adaboost 的人臉偵測技術，並且以此人臉的資訊取得新的上半身範圍影像，我們將得到的新影像做後續的手勢辨識實驗，本實驗的環境分為三排，如圖 4-1 所示，從教室的環境中可以觀察得到，因為網路攝影機角度的關係，第二排的情況會受到前後的影響，此情況會影響到後續的辨識結果。並且假若在訓練背景前桌上是空無一物，在測試時桌上擺有水瓶或其他雜物時，此類狀況也會影響到後續的辨識結果，因此本研究會針對這類的問題做討論並且解決。

本實驗最主要分為兩個部分作探討，第一個為單純使用高斯混合模型(Gaussian Mixture of Background Model)，第二個為連續影像相減(temporal differencing)方法再加上高斯混合模型(Gaussian Mixture of Background Model)，其目的是為了要解決上述的問題。

在實驗中是以教室的環境作為實驗環境，使用網路攝影機拍攝不同時間點的實驗影片，該系統測試的硬體設備為 Intel Core i5-2500S CPU @ 2.70GHz 、記憶體為 8.00GB。

圖 4-1 實驗環境

## 4.1 實驗資料庫

本論文使用的資料庫皆是在相同的教室上拍攝的影片，該實驗架設三台網路攝影機每台攝影機對準 6 個座位，共 18 個座位，總共有三組不同時間點拍攝的影片，這三種影片都有自己的訓練資料(training data)稱為 M1、M2、M3，而每組影片都可當作測試資料(testing data)稱為 V1、V2、V3。

此三組資料所擁有的影像張數都不一樣，舉左手、正常姿勢、舉右手的次數也不盡相同，通常正常姿勢的影像張數會比較多，表 4-1 是此三組從拍攝的影片中得到的資料內容。

表 4-1 不同組資料來源的內容

| V1 | 舉左手 | 正常姿勢 | 舉右手 | 總共 |
|---|---|---|---|---|
| 影像張數 | 1472 | 4718 | 1546 | 7736 |

| V2 | 舉左手 | 正常姿勢 | 舉右手 | 總共 |
|---|---|---|---|---|
| 影像張數 | 2750 | 13900 | 3486 | 20136 |

| V3 | 舉左手 | 正常姿勢 | 舉右手 | 總共 |
|---|---|---|---|---|
| 影像張數 | 595 | 3132 | 672 | 4399 |

本實驗所建立的高斯混合模型所使用的模型參數如表 4-2 所示。

表 4-2 高斯混合模型使用的參數

| 參數 | 值 | 參數 | 值 |
|---|---|---|---|
| Standard deviation σ | 50 | N | 2.5 |
| Weight ω | 0.001 | α | 0.05 |
| Number of Gaussian K | 4 | 門檻值 T | 0.04 |

## 4.2 實驗流程

本實驗將教室中分為三排，每排都有自己的訓練資料(training data)，因此每一組資料會分別得到這三排個別的辨識率，本實驗使用兩種不同的方法，一種是第三章中所提及的方法，另一種則是將原本方法的連續影像差異(temporal differencing)拿掉，後續的動作都沒有變，使用的訓練資料(training data)也會跟著改變，而測試資料(testing data)則都沒變。

這兩種不同的方法，我們的訓練資料(training data)以及測試資料(testing data)來源是來自於這三組資料中，每組資料都是取一半的資料當作訓練資料(training data)另一半的資料當作測試資料(testing data)，如此一來對同一組資料來說訓練資料(training data)以及測試資料(testing data)會是不盡相同的。

### 4.2.1 不同時間點的辨識率

在此將以資料庫中 V1、V2、V3 的測試影片進行實驗，我們將針對 V1 和 V2 的資料作交叉比較(cross-validation)以及 V1 和 V2 分別對 V3 的辨識率做統計，本節將針對兩種不同的方法做測試，分別是第三章所介紹的方法拿掉連續影像差異(temporal differencing)，後續的動作都沒改變以及第三章所提及的方法，我們將以方法一和方法二來表示：

**I. 方法一：**

針對此方法使用 V1、V2 的資料做交叉比較所以總共有四組數據，數據一，我們將使用 M1 當作訓練模型(training model)、V1 當作測試資料(testing data)，結果如下表 4-2 所示：

表 4-2 資料來源為 M1、V1

M1→V1

| 第三排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 323 (99%) | 2 (1%) | 2 (1%) |
| 正常姿勢 | 41 (6%) | 564 (89%) | 26 (4%) |
| 舉右手 | 5 (1%) | 15 (5%) | 316 (94%) |

| 第二排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 199 (91%) | 15 (7%) | 5 (2%) |
| 舉右手 | 26 (3%) | 655 (89%) | 193 (22%) |
| 正常姿勢 | 3 (2%) | 41 (21%) | 155 (78%) |

| 第一排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 173 (92%) | 4 (2%) | 11 (6%) |
| 正常姿勢 | 33 (4%) | 808 (94%) | 17 (2%) |
| 舉右手 | 7 (3%) | 21 (10%) | 226 (87%) |

表格內容的意義為張數(所占百分比)

數據二，我們將使用 M1 當作訓練模型(training model)、V2 當作測試資料(testing data)，結果如下表 4-3 所示 ：

表 4-3 資料來源為 M1、V2

M1→V2

| 第三排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 422 (94%) | 22 (5%) | 2 (1%) |
| 正常姿勢 | 40 (2%) | 2157 (96%) | 40 (2%) |
| 舉右手 | 50 (8%) | 79 (12%) | 530 (80%) |

| 第二排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 295 (72%) | 97 (24%) | 19 (5%) |
| 舉右手 | 423 (18%) | 1612 (67%) | 372 (15%) |
| 正常姿勢 | 58 (11%) | 160 (30%) | 318 (59%) |

| 第一排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 335 (65%) | 119 (23%) | 58 (11%) |
| 正常姿勢 | 440 (19%) | 1672 (72%) | 201 (9%) |
| 舉右手 | 130 (25%) | 136 (26%) | 263 (50%) |

表格內容的意義為張數(所占百分比)

數據三，我們將使用 M2 當作訓練模型(training model)、V2 當作測試資料(testing data)，結果如下表 4-4 所示：

表 4-4 資料來源為 M2、V2

M2→V2

| 第三排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 438 (98%) | 7 (2%) | 2 (1%) |
| 正常姿勢 | 20 (1%) | 2167 (97%) | 50 (2%) |
| 舉右手 | 0 (0%) | 20 (3%) | 638 (97%) |

| 第二排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 382 (93%) | 22 (5%) | 7 (2%) |
| 舉右手 | 86 (4%) | 2027 (84%) | 298 (12%) |
| 正常姿勢 | 16 (3%) | 48 (8%) | 477 (89%) |

| 第一排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 492 (96%) | 3 (1%) | 17 (3%) |
| 正常姿勢 | 213 (9%) | 1964 (85%) | 136 (6%) |
| 舉右手 | 18 (3%) | 79 (15%) | 431 (81%) |

表格內容的意義為張數(所占百分比)

數據四，我們將使用 M2 當作訓練模型(training model)、V1 當作測試資料(testing data)，結果如下表 4-5 所示 ：

表 4-5 資料來源為 M2、V1

M2→V1

| 第三排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 323 (99%) | 0 (0%) | 4 (1%) |
| 正常姿勢 | 53 (8%) | 542 (85%) | 40 (6%) |
| 舉右手 | 7 (2%) | 2 (1%) | 326 (97%) |

| 第二排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 188 (86%) | 19 (9%) | 12 (6%) |
| 舉右手 | 185 (21%) | 339 (39%) | 349 (40%) |
| 正常姿勢 | 8 (4%) | 46 (23%) | 145 (73%) |

| 第一排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 103 (54%) | 7 (4%) | 78 (42%) |
| 正常姿勢 | 160 (19%) | 232 (27%) | 466 (54%) |
| 舉右手 | 11 (5%) | 20 (9%) | 194 (86%) |

表格內容的意義為張數(所占百分比)

## II. 方法二:

針對此方法使用 V1、V2 的資料做交叉比較以及 V1、V2 對 V3，所以總共有六組數據，數據一，我們將使用 M1 當作訓練模型(training model)、V1 當作測試資料(testing data)，結果如下表 4-6 所示 :

M1→V1

表 4-6 資料來源為 M1、V1

| 第三排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 324 (99%) | 2 (1%) | 0 (0%) |
| 正常姿勢 | 17 (3%) | 603 (95%) | 15 (2%) |
| 舉右手 | 0 (0%) | 0 (0%) | 336 (100%) |

| 第二排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 216 (99%) | 3 (1%) | 0 (0%) |
| 舉右手 | 42 (5%) | 695 (80%) | 127 (15%) |
| 正常姿勢 | 15 (7%) | 9 (4%) | 186 (88%) |

| 第一排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 166 (88%) | 2 (1%) | 21 (11%) |
| 正常姿勢 | 27 (3%) | 775 (90%) | 56 (7%) |
| 舉右手 | 8 (4%) | 17 (8%) | 200 (89%) |

表格內容的意義為張數(所占百分比)

數據二，我們將使用 M1 當作訓練模型(training model)、V2 當作測試資料(testing data)，結果如下表 4-7 所示：

M1→V2

表 4-7 資料來源為 M1、V2

| 第三排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 401 (90%) | 45 (10%) | 1 (0%) |
| 正常姿勢 | 4 (0%) | 2209 (98%) | 35 (2%) |
| 舉右手 | 98 (15%) | 43 (7%) | 518 (79%) |

| 第二排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 307 (75%) | 64 (16%) | 39 (10%) |
| 舉右手 | 197 (8%) | 1841 (76%) | 368 (15%) |
| 正常姿勢 | 26 (5%) | 112 (21%) | 399 (74%) |

| 第一排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 442 (86%) | 29 (6%) | 44 (9%) |
| 正常姿勢 | 251 (8%) | 1740 (78%) | 252 (11%) |
| 舉右手 | 37 (7%) | 49 (9%) | 459 (84%) |

表格內容的意義為張數(所占百分比)

數據三，我們將使用 M2 當作訓練模型(training model)、V2 當作測試資料(testing data)，結果如下表 4-8 所示 ：

M2-→V2

表 4-8 資料來源為 M2、V2

| 第三排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 439 (98%) | 3 (1%) | 5 (1%) |
| 正常姿勢 | 20 (1%) | 2208 (98%) | 20 (1%) |
| 舉右手 | 9 (1%) | 11 (2%) | 639 (97%) |

| 第二排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 407 (99%) | 1 (0%) | 2 (1%) |
| 舉右手 | 61 (3%) | 2195 (91%) | 150 (6%) |
| 正常姿勢 | 0 (0%) | 45 (8%) | 492 (92%) |

| 第一排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 444 (86%) | 51 (10%) | 20 (4%) |
| 正常姿勢 | 75 (3%) | 2083 (90%) | 150 (7%) |
| 舉右手 | 6 (1%) | 31 (6%) | 493 (93%) |

表格內容的意義為張數(所占百分比)

數據四，我們將使用 M2 當作訓練模型(training model)、V1 當作測試資料(testing data)，結果如下表 4-9 所示：

M2→V1

表 4-9 資料來源為 M2、V1

| 第三排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 293 (90%) | 18 (5%) | 16 (5%) |
| 正常姿勢 | 23 (4%) | 579 (91%) | 33 (5%) |
| 舉右手 | 1 (0%) | 19 (6%) | 316 (94%) |

| 第二排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 193 (88%) | 19 (9%) | 6 (3%) |
| 舉右手 | 29 (3%) | 623 (72%) | 211 (12%) |
| 正常姿勢 | 4 (2%) | 3 (2%) | 203 (96%) |

| 第一排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 148 (79%) | 3 (2%) | 37 (20%) |
| 正常姿勢 | 111 (13%) | 611 (71%) | 136 (16%) |
| 舉右手 | 23 (10%) | 13 (6%) | 189 (84%) |

表格內容的意義為張數(所占百分比)

數據五，我們將使用 M1 當作訓練模型(training model)、V3 當作測試資料(testing data)，結果如下表 4-10 所示 ：

M1→V3

表 4-10 資料來源為 M1、V3

| 第三排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 32 (84%) | 2 (7%) | 3 (9%) |
| 正常姿勢 | 0 (0%) | 274 (87%) | 40 (13%) |
| 舉右手 | 3 (6%) | 2 (4%) | 41 (89%) |

| 第二排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 101 (76%) | 21 (24%) | 1 (1%) |
| 舉右手 | 47 (7%) | 677 (93%) | 3 (0%) |
| 正常姿勢 | 18 (13%) | 22 (16%) | 96 (70%) |

| 第一排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 90 (72%) | 18 (14%) | 17 (14%) |
| 正常姿勢 | 7 (1%) | 425 (81%) | 89 (17%) |
| 舉右手 | 1 (1%) | 40 (27%) | 110 (73%) |

表格內容的意義為張數(所占百分比)

數據六，我們將使用 M2 當作訓練模型(training model)、V3 當作測試資料(testing data)，結果如下表 4-11 所示 ：

M2→V3

表 4-11 資料來源為 M1、V3

| 第三排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 30 (79%) | 3 (8%) | 5 (13%) |
| 正常姿勢 | 0 (0%) | 272 (86%) | 43 (14%) |
| 舉右手 | 0 (0%) | 0 (0%) | 46 (100%) |

| 第二排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 113 (84%) | 20 (15%) | 1 (1%) |
| 舉右手 | 205 (28%) | 516 (71%) | 7 (1%) |
| 正常姿勢 | 34 (25%) | 1 (1%) | 102 (74%) |

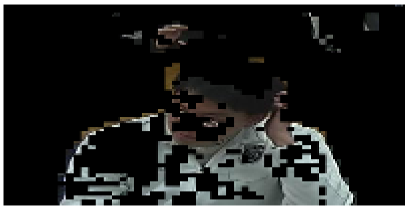
| 第一排 | 舉左手 | 正常姿勢 | 舉右手 |
|---|---|---|---|
| 舉左手 | 91 (73%) | 24 (19%) | 10 (8%) |
| 正常姿勢 | 11 (2%) | 422 (81%) | 89 (17%) |
| 舉右手 | 0 (0%) | 42 (28%) | 109 (72%) |

表格內容的意義為張數(所占百分比)

### 4.3 實驗結果分析

經由上面幾個實驗我們可以觀察出在背景複雜的情況下會降低本研究的辨識率，例如人物的重疊以及桌面上有其他雜物如圖 4-2 (a)所示，因此本研究提出連續影像差異(temporal differencing)方法來解決此類問題，我們比較兩種方法，結果如 4.3 節，我們由實驗數據可以驗證，本研究加入了連續影像差異(temporal differencing)方法後有達到消除人物重疊以及桌面上有雜物的影響，如圖 4-2(b)，圖中背景有人物重疊以及桌上有雜物的情況，加入連續影像差異(temporal differencing)方法後可以達到消除以上干擾物的問題，也因此提升了辨識率。

(a)

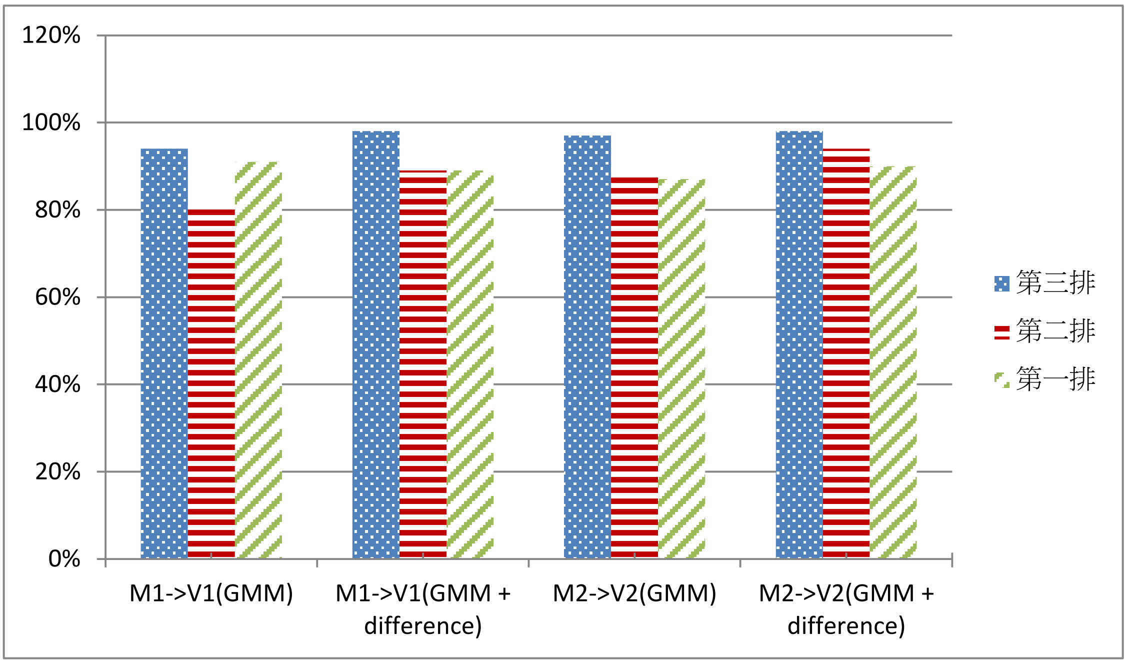


(b)

圖 4-2 使用連續影像差異(temporal differencing)後的效果

本小節將分兩種情況來討論，一種是同一組的資料作實驗結果分析，其結果在 4.3 節有更詳細的資料，我們將 M1 當作訓練資料(training data) 、V1 當作是測試資料(testing data)以及 M2 為訓練資料(training data)、V2 為測試資料(testing data)，因為訓練資料(training data)與測試資料(testing data)都來自同一組資料，所以，在此情況下，辨識率基本上都有達到八成以上，如表 4-12 所示。針對此實驗數據是想驗證本研究的方法是可行的，因此在後續的討論中，資料來源將會是不同組的資料做交叉比較。

表 4-12　M1、V1 和 M2、V2 的辨識率



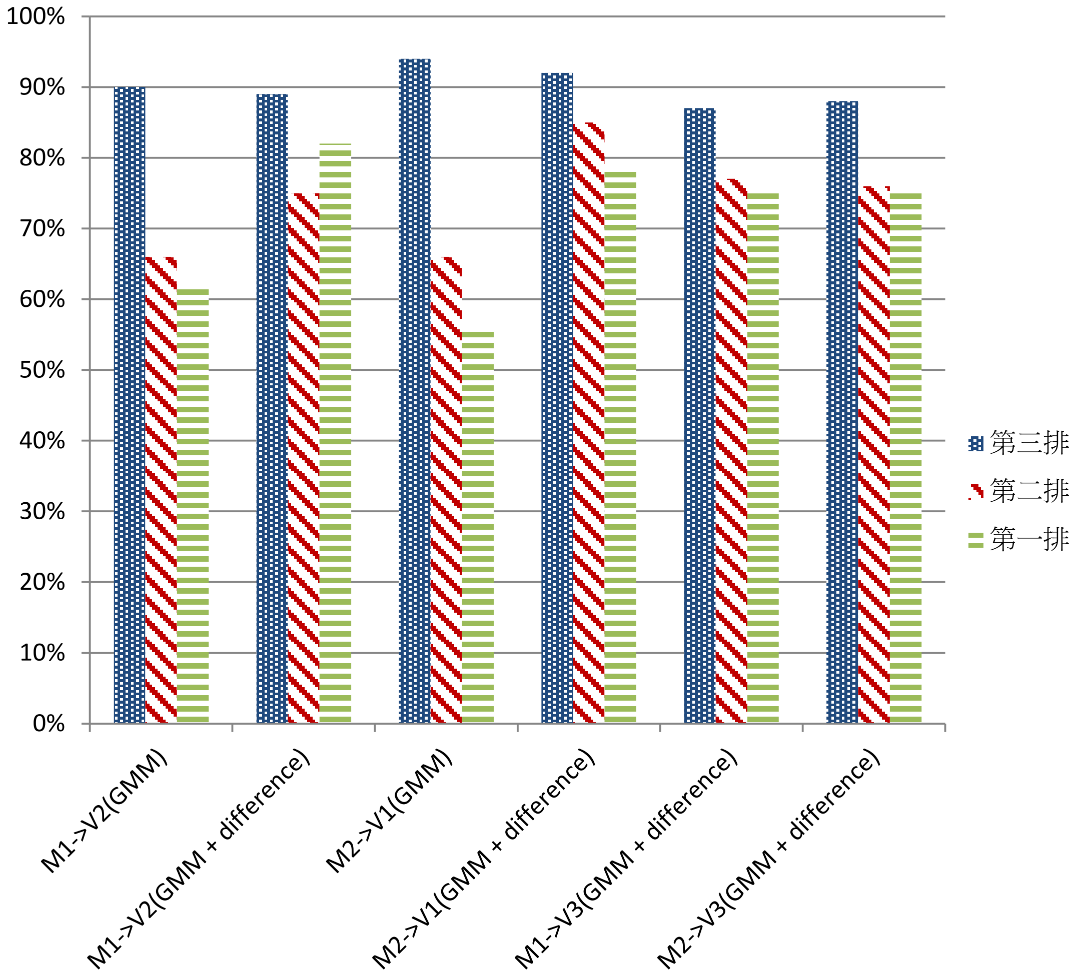

針對另一種實驗，將不同組的資料來源做交叉比較，其實驗結果在 4.3 節有更詳細的資料內容，根據實驗數據可以觀察出，如表 4-13 所示，在很多不同資料來源的交叉比較下辨識率都有提升到 70~80%以上，因此在此本研究只拿其中一組來討論，以 M1 為訓練資料(training data)、V2 為測試資料(testing data)的情況下，其方法在未使用連續影像差異(temporal differencing)時，第二排以及第一排都會受到前後的影響而降低辨識率，如圖 4-3 所示，圖 4-4 為單純只使用高斯混合模型(Gaussian Mixture of Background Model)取得前景影像的結果，用此結果會影響到辨識的結果，因此圖 4-5 為加入連續影像差異(temporal differencing)後的結果，此結果降低了前後排影響也提升辨識率。

圖 4-3 原始圖

圖 4-4 單純使用高斯混合模型取得的前景

圖 4-5 加入連續影像差異的結果

表 4-13 在不同資料來源交叉比較的辨識率



此方法加入連續影像差異(temporal differencing)後辨識率提升不少，但是依然會有錯誤的情況，辨識錯誤的例子如圖 4-6 所示。

(a)

(b)

圖 4-6 本方法判斷錯誤的例子

圖 4-6 (a)為原圖，圖 4-6 (b)為使用高斯混合模型(Gaussian Mixture of Background Model)後擷取前景的影像，觀察其結果會發現左上角有一塊未被去除的背景，此結果可能是因為高斯混合模型(Gaussian Mixture of Background Model)受到光線閃爍的影響，造成系統判斷為舉右手，經過一段時間後左上角那塊雜訊會被連續影像差異(temporal differencing)視為未移動的物件，因此之後進來的影像可以修正高斯混合模型(Gaussian Mixture of Background Model)的結果，如圖 4-7 所示，此結果為加入連續影像差異(temporal differencing)後再利用高斯混合模型(Gaussian Mixture of Background Model)去除背景的結果，因此根據實驗數據在同一組資料來源或是不同資料來源下，此方法都可以提升辨識率。

圖 4-7 使用連續影像差異後的效果

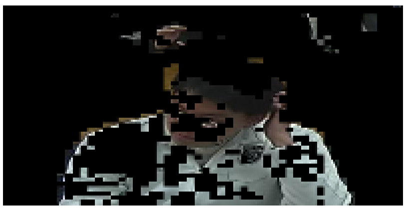
# 第五章　結論與未來研究

## 5.1 結論

本研究提出一個基於高斯混合模型(Gaussian Mixture of Background Model)之課堂舉手辨識研究，探討在複雜背景情況下可維持一定的辨識率，並且可以在教室中作為實驗的環境來測試。在這樣的真實環境中會有許多無法預估的干擾，例如前後重疊、桌面上放有雜物，實驗環境光線閃爍等等，辨識率會因此而降，背景影像的好壞間接影響研究的結果，因此本研究要減少這類的影響來提升辨識率。

在研究中利用兩種不同的方法來驗證本論文的研究方法，由此結果可以得知辨識率都有提升，其辨識率都可以達到 70~80%，若只單純使用高斯混合模型(Gaussian Mixture of Background Model)，可能會受到前後影響如圖 5-1(a) 所示，但是在加入連續影像差異(temporal differencing)後可以修正以上的問題如圖 5-1 (b) 所示，由實驗數據可以得知本研究方法是有助於提升辨識率，並且在多組不同的資料來源都可以維持一定的辨識率。

(a)



(b)

圖 5-1

## 5.2 未來研究

本研究所使用的高斯混合模型(Gaussian Mixture of Background Model)建立方法，提供的前景影像給後續的應用，但是此方法並非是完美，在環境太過複雜、強光等條件的干擾，都會影響到辨識的結果。改良的方法則可以考慮使用隱藏式馬可夫模型(Hidden Markov Model)來取代支持向量機(Support Vector Machine，SVM)，讓此系統更加完善。

# 參考文獻


1. D. M. Gavrila, “The visual analysis of human movement: A survey,”Comput. Vis. Image Understanding, vol. 72, pp. 82–98, 1999.

2. T. C. C. Henry, E. G. R. Janapriya, and L. C. deSilva, “An automatic system for multiple human tracking and actions recognition in office environment,” in Proc. ICASSP, 2003, vol. 3, pp. 45–48.

3. J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer, Multi-camera multi-person tracking for easy living,” in Proc. 3rd IEEE Int. Workshop Visual Surveillance, Jul. 2000, pp. 3–10.

4. S. Dagtas, W. A. Khatib, A. Ghafoor, and R. L. Kashyap, “Models for motion-based video indexing and retrieval,” IEEE Trans. Image Process., vol. 9, no. 1, pp. 88–101, Jan. 2000.

5. Wren C. R., Azarbayejani A., Darrell T. and Pentland A. P., “Pfinder: Real-Time Tracking of the Human Body”, IEEE Transactions on Pattern Analysis and Machine Intelligence , Vol. 19, No. 7, pp. 780-785, July 1997.

6. N. Dalal, and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, pp. 886-893, 2005.

7. R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Detecting moving objects, ghosts, and shadows in video streams,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 10, pp. 1337-1342, 2003.

8. B. Shoushtarian, and H. E. Bez, “A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking,” Pattern Recognition Letters, vol. 26, no. 1 pp. 5-26, 2005.

9. C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “Pfinder: Real-time tracking of the human body,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 780-785, 1997.

10. Qi Zang,and Reinhard Klette,“Parameter analysis for mixture of Gaussians model,” Communication and Information Technology Research Technical Report 188, 2006.

11. C. Stauffer, and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 246-252, 1999.

12. P. KaewTraKulPong, and R. Bowden, “An improved background mixture model for real-time tracking with shadow detection,” Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, vol. 25, 2001.

S. E. Chen, “QuickTime VR – An image based approach to virtual environment navigation,” Proc. SIGGRAPH 95, pp. 29-38, 1995.

13. Y. Ren, C. S. Chua, and Y. K. Ho, “Statistical background modeling for non-stationary camera,” Pattern Recognition Letters, vol. 24, pp. 183-196, 2003.

14. B.-W. Min, H.-S. Yoon, J. Soh, Y.-M. Yang, and T. Ejima, “ Hand gesture recognition using hidden Markov models,” IEEE International Conference on Systems, Man, and Cybernetics, 'Computational Cybernetics and Simulation'., Florida, USA, pp. 4232-4235, Oct. 1997.

15. J. Lafferty, A. McCallum, and F. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” International Conference on Machine Learning, Williams College, Williamstown, MA, USA, pp. 282-289, June 2001.

16. S. Yuping, and H. Foroosh, “View-invariant action recognition from point triplets,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 10, pp. 1898-1905, Oct. 2009.

17. Dailey, D.J., Cathey,F.W.,&Pumrin,S.(2000).An algorithm to estimate mean traffic speed using uncalibrated cameras. IEEE Trans. on Intelligent Transportation Systems, Vol. 1,Issue:2,June,98-107.

18. Q. Zang, R. Klette,"Parameter Analysis of Mixture of Gauissams Model,"CITR Technical Report 188, Auckland University,2006.