

第一章 序論

1.1 研究動機與背景

在我們日常生活中週遭，不管是在室內還是室外環境，隨處可見到各式各樣的監視系統，像是辦公大樓的防盜系統、停車場的車位監控系統、家庭的看護監視系統...等，其中又以攝影機的視覺監視系統(Vision Surveillance)居多，此種監視系統已從傳統的類比式攝影技術轉變到現在的數位式監視系統，而數位式監視系統的好處是在於它能對於所拍攝的內容加以分析，並對分析結果做出適當的判斷，監視過程中，隨時把資料儲存在資料庫裡，對於資料做長期性的追蹤統計、運算、分析及判斷，最後綜合得到結論，回報給在控制室的監控者，讓監控者能做適當的處置。

在自動化的數位監視系統中，如何有效率且精確的達到監控的目的，有一個很重要的前處理步驟，就是如何將前景完整且有效率的從影像中分離出來，所謂的前景是在各式各樣環境中我們有興趣的物件，如進出辦公大樓的人員、停車場內的車輛、家裡的成員...等，而將這些物件從影像中分離出來的機制，我們稱之為背景濾除(Background Subtraction)。背景濾除如果處理的好，相對地能提高後端高階處理的效果，例如因背景濾除增加物件定位的準確性，而減少追蹤時的搜尋區域；因背景濾除能將移動中的物件準確地被紀錄，而提高辨識的準確度。背景濾除的應用非常廣泛，除了在一般視覺監視系統中，還應用在工廠中產品的自動檢測、以物件為依據的編碼、視訊會議、人機介面...等各種領域。

在背景濾除的系統中，有兩個很重要的課題，第一個是如何根據當時的環境(不管是靜態或是動態)去建立最適合的背景模型，我們簡稱為背景建立(Background Construction)，這部分包含了隨著環境的改變，如何去維持以及更新背景模型；第二個部分是光影的變化，包括前景物體投射的陰

影，或是前景物體所造成周圍亮度的增加，這些變化對視覺監視系統來說，往往容易產生很大的誤判，直接影響背景模型建立的好壞。所以如何濾除光影的變化以及適應性地建立背景模型，這兩個處理的好壞，決定了整個背景濾除的效能。

至於在它的應用上，近年來由於寬頻及無線網路的普及，使得『數位家庭』(Digital Home)不斷地推廣，但大部分的目光均著重於娛樂(如網路化多媒體、數位電視等方面)，若從人類的基本需求來探討，如何達到家人安全與安心的需求，如褓母照顧小孩的狀況、家中老人的身體狀況、有無歹徒的入侵、家中環境的變動...等，『居家照護系統』是重要的一環。所以如何在數位式視覺監視系統下，結合影像背景濾除、前景的追蹤與辨識、網際網路...等數位化的技術，來提昇『居家照護』的品質，必定是未來『數位家庭』建設蓬勃發展後的一項重要的應用

基於上述的各種理由，並參考近年來的學術研究，我們去分析在室內環境中，可能會發生的各種狀況，並提出了一個結合了陰影濾除的背景模型建立演算法，達到背景濾除效果，由找出的前景資訊，提高後端所進行的高階處理之效能。

1.2 相關研究回顧

近幾年有大量的背景濾除的方法被提出，其基本的概念在最早期 Leung [1]已經明白的說明，一般來說，可以大略分成三個部份：

- 背景模型化(Background Model):如何及時且有效地建立一個強健的模型，且不定時地更新來當作背景，是背景濾除研究領域中最核心的技術所在。
- 像素分類(Pixel Classification)或相減運算(Subtraction Operation):將目前的影像與之前建立的背景模型經過此步驟的運算，將移動的物體(可能是前景)與背景的部份分開。

- 臨界點選擇(Threshold Selection):最後再對像素分類運算後的影像通過臨界點的二值化，偵測出我們所要的前景，在這部份，如何找到最合適的臨界點，使得系統能精確的偵測出前景，也是重要的目的之一。

1.2.1 背景模型化

建立背景模型是背景濾除系統中最重要也是必須先被完成的部份，所以在這方面的研究也算是最多最廣泛的，建立模型的方式有很多，不同的影像特徵值，不同的數學統計模型，就有不同的背景模型，下面歸納出到目前為止，所有曾經提出的背景模型的方法，並簡單介紹它們的適用情況與各別的優缺點。

1. 時間平均 (Time Averaging)

最基本的背景濾除模型我們可以追溯到 19 世紀，將底片經過一段時間的曝光而得到背景影像，所以建立最基本且簡單的背景模型就是利用長時間平均法(Long-Term Time Average)，由一段時間 t 內所有的影像平均而得到背景模型，數學表示式如下：

$$B(x, y, t) = \frac{1}{t} \sum_{t'=1}^t I(x, y, t') \quad (1.1)$$

其中 $I(x, y, t)$ 是在時間 t 像素 (x, y) 的強度值， $B(x, y, t)$ 是收集 t 時間後得到的參考影像，此種計算背景影像的方法為標準平均法(Standard Average)，但此種方法必須記錄所有的影像，並做加合的運算，耗時且浪費記憶體，所以一般採用移動平均法(Moving Average)，即當有新的影像 $I(x, y, t+1)$ 進來時，利用下式來更新背景模型：

$$B(x, y, t+1) = \frac{t}{t+1} B(x, y, t) + \frac{1}{t+1} I(x, y, t+1) \quad (1.2)$$

此方法有一個很大的缺點，就是一旦前景移動的很慢，或是前景在影像中停留過久，經過長時間的平均演算後，就可能被誤判成背景，

所以 Friedman [16]中提出了 Exponential Forgetting 概念，修改背景更新方式，其數學式如下：

$$B(x, y, t+1) = (1-\alpha)B(x, y, t) + \alpha I(x, y, t+1) \quad (1.3)$$

藉由調整 α 的大小，可以決定目前這張影像是否要被拿來當作背景，所以經由這種可自由調整 α 的機制，來解決前面的問題，但是效果仍然不是非常理想。

2. 統計模型(Statistical Model)

在背景濾除的方法裡，對每一個像素的強度值以統計的數學分布來代表它，我們稱之為統計模型，在這類的模型中，我們又可將其分為有參數模型(例：單一高斯模型、高斯混合模型)與無參數模型(例：核心密度估測)兩種類別：

(1)有參數模型(Parametric Model)

根據所收到的影像資訊，以及其估測的參數，使用「特定」的統計分佈來近似其真正的分布，主要有：

- 單一高斯模型(Single Gaussian Model)

在靜態的背景影像中，因為每一個像素的強度值的變化不大，所以以一個偏差量不大，且平均值為其強度的高斯分佈來代表像素的強度分佈，Azerbyjani [2]利用此法來建立模型。但是對大多數的現實生活中，背景像素的強度變化很難只用一個高斯分佈就能近似它，所以有必要將它增加到多個高斯分佈。

- 高斯混合模型(Mixture Gaussian Model-GMM)

一般在交通路況的監視系統中[16]，每一個特定的像素，在某些畫面是道路、在其它畫面又變成是車子、或是車子的陰影，所以每一個像素的顏色強度分佈，我們必須用多個高斯分佈的組合來代表它，這就是高斯混合模型的概念。

在許多過去研究中，像 Stauffer and Grimson [3]、Friedman[16]...等，都採取 GMM 的方式來建立背景的模型，除此之外，Gao et.al [4] 比較單一(single)與混合(mixture)高斯分佈的背景模型，證明出混合模型確實比單一模型在效能上來的好，即使在靜態的場景中也是如此，所以目前做背景濾除方面的研究，鮮少拿單一的高斯分佈來做背景模型。

(2)無參數模型(Non-Parametric Model)

不需要任何的理論假設，直接去預測它的密度函數去做近似真正分佈的動作，主要的方式像Elgammal [7]提出的方法，就是一種無參數的背景濾除的方法。它們利用無參數核心密度估測(nonparametric kernel density estimation)的方法，簡稱KDE，可以在凌亂且有一些微小的物體移動的環境下，做連續不斷且適應性的背景建立與濾除。而此種KDE的方法，其簡單來說，就是針對一個像素，利用此像素先前的N個顏色資訊，並配合現在觀測到的顏色資訊以及一個具有頻寬係數 σ 的核心函數 K_σ (例：Gaussian、Uniform、Triangle...等)，來算出此像素現在觀測到的顏色資訊的機率值，其數學方程式為：

$$P_r(x_t) = \frac{1}{N} \sum_{i=1}^N K_\sigma(x_t - x_i) \quad (1.4)$$

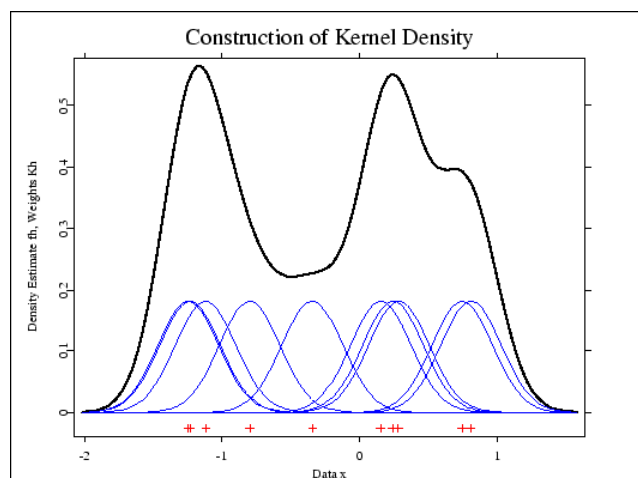


圖 1-1 利用 KDE 近似真實資料的統計分佈

圖 1-1 是在每一個資料點(紅色+)為中心，給予一個高斯的核心函數，並做加和及取權重，所得到的機率分佈函式。核心函式(Kernel Function)的種類有很多，下面表 1-1 列出常見的種類及其數學表示式。

Kernel Name	$K(u)$
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$
Uniform	$\frac{1}{2} I(u \leq 1)$
Triangle	$(1- u)I(u \leq 1)$
Epanechnikov	$\frac{3}{4}(1-u^2)I(u \leq 1)$
Quartic	$\frac{15}{16}(1-u^2)^2 I(u \leq 1)$
Triweight	$\frac{35}{32}(1-u^2)^3 I(u \leq 1)$
Co sinus	$\frac{\pi}{4} \cos(\frac{\pi}{2}u)I(u \leq 1)$

表 1-1 常見的核心函式

值得注意的是，在核心函數 K 裡的頻寬係數 σ ，若其越小，則會造成估測的機率曲線過於鋸齒化，加入了很多高頻的雜訊；反之，若其越大，則會造成估測的機率曲線太過平滑，而忽略了資料分佈中較精細的部分，所以如何選取一個適當的頻寬係數 σ ，使估測的機率曲線最接近實際的情況，是KDE中重要的一項工作，依序如下圖所示：

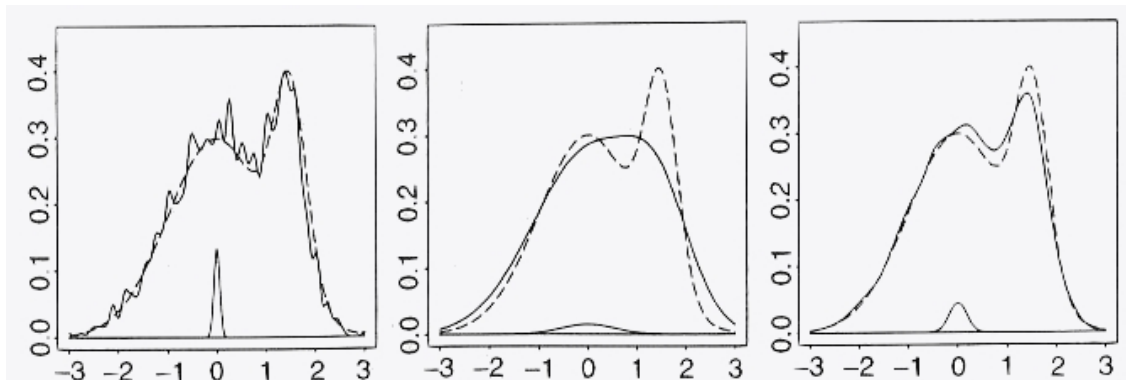


圖 1-2 Band-width 對於 KDE 的影響

3. 邊緣模型 (Edge Model)

到目前所提的所有方法，都是根據影像像素的顏色資訊(color)來建立模型，在 Yang et .al 於[5]裡提出了以影像的邊緣特徵來建立背景模型，利用每張新的影像中所有的邊緣與背景的邊緣做比對，而得到前景的邊緣，此方法稱為背景的 "Primal Sketch"。雖然利用邊緣特徵可以克服因整體影像照度變化，所造成背景被誤判的影響，提升背景濾除的效果，但卻忽略了影像區塊的連續性，所以在 Jabri et. al[6]提出了結合顏色與邊緣特徵的資訊來建立背景模型，這種方法對於複雜凌亂的背景、照度緩慢的變化、照相機的雜訊...等情況，都有不錯的效能。

但它不能克服突然的亮度變化，所造成顏色和邊緣都有大量的變化，仍會有誤判的前景；以及因其背景的邊緣資訊未做統計化的模型，所以不能解決背景物體移動的問題。

4. 以區塊為基礎之模型 (Block-based model)

前面所提出的方法幾乎都是針對每一個像素來建立模型，但是背景若是一大片幾乎不會有變化的白牆，只要以一個區塊來代表這些所有是牆的像素，建立區塊的模型，就不用再對每一個像素建立模型。根據此想法，將整張影像區塊化，分成若干個區塊，經由區塊的比對(Block Matching)來偵測前景，此以區塊為依據的背景濾除將會降低比對的運算量、資訊的儲存量，因而提高了系統的效率，如在 Hsu et. al[12]中，就用一個二元二次多項式來代表一個區塊，並使用 likelihood test 來對這些區塊進行偵測。此種方法最主要的缺點，就是因為它是以區塊為偵測單位，所以算是較不精確的偵測，加上如果處理的影像較複雜，導致一開始作影像區塊化時就有問題，反而降低了背景濾除的效能。

1.2.2 背景濾除的相關技術

背景濾除的技術除了上述的基本三項外，還有一些相關的技術被不斷地發展，這些技術能使背景濾除的效果更穩健、更有效率，接下來分別介紹這些技術及其相關研究：

1. 陰影偵測與濾除

一般在現實的生活環境中，任何物體都會伴隨著陰影的產生，大致上可分成 3 種類型的陰影：

- (1)物件的陰影投影到背景；
- (2)物件的陰影投影到本身；
- (3)物件的陰影投影到其它物件上；

在許多的影像處理應用中，包括物件的偵測、物件的分類...等，陰影將對大幅降低系統的效能，所以偵測並去除出陰影有其必要性。至於在之前所提到的背景模型中，因為沒有考慮到陰影的特性，因此系統很容易將前景物體的陰影混淆誤判斷成前景，但這並不是我們所希望的，所以偵測陰影就成為背景濾除中必須被考慮的一項重要因素。

在 1972 年時，Stockham[14]提出影像可以分成兩個部分，本身亮度的變化量和反射亮度的能力，每一個影像的強度以 $I_{x,y}$ 來表示，

$$I_{x,y} = r_{x,y} * L_{x,y} \quad (1.5)$$

其中 $L_{x,y}$ 是本身的亮度源， $r_{x,y}$ 則是反射的函數，而被陰影投影到的地方，則是將強度值乘上一個衰減函數 $\alpha_{x,y}$ ，

$$I'_{x,y} = \alpha_{x,y} * r_{x,y} * L_{x,y} \quad (1.6)$$

為了達到移除陰影的目的，首先必須先求出衰減函數 $\alpha_{x,y}$ ，再將 $I'_{x,y}$ 轉換回 $I_{x,y}$ 。

到 1994 年，Rosin etc.al[15]提出了另一種陰影偵測的方式，首先他覺得一般陰影對影像來說，類似一個半透明(semi-transparent)的區域，會使原本該區域的影像強度變小，所以又提出了陰影應該具備兩種光度上的特性：

- (1)每個屬於陰影的像素皆有一個小於 1 的光度比，此光度比是被陰影覆蓋時的強度和未覆蓋時的比值；
- (2)在陰影區域內，除了在邊緣處可能有所差異外，每一個像素的光度比皆相同。

根據上面這兩個特性，利用區域成長(region growing)的方式，以一個偵測種子(seed)為出發點，向周圍鄰近的 8 個點擴張，如果光度比相同則表示是陰影的一部份，成為下一個偵測種子，依此不斷的向外擴張偵測，而找出屬於陰影的區塊範圍。

此方法雖然在陰影偵測上有不錯的效能，但是有一些現實生活中可能會發生的狀況，可能就不能適用此方法，如可能有些物體本身就比较暗，加上它有相同的光度比，且跟其它物體產生遮蔽的效果，則就有可能會被誤判成陰影，所以在 Rosin etc.al[15]裡，他把這些狀況當作是幾乎很少發生的意外，而不去考慮它，如此的假設，在真實的環境中似乎是不可行的。

通常陰影會造成該區塊的顏色照度下降，所以透過色彩空間的轉換來偵測陰影是最普遍的方法，在 Levine [13]中是將原本 RGB 三個顏色變數轉換到 rgb 的色度座標(chromaticity coordinates)，其轉換公式如下：

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B}, r+g+b=1 \quad (1.7)$$

這種 rgb 的座標系，對於陰影所產生的照度變化沒有很大的敏感度，也就不會將陰影誤判成前景。但是這種轉換將會失去亮度的資訊，所以可能發生誤判，例如一個人穿著白色的衣服走在灰色的牆前，因為

沒有顏色的資訊，所以白色和灰色在色度座標下是一樣的，造成人無法被偵測出來，因此，需要多加了一個亮度的資訊 S 來進行補償， $S=R+G+B$ 。因此為了偵測陰影，可以將每一個像素先從 $\langle R, G, B \rangle$ 座標轉換成 $\langle r, g, S \rangle$ 座標，再進行背景濾除，如此陰影就不會被誤判成前景。

在 R. Cucchiara[20]裡提出另一種透過 HSV 顏色的轉化，來做陰影偵測的方法，轉成 HSV 顏色空間的最主要的原因，是因為它們所代表的物理涵義，與人類對顏色的感知較為接近，所以相對地有較高的辨識陰影能力。投影到背景的陰影，其色調(Hue)變化不大，亮度值(Bright Value)變小，並由實驗發現有較小的飽和度(Saturation)，根據陰影的這三樣特質，從屬於前景的像素點中偵測出陰影，其判斷的方程式如下所示：

$$SP_t(x, y) = \begin{cases} 1, & \text{if } \alpha \leq \frac{I_t^V(x, y)}{B_t^V(x, y)} \leq \beta, |I_t^H(x, y) - B_t^H(x, y)| \leq \tau^H, |I_t^S(x, y) - B_t^S(x, y)| \leq \tau^S \\ 0, & \text{otherwise} \end{cases} \quad (1.8)$$

上式中的 $I_t(x, y)$ 和 $B_t(x, y)$ 分別代表第 t 時間點，座標 (x, y) 新影像與背景模型的像素值，HSV 三個符號各別代表顏色的色調(H)、飽和度(S)和亮度值(V)，這裡共有四個門檻值須做決定，一般都是透過實驗上的經驗來決定其數值，所以門檻值決定的好壞直接影響到陰影偵測的結果。

2. 背景模型的更新與維持

前面章節提到背景模型的建立技術，能夠利用一連串의影像序列來建構初始的背景模型，但是當系統開始運作之後，背景會隨著時間而改變(例如室外環境由早到晚的陽光變化會有所不同，風向的改變造成背景中的成員動態改變，室內環境的燈光、窗簾開關以及桌椅的移動等等...)，因此如何持續的維持背景模型之正確性就成為一項很重要的研究課題了。

Elgamal[9]在 2000 年就提出了短期時間模型(Short-Term Model)與長期時間模型(Long-Term Model)的概念，所謂的短期時間模型是用來描述背景於短時間之內的變化，因此該模型的更新機制是採用選擇性的更新(Selective Update)，依據原始背景模型的分類結果來決定各個像素點的短期背景模型是否需要更新，而長期時間模型則是用來描述背景於長時間的變化情況，更新機制是利用無選擇性的更新(Blind Update)，不管背景分類結果與否，各個像素點的長期背景模型都需要更新，而 Elgamal 利用這兩個模型的結果來解決背景模型更新的問題，例如：停車場背景模型一開始建立好之後，當有一台車離開該背景視野，則長期背景模型會持續的更新以至於整體背景模型不至於誤判，而當畫面中出現短暫的擾動時，雖然長期背景模型會受到影響，但是短期背景模型依然保持正確，因而整體背景濾除的效果仍然可以保持正確性。

3. 預測 (Predictor) 的機制

還有一些背景濾除系統中會搭配預測的機制，經由一些過去觀察到的影像，來對目前的影像做預測，利用預測結果與真正觀測的影像之誤差來更新修正預測的模型，此預測的機制可以預測移動的前景位置，間接減少背景濾除的運算量，Peter et. al[8]、Koller et. al[9]已經提出了以卡式濾波器(Kalman Filter)來估測動態物體的軌跡，而 Toyama et. al[10]則是採用較卡式濾波器簡單的預測機制-韋納濾波器(Weiner filter)，而目前有越來越多較複雜的模型被提出，例如在 Zhong et. al[11]先是利用自動回歸移動平均模型(Autoregressive Moving Average Model-AEMA)來描述動態場景，再利用一個穩健的卡式濾波器來預測前景。

1.3 論文主題

除了以上所述的幾種方法外，在已知的文獻裡，綜合歸納大致上還有幾種方法，例：特徵背景法(Eigenbackgrounds)、連續核心密度近似法(Sequential KD Approximation-SKDA)... 等，把一些主要的方法拿來一起比較，我們可以得到以下的結論：

◇ 執行速度：

快速：標準平均法、中位數法(Median)、移動平均法

中速：高斯混合模型、核心密度估測、特徵背景法、最佳化平均移動法

慢速：標準平均法

◇ 記憶體需求：

高：標準平均法、中位數法、核心密度估測、平均移動法

中：高斯混合模型、特徵背景法、連續核心密度近似法

低：移動平均法

◇ 精確度：

高斯混合模型和特徵背景法能夠提供比較好的精確度，而一些較簡單的方法，例如：標準平均法、平均移動法、中位數法，在一些特殊的情況時，也會有不錯的精確度。

所以，由上述的比較可知，GMM 的方法精確度不錯，所需要的記憶體與執行速度也還算中上可以接受，所以在本論文中，我們使用 GMM 的方法做為背景濾除的主要架構，而在本論文中所做的實驗，都是針對室內環境所做的背景濾除，所以我們考慮的因素主要可以歸納成下列幾點：

- 照度上有劇烈的變化：像電燈的開關、電視或電腦螢幕畫面的變化、窗簾的開關或是天氣變化...等，所造成室內光線亮度的改變。這些可能會與原本建立的背景模型有很大的差距，而將應該是背景的部分誤判成前景。

- 背景物體的移動：像家裡的家俱的位置可能會有所變動、周期性運轉的物體(如電風扇的左右旋轉、鐘擺的擺盪、搖擺中的搖椅)，都可能被誤判成前景。
- 動態環境的建立：因為室內環境的背景可能會改變，所以要不斷地更新背景模型，且如何在有前景(包括人或移動的物體)的狀態下，對背景進行更新的動作，而不至於因前景的移動，造成部分背景被誤判成前景。
- 光影變化的濾除：在室內環境，很容易因前景物體的進入或移動，造成前景物體周圍的光影產生巨大的變化，變亮稱 Highlight，變暗則稱 Shadow。Highlight 的產生往往是因為前景物體的進入，造成光線透過物體表面產生反射，使得周圍的亮度值增加，這種情形在前景物體佔整個畫面越大時，Highlight 所造成的影響越大。由下圖 1-3 可以清楚的看到，圖(a)是背景的背景影像，圖(b)是放了一個玩偶，由 2 張影像做比較，可以明顯地看出，玩偶周圍的像素整個變亮。至於陰影的產生，在前面的章節 1.2.2 已有詳細的說明，因此不管是 Highlight 或是 Shadow 這些光影的變化都不是真正的前景，我們必須消除這些誤判。



(a)

(b)

圖 1-3 (a)沒有物體時的背景影像 (b)放了前景後周圍變亮的測試影像

總而言之，本文的主題就是針對這些問題，去建立結合顏色及邊緣資訊之 GMM，再加上一些輔助的技術，包括陰影的濾除、階層式的判斷...等，達到精確(accuracy)、穩定(robust)且適應性(adaptive)的背景濾除效果，進而透過各種室內環境的模擬實驗，來驗證其結果。

1.4 本論文的貢獻

本論文提出了一個完整的背景濾除系統，針對動態的環境變化，利用 GMM 的方法建立背景模型，且可以藉由連續影像之間活動的程度，來調整更新背景模型的速度；在前景與背景的判斷部分，因採用結合顏色與梯度的 GMM 統計資訊，做階層式的處理，可降低單純以顏色做判斷所造成的誤差。本論文中，提出了結合短期和長期模型的陰影濾除機制，減少因光影變化造成的前景誤判，提高階層式處理的正確性，最後由實驗的結果，證明此系統的確具有穩定且適應性的效能。

1.5 論文架構

本論文的組織架構簡述如下

第二章：以高斯混合模型、K 平均值分類法、期望值最大演算法來建立背景模型。

第三章：針對室內環境，提出一個結合顏色與梯度資訊的階層式(Hierarchical)背景濾除演算法，並提出一個新的陰影濾除的方法，能將 Shadow 和 Highlight 從前景中分離出來。

第四章：去模擬室內環境中常發生的情況，以我們的演算法對這些情況進行背景濾除的實驗及討論。

第五章：對於本論文做一個總結，並提出未來可以繼續發展的方向。