十三、研究計畫內容：

（二）研究計畫之背景及目的。

我們的研究計畫主要是實踐於智慧型電視(Smart TV)上，透過 Android 平台開發一套智慧型視訊監控系統(Intelligent Video Surveillance APP)。根據我們與業界全球最大的個人電腦監視器及第四大液晶電視製造商-冠捷科技公司的技術移轉經驗，以及觀察到智慧型手機(Smart Phone)的成功範例，未來智慧型電視(Smart TV)及電視專用應用程式 APP 都會有龐大的商業市場以及技術專業人才的需求。智慧型電視(Smart TV)將複製智慧型手機(Smart Phone)的 APP Store 成功商業模式和銷售量的快速成長，根據市場研究機構 Gartner 統計顯示，全球智慧型手機(Smart Phone)銷量呈現直線上升的趨勢，2013 年第二季智慧型手機(Smart Phone)銷售量達到 2.25 億支，比起去年同季的 1.54 億支還要成長 46.5% ，形成其銷售量超越功能型手機的情況，明顯表示智慧型手機(Smart Phone)已成為全球手機產品的發展重心。由於智慧型手機(Smart Phone)的逐漸普及，也同時帶動智慧型手機(Smart Phone)上應用程式 APP 的需求量越來越大，以手機軟體市場 Apple App Store 與 Google Play 這兩個雲端軟體平台來觀察，Apple App Store 於今年五月已達到了 500 億次的下載量；而 Google Play 也達到 480 億次的下載量，顯示出雲端軟體商店的商業模式逐漸的穩定成熟且已成為未來趨勢。雖然目前 Android 平台上 Google Play 雲端軟體平台已有大量各類型的應用程式 APP，但是目前都只適用在智慧型手機(Smart Phone)上，大多數無法在智慧型電視(Smart TV)使用，其主要問題可分為下列幾項要點：

1) 智慧型手機(Smart Phone)和智慧型電視(Smart TV)的控制方式不同

智慧型手機(Smart Phone)的操作目前有螢幕多點觸碰(Multi-Touch)方式，以及包含重力加速計(G-sensor)、電子羅盤(M-sensor)與陀螺儀(Gyroscope)等，感測器方式來做為使用者操控；而智慧型電視(Smart TV)主要透過無線遙控器、語音操縱、手勢辨識等方式，與使用者操控與互動，也因為兩者的控制方式有很大的不同，因此在智慧型手機(Smart Phone)上所設計的應用程式APP無法順利移植到智慧型電視(Smart TV)上來使用。

2) 智慧型手機(Smart Phone)和智慧型電視(Smart TV)的螢幕畫面的尺寸差異

目前市面上的智慧型手機(Smart Phone)的螢幕尺寸普遍約3.5吋至6.9吋左右之間；而智慧型電視(Smart TV)的螢幕尺寸則介於32吋至60吋之間。而Android中的Layout主要有LinearLayout、FrameLayout、RelativeLayout與TableLayout等不同的介面佈局方式，再加上智慧型手機(Smart Phone)與智慧型電視(Smart TV)兩者的螢幕尺寸與解析度的差異極大，因此在智慧型手機(Smart Phone)上所設計的使用者操作介面(User Interface, UI)，若移植到智慧型電視(Smart TV)，會因為螢幕尺寸與解析度的有所不同，則可能導致智慧型電視(Smart TV)上無法顯現合適正確的UI畫面。

3) 智慧型電視(Smart TV)與智慧型手機(Smart Phone)兩者的隱私性與使用性質是不同

目前智慧型手機(Smart Phone)被定位於單一使用者使用，其儲存與顯示的內容以及使用方式算屬於個人隱私的範疇；而智慧型電視(Smart TV)則普遍被定位在多人共同使用，因此被設計於智慧型手機(Smart Phone)中較屬於隱私類型的應用程式APP(例如：聊天視訊、社群網路、信箱等軟體)，若將如此類型之應用程式APP移植在智慧型電視(Smart TV)上，則可能發生個人隱私資訊洩漏的問題。因此，本研究計畫將基於Android平台的智慧型電視(Smart TV)上開發一套智慧型視訊監控系統(Intelligent Video Surveillance APP)，其主要應用範圍包含居家老人與小孩的安全監控系統、居家防盜系統、社區與大樓保全…等私人區域的視訊監控系統，並針對我們上述所討論的三個要點問題來進行考量與解決，

而我們智慧型視訊監控系統(Intelligent Video Surveillance APP)的主要特點可分為下列幾項：

1) 基於智慧型電視(Smart TV)上對智慧型手機(Smart Phone)設計有效操控 APP

目前智慧型電視(Smart TV)的控制方式包含有紅外線遙控器、語音操縱、體感控制以及手勢辨識等方式。但根據我們和冠捷科技公司的實務技術移轉經驗，目前操控電視仍是紅外線遙控器為主，而語音操縱的缺點在於容易遭受電視上的聲音影響，並可能造成控制上的誤判，至於手勢與體感辨識的缺點則是在於敏銳度不高，反應時間過長。紅外線遙控器，是目前市場接受度最高，同時也是許多人習慣的控制方式，並具有反應速度快的特性，所以我們設計在智慧型電視(Smart TV)上的智慧型視訊監控系統(Intelligent Video Surveillance APP)的介面，將會對紅外線遙控器做最佳化的操作介面設計。此外，由於智慧型電視(Smart TV)上每個APP界面的差異性太大，一般的紅外線遙控器上的固定按鈕，可能無法有效地快速操控每個智慧型電視(Smart TV) APP的介面按鈕。因此，我們將在智慧型手機(Smart Phone)上，針對不同智慧型電視(Smart TV) APP來設計專屬的遙控器APP，讓每個智慧型電視(Smart TV)上的APP都有最佳化的遙控UI介面。例如：智慧型電視(Smart TV)播放KTV APP畫面時，一般的紅外線遙控器對於選歌紀錄、歌手查詢與插歌等動作的情況會難以有效的操控。因此，如果在智慧型手機(Smart Phone)上，設計專屬的遙控器APP，即可輕易地進行歌曲與歌手查詢、歌曲紀錄與插歌等操作，如下圖示意圖所示：

圖一、 Smart Phone 上設計專屬的遙控器 APP，即可輕易的控制 Smart TV 上的 KTV APP

2) 能根據任何畫面尺寸的智慧型電視(Smart TV)來自動地達到最佳化的UI介面

我們智慧型視訊監控系統(Intelligent Video Surveillance APP)將會針對各智慧型電視(Smart TV)的畫面尺寸與解析度來進行判斷，以及根據畫面尺寸大小計算出長與寬的比例權重，並再透過比例權重對Android中的LinearLayout、FrameLayout、RelativeLayout與TableLayout等介面佈局方式來調整，針對不同尺寸的智慧型電視(Smart TV)畫面，調整出最佳化的介面佈局。

3) 將智慧型視訊監控系統(Intelligent Video Surveillance APP)結合人臉辨識技術，達到隱私性與便利性共存之目的

隱私性與便利性在我們生活中是非常被受關注的議題，但往往這兩種議題一直都是相互衝突。而我們同時基於隱私性與便利性這兩者議題的考量，因此我們將在智慧型視訊監控系統(Intelligent Video Surveillance APP)上結合人臉辨識技術，其透過人的臉部特徵進行身份辨識，並且具有直覺、友善、迅速與便利等特性，以達到隱私防護性與迅速便利性的目的。根據我們與冠捷科技公司的技術移轉經驗，在智慧型電視(Smart TV)的家庭應用APP類型當中，智慧型視訊監控系統(Intelligent Video Surveillance APP)的需求量是排行第一，此外在雲端軟體平台上，目前也還沒有一個好的智慧型家庭監控APP能符合市場的需求。 因此，我們將於智慧型電視(Smart TV)上，透過Android平台開發一套智慧型視訊監控系統(Intelligent Video Surveillance APP)，主要目的在於強化每個家庭的生活安全，進而有效提升防盜與小孩和老人居家安全監控的防護需求。甚至冠捷科技公司表示未來所生產的每台智慧型電視

(Smart TV)上都內建一套智慧型視訊監控系統(Intelligent Video Surveillance APP)的強烈需求。而目前著名的視訊監控演算法有Sigma Difference Estimation (SDE)[1]、Multiple Sigma Difference Estimation (MSDE)[2]、Gaussian Mixtures Models (GMM)[3]、Simple Statistical Difference (SSD)[4]、Multiple Temporal Difference (MTD)[5] 、 Self-Organizing Background Subtraction (SOBS)[6] 、 Codebook Background Subtraction (CBS) [7]，以下我們將簡短地描述這幾個主要方法[1]-[7]：

1. Sigma Difference Estimation (SDE) [1]

一種漸進式適應法(Sigma Difference Estimation)被提出來估測原始輸入影像序列中每一個像素隨著時間之變化情形(temporal statistics)。在第一階段的估測中，漸進式適應法(Sigma Difference Estimation)使用sgn函數建立出自適應背景模型。sgn函數可依下列方式來表示：

$$\text{sgn}(a)=\begin{cases}1, & a>0\\ 0, & a=0\\ -1, & a<0\end{cases} \quad (1)$$

其中a代表一個實數。而漸進式適應法中的自適應背景模型可依下列 recursive filter 進行建構：

$$B_t(x,y) = B_{t-1}(x,y) + \text{sgn}(I_t(x,y) - B_{t-1}(x,y)) \quad (2)$$

其中$B_t(x,y)$代表自適應背景模型之第 t 張背景影像，而$I_t(x,y)$代表第 t 張原始輸入影像。使用sgn函數之後，此方法之背景模型的各個像素將於每個畫面進行±1的調整而模擬出真實的背景影像。則各個監視畫面$I_t(x,y)$與背景影像$B_t(x,y)$的絕對差值影像可表示為：

$$\Delta_t(x,y) = |I_t(x,y) - B_t(x,y)| \quad (3)$$

而在第二階段的估測中，類似於自適應背景模型的建立方式，一種時間活動量的衡量(measure of motion activity)被用來偵測各個監視畫面$I_t(\text{x,y})$是否包含移動物體，此衡量方式可依下列式子來表示：

$$V_t(x,y) = V_{t-1}(x,y) + \text{sgn}(N(\Delta_t(x,y)) - V_{t-1}(x,y)) \quad (4)$$

其中$V_t(x,y)$代表現在畫面之time-variance，$V_{t-1}(x,y)$代表前一張畫面之time-variance，$\Delta_t(x,y)$代表現在畫面之絕對差值影像，而N為預先定義的參數，定義之實數範圍為1至4。最後，每個畫面下判斷移動物體之 binary motion detection mask $D_t(x,y)$可被表示為：

$$D_t(x,y) = \begin{cases}1, \Delta_t(x,y) > V_t(x,y)\\ 0, \Delta_t(x,y) \le V_t(x,y)\end{cases} \quad (5)$$

2. Multiple Sigma Difference Estimation (MSDE) [2]

在現實環境中通常包含一些較複雜的監視場景，像是移動物體會突然放慢速度、突然停止，或是移動物體過多的情形。這些情形導致漸進式適應法(Sigma Difference Estimation)建立背景模型的失敗。因此，一種多重漸進式適應法(Multiple Sigma Difference Estimation)被提出來建立複合式背景模型來解決上述問題。則此複合式背景模型可被表示為：

$$b_t^i(x,y) = b_{t-1}^i(x,y) + \text{sgn}(b_t^{i-1}(x,y) - b_{t-1}^i(x,y)) \quad (6)$$

其中$b_t^i(x,y)$代表現在畫面之第i層參考背景，b_{t-1}^i代表前一張畫面之第i層參考背景，b_t^{i-1}代表現在畫面之第$i-1$層參考背景。類似地，reference difference $\Delta_t^i(x,y)$和reference time-variance $v_t^i(x,y)$也是以相似的方式被計算出來。

$$v_t^i(x,y) = v_{t-1}^i(x,y) + \text{sgn}\left(N\left(\Delta_t^i(x,y)\right) - v_{t-1}^i(x,y)\right) \quad (7)$$

其中 $\Delta_t^i(x,y) = \left|I_t(x,y) - b_t^i(x,y)\right|$。而在所有的$b_t^i(x,y)$和$v_t^i(x,y)$被得到之後，每個畫面下最可以被信賴的背景影像$B_t(x,y)$可依下列式子被定義：

$$B_t(x,y) = \frac{\sum_{i\in[1,K]}\frac{\alpha_i(b_t^i(x,y))}{v_t^i(x,y)}}{\sum_{i\in[1,K]}\frac{\alpha_i}{v_t^i(x,y)}} \tag{8}$$

其中α_i代表各層計算預先定義的變數，i變數代表背景影像$B_t(x,y)$所需要的計算階層，K代表計算階層i的總層數。

3. Gaussian Mixture Model (GMM)[3]

高斯混合模型法(Gaussian Mixture Model)使用混合的K個高斯分佈來將一段期間內的背景模型化並儲存，其當前像素的機率可用下式表示：

$$P(X_t) = \sum_{i=1}^{k}\omega_{i,t}\,\eta(X_t,\mu_{i,t},\Sigma_{i,t}) \tag{9}$$

其中X_t為第 t 個輸入影像的每個像素值，$\omega_{i,t}$為一個權重向量，其對應的高斯模型可以用下式表示：

$$\eta\left(X_t,\mu_{i,t},\Sigma_{i,t}\right) = \frac{\exp[-((X_t-\mu_t)^T/2)\Sigma^{-1}(X_t-\mu_t)]}{2\pi^{n/2}|\Sigma|^{1/2}} \tag{10}$$

其中$\mu_{i,t}$為第 i 個高斯分佈的平均值，$\Sigma_{i,t}$爲第 i 個高斯分佈的協方差矩陣(covariance matrix)。為了計算上的方便，$\Sigma_{i,t}$可假設為：

$$\Sigma_{k,t} = \sigma_k^2 I \tag{11}$$

如果像素值在2.5倍標準差之內，該像素的狀態可被判為匹配，反之像素值會和已存在的K個高斯分佈去比對。匹配的高斯混合模型的適應參數以下式更新：

$$\omega_{k,t} = (1-\alpha)\omega_{k,t-1} + \alpha M_{k,t} \tag{12}$$

$$\mu_t = (1-\rho)\mu_{-1} + \rho X_t \tag{13}$$

$$\delta_t^2 = (1-\rho)\delta_{t-1}^2 + \rho(X_t-\mu_t)^T(X_t-\mu_t) \tag{14}$$

和模型匹配的像素，參數$M_{k,t}$設為1，反之設為0。接下來，用每一個高斯分佈的ω/σ值得到計算出背景模型，初始的B分佈可用下式得到：

$$\mathrm{B} = \mathrm{argmin}(\textstyle\sum_{k=1}^{b}\omega_k > T_2) \tag{15}$$

其中T_2為判斷像素為背景的最小值。

4.Simple Statistical Difference(SSD)[4]

基於統計學的平均值和標準差理論，統計估測法(Simple Statistical Difference)是計算出輸入影像序列每個像素於時間軸上的平均值以及標準差來偵測移動物體。則於時間間隔$[t_0,\ t_{\mathrm{k}-1}]$中每個輸入畫面所得到的平均值二維陣列μ_{xy}和標準差二維陣列σ_{xy}可被表示為：

$$\mu_{xy} = \frac{1}{K}\sum_{k=0}^{K-1} I_k(x,y) \tag{16}$$

$$\sigma_{xy} = \sqrt{\frac{1}{K}\sum_{k=0}^{K-1}\left(I_k(x,y)-\mu_{xy}\right)^2} \tag{17}$$

基於時間軸上的平均值和標準差，每個畫面下判斷移動物體之binary motion detection mask $D_t(x,y)$可被表示為：

$$D_t(x,y) = \begin{cases} 1, \left|I_t(x,y)-\mu_{xy}\right| > \lambda\sigma_{xy} \\ 0, \left|I_t(x,y)-\mu_{xy}\right| \le \lambda\sigma_{xy} \end{cases} \tag{18}$$

其中λ代表人為調整的可適應性參數。根據文獻Simple Statistical Difference(SSD)，當λ設為3時可以得到較為精確的移動物體偵測結果。

5. Multiple Temporal Difference (MTD)[5]

時間域差值法是透過計算連續影像畫面的差值來偵測移動物體，但其偵測到的物體常形狀不完整，特別是移動物體在畫面中停留或移動緩慢時。因此多重差值影像法(Multiple Difference Images)被提出來解決這個問題，其偵測結果可藉由下式(19)和(20)得到：

$$I_d^n(t) = I(t) - I(t-n) \tag{19}$$

$$I_d^s(t) = \sum_{n=1}^{m} I_d^n(t) \tag{20}$$

$I(t)$為當前的影像，$I(t-n)$為n張前的影像，$I_d^n(t)$爲$I(t)$和$I(t-n)$的差值影像，$I_d^s(t)$為m張差值影像的總合，最後的二值化偵測結果可透過一門檻值判斷$I_d^s(t)$來得到，門檻值和影像張數m皆是可調整的參數。傳統的方法只計算兩張連續影像$I(t)$和$I(t-1)$的差值，而多重差值影像法透過多張差值的計算來改善傳統的方法。

6.Self-Organizing Background Subtraction (SOBS)[6]

方法包含兩個基本步驟。第一步，把原始影像中的每個像素映射在一個3×3的神經網路矩陣(matrix of neuronal map)上，用來建構初始的背景模型。第二步，使用一個固定的門檻值，在3×3矩陣中針對每個輸入像素找出和背景最相符的矩陣候選：

$$d\big(c_m, p_t(x,y)\big) = \min_{i=1,\dots,9} d\left(c_i, p_t(x,y)\right) \leq \in \tag{21}$$

其中$p_t(x,y)$是輸入的像素，c_i是3×3矩陣中的第i個候選值，c_m是最佳的配對。如果輸入像素 $p_t(x,y)$中找不到最佳配對，$p_t(x,y)$就被當作是移動物體的一部分；反之，$p_t(x,y)$就被當作是背景的像素。如果最佳配對c_m是在背景模型的(x,y)座標位置上，背景模型就會修正成下列：

$$A_t(i,j) = \Big(1 - \alpha_{i,j}(t)\Big) A_{t-1}(i,j) + \alpha_{i,j}(t) p_t(x,y) \tag{22}$$

其中的$i = \bar{x} - 1 \dots \bar{x} + 1$，而$j = \bar{y} - 1 \dots \bar{y} + 1$，$A$是神經網路背景模型，$\alpha$是學習速率。

7.Codebook Background Subtraction(CBS) [7]

編碼簿背景相減法(Codebook Background Subtraction)是經由觀察一段長時間後，使用量化/群集技術來建構背景模型。編碼表中有一或多個編碼，是由每個像素生成的，如下：

$$C = \{c_1, c_2, \dots, c_L\} \tag{23}$$

其中C是編碼表，有L筆編碼資料，c_i是第i筆編碼資料，每筆編碼資料含有RGB向量v_i和位元組u_i，並分別可以表示成(R_i, G_i, B_i)和$(I_i^{min}, I_i^{max}, f_i, \lambda_i, p_i, q_i)$，其中的$I^{min}$、$I^{max}$分別代表編碼的最小亮度值和最大亮度值，$f$是編碼發生的頻率，$\lambda$是測試時間的區間，表示編碼沒有出現的最長時間區間，p和q則是第一次和最後一次的取得時間。當輸入像素 $x_t = (R, G, B)$，編碼c_i的RGB向量v_i可以表示成如下：

$$\|x_t\|^2 = R^2 + G^2 + B^2 \tag{24}$$

$$\|v_i\|^2 = \overline{R_i^2} + \overline{G_i^2} + \overline{B_i^2} \tag{25}$$

$$\langle x_t, v_i \rangle^2 = (\overline{R}_i R + \overline{G}_i G + \overline{B}_i B)^2 \tag{26}$$

因此，色彩失真δ可以由以下公式得到：

$$p^2 = \|x_t\|^2 cos^2\theta = \frac{\langle x_t, v_i \rangle^2}{\|v_i\|^2} \tag{27}$$

$$\text{color dist}(x_t, v_i) = \delta = \sqrt{\|x_t\|^2 - p^2} \tag{28}$$

其中v_i是編碼c_i的RGB向量，δ是x_t和v_i中間的色彩失真。邏輯亮度函數(logical brightness function)定義如下：

$$\text{brightness}\left(I, \langle \check{I}, \hat{I} \rangle\right) = \begin{cases} \text{true}, & if\ I_{low} \le \|x_t\| \le I_{hi} \\ \text{false}, & \text{otherwise} \end{cases} \tag{29}$$

接著，每筆編碼的範圍$[I_{low}, I_{hi}]$ 又可以定義成如下：

$$I_{low} = \alpha \hat{I} \tag{30}$$

$$I_{hi} = min\left\{\beta \hat{I}, \frac{\check{I}}{\alpha}\right\} \tag{31}$$

其中α和β是預先定義的參數。一般來說，α的範圍在0.4到0.7之間，β的範圍在1.1到1.5之間。最後，偵測結果可以經由下列兩個情況得到：

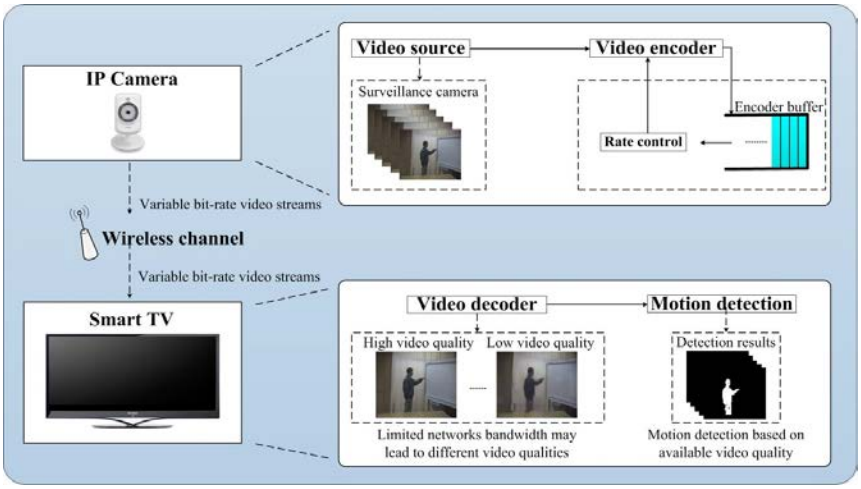
$$\text{color dist}(x, c_m) \le \varepsilon_2 \tag{32}$$

$$\text{brightness}\left(\text{I}, \langle \check{I}_m, \hat{I}_m \rangle\right) = \text{true} \tag{33}$$

其中ε_2是偵測的門檻值，c_m是背景的編碼。如果輸入像素符合上述的兩個條件，它就被視為是背景，反之，它就是移動物體。而根據先前的實驗結果，每筆參考背景的編碼約是6.5左右。

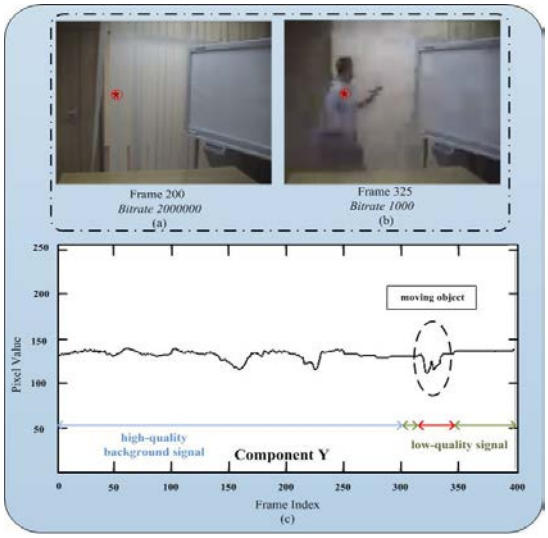
因此，本研究計畫智慧型視訊監控系統(Intelligent Video Surveillance APP)，將透過Android平台並實踐於智慧型電視(Smart TV)上，研究與開發期主要劃分為三部分，其分別為：第一部分，智慧型視訊監控系統(Intelligent Video Surveillance APP)設計自動化偵測感興趣的移動物體，並可支援擴充IP Camera的無線視訊影像傳輸。第二部分，智慧型視訊監控系統(Intelligent Video Surveillance APP)針對感興趣的移動物體，設計人臉和物體辨識技術。第三部分，智慧型視訊監控系統(Intelligent Video Surveillance APP)基於雲端伺服器之分散式資料庫，對人臉和物體的資料做有效快速的管理。

因為無線IP Camera的視訊影像傳輸，具有即時、高度擴充性與方便性，以及無實體線路困擾等特性，因此無線IP Camera已成為市場主流與技術發展的重點。我們計畫在智慧型視訊監控系統(Intelligent Video Surveillance APP)支援IP Camera的無線視訊影像傳輸。而如下圖二表示為在當無線IP Camera透過無線網路傳輸傳遞視訊影像至Smart TV上時，IP Camera上的視訊編碼晶片會依據目前的網路狀況進而調整至目前視訊影像最佳的位元傳輸率(Bit-rate)，同時並產生可變位元率(Variable Bit-rate)的影像串流來進行傳輸，因此，在Smart TV端的Video decoder將會解出不同品質的視訊影像。



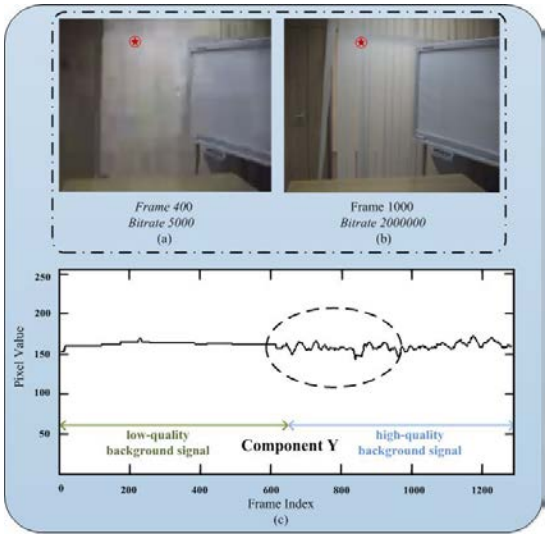
圖二、在真實情況的無線網路傳輸中，IP Camera透過無線視訊串流傳遞至Smart TV之系統架構示意圖

然而，如上述所提到目前七個著名的視訊監控演算法，其分別有Sigma Difference Estimation(SDE)[1]、Multiple Sigma Difference Estimation(MSDE)[2]、Gaussian Mixtures Models(GMM)[3]、Simple Statistical Difference(SSD)[4]、Multiple Temporal Difference(MTD)[5]、Self-Organizing Background Subtraction (SOBS)[6]、Codebook Background Subtraction(CBS)[7]，雖然這些知名的演算法在有線的USB Camera視訊串流傳輸下能有效地偵測到移動物體。然而，我們發現上述七個著名的視訊監控演算法，在無線網路IP Camera的視訊串流傳輸下，會造成嚴重的雜訊、鬼影等情況產生，導致這些著名的演算法無法有效的偵測移動物體。而主要的原因在於無線的網路傳輸下，通常會造成有限且不穩定的頻寬問題。而為了在無線網路下能達到即時的視訊傳輸，速率控制(Rate Control)被採用來依據目前的網路狀況進而調整目前影像的位元傳輸率(Bit-rate)，同時並產生可變位元率(Variable Bit-rate)的影像串流來進行傳輸，因此，在Video Decoder會解出不同品質的視訊影像。然而，使用這些不同品質的視訊影像來偵測移動物體，如上述這些著名的演算法皆會產生嚴重的誤判，導致無法有效的偵測移動物體。圖三為我們分析移動物體的偵測在無線網路情況中為何失效，我們分析無線網路透過無線IP Camera視訊傳輸下的像素強度訊號，並同時描述從高品質頻寬(高品質影像)至低品質頻寬(低品質影像)的傳輸情況。其中圖三(a)為高品質視訊傳輸下的清晰影像，且具有較大像素強度的浮動訊號；而圖三(b)為低品質視訊傳輸下的模糊影像，且具有較低強度的浮動訊號。然而，像是上述所提的知名的視訊監控演算法，在高品質頻寬(高品質影像)的無線視訊傳輸時，已經適應這些較大像素強度的浮動訊號為背景訊號。因此，在無線的低品質視訊傳輸下，會將移動物體的浮動訊號誤判為背景訊號，導致錯誤的偵測結果。



圖三、在無線視訊傳輸下，高位元頻率(高品質影像)逐漸調整至低位元頻率(低品質影像)的像素強度的變動情況示意圖

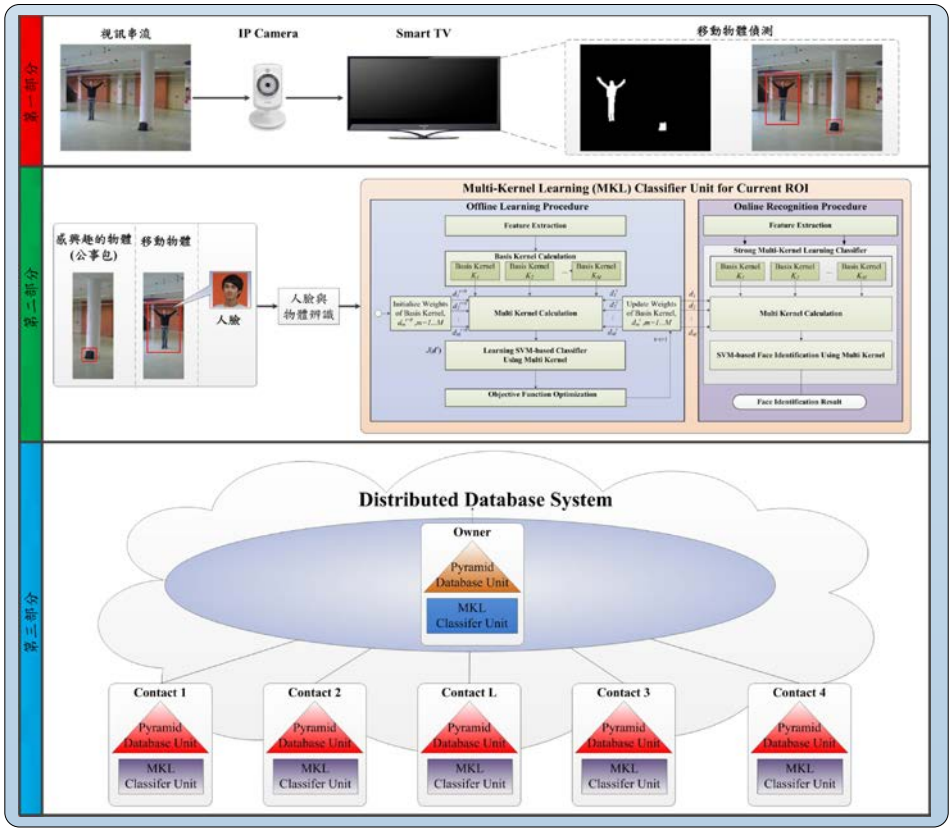
此外，下圖四也是我們分析於無線網路情況當中，並透過無線IP Camera視訊傳輸下的像素強度訊號分析，以及描述從低品質頻寬(低品質影像)至高品質頻寬(高品質影像)的傳輸情況。其中圖四(a)為低品質視訊傳輸下的模糊影像，具有較低像素強度的浮動訊號；而圖四(b)為高品質視訊傳輸下的清晰影像，而具有較大像素強度的浮動訊號。在無線傳輸的低品質頻寬(低品質影像)情況時，將會把較低強度浮動訊號視為背景訊號，如果無線網路的頻寬變好，速率控制(Rate Control)調整目前影像的產生訊號浮動較大的品質傳輸(高品質影像)時，上述所提到的七個著名視訊監控演算法，容易誤判訊號浮動較大的高品質影像為移動物體訊號，導致錯誤的偵測結果。



圖四、在無線視訊傳輸下，低位元頻率(低品質影像)逐漸調整至高位元頻率(高品質影像)的像素強度的變動情況示意圖

（三）研究方法、進行步驟及執行進度。

圖五為本研究計畫智慧型視訊監控系統(Intelligent Video Surveillance APP)的整體架構圖：



圖五、本研究計畫之整體架構圖

第一部分：智慧型視訊監控系統(Intelligent Video Surveillance APP)設計自動化偵測感興趣的移動物體，並可支援擴充 IP Camera 的無線視訊影像傳輸

為了有效改善因為IP Camera在無線網路視訊傳輸中的速率控制(Rate Control)所造成不同位元頻率視訊串流(Different Bit-rate Video Streams)的問題，以及達到精確地監控視訊的偵測結果。因此，我們提出一個能在低位元頻率或是高位元頻率的視訊串流情況下都能達到精確的監控視訊演算法，我們

監控視訊演算法將基於類神經網路理論的反傳遞網路(CPN)技術，以反傳遞網路為基底所建構移動物偵測(CPNMD)的兩個重要的模組：多重背景產生模組和移動物體擷取模組。而多重背景產生模組與移動物體擷取模組的主要步驟說明分別如下：

步驟一) 多重背景產生模組

為了支援各種不同的影片串流，這裡列出三個變數Y(照度)和C_b(藍色濃度差)和C_r紅色濃度差用來建構輸入影像的YC_bC_r色彩空間。I_t表示輸入的frame輸入影像，而每個輸入像素用$p_t(x,y)$表示，(Y,C_b,C_r)分別代表像素的照度、藍色濃度差與紅色濃度差。適應型背景模型可以支援有位元頻率變化的視訊串流，而根據贏者全拿學習規則，適應型背景模型是在候選資料中找出比贏的數值，用來將模型很快的建構出來。為了決定正確的數值，我們使用歐幾里德距離法來計算在第t張frame I_t上每個輸入的像素 $p_t(x,y)$和對應的候選背景$B(x,y)_1$到$B(x,y)_k$的距離值，其公式如下：

$$d\big(p_t(x,y),B(x,y)_k\big)=\left\|p_t(x,y)\text{-}B(x,y)_k\right\|_2^2 \ , \ where\ k=1,2,\dots,N \tag{34}$$

算出距離之後，勝利的數值$B(x,y)_{win}$就是背景候選值和輸入值$p_t(x,y)$的最短距離，表示如下：

$$d\big(p_t(x,y),B(x,y)_{win}\big)=\min_{k=1\sim N} d\big(p_t(x,y),B(x,y)_k\big) \tag{35}$$

當最小值超過ε值的時候，對應的輸入像素$p_t(x,y)$ 就會成為新的背景候選之一，並表示如下：

$$p_t(x,y)\begin{cases}\notin B(x,y)_{win}\ , if\ d\big(p_t(x,y),B(x,y)_{win}\big)>\varepsilon\\ \in B(x,y)_{win}\ , otherwise\end{cases} \tag{36}$$

其中ε值是一個容忍值，用來決定輸入像素$p_t(x,y)$是否是屬於勝利值$B(x,y)_{win}$。而當勝利值$B(x,y)_{win}$是從輸入的像素$p_t(x,y)$之中選出時，下一張frame的勝利值$B(x,y)_{win}$和對應的權重$\pi(x,y)$值就可以用藉由下列的公式算出：

$$B(x,y)'_{win}=B(x,y)_{win}+\alpha[p_t(x,y)-B(x,y)_{win}] \tag{37}$$

$$\pi(x,y)'_{win}=\pi(x,y)_{win}+\beta[Y_t(x,y)-\pi(x,y)_{win}] \tag{38}$$

其中$Y_t(x,y)$代表第t張frame的Grossberg layer輸出值，α和β是學習速率，且所有權種的初始值都是1。而如果$B(x,y)_{win}$在第t張frame I_t中和輸入像素$p_t(x,y)$非常相近時，對應的權重$\pi(x,y)$要跟著遞增，而其他的權重則要遞減。

步驟二) 移動物體擷取模組

1）移動物體偵測：

多重背景產生模組建構完之後，輸入像素$p_t(x,y)$的YC_bC_r色彩成分會在Kohonen layer $W(x,y)_1$到$W(x,y)_i$中被跟著算出來，而在這裡我們將介紹高斯歸屬函數中的移動物體偵測步驟，並加上歐幾里德距離法，用來計算當下的輸入像素$p_t(x,y)$到第i個Kohonen layer W的相似度，其高斯歸屬函數：

$$S_i(p_t,W_i)=\exp\left(\frac{-\|p_t-W_i\|^2}{2\Delta^2}\right) \tag{39}$$

而其中的$\mathrm{i}=1\sim M$，M是Kohonen neurons，Δ是經驗上的容忍值，$\mathrm{S_i}\in[0,1]$。輸入的frame被分割為$N\times N$大小，之後每個$N\times N$的區塊經由高斯歸屬函數計算，每個$N\times N$的區塊的計算結果總和如下：

$$\gamma=\sum_{p_t\in\mu}\sum_{i=1}^{M}S_i(p_t,W_i) \tag{40}$$

區塊大小N設定為4，p_t是每個像素對應區塊的ρ值，M是一個Kohonen neurons數值。而當區塊(i,j)沒有超越門檻值τ值，$A_s(i,j)$就設為1，代表的是區塊有較高的機率是移動物體，設為0，則是代表有較高

的機率是背景資訊，$A_s(i,j)$可以表示為：

$$A_s(i,j)=\begin{cases}1, if\ \gamma \le 0\\ 0, otherwise\end{cases} \tag{41}$$

2）移動物體擷取：

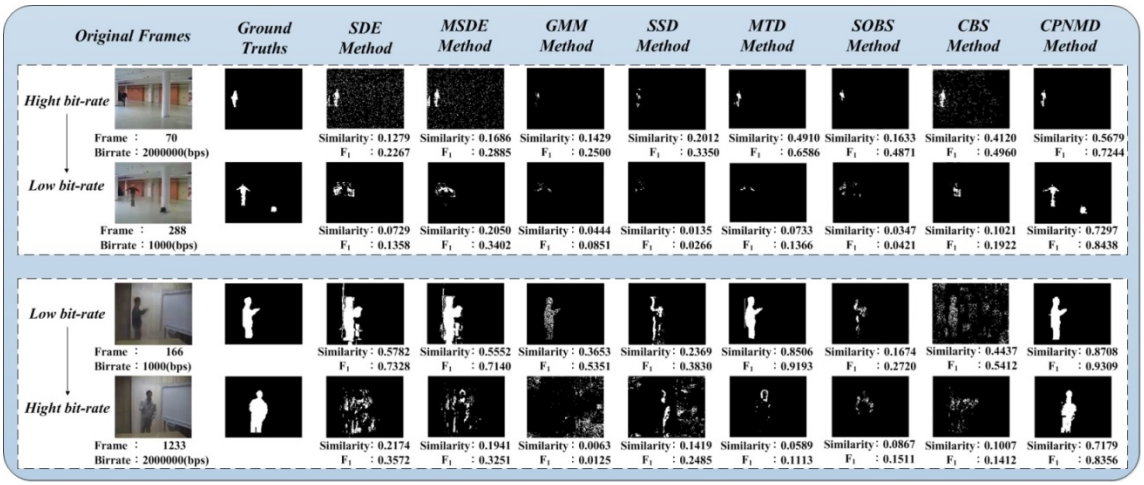
移動物體偵測步驟會有效地排除含有較高機率的背景資訊區塊之後，移動物體擷取步驟只會檢查含有較高機率是目標物體的區塊。最後，再透過CPN的輸出值來產生一個二維移動物體偵測遮罩。而這種方法是利用Kohonen layer和Grossberg layer的權重線性組合，其表示如下：

$$\mathrm{Y}=\sum_{i=1}^{M} S_i\pi_i \tag{42}$$

其中的S_i是第i筆Kohonen neuron的輸出果，π_i是Kohonen neuron和Grossberg layer之中第i筆資料的權重，M則是一個Kohonen neurons的值。這個二值化的移動物遮罩可以寫成：

$$E(x,y)=\begin{cases}1, if\ Y(x,y) \le \omega\\ 0, otherwise\end{cases} \tag{43}$$

其中$E(x,y)$設為代表一個移動物體中的像素，設為 1 代表前景像素，若為 0 則是代表背景像素；而ω為一個容忍的門檻值。下圖六為我們的視訊監控演算法 CPNMD 與上述七個演算法的初步實驗結果圖：



圖六、初步實驗結果

在無線視訊串流傳輸下，我們同時考慮高位元頻率(高品質影像)調變至低位元頻率(低品質影像)與低位元頻率(低品質影像)調變至高位元頻率(高品質影像)這兩種情況，而初步的實驗結果如上圖六所示。其中的*F1*和*Similarity*表示偵測精確度的評估方法(即數值越大代表精確度越高)。我們能藉由圖六來做主觀影像分析與客觀數據分析兩種方式來進行比較，從我們初步實驗結果能證明我們所提出的CPNMD演算法比上述著名的演算法更能精確偵測到移動物體，初步研究成果已經發表在2012年ACM Multimedia的短篇論文(四頁，接受率小於30%)[8]，我們計畫將做更深入的探討和實驗，並投到優良國際期刊。目前預計將此監控視訊演算法的研究成果透過智慧型電視(Smart TV)與無線傳輸IP Camera，實踐於智慧型視訊監控系統(Intelligent Video Surveillance APP)上。而為了讓智慧型視訊監控系統(Intelligent Video Surveillance APP)達到最佳的視訊監控穩定性，我們將會實測不同的無線網路浮動頻寬情況，同時記錄相關測試結果。

第二部分：智慧型視訊監控系統(Intelligent Video Surveillance APP)針對感興趣的移動物體，設計人臉和物體辨識技術

在擷取感興趣的移動物體之後，我們將對該感興趣移動物體進行辨識的動作。而如果是辨識到人

臉資訊的情況時，同時考量家庭內成員的身分辨識，以及對家庭成員進行有效的身分標記，並可透過身分資訊以提供直覺、友善、迅速與便利等特性，以達到個人的隱私防護性與迅速便利性為目的。因此，我們將使用Support Vector Machine(SVM)分類演算法與Multi-Kernel Learning演算法的技術，來設計具有高精確度的人臉與物體辨識技術。其中，Support Vector Machine(SVM)是一種基於統計學習理論的一種預測分類演算法，其主要精神是將資料投射在高維度空間或特徵空間上，並對兩類不同的訓練資料做最大距離(Max margin)的線性分割，以及分隔區域最大化的特性，進而達到分類的目的。而Multi-kernel學習演算法能夠藉由多種不同kernel functions組合出適應當前情況，並動態地決定一組最佳Kernel的權重值的特性，因此對於人臉和物體的辨識結果是非常有效的。其主要研究方法步驟分別說明如下：

步驟一) Offline Learning Procedure

1) Initialization for weights-of-basis kernels

我們採用Support Vector Machine(SVM)分類器與Multi-Kernel Learning 演算法來提出一個高可靠的MKL分類器。在學習程序中，我們將初始權重向量設為$d^{t=0}=\{{d_m}^{t=0}\}_{m=1}^{M}$，其中每一個basis kernel的初始權重值都是一致的，並且定義如下：

$${d_m}^{t=0}=\frac{1}{M}\ \forall m \tag{44}$$

而為了實現一個更有效能的分類器，通常會使用兩種常見的kernel function：Gaussian RBF kernel function與Polynomial kernel function，來精確反映任兩個樣本x,x'的相似度，如下所示：

- Gaussian RBF kernel function with parameter σ：

$$ker(x,x')=\exp\left(-\frac{\|x-x'\|^2}{\sigma}\right)\text{where }\sigma\in\mathbb{R} \tag{45}$$

- Polynomial kernel function with parameter s：

$$ker(x,x')=(x^Tx'+\sigma)^s\text{ where }\sigma\text{ is a constant},s\in\mathbb{N} \tag{46}$$

其中一組的多個basis kernels $\{K_m\}_{m=1}^{M}$ 能被表示成如下：

$$K_m=[k_m(i,j)]_{N_s\times N_s}\forall\text{m} \tag{47}$$

其中公式(47)中的 $k_m(i,j)=ker(vx_i,vx_j)$則套用公式(45)和公式(46)。

2) Multi kernel calculation

接下來，multi kernel K能被定義為多個 basis kernels $\{K_m\}_{m=1}^{M}$的線性組合，並表示如下：

$$K=\sum_{m=1}^{M}{d_m}^tK_m\ \ with\ {d_m}^t\geq 0,\sum_{m=1}^{M}{d_m}^t=1 \tag{48}$$

K_m表示第m個 basis kernel，${d_m}^t$表示第m個 basis kernel 之權重值且數值大於等於零，並將同一時間點t所有 basis kernel 之權重值${d_m}^t$加總的結果限制為 1。在此學習程序中，當第m個 basis kernel K_m達到較佳的分類效能時，則對應的 basis kernel 權重值${d_m}^t$將會分配至更高。

3) Learning SVM-based classifier using multi kernel

不同於標準的 SVM 的多核學習分類器。首先，它會反覆學習出適合當前使用者的最佳 basis kernel 線性組合。接下來，採用 multi kernel K來最佳化 SVM-based 分類器的精確度。更具體的說，基於

SVM-based classifier 的 MKL classifier unit，可以在較高維度的特徵空間中找到最佳分割超平面，並達到分隔區域最大化的特性。藉由最佳化不同的 multi kernel K，而將分類錯誤降到最低，並產生具有較高分類效能的 MKL 決策函數。MKL 主要問題可以透過下面的最佳化公式解決：

$$\max_{\alpha}\left\{\sum_{i=1}^{N_s}\alpha_i-\frac{1}{2}\sum_{i,j}^{N_s}\alpha_i\alpha_j y_i y_j\sum_{m=1}^{M}{d_m}^t K_m\right\}\text{ subjected to }0\le\alpha_i\le C,\sum_i\alpha_i y_i=0,\forall i$$

$$\sum_{m=1}^{M}{d_m}^t=1,{d_m}^t\ge 0,\forall m \tag{49}$$

而${d_m}^t$是一個由第m個 basis kernel 對應權重值之參數向量；最佳的參數向量$\boldsymbol{\alpha}=[\alpha_1\ \ \cdots\ \ \alpha_{N_S}]$可以透過任何 SVM 分類器學習到；$y_i$表示是對應的身份標籤；最後，SVM-based classifier 可以藉由超參數C控制，其中超參數C為固定常數。因為一個極大的超參數C將導致較高的誤判錯誤的損失，所以超參數C是取決於目前使用者的資料庫。

如同公式(49)的 MKL 主要問題等同於下列的限制性的目標最佳化問題：

$$\min_{d^t}J(d^t)\ such\ that\ \sum_{m=1}^{M}{d_m}^t=1,{d_m}^t\ge 0 \tag{50}$$

$$J(d^t)=\sum_{i=1}^{N_s}\alpha_i-\frac{1}{2}\sum_{i,j}^{N_s}\alpha_i\alpha_j y_i y_j\sum_{m=1}^{M}{d_m}^t K_m \tag{51}$$

其中當公式(49)的$\boldsymbol{\alpha}=\{\alpha_i\}_{i=1}^{N_S}$為最佳解時，則對於時間點$t$上的目標值$J(d^t)$可以獲得。公式(50)中的最佳化問題可以藉由梯度下降演算法解決，且達到一個更適合且最佳化的 multi kernel K。因此，衍生出在高維度的特徵空間中的 MKL 決策函數df_{MKL}，如下所示：

$$df_{MKL}(vx)=\sum_{i=1}^{N_S}\alpha_i\sum_{m=1}^{M}d_m K_m(vx,vx_i)+b \tag{52}$$

4) Objective function optimization for personalized MKL classifier

為了讓 MKL Classifier Unit 適用於感興趣物體，必須在反覆地學習程序中，藉由將目標函數值$J(d^t)$縮減到最低，而決定一組最佳化的權重$\{d_m\}_{m=1}^{M}$。而為了處理公式(50)中提到的問題，基於 Simple MKL 方法的 MKL 分類器單元特別使用梯度下降演算法試圖使目標值的下降率達到最大值。

起初，計算出與$\{{d_m}^t\}_{m=1}^{M}$相關的梯度$[\nabla J]_m$、化簡梯度$[\nabla J_{red}]_m$，如下所示：

$$[\nabla J]_m=\frac{\partial J}{\partial\ {d_m}^t}=-\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j K_m(i,j)\ \forall m \tag{53}$$

$$[\nabla J_{red}]_m=\begin{cases}\frac{\partial J}{\partial\ {d_m}^t}-\frac{\partial J}{\partial\ {d_\mu}^t} & \forall m\ne\mu\\ \sum_{m\ne\mu}-[\nabla J_{red}]_m=\sum_{m\ne\mu}\frac{\partial J}{\partial\ {d_\mu}^t}-\frac{\partial J}{\partial\ {d_m}^t} & for\ m=\mu\end{cases} \tag{54}$$

其中$\mu=\arg\max_m {d_m}^t$，目標函數最佳化程序會重複學習下列步驟直到目標值無法繼續收斂為止，即：目標值無法再減少。接下來，下降方向$D^t=\{{D_m}^t\}_{m=1}^{M}$的計算方式如下：

$${D_m}^t=\begin{cases}-[\nabla J_{red}]_m & \text{if}\quad m\ne\mu\text{ and }{d_m}^t>0\\ 0 & \text{if}\quad {d_m}^t=0\text{ and }[\nabla J_{red}]_m>0\\ \sum_{m\ne\mu}-{D_m}^t & \text{for}\quad m=\mu\end{cases} \tag{55}$$

其中的$-[\nabla J_{red}]_m$ 表示為下降方向(the descent direction)，試圖使目標函數$J(d^t)$達到最小化。然而，為了滿足公式(50)的第二個限制條件成立，當特定 basis kernel 的權重值 ${d_m}^t$等於零且對應 basis kernel

的化簡梯度 $[\nabla J_{red}]_m > 0$時，則對應 basis kernel 的下降梯度會設為零。同樣地，為了確保公式(50)的第一個限制條件成立，當所有 basis kernel 中有特定的權重值是最大時，則對應 basis kernel 的下降梯度設為$\sum_{m\neq\mu} - {D_m}^t$。在最後一個步驟中，需要透過下降方向 $D^t = \{{D_m}^t\}_{m=1}^M$找到一個最佳的$\gamma^t$，並充分的將目標函數縮到最小，如下所示：

$$\gamma^t = \min -\frac{{d_m}^t}{{D_m}^t} \ \forall {D_m}^t < 0 \tag{56}$$

為了將 basis kernels 的權重值更新，透過下降方向${D_m}^t$小於零的條件找到適當的γ^t，來將具有負下降方向的最大化簡梯度所對應的權重更新為 0。

5) Update weights of basis kernels

接著，權重值的更新方式如下所示：

$${d_m}^{t+1} = {d_m}^t + \gamma^t {D_m}^t \ \forall m \tag{57}$$

其中D^t由公式(55)產生、最佳步階的計算如公式(56)所示。將更新後的所有權重集合送入 Multi kernel calculation，其能夠將 basis kernel 與更新後的權重進行組合。

6) Termination for multiple-kernel learning

在學習程序中，我們採用適合的終止學習法則來判定是否停止學習。也就是說，為了確保全域收斂，當滿足終止學習法則時，學習程序就會停止學習。同樣地，如果還沒滿足終止學習法則，還沒符合全域收斂時，分別透過公式(55)跟公式(56)再次計算梯度和化簡梯度。在結束學習程序後，透過學習了一組最佳的 $\{\alpha_i\}_{i=1}^{N_S}$ 與 $\{{d_m}^t\}_{m=1}^M$完成了MKL強分類器單元。

步驟二) Online Recognition Procedure

1) Basis kernels calculation

我們分別使用查詢影像集Q和比對特定影像集R'計算出一組 basis kernel，表示如下：

$$K_m = [k_m(i,j)]_{N_Q \times N_{R'}} \forall m \tag{58}$$

其中第 m 個 basis kernel 中的每一個成分 $k_m(i,j) = ker(vq_i, vx_j)$ 表示 $k_m(i,j)$能夠藉由查詢影像集Q中的查詢特徵向量$vq_i \in V_Q$和比對特定影像集R'的比對特徵向量 $vx_j \in V_R$之間的內積得到。特別注意到，為了達到有效率的辨識，所採用的比對特定影像集R'是比對影像集R中的支持向量集合，其中支持向量所對應的參數 $\alpha_j > 0$。

2) Multi kernel calculation

在 Multi kernel calculation 步驟中，將多個 basis kernels $\{\mathrm{K_m}\}_{\mathrm{m=1}}^{\mathrm{M}}$與對應的最佳權重值$\{\mathrm{d_m}^{\mathrm{t}}\}_{\mathrm{m=1}}^{\mathrm{M}}$進行線性組合，來獲得最佳相似度，其中 multi kernel K的組合如同公式(48)所示。

3) SVM-based face identification using multi kernel

在最後的辨識階段中，使用目前的 MKL 分類器 $\mathcal{H}_{ROI}$，在特定角度p下的查詢子集合$Q' \subseteq Q$的身分是從比對影像集R中的身分所定義。因此，目前 MKL 分類器辨識出的類別向量： $class_{ROI}{}^{*} = \left[class_{q,ROI}{}^{*}\right]^{T}_{1\times N_{Q'}}$ 是由 $N_{Q'}$ 個類別 $class_{q,ROI}{}^{*}$ 所組成，且對應的分數向量 $score_{ROI} = \left[score_{q,ROI}\right]^{T}_{1\times N_{Q'}}$ 是由$N_{Q'}$個$score_{q,ROI}$ 所組成，其分別表示如下：

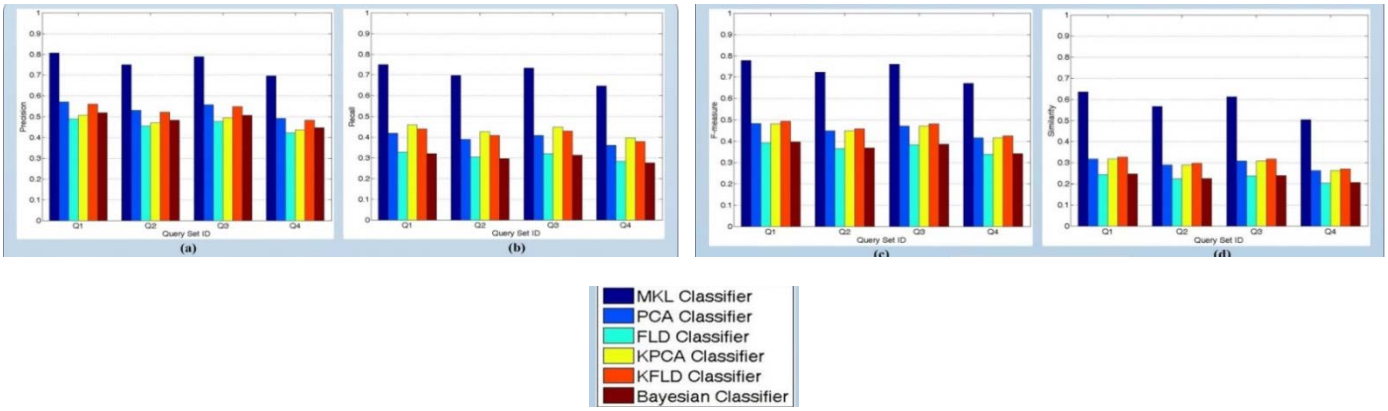
$$class^{*} = \mathcal{H}_{ROI}(Q', R) = \underset{class\in\Omega_{l-layer}}{\arg\max}\ df_{class}(Q', R) \tag{59}$$

$$score_{ROI} = \underset{class\in\Omega_{l-layer}}{\max}\ df_{class}(Q', R) \tag{60}$$

其中，df_{class}表示 對應到的特定類別的 MKL 決策函數。最後，傳回查詢子集合Q'對應的身份向量 $identity_{ROI} = \left[identity_{q,ROI}\right]^{T}_{1\times N_{Q'}}$，可以被表示成如下：

$$identity_{ROI} = \mathrm{I}\left(\mathcal{H}_{ROI}(Q', R)\right) = \mathrm{I}(class_{ROI}{}^{*}) \tag{61}$$

其中，I(·)表示身分函數，即：當得到查詢子集合Q'所對應的辨識類別時，系統會回傳一組相應的身分標籤，而T表示向量的轉置。



圖七、人臉和物體辨識的初步實驗結果圖

而上圖七為我們的人臉與物體辨識演算法 Multi-Kernel Learning 與 Principal Component Analysis(PCA)[9]、Fisher linear discriminant analysis(FLD)[10]、kernel principal component analysis(KPCA)[11]、kernel Fisher linear discriminant analysis (KFLD)[12]以及 Bayesian analysis [13]的初步比較實驗結果。其中 Precision、Recall、F-measure 與 Similarity 表示為評估辨識率程度的方法。而我們使用四組的 test bench 並分別為 Q1、Q2、Q3 與 Q4。從圖七初步實驗結果的辨識率來比較分析，可以證明我們所提出的 Multi-Kernel Learning 演算法的辨識率皆高於上述著名的演算法。

第三部分：智慧型視訊監控系統(Intelligent Video Surveillance APP)基於雲端伺服器之分散式資料庫，對人臉和物體的資料做有效快速的管理

我們考量到智慧型視訊監控系統(Intelligent Video Surveillance APP)上，將有越來越多人臉與物體影像資料上傳至雲端伺服器中建立與儲存。因此，我們提出基於雲端伺服器之分散式資料庫，並使用階層式金字塔單元(Pyramid Database Unit)架構方式，對人臉和物體的資料達到快速且有效的資料管理方式。其主要研究步驟，將分別說明如下：

1) Pyramid Database Unit

為了建立個人化金字塔資料庫，我們提出兩個步驟，其分別為評估正規化連結分數($\mathrm{CS_n}$)與選擇適合的身分(相關類別)。接下來，我們建立所有相關的身分標籤子集合 $\boldsymbol{\Omega}_{bottom} = \{\ell_{COI}, \ell_1, \dots, \ell_{\boldsymbol{L}}\} = \{\boldsymbol{\Omega}_g\}_{g=1}^{\boldsymbol{G}}$，而$\boldsymbol{\Omega}_{bottom}$可以依據使用者定義分組為$\boldsymbol{G}$個群組，來支援後續其它階層的資料庫結構的建立。然而，使用所有相關的身分標籤子集合 $\boldsymbol{\Omega}_{bottom}$來辨識人臉和物體的身分是非常消耗時間，且為了增強人臉識別的效能，我們定義所有相關的身分標籤子集合$\boldsymbol{\Omega}_{bottom}$在整個 Pyramid Database Unit(金字塔資料庫單元)中擁有最低優先的存取權。接著，資料庫單元會建立基於群組資訊(group context)的特定群組相關類別的身分標籤子集合$\boldsymbol{\Omega}_{group} = \{\boldsymbol{\Omega}_g\}_{g=g1}^{gk}$，其相應於所有相關的身分標籤子集合 $\boldsymbol{\Omega}_{bottom}$中的特定群組，且此群組資訊(group context)資訊是在第一批存取控制程序結束後獲得，且特定群組相關類別的身分標籤子集合$\boldsymbol{\Omega}_{group}$。為了建立身分標籤子集合 $\boldsymbol{\Omega}_{time}$ 和 $\boldsymbol{\Omega}_{time,g}$，選擇適合的身分是重要的關鍵，且此兩個身分標籤子集合 $\boldsymbol{\Omega}_{time}$ 和 $\boldsymbol{\Omega}_{time,g}$ 分別是基於時間資訊(temporal context)和時間-群組資訊(temporal-group context)，此兩種資訊皆來自於目前 ROI 蒐集的所有資料所取得。

第一步驟：我們評估每一個感興趣物體$\{\ell_n\}_{n=1}^{\boldsymbol{L}}$和目前感興趣物體$\ell_{COI}$之間的近期連結關係強度，並將近期連結關係強度正規化連結分數($\mathrm{CS_n}$)為下列：

$$\mathrm{CS}_n(\ell_n, \ell_{COI}|\Delta\mathrm{t}) = \frac{\mathrm{C}_n - \mathrm{C}_{min}}{\mathrm{C}_{max} - \mathrm{C}_{min}} \ \forall \mathrm{n} \tag{62}$$

其中，連結分數$\mathrm{C}_n = \exp\big(\mathrm{E}(\ell_n, \ell_{COI}|\Delta\mathrm{t})\big)\ \forall \mathrm{n}$。而$\mathrm{C}_{max}$和 C_{min}標示連結分數$\mathrm{C}_n(n = 1, \dots, \boldsymbol{L})$的最大和最小值。而上述所提到期間的時間$\Delta\mathrm{t}$下反映連結關係的強度。而評估的表示方式如下：

$$\mathrm{E}(\ell_n, \ell_{COI}|\Delta\mathrm{t}) = \phi_1(\ell_n, \ell_{COI}|\Delta\mathrm{t}) + \phi_2(\ell_n, \ell_{COI}|\Delta\mathrm{t}) \tag{63}$$

為了獲得，我們考量下列的單向連結函數ϕ_1和雙向連結函數ϕ_2來評估$\mathrm{E}(\ell_n, \ell_{COI}|\Delta\mathrm{t})$：

- 單向連結函數ϕ_1評估每一個感興趣物體$\{\ell_n\}_{n=1}^{\boldsymbol{L}}$出現在目前感興趣物體$\ell_{COI}$所蒐集的$\mathrm{N}_{\mathrm{A}_{ROI}|\Delta\mathrm{t}}$張近期照片中的機率分布。
- 雙向連結函數ϕ_2評估每一個感興趣物體$\{\ell_n\}_{n=1}^{\boldsymbol{L}}$與目前感興趣物體$\ell_{COI}$共同出現在所有感興趣物體$\{\ell_{COI}, \ell_1, \dots, \ell_{\boldsymbol{L}}\}$組成的$N_{\mathrm{A}|\Delta\mathrm{t}}$張近期照片中的機率分布。

而單向連結函數ϕ_1和雙向連結函數ϕ_2表示如下：

$$\phi_1(\ell_n, \ell_{COI}|\Delta\mathrm{t}) = \frac{\delta_1}{N_{\mathrm{A}_{COI}|\Delta\mathrm{t}}}\Big(\sum_{photo \in (\mathrm{A}_{COI}|\Delta\mathrm{t})} IND_1(\ell_n, photo)\Big)\ \forall n \tag{64}$$

$$\phi_2(\ell_n, \ell_{COI}|\Delta\mathrm{t}) = \frac{1-\delta_1}{N_{\mathrm{A}|\Delta\mathrm{t}}}\Big(\sum_{photo \in (\mathrm{A}|\Delta\mathrm{t})} IND_2(\ell_n, \ell_{COI}, photo)\Big)\ \forall n \tag{65}$$

其中：單指標函數IND_1表示，當第n個感興趣物體的身分ℓ_n是標記在目前感興趣物體ℓ_{COI}所蒐集的近期照片$photo \in (\mathrm{A}_{COI}|\Delta\mathrm{t})$ 時，該函數則回傳值 1；反之，該函數則回傳值 0。雙指標函數IND_2表示，當第n個感興趣物體的身分ℓ_n 與目前感興趣物體ℓ_{COI}兩者都標記在所有感興趣物體$\{\ell_{COI}, \ell_1, \dots, \ell_{\boldsymbol{L}}\}$組成的近期照片$photo \in (\mathrm{A}|\Delta\mathrm{t})$中時，該函數則回傳值 1；反之，該函數則回傳值 0。$\Delta\mathrm{t}$表示資料庫蒐集近期照片的時間。

第二步驟：當較大容量的特定資訊相關類別的身分標籤子集合，依據相應的正規化連結分數($\mathrm{CS_n}$)進行排序，然後我們必須決定從較大容量的身分標籤子集合(如：$\boldsymbol{\Omega}_{bottom}$ 和 $\boldsymbol{\Omega}_{group}$)中所選擇適當身分的數量，進而產生較小容量的近期相關的身分標籤子集合，如：$\boldsymbol{\Omega}_{time}$ 和 $\boldsymbol{\Omega}_{time,g}$。尤其是，選擇適當身分數量的方式是基於我們提出的最佳門檻值方法，其利用不同種類的相依性資訊，如：時間資訊和時間-群組資訊等。

我們提出的短期間最佳門檻值方法是基於兩個概念，即：資料庫中的歷史統計資訊和目前狀態資訊。因此，基於此兩個概念的最佳門檻值Th_{CS}，整合了先前的最佳連結分數 $\mathrm{CS}_{pred}{}^{\Delta \mathrm{t}-1}$和目前平均連結分數$\mathrm{CS}_{mean}{}^{\Delta \mathrm{t}}$，如下所示：

$$Th_{\mathrm{CS}} = \delta_2 \cdot \mathrm{CS}_{pred}{}^{\Delta \mathrm{t}-1} + (1-\delta_2) \cdot \mathrm{CS}_{mean}{}^{\Delta \mathrm{t}} \tag{66}$$

其中：參數δ_2 $(0 \le \delta_2 \le 1)$表示對先前的最佳連結分數 $\mathrm{CS}_{pred}{}^{\Delta \mathrm{t}-1}$的重大影響；最佳連結分數 $\mathrm{CS}_{pred}{}^{\Delta \mathrm{t}-1}$是依據先前時間$\Delta \mathrm{t}-1$更新的統計資料來預測得到，如下所示：

$$\mathrm{CS}_{pred}{}^{\Delta \mathrm{t}-1} = \mathrm{CS}_{opt}{}^{\Delta \mathrm{t}-1}(1+\beta^{\Delta \mathrm{t}}) \tag{67}$$

其中：變化率$\beta^{\Delta \mathrm{t}} = \frac{\mathrm{CS}_{median}{}^{\Delta \mathrm{t}} - \mathrm{CS}_{median}{}^{\Delta \mathrm{t}-1}}{\mathrm{CS}_{median}{}^{\Delta \mathrm{t}-1}}$定義為先前時間$\Delta \mathrm{t}-1$和當前時間$\Delta \mathrm{t}$之間，分別對所有正規化連結分數($\mathrm{CS_n}$)取中位數的變化率；$\mathrm{CS}_{mean}{}^{\Delta \mathrm{t}}$反映出，在較大容量的身分標籤子集合中，整體正規化連結分數($\mathrm{CS_n}$)的當前趨勢。

2) First Batch of Access Control (FBAC) Procedure

當輸入第一批查詢影像集時，此程序能夠存取特定階層的金字塔資料庫單元送入目前感興趣物體ℓ_{COI}的專屬 MKL 分類器進行人臉和物體辨識，而金字塔資料庫單元的存取特定階層的優先順序：$DB_{time}$and DB_{bottom}。具體來說，首先存取時間相關的身分標籤子集合$\boldsymbol{\Omega_{time}}$，並送入目前感興趣物體$\ell_{COI}$的專屬 MKL 分類器進行辨識。當傳回辨識結果的分數向量$\boldsymbol{score_{COI}}$後，如果此分數向量$\boldsymbol{score_{COI}}$中對應的成分超過已辨識特定類別的門檻值Th_{class^*}時，則此存取控制$\mathbf{AC}_{\mathrm{first}}$標示為 1，表示對應的查詢影像為正確辨識結果，因此該查詢影像需進行標記。相反地，存取控制$\mathbf{AC}_{\mathrm{first}}$標示為 0，表示 FBAC 程序需要存取下一層的身分標籤子集合，進一步提供給目前感興趣物體ℓ_{COI}的專屬 MKL 分類器再次進行辨識，即：$\boldsymbol{\Omega}_{bottom}$。因此，第一批存取控制$\mathbf{AC}_{\mathrm{first}}$表示如下：

$$\mathbf{AC}_{\mathrm{first}} = \begin{cases} 1, \text{if } \boldsymbol{score_{COI}} \ge Th_{class^*} \\ 0, \text{otherwise} \end{cases} \tag{68}$$

其中：每一個感興趣物體(類別)的門檻值皆由學習分類器計算得到。然而，如果該程序已經存取所有相關的身分標籤子集合$\boldsymbol{\Omega}_{bottom}$且存取控制$\mathbf{AC}_{\mathrm{first}}$仍標示為 0 時，此查詢影像視為分類錯誤，因此將其定義為無法識別。由於在進行第一批人臉標記後，此 FBAC 程序將獲得第一批查詢人臉所對應的特定群組資訊和重複出現的資訊，因此當 FBAC 程序完成後，能夠建立完整的金字塔資料庫單元。

3) Non-First Batch of Access Control (NFBAC) Procedure

為了使目前感興趣物體ℓ_{COI}的專屬 MKL 分類器能夠有效識別第二批查詢影像集，我們提出 NFBAC 程序著重於存取高可能性將會辨識正確且較少身分數的比對集，並進一步提供給分類器。而整個專屬金字塔資料庫單元的存取特定階層的優先順序為下列：$DB_{recurrence}$、$DB_{time,g}$、DB_{group}和 DB_{bottom}。接著，此 NFBAC 程序的執行方式，相似於 FBAC 程序，即：當傳回辨識類別時，對應的

存取控制$\mathbf{AC}_{\text{non-first}}$若標示為 1，表示獲得的查詢身分為正確結果；相反地，對應的存取控制$\mathbf{AC}_{\text{non-first}}$仍然是標示為 0。

Collaborative Face Annotation Approach With Pyramid Database Unit	Baseline：Without Pyramid Database Unit			
	Query Set ID			
	Q1	Q2	Q3	Q4
MKL	0.1832	0.1162	0.1857	0.1899
BDRF[14]	0.1933	0.1939	0.1956	0.1932
CMVF[14]	0.2133	0.2140	0.2158	0.2131

表八、有無使用階層式金字塔單元架構的初步實驗結果

表八為我們對於有無使用階層式金字塔單元(Pyramid Database Unit)架構的初步實驗結果。我們採用我們所設計的人臉與物體辨識演算法，和 2011 IEEE Transactions Multimedia(TMM)所提出的 Bayesian Decision Rule(BDRF)[14]與 Confidence-based Majority Voting(CMVF)[14]的人臉與物體辨識演算法，並在四個 test bench 下來進行分析。我們將利用這三個人臉與物體辨識演算法來分析在未加入我們所設計的階層式金字塔單元(Pyramid Database Unit)架構以及加入有使用我們所設計的階層式金字塔單元(Pyramid Database Unit)架構的時間比。我們發現這三個人臉與物體辨識演算法在有加入階層式金字塔單元(Pyramid Database Unit)架構後，比未加入階層式金字塔單元(Pyramid Database Unit)架構的時間速度還快約 4.69 到 8.61 倍左右。而從上表八的初步實驗結果，可以證明使用階層式金字塔單元(Pyramid Database Unit)架構可以有效加快辨識速度以及縮短辨識時間的效能[25]，計劃期間我們預計將第二和三部分彙整並撰寫成論文，投稿至知名國際期刊發表。

（四）預期完成之工作項目、成果及績效。

在此部份，我們將本研究計畫預期進度以甘特圖(Gantt Chart)來表示：

月次 / 工作項目	一	二	三	四	五	六	七	八	九	十	十一	十二
第一部分、智慧型視訊監控系統(Intelligent Video Surveillance APP)設計自動化偵測感興趣的移動物體，並可支援擴充 IP Camera 的無線視訊影像傳輸												
於 Smart Phone 設計與開發有效操控 APP 於 Smart TV 上	ˇ											
透過 Smart TV 上開發與設計 Intelligent Video Surveillance APP		ˇ	ˇ									
開發 Intelligent Video Surveillance APP 於不同尺寸的 Smart TV 畫面來調整最佳化介面佈局設計				ˇ								

在不同的無線網路浮動頻寬情況下實測各視訊監控演算法與本研究方法					ˇ	ˇ	ˇ					
透過 Smart TV 的 Intelligent Video Surveillance APP，以及考量無線視訊傳輸環境下，實測與驗證本計畫之視訊監控演算法								ˇ	ˇ	ˇ		
第二部分、智慧型視訊監控系統(Intelligent Video Surveillance APP)針對感興趣的移動物體，設計人臉和物體辨識技術												
相關人臉與物體辨識文獻研讀探討	ˇ	ˇ										
模擬與驗證本研究之人臉與物體辨識演算法			ˇ	ˇ	ˇ							
比較探討著名的人臉與物體辨識演算法						ˇ	ˇ					
第三部分、智慧型視訊監控系統(Intelligent Video Surveillance APP)基於雲端伺服器之分散式資料庫，對人臉和物體的資料做有效快速的管理												
相關分散式資料庫管理文獻研讀以及探討	ˇ	ˇ										
研究與設計分散式資料庫管理演算法			ˇ	ˇ	ˇ							
驗證與模擬分散式資料庫管理演算法						ˇ	ˇ					
研究成果進行整合								ˇ	ˇ	ˇ		
撰寫研究計畫報告											ˇ	ˇ
預計累計研究進度百分比(%)	5	10	15	25	30	40	50	60	70	80	90	100

本研究計畫主要是開發一套智慧型視訊監控系統(Intelligent Video Surveillance APP)，透過 Android 環境平台實踐於智慧型電視(Smart TV)上，且能在低位元頻率或是高位元頻率的無線視訊串流情況下達到精確的視訊監控。並結合人臉和物體辨識技術，對於感興趣移動物體達到有效的身分的辨識與標記，以及具有分散式資料庫之特性，並且對人臉和物體的資料，提供有效且快速的管理模式。預期完成工作之項目如下：

1. 基於智慧型電視(Smart TV)上對智慧型手機(Smart Phone)設計有效操控 APP。
2. 開發與實踐 Intelligent Video Surveillance APP 能根據任何畫面尺寸的智慧型電視(Smart TV)來自

動地達到最佳化的 UI 介面。

3. 研究與實踐一套高精確性智慧型即時視訊監控系統。
4. 建立與實現準確地身分標記之人臉與物體辨識系統。
5. 建立與實現有效且快速之分散式資料庫管理系統。
6. 討論對於無線視訊傳輸下高精確的即時視訊監控技術之多元化應用，以及開拓新的學術議題。

由於本人過去在嵌入式軟體演算法與系統設計和視訊多媒體系統設計擁有許多豐富的經驗，並且擁有多項美國與台灣專利，並將關鍵技術與相關專利成功的技術移轉給台灣的軟體和晶片系統設計公司，本人也是其中的重要技術顧問，此對未來此領域的發展趨勢及走向亦能完全掌握。本計劃的重點，除了全方位的考量之外，更重要的是把學識、經驗與技術傳承下去，進而培養更多此一領域的人才，進而提升國家在學術領域之國際競爭力。

參考文獻

[1] A. Manzanera, J. C. Richefeu, “A Robust and Computationally Efficient Motion Detection Algorithm Based on Σ-Δ Background Estimation,” In Proc. ICVGIP’04, pp. 46-51, 2004.

[2] A. Manzanera and J. C. Richefeu, “A New Motion Detection Algorithm Based on Σ–Δ Background Estimation,” Pattern Recognit. Lett., pp. 320-328, 2007.

[3] C. Stauffer and W. E. L. Grimson, “Learning patterns of activity using real-time tracking,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 747-757, Aug. 2000.

[4] M. Oral and U. Deniz, “Centre of mass model - a novel approach to background modelling for segmentation of moving objects,” Image Vis. Comput., vol. 25, no. 8, pp. 1365–1376, Aug. 2007.

[5] J. E. Ha and W. H. Lee, “Foreground objects detection using multiple difference images,” Optical Engineering, vol. 49, no. 4, pp. 047201 Apr. 2010.

[6] L. Maddalena and A. Petrosino, “A self-organizing approach to background subtraction for visual surveillance applications,” IEEE Trans. Image Process., vol. 17, no. 7, pp. 1168–1177, Jul. 2008.

[7] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, “Background modeling and subtraction by codebook construction,” in Proc. ICIP, vol. 5. 2004, pp. 3061–3064.

[8] J. Y. Yen, B. H. Chen, and S. C. Huang "Enhanced Extraction of Moving Objects in Variable Bit-Rate Video Streams," ACM Multimedia, pp. 717 - 720, Nara, Japan, Oct. 29 - Nov. 2, 2012.

[9] M. Turk and A. Pentland, “Eigenfaces for recognition,” Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[10] P.N. Belhumeur , J.P. Hespanha and D.J. Kriengman, “Eigenfaces vs. Fisherfaces： Recognition using class specific linear projection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, July. 1997.

[11] J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang and Z. Jin, “KPCA plus LDA： A complete kernel fisher discriminant framework for feature extraction and recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 2, Feb. 2005.

[12] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Muller, “Fisher discriminant analysis with

kernels,” Proc. IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing IX, pp. 41-48, Aug. 1999.

[13] B. Moghaddam, T. Jebara and A. Pentland, “Bayesian face recognition,” Pattern Recognition, vol. 33, no. 11, pp. 1771-1782, Nov. 2000.

[14] Y. Choi, W. D. Neve, K. N. Plataniotis, and Y. M. Ro, “Collaborative face recognition for improved face annotation in personal photo collections shared on online social networks,” IEEE Transactions on Multimedia, vol. 13, no. 1, pp. 14-28, Feb. 2011.

[15] M. Saini, X. Wang, P.K. Atrey, M. Kankanhalli, “Adaptive Workload Equalization in Multi-Camera Surveillance Systems,”IEEE Trans. Multimedia,, vol. 14, no. 3, pp. 555-562, Jun. 2012.

[16] M. Oral and U. Deniz, “Centre of mass model-A novel approach to background modelling for segmentation of moving objects,” Image Vis. Comput., vol. 25, pp. 1365-1376, Aug. 2007.

[17] N. Li, J.J. Jain, and C. Busso, “Modeling of Driver Behavior in Real World Scenarios Using Multiple Noninvasive Sensors,” IEEE Trans. Multimedia,, vol. 15, no. 5, pp. 1213-1225, Aug. 2013.

[18] M. Turk and A. Pentland, “Eigenfaces for recognition,” Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

[19] P.N. Belhumeur , J.P. Hespanha and D.J. Kriengman, “Eigenfaces vs. Fisherfaces： Recognition using class specific linear projection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, July. 1997.

[20] J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang and Z. Jin, “KPCA plus LDA： A complete kernel fisher discriminant framework for feature extraction and recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 2, Feb. 2005.

[21] Y. Choi, W. D. Neve, K. N. Plataniotis, and Y. M. Ro, “Collaborative face recognition for improved face annotation in personal photo collections shared on online social networks,” IEEE Transactions on Multimedia, vol. 13, no. 1, pp. 14-28, Feb. 2011.

[22] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, “Simple MKL,” Journal of Machine Learning Research, vol.9, pp. 2491–2521, 2008.

[23] F.E.H. Tay, L.J. Cao, “Improved financial time series forecasting by combining support vector machines with self-organizing feature map,” Intelligent Data Analysis, vol. 5, no. 4, pp.339 -354, 2001.

[24] G. Camps-Valls and L. Bruzzone, “Kernel-based methods for hyperspectral image classification,” IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 6, pp. 1351-1362, 2005.

[25] Y. H. Jian, M. K. Jiau and S. C. Huang, "Automatic Face Annotation System Used Pyramid Database Architecture for Online Social Networks," IEEE International Conference on Ubiquitous Intelligence and Computing, Vietri sul Mare, Italy, Dec. 18-20, 2013.