智慧型視訊分析於行動網路視覺監控之應用

王元凱　胡永祥　張惟清
輔仁大學電子工程學系
ykwang@mails.fju.edu.tw, ackbarci@yahoo.com.tw, nelsonchang66@gmail.com

摘要

本論文提出一整合式智慧型監控系統，使用高斯模型、型態濾波和連通元件標記法去偵測移動物件。在即時關鍵圖框選擇出最佳的移動物件影像後，系統將自動發送多媒體訊息(MMS)的警告和上網連結的簡訊通知(WAP push)給使用者的行動電話裝置，並將含有移動物件的靜態圖框影像轉換成視訊串流媒體檔案加以儲存。如此，使用者可即時經由 3G 行動通訊網路透過視訊串流來監看已錄製完成的移動物件監控視訊。由實驗結果顯示，系統可正確偵測出移動物件和採用物件式監控比全程式監控更能有效節省儲存容量。

關鍵詞：智慧型監控、視訊串流、3G 行動通訊網路。

Abstract

This paper proposes an integrated intelligent surveillance system that uses Gaussian modeling, Morphology filters and Connected Component Labeling to detect moving objects. The system will automatically deliver Multimedia Messaging Service (MMS) alarms and WAP push notifications to user's mobile phone, and then records and transcodes those frames of moving objects into video streaming file format after the real time key frame selection decided the most clearly frame of moving object. Users can real time monitor the recorded surveillance video streaming through 3G mobile communication networks. Experimental results show that the system detects moving objects successfully, and the storage of object-based recording is more efficiently than full-time recording.

Keywords: intelligent surveillance、video streaming、3G mobile communication networks.

1. 前言

現今監視系統的普及率已大幅增加，隨地皆可發現監視器的安裝，其監視系統使用功能也因安裝的地方而有所不同，如一般保全、居家看護、交通監控等。但一般的監控系統皆將監視器拍攝到的視訊傳送至監控主機後，就將監控視訊資料全程予以紀錄與儲存。但如遇到異常事件發生就只能從資料庫中去尋找已錄製完成的監控視訊資料，如此大大降低了對異常事件監控的即時性。

因此近年來有許多在智慧型遠端監控的學術研究，其遠端監控方式大至可區分為兩種：以有線或無線網際網路的架構進行遠端監控，在此稱為"網際網路式"遠端監控；另一種方式為結合行動通訊網路的架構以進行遠端監控，在此稱為"行動網路式"遠端監控。

在"網際網路式"遠端監控方面，如 Brown et al.[7] 研發出一套具有開放式架構與可擴充監控功能的系統，但此監控系統只提供有線網際網路架構的遠端監控方式，降低了使用者進行遠端監控的便利性。Stringa and Regazzoni[3]以有線區域網路架構的遠端監控為主，研究在室內環境中以視訊監控自動偵測出遺留的物件，並使用視訊索引建立懷疑物件和遺留物件之間在邏輯上關聯性，讓使用者能在龐大的視訊資料庫中容易取得需要的影像或視訊。Lee et al.[9]研究經由有效率的資料庫管理對網路攝影機進行控制和紀錄視訊影像，是一套透過有線網際網路架構進行遠端監看的監控系統。Tsutsumi[4] 提出即時儲存和擷取的工具，此系統可透過無線網際網路架構進行遠端監看。

上述這些"網際網路式"遠端監控技術雖然不盡相同，但其共通架構如圖 1 所示。監視器將監視視訊傳送至監控主機後，進行視訊分析與將監控時間內的視訊或影像全部予以紀錄和加上索引之後，儲存於資料庫內以方便監控資料的檢索。因只提供網際網路架構進行遠端監控，所以必須在具有有線區域網路或無線區域網路環境中才可使用。如欲透過行動通訊網路進行遠端監控，因礙於行動通訊網路所能提供的頻寬有限，因此仍需再將監控的視訊進行即時的轉碼處理，以期能符合行動通訊網路在進行資料傳輸時所採用的動態頻寬調整特性，方能將監控視訊成功傳輸至遠端監控裝置上進行監控。

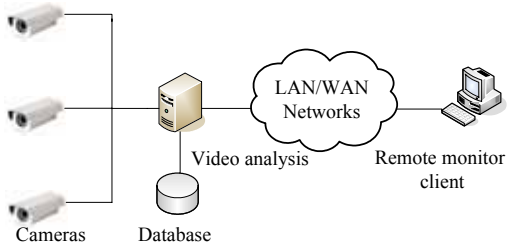


圖 1 "網際網路式"遠端監控技術

在“行動網路式”遠端監控方面，如 Steiger et al.[10]將監控視訊以壓縮的媒體檔案格式(如 MPEG-1)依據行動通訊網路的頻寬，使用物件式動態調整位元傳輸率的方式傳送至行動電話上，但卻無提供物件偵測功能只是單純將監視器的視訊透過行動通訊網路傳送至行動電話裝置。而目前應用在“行動網路式”遠端監控方面的產品，如 3G 行動通訊業者威寶電信的 ZTE MF58[16] 3G 監控設備，為利用影像電話的技術將攝影機結合行動電話用戶識別模組(USIM)卡片，讓使用者以撥打影像電話的方式透過 3G 行動通訊網路之電路交換網路(circuit-switched)撥打到遠端的監控攝影機進行觀看攝影機的監控視訊影像，缺點為無異常物件偵測亦無主動通知使用者功能。另如歐特斯科技的產品 AES-M800[11] 數位式網路攝影機，透過行動通訊網路之封包交換網路(packet-switched)進行觀看攝影機的監控視訊影像，唯此種方式亦無主動異常物件偵測。

上述“行動網路式”遠端監控技術，其共同架構如圖 2 所示。攝影機的監控視訊透過行動通訊網路之電路交換網路架構或透過即時視訊串流的方式經由行動通訊網路之封包交換網路架構，傳送至使用者行動電話裝置上以進行遠端監控。這些技術皆是直接將監視器的監控視訊傳送到遠端監控端的行動電話裝置上，皆未考慮到智慧型物件偵測技術與監控視訊的儲存。以一般家居環境監控應用而言，約只有 0.3%的監控視訊認為是有異常物件出現的視訊。若如這些“行動網路式”遠端監控系統，只能執行即時監控而無法透過智慧型視訊分析技術僅將有意義的視訊傳送至行動終端裝置上，如此在行動終端裝置上將有 99.7%的機率會監控到無意義的遠端監控視訊。

因此本論文所提出的智慧型監控系統是以改進“網際網路式”遠端監控與“行動網路式”遠端監控之不足並整合 3G 行動通訊網路應用在視訊串流之上，系統架構如圖 3 所示。不同於一般監控系統的全程式監控，本論文提出之系統以物件式監控為主，監控系統只將含有偵測出有移動物件之視訊予以紀錄與儲存，如此讓使用者可只監看移動物件之異常事件視訊。不需如全程式監控，需要花時間在全部監控時間內已錄製完成的監控視訊中，去搜尋已發生的異常事件視訊。這讓使用者可更有效率去監看整段異常事件的視訊，也節省監控系統儲存監控視訊所需的儲存容量。

本論文之結構，第二段說明本系統架構與系統整合所使用的相關技術，第三段討論採用物件式監控的儲存效率實驗結果，最後第四段為結論和未來研究方向。

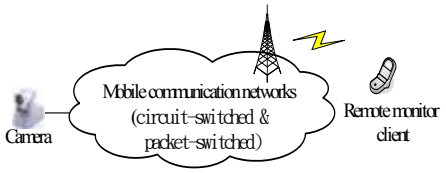


圖 2 “行動網路式”遠端監控技術

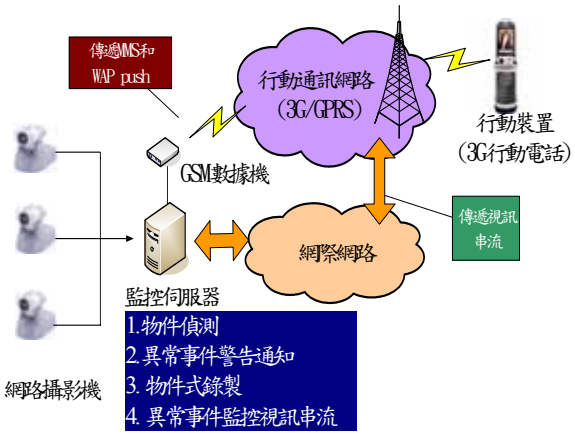


圖 3 本系統之架構圖

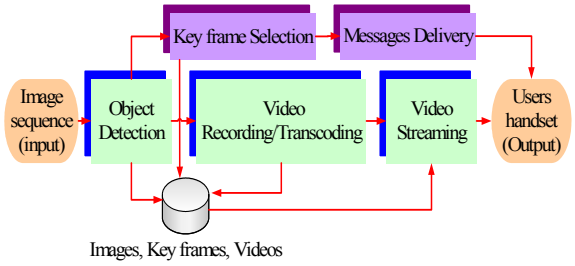


圖 4 監控伺服器內之系統模組與處理流程

2. 智慧型監控系統

本系統提出以智慧型視訊分析為主的監控系統。本系統之監控伺服器內包括物件偵測模組、關鍵圖框選擇模組、視訊錄製/轉碼模組、視訊串流模組和訊息傳送模組。物件偵測模組為在連續靜態影像中偵測物件，關鍵圖框選擇模組會在移動物件影像中決定一張關鍵圖框影像，並經由訊息傳送模組將此關鍵圖框影像傳送至行動裝置端。含有移動物件的連續圖框影像會予以紀錄和轉碼成視訊資料，並經由動態編碼方式傳送至行動網路上。圖 4 為監控伺服器內之系統模組與處理流程。

2.1 移動物件偵測

移動物件偵測可歸類為邊緣偵測和影像相減兩種方法。邊緣偵測為處理物件外形向量資料，經由找出影像中強度劇烈變化的地方，以找出移動物件的形狀邊緣。影像相減則為將兩張影像的像素進行相減，找出兩張影像內像素的不同處，以得到移動物件。影像相減比邊緣偵測有效率，且更適合應用於即時監控上。

影像相減有兩種方法，時間差異法(Temporal differencing)如 Hu et al. [5]為將連續兩張影像進行相減，所得到移動物件呈現破碎不完整的外型，會增加移動物件偵測的複雜度。背景相減法(Background subtraction)如 Tiehan et al. [8]為先建立

背景影像後，再將此影像與即時影像進行相減以得到移動物件。由此所得到的移動物件外型可較完整，並減少移動物件偵測的複雜度。

本系統採用的移動物件偵測，使用遞迴方式建立背景影像：

$$B_c^{k+1}(x,y)=\frac{T}{T+1}B_c^k(x,y)+\frac{1}{T+1}P_c^{k+1}(x,y) \quad (1)$$

其中 $B_c^{k+1}(x,y)$ 為新建立的第 k+1 張色彩模型 c 之背景影像在位置 (x,y) 的像素，$B_c^k(x,y)$ 為已建立的第 k 張色彩模型 c 之背景影像在位置 (x,y) 的像素，$P_c^{k+1}(x,y)$ 為第 k+1 張色彩模型 c 之監視影像在位置 (x,y) 的像素。其中 T 為一段取樣時間區間內的影像總張數，c 為色彩模型 R，G，B。使用式(1)做為建立即時更新的背景影像以便進行影像相減，如此可降低計算時的複雜度與有效使用系統記憶體空間。

為找出二值化臨界點值(Threshold)，使用高斯模型法(Gaussian modeling)求得高斯分佈之標準差 σ：

$$\sigma_c^k(x,y)=\left[\frac{1}{T}\sum_{i=1}^{T}\left(P_{i,c}^k(x,y)-B_c^k(x,y)\right)^2\right]^{1/2} \quad (2)$$

$$\sigma_c^k(x,y)=\left[\frac{T-1}{T}\left(\sigma_c^{k-1}(x,y)\right)^2+\frac{1}{T}\left(P_c^k(x,y)-B_c^k(x,y)\right)^2\right]^{1/2} \quad (3)$$

其中 $\sigma_c^k(x,y)$ 為第 k 張色彩模型 c 之標準差影像在位置 (x,y) 的像素。採用式(2)和式(3)做為適應性二值化臨界點值(Adaptive threshold)以判斷出前景影像。

建立即時更新的背景影像後，使用此背景影像 $B_c^k(x,y)$ 與即時取得的監視影像 $P_c^{k+1}(x,y)$ 相減，並使用適應性二值化臨界點判斷方式以得到前景的灰階影像：

$$D_c^{k+1}(x,y)=\left|P_c^{k+1}(x,y)-B_c^k(x,y)\right| \quad (4)$$

$$\begin{cases} H^{k+1}(x,y)=255, & D_c^{k+1}(x,y)>2.5\sigma_c^k(x,y) \\ H^{k+1}(x,y)=0, & \text{else} \end{cases} \quad (5)$$

其中 $D_c^{k+1}(x,y)$ 為相減後第k+1 張色彩模型c之前景影像在位置 (x,y) 的像素，$H^{k+1}(x,y)$ 為取適應性二值化臨界點判斷後第k+1 張之灰階前景影像在位置 (x,y) 的像素。

得到前景灰階影像後使用型態濾波(Morphology)去除影像雜訊和連通元件標記法(Connected Component Labeling)將前景影像內之物件有連接的像素定成群組和定義標籤，並使用最

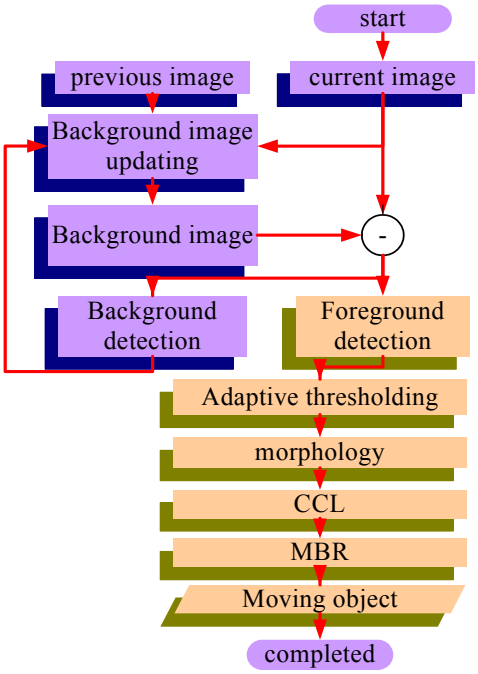


圖 5 本系統之移動物件偵測影像處理流程圖

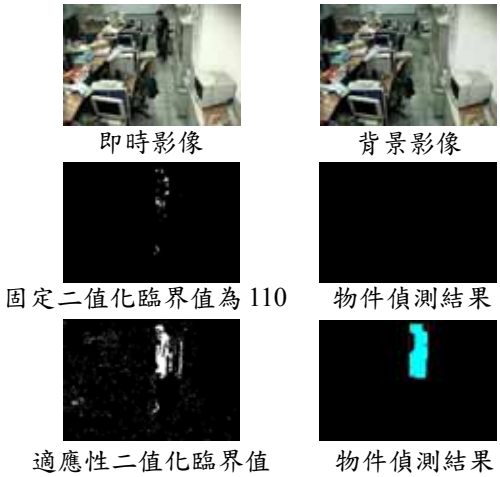


圖 6 固定與適應性二值化移動物件偵測結果比較圖

小矩形框(Minimum Bounding Rectangle)去計算前景影像內之物件，作為判斷移動物件和選擇關鍵圖框影像(Key frame)之用，圖 5 為系統所使用的移動物件偵測影像處理流程圖。圖 6 為使用固定二值化臨界點與適應性二值化臨界點進行移動物件偵測的結果，其中使用固定二值化時結果偵測不出移動物件，而使用適應性二值化則可正確偵測出移動物件。

2.2 關鍵圖框影像選擇

在移動物件偵測後，需選擇適當的圖框影像作為傳送 MMS 告警之用。一般關鍵圖框影像選擇的方法為非即時性且需要從視訊影像中去分析各個

圖框影像(frame)來找出關鍵圖框，如 Durucan et al.[2]和 Kim and Hwang[1]。

本系統所提出的方法是以最小矩形(MBR)面積和圖框影像(frame)面積去求出面積比率值，用來判斷出關鍵圖框影像(Key frame)：

$$\begin{cases} DF^k = 1, & \max\{M_{1,\ldots,} M_L\} / F^k > Threshold \\ DF^k = 0, & \text{else} \end{cases} \quad (6)$$

其中圖框影像內最小矩形(MBR)的面積為 $M_i, i \in \{1,\ldots,L\}$ L 為使用連通元件標記法所標定出的數目，F^k 為第 k 張圖框影像的面積。DF^k 如為 1，則 F^k 被選擇為關鍵圖框。

2.3 視訊轉碼與串流

通常要在行動電話上觀看監控視訊，需先透過網路伺服器讓行動電話裝置下載後，再從行動電話裝置上播放。往往由於行動電話記憶體容量的限制，如果監控的視訊檔案過大時會造成無法全部下載。

本系統是將監控視訊經由視訊串流伺服器使用串流方式傳送，所以就不會有行動電話記憶體容量的限制和等待視訊檔案下載的時間。為求達到讓行動電話裝置能播放的視訊媒體格式[6][17]，必須先進行視訊串流轉碼。將監控的視訊檔案媒體格式轉換成 3gp 後，再使用視訊串流伺服器將監控視訊傳送至行動電話裝置端。進行視訊串流轉碼主要的參數有：編碼格式參數、圖框影像大小參數(frame size)、影像傳輸率參數(frame rate)和位元傳輸率參數(bit rate)。

編碼格式參數為編碼使用的視訊壓縮標準如 MPEG-2、MPEG-4...等，本系統是採用 MPEG-4 Simple Profile 的視訊壓縮標準。圖框影像大小參數為顯示視訊時的大小，如 CIF(352×288 像素)、QCIF(176×144 像素)...等。本系統為使行動電話裝置可以正常播放，需轉換圖框影像大小為 QCIF。

影像傳輸率參數為每秒播放的靜態影像，採用較低的影像傳輸率在播放視訊時可降低使用頻寬和減少中央處理器處理程式的需求；採用較高的影像傳輸率則會有較好的播放效果，一般行動電話裝置允許的影像傳輸率為 10 到 15 fps。因此需將監控視訊的影像傳輸率，轉換成行動電話裝置允許的影像傳輸率。位元傳輸率參數為每秒傳送的位元數，通常較高的位元傳輸率會有較好的播放品質。一般行動電話裝置允許的位元傳輸率為 32 到 64 Kbps，所以進行視訊串流轉碼時，需將位元傳輸率轉換為符合行動電話裝置的位元傳輸率。

圖 7 行動電話接收移動物件的 MMS 告警

圖 8 3G 行動電話觀看移動物件的視訊串流

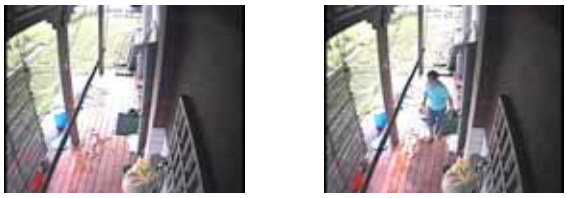
2.4 系統整合

本系統透過網路攝影機即時擷取監控的靜態影像並傳送到監控伺服器內進行影像分析。當偵測到移動物件後，安裝在訊息傳送模組的 NOW SMS/MMS Gateway[15]將透過本系統之 GSM 數據機發送 MMS 告警和 WAP push 通知給行動電話端，圖 7 為行動電話裝置接收到多媒體訊息(MMS)的告警。

為使用行動電話裝置透過視訊串流觀看監控的視訊，視訊錄製/轉碼模組使用 JMF(Java Media Framework)將含有移動物件的連續靜態影像轉換成 AVI 的視訊媒體格式，再使用 FFMPEG[13]轉碼為 3gp。因安裝在視訊串流伺服器模組內的 Darwin Streaming Server(DSS)[12]需要提示(hinting)作為標示串流媒體的串流記號，所以還需在欲進行串流的視訊媒體資料中使用 MPEG4IP [14]加入提示軌(hint track)以供進行串流之用。圖 8 為使用者透過 3G 行動電話使用 RTSP 監看移動物件的視訊串流。

3. 實驗

本系統的監控伺服器實驗平台，使用中央處理器 Pentium 4 2.8GHZ、記憶體容量 1GB 的桌上型電腦。應用程式為使用 j2sdk1.4.2 程式語言所設計與開發本論文所提出的監控系統程式。實驗來源的輸入靜態影像為固定於民宅屋簷的 CCTV 監視器以每秒 30 張影像所拍攝的實際居家生活，每張彩色色階影像解析度為 360×240 像素，監控的時間為一天早上六時至晚上七時，總共為 1404000 張監控影像用以驗證本系統的物件式監控。在其監控的內容中，將發生移動物件、留置物件與取出物件的情形作為異常事件處理，實驗來源的輸入靜態影像如圖 9 所示。



正常情形之監控影像　　異常事件之監控影像

圖 9 實驗輸入靜態影像範例

本實驗中若系統連續偵測出異常事件時，就將此異常事件連續出現之第三張靜態影像作為轉換該段異常事件監控視訊時所需的第一張異常事件靜態影像。若系統連續無偵測到異常事件時，就將此靜態影像作為轉換該段異常事件監控視訊時所需的最後一張異常事件靜態影像。

系統會將第一張到最後一張的異常事件靜態影像轉換成傳輸率（frame rate）為 10fps，位元傳輸率(bit rate)為 64Kbps 和圖框影像大小(frame size)為 QCIF(176×144 像素)的 3gp 視訊媒體檔案壓縮格式，提供給視訊串流之用。

本測試資料總共 1404000 張影像中發生了 33 次的異常事件，在其真實發生的異常事件中所包含的影像張數最多為 421 張，最少為 12 張，平均為 134 張，詳細的資料如圖 10 中的虛線所示。而系統偵測出有發生異常事件中所包含的影像張數最多為 374 張，最少為 9 張，平均為 117 張，詳細的資料如圖 10 中的實線所示。

實驗結果中顯示本論文提出之物件偵測方法，所偵測出發生異常事件的次數與實際狀態下發生異常事件的次數相同；但每次偵測出有異常事件的影像張數與真實發生有異常事件的影像張數會有所差距。是因為本系統的物件偵測方法是依據一個高斯模型進行異常事件偵測，若異常事件之實際物件移動過慢或在原地停留過久時，其實際物件便會隨著背景影像的更新而漸漸成為背景影像的一部份，導致影響到系統異常事件偵測的準確性，但此並不會對本系統的物件式監控之整體監控效能造成影響。以單一件異常事件為例，其方式為找出開始發生此異常事件到此異常事件結束的影像，再將此異常事件從開始到結束的所有監控影像重新轉碼成為此異常事件發生經過的監控視訊，目的為讓使用者能直接觀看到此異常事件從發生到結束的經過。

採用全程式監控與物件式監控所需儲存的靜態影像張數比較，如圖 11 所示。其中全程式監控所需儲存的靜態影像張數，為監控時間內所有靜態影像張數的總和。物件式監控所需儲存的靜態影像張數，則為監控時間內系統偵測出有發生異常事件時的靜態影像張數總和。從實驗結果得知，採用全程式監控需儲存 1404000 張監控影像；但採用物件式監控在不影響整體監控異常事件下，在相同監控時間內只需儲存含有異常事件的 3760 張監控影像即可。

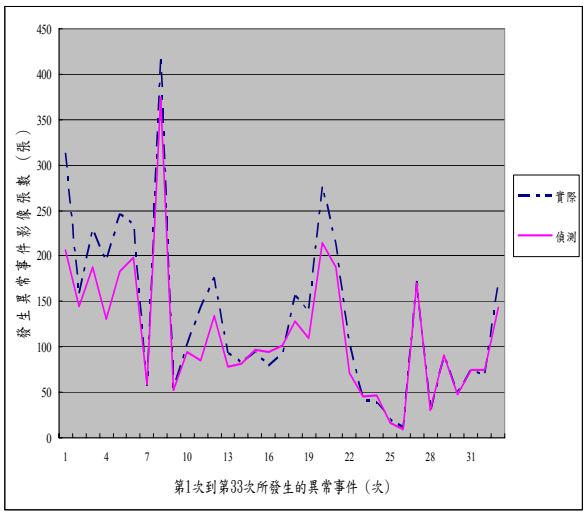


圖 10 實際與偵測之發生異常事件出現張數比較圖

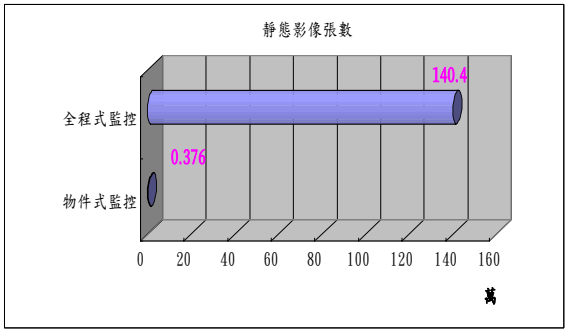


圖 11 全程式監控與物件式監控實驗結果比較圖

在單一異常事件下本系統進行監控時各模組所需的平均時間，如表 1 所示。其中單一異常事件之物件偵測平均時間 269.7 秒，為單張監控影像偵測平均時間 2.31 秒/張乘以單一異常事件之影像張數 117 張，所得到的單一異常事件進行物件偵測需要的時間。單一異常事件之關鍵圖框影像選擇平均時間為關鍵圖框影像選擇模組平均於單一異常事件下，決定出異常事件靜態監控影像所需要的時間。單一異常事件之視訊錄製與轉碼平均時間為視訊錄製/轉碼模組平均於單一異常事件下，將含有移動物件之連續靜態監控影像進行 3gp 視訊媒體格式轉碼所需要的時間。單一異常事件之多媒體訊息傳送平均時間為訊息傳送模組平均於單一異常事件下，經由 3G 行動通訊網路發送多媒體訊息(MMS)警告到使用者行動電話裝置成功接收此告警所需要的時間。

從實驗結果可得知，採用物件式監控在不影響整體異常事件偵測準確率下，比全程式監控可節省大量的監控資料儲存容量。由於本系統只儲存發生異常事件時間內的監控視訊資料，所以使用者不需再由資料庫中去尋找已錄製的監控資料，可直接選擇要監看已錄製的異常事件監控視訊並經由 3G 行

動通訊網路的視訊串流直接進行監看，進而達到即時監控的目。

表1 本系統各模組所需的時間

Modules / Average Time (Sec)	Object detection	Key frame selection	Video recording/ transcoding	Message delivery
An Event	269.7	0.03	3.54	26.5
A Frame	2.31	0.26×10^{-3}	0.30×10^{-1}	

4. 結論

本研究提出一套結合 3G 行動網路的智慧型監控系統。本系統經由網路攝影機擷取監控影像，一旦偵測出移動物件後會發送含有移動物件影像的多媒體訊息(MMS)告警與發送上網連結的簡訊(WAP push)通知給使用者，並將含有移動物件時間區段內的影像，轉換為視訊串流媒體格式檔案予以儲存。使用者可使用 3G 行動電話裝置經由視訊串流監看已錄製的移動物件視訊。藉由本系統所提出的物件式監控和結合 3G 行動網路之應用，可節省監控視訊檔案儲存的空間，並讓使用者可利用 3G 行動電話裝置觀看即時發生移動物件的監控視訊。未來繼續研究方向為使用偵測率與辨識率更高的影像處理技術，如高斯混合模型、機率辨識方法進行提升系統準確率。

參考文獻

[1] C. Kim and J. N. Hwang, “Object-based video abstraction for video surveillance systems,” *IEEE Trans. On Circuits and Systems for Video Technology,* Vol. 12, pp. 1128-1138, Dec. 2002.

[2] E. Durucan, F. Ziliani, and O. N. Gerek, “Change detection with automatic reference frame update and key frame detector,” *Proc. of IEEE-EURASIP. Workshop Nonlinear Signal and Image Processing (NSIP),* pp. 57-60, 1999.

[3] E. Stringa, and C. S. Regazzoni, “Real time video-shot detection for scene surveillance applications,” *IEEE Trans. on Image Processing,* Vol. 9, pp. 69-79, Jan. 2000.

[4] F. Tsutsumi, “Real time storage and simultaneous retrieval for surveillance and patrol video,” *IEEE Int. Conf. Proc. of Multimedia and Expo,* Vol.2, pp. 1175-1178, 2000.

[5] F. Y. Hu, Y. N. Zhang, and L. Yao, “An effective detection algorithm for moving object with complex background,” *Proc. of Machine Learning and Cybernetics,* Vol.8, pp. 5011-5015, Aug. 2005.

[6] I. Elsen, F. Hartung, U. Horn, M. Kampmann, and L. Peters, “ Streaming technology in 3G mobile communication systems,” *IEEE Computer,* Vol.34, pp. 46-52, Sep. 2001.

[7] L. Brown, A. Hampapur, J. Connell, M. Lu, A. Senior, C. F. Shu, and Y. Tian, “IBM Smart surveillance system(S3):an open and extensible architecture for smart video surveillance,” *IEEE Conf. Proc. of Advanced Video and Signal Based Surveillance,* pp. 318-323, Sep. 2005.

[8] L. Tiehan, B. Ozer, and W. Wolf,“A real-time background subtraction method with camera motion compensation,” *IEEE Int. Conf. Proc. of Multimedia and Expo,* Vol.1, pp. 331-334, Jun. 2004.

[9] M. O. Lee, C. W. Lee, and Y. C. Kim, “Network accessable security web-based cameras server system using index search algorithm,” *Proc. of Info-tech and Info-net,* Vol.5, pp. 60-65, 2001.

[10] O. Steiger, T. Ebrahimi, and A. Cavallaro, “Surveillance video for mobile devices,” *IEEE Conf. Proc. of Advanced Video and Signal Based Surveillance,* pp. 620-625, Sep. 2005.

[11] Autotools homepage. [Online].Available: http://www.autotools.net/portal/media-type/html/language/en/user/anon/page/default.psml/js_pane/Homepage, 2007.

[12] Darwin Streaming Server homepage. [Online]. Available: http://developer.apple.com/opensource/server/streaming/index.html, 2007.

[13] FFMPEG homepage. [Online]. Available: http://ffmpeg.mplayerhq.hu/, 2007.

[14] MPEG4IP homepage. [Online].Available: http://mpeg4ip.sourceforge.net/, 2007.

[15] Now SMS/MMS Gateway homepage. [Online]. Available: http://www.nowsms.com/, 2007.

[16] Vibo homepage. [Online].Available: http://www.vibo.com.tw/CWS/index.html, 2007.

[17] 3GPP – TSG - SA4 PSM SWG internal working draft,“Transparent end to end packet switched streaming service (PSS); protocols and codecs,” Release 6, 2004.