

以視訊檢索技術協助年長者找尋遺忘物品

黃金本

銘傳大學 資訊傳播工程學系 教授

hcptw@mail.mcu.edu.tw

彭胤峯、林綉庭、潘澄玉、黃培恩、王致傑

銘傳大學 資訊傳播工程學系 學生

05166146@mail.mcu.edu.tw

摘要

由於年長者迫切的需求，本文發展一套適合年長者使用的視訊技術。透過硬體裝置拍攝年長者日常生活的視訊並存檔，再經由發展視訊檢索技術，協助年長者回憶起遺忘物品之擺放位置或遺忘物品之情景，讓年長者快速找到遺忘的物品，如此可節省許多時間，進而造福廣大的銀髮族群。為有效率的觀看視訊，本文提出的有效視訊關鍵幅(Key-frames)提取方法，即熵值(Entropy)演算法。經由計算出每一像幅之熵值，及設定適當的門檻值，若該幅的熵值超過設定之門檻值，則該幅為關鍵幅，予以保留；否則將該幅丟棄。並建立適合年長者的圖形化使用者互動介面(Graphical User Interface, GUI)及結合索引(Index)的技術，使年長者可節省瀏覽的時間及快速的找到遺忘的物品。

關鍵詞：圖形化使用者互動介面；關鍵幅；視訊索引

Abstract

Motivated by the urgent needs of senior citizens, we develop a set of video technology suitable for use by the elderly. First, through a video hardware device, the daily lives of seniors are recorded and archived. Then a video retrieval scheme is implemented to assist the elderly to recall forgotten items placed or forgotten objects scenario. As a result seniors quickly find their forgotten items and thus gain a benefit of saving a lot of time. In order to efficiently watch the video, we propose an effective key-frame extraction method, namely entropy algorithm, to summarize the video. We calculate the entropy of every frame and set an appropriate threshold value for the video. Then we compare the calculated entropy with the specified threshold value. If the entropy of a frame exceeds the threshold value, that frame is assigned as a key-frame and will be retained; otherwise, that frame is discarded. Furthermore, we employ graphical user interface (GUI) and index techniques, to improve using environments,

save browsing time, and find the forgotten items quickly for the elderly.

Keywords: Graphical User Interface, key-frames, video indexing

1. 前言

目前台灣醫療水準的提昇，年長者平均壽命逐年攀升，為了因應高齡化社會到來，已有許多科技產品的研發，針對年長者的需求設計與生活相關的輔助工具，然而較少對年長者日常生活所需的方便性進行探討。因此，我們以年長者居家場景為基礎，擺放日常生活物品（目前考慮手機、眼鏡、書本、手錶與藥罐）模擬居家生活習慣，並錄製模擬視訊，透過熵值演算法，找出其中的關鍵幅，因為找出關鍵幅，能讓使用者不必花冗長的時間去瀏覽完整視訊，只需透過關鍵幅，就能快速得知視訊重要的內容，其優點是省時，可以快速找出所需的視訊，再利用索引技術於必要可以對應到原始視訊，觀看完整視訊片段。以這種方式，可以協助年長者快速回憶所遺忘物品的情景，找回該遺忘之物品。

2. 文獻探討

本節為相關文獻的探討，分成視訊濃縮技術與關鍵幅提取技術兩部分說明。

2.1 視訊濃縮技術

視訊濃縮[10] (Video Synopsis) 簡單來說就是擷取視訊的精華版，讓整個視訊能有效率的被瀏覽，也就是將不同時間所發生的事件，濃縮成一個視訊的概念，而它主要改變了時間而非空間，而擷取後的視訊也必定比原始視訊還短，取出的片段也是連續的並非斷斷續續。在視訊濃縮部分，首先必須知道視訊中的最重要部分，來進行擷取，進而用很短的時間看完整段冗長的視訊，讓同一場景能在同一時間同時呈現（顯示）。而視訊濃縮有一個尚待改進的缺點，就是碰撞的產生，將影響視訊事件發生的前後順序，而導致前後錯置的結果，這也是

必須突破的一環。

2.2 關鍵幅的提取方法

近年來已有許多關鍵幅的提取方法被提出，有透過物件比對方法、相鄰幅的特徵差異（如顏色、紋理、線、邊與角等）、場景偵測與鏡頭偵測的方法提取等。

2.2.1 物件比對提取方法

文獻[9]提出將年長者與日常生活中取放物品影像予以錄製，放入資料庫，再進入資料庫利用物件特徵點的比對方式與資料庫中預先建立好的物件圖片做比對，當比對符合即可搜尋到此物件及其所在的場景，但是物件比對方式容易受物品擺放角度的影響，若擺放角度不同相對結果也可能不同，因此導致比對的準確度降低。

物件比對方式[7][2]，主要可以分為兩類，第一類為點的比對方法，第二類為特徵的比對方法。點的比對方法是直接採用影像像素的灰階值來計算同一物件在兩張影像中的像素點匹配關係，其優點在於可以獲得高密度的視差影像，但是這類方法依賴影像的灰階值，當影像的差異太大時，比對結果可能不夠準確。而特徵的比對方法，通常是先求取影像中的物件特徵，如線段、輪廓、邊緣、角點...等，再利用形狀、灰階度等限制條件進行特徵量的比對。此類方法得到的視差影像密度較低，必須透過內插手法改善，但是會因此而提高運算的複雜度。

2.2.2 相鄰幅提取方法

以相鄰幅間的相似度[4]為特徵為評估的標準，當相似度低於設定的門檻值（如 50%），此像幅即為關鍵幅，予以保留，否則該幅予以丟棄，此種方式找出的關鍵幅準確性較低，假使在兩個相似場景中只是物品擺放位置不同，會認定為關鍵幅，此方法的門檻值的設定亦會影響關鍵幅提取的幅數與準確度。

2.2.3 鏡頭邊界偵測方法

利用鏡頭邊界（Shot Boundary）[11]的方式來提取關鍵幅，此方法的優點是簡單快速；而缺點是取出的關鍵幅可能不是最能代表視訊的內容，因為其方法是取第一幅與最後一幅當作關鍵幅，這樣的方式，容易忽略掉視訊中間更有代表性的像幅。

2.2.4 以顏色為基礎提取

色彩直方圖對視訊進行分割找出關鍵幅，視訊分割最簡單方法就是將視訊中相鄰幅之RGB灰階度值相減，差值越小代表這兩個幅屬於同一個場景的機率越高，但視訊動作（可能來自於物體移動、人物或是鏡頭本身動作等等）越激烈，RGB灰階度差值也會跟著越大，導致本來應在同一個場景畫面被誤認為是不同場景，但在緩慢場景變化時，雖然比在同一個場景中相鄰幅之色彩直方圖差值大，但還是不足以大到被辨識出場景改變，所以用兩個門檻值（Threshold），高門檻值用來偵測劇烈場景變化；低門檻值用來偵測緩慢場景變化。此方法的缺點演算過程較複雜且須花較多的時間，找出的關鍵幅，也可能因為相鄰幅顏色相似，而誤判成同一場景。

2.2.5 場景偵測

在取關鍵幅中，場景的變化[6][5]扮演著重要的角色，許多相關的應用也廣泛的被運用在視頻索引與視訊追蹤上。由此可見，場景的變化檢測在視訊提取的範疇下是不容忽視的一環。

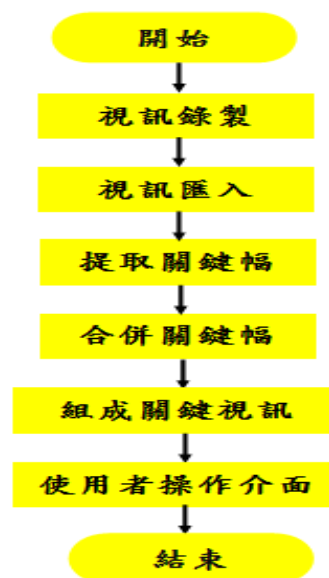


圖 1.系統流程圖

3. 研究方法

首先提取關鍵幅與合併關鍵幅，緊接著再利用視訊索引的概念，在圖形化使用者互動介面呈現整體系統。整體系統流程，如圖 1.所示。

3.1 熵值演算法

熵值愈大表示幅中的平均消息量愈多[7]，以下為我們提出方法的說明。先計算每幅幅中各個灰階度（0~255）出現的機率值 $p_f(k)$ ，如公式（1）所示

$$p_f(k) = h_f(k) / (M \cdot N) \quad (1)$$

$h_f(k)$ 代表累計每幅的灰階度（ $k=0\sim 255$ ）的直方圖， M, N 為幅的大小，將公式（1）的機率值倒數後，再以2為基底取 $\log$ 值，得到消息量，以 $Q_f(k)$ 表示，如公式（2）

$$Q_f(k) = \log_2\left(\frac{1}{p_f(k)}\right) = -\log_2(p_f(k)) \quad (2)$$

透過產生的消息量乘以機率值得到熵值，如公式（3）所示

$$e_f(k) = p_f(k) \cdot Q_f(k) \quad (3)$$

$$E_{total} = \sum_{k=0}^{k_{max}} e_f(k) \quad (4)$$

公式（4）是每一幅所有灰階度（0~255）的熵值加總，其中 k 表示灰階度， k_{max} 為最大灰階度。

$$E_{Th} = \sum_{k=z}^n e_m(k) \geq 0.05 \cdot E_{total} \quad (5)$$

公式（5）中將 e_f 做降冪排列後以 e_m 表示，降冪排列是為了計算熵值時能有序的從大至小分佈，而不是參差的排列； $k=z$ 時代表 $e_m(z)$ 具有最大的熵值，當較大熵值 k 從 z 至 n 時的總和大於等於全部熵值之5%，此時的總和數值設定為門檻值，此值為經實驗所得之經驗值。

$$Diff(f, f+1) = \left| \sum_{k=z}^n e_{m,f}(k) - \sum_{k=z}^n e_{m,f+1}(k) \right| \quad (6)$$

$$Diff(f, f+1) \geq 0.1 \quad (7)$$

公式（6）計算相鄰兩幅經降冪排列後 $e_{m,f}$ 的 k 從 z 至 n 時的總和與 $e_{m,f+1}$ 的 k 從 z 至 n 時的總和相減取絕對值，得到 $Diff(f, f+1)$ 。公式（7）表示當差值大於等於10%，則認定該幅（ $f+1$ ）

為關鍵幅。

3.2 合併關鍵幅

取出關鍵幅後，為了防止取出的關鍵幅中，有些很明顯就能看出差異；有些關鍵幅卻因光線與物體移動等因素，造成系統取出極為相似的關鍵幅，使老年人更不易找回遺忘物品與相對應之影像實際內容。為避免取出相似的關鍵幅，產生資訊淹沒的情況發生，我們也針對這點研究相關資料與文獻，藉以減少這個缺點，以下為我們提出的方法。

先利用熵值演算法，將原始影片利用實驗所得門檻值，取出影片關鍵幅，此時關鍵幅可能有許多因為光線等因素，造成部份取出的關鍵像幅有相似的情況發生，因此，在熵值演算法運算後，再利用門檻值的增設，將取出的關鍵幅值接近且像幅位置相近的像幅，設定當相似幅低於五幅時，則合併為一關鍵幅影像；我們之所以用以五幅為基準，合併關鍵幅是有依據的。經過不斷的實驗，我們實驗3幅、5幅、8幅以及10幅作為合併關鍵幅的基準，我們發現以3幅為基準時，會造成取出關鍵幅有許多相似之處，已造成資訊冗餘；而以8幅、10幅為合併關鍵幅基準時，也會造成許多真正的關鍵像幅有被淹沒的情況發生，因此經由實驗我們取中間最佳值5幅作為合併關鍵幅的基準。

3.3 視訊索引

Index 即檢索[3]，通常出現在報紙、雜誌、書刊的前面，告訴讀者這份刊物包括哪些內容，什麼東西在第幾頁，而將這個概念應用在這次的研究中，藉由檢索可以快速找到所需的影片的概略位置，找到影片的概略位置後，再仔細尋找準確的位置，而此做法可以減少大量花費在檢視影片的時間，進而達到所要的效果。

3.4 圖形化使用者互動介面

在圖形化使用者介面中，介面上會顯示視窗、圖示以及按鈕等預設圖形，用以表示不同需求的動作，使用者可透過滑鼠事件來對自己所需要的功能，進行選擇畫面個數（二個畫面及四個畫面），不同個數的畫面，大小會有差異（二個的畫面較大；四個的畫面較小），從選擇顯示畫面的數量來看，分為兩方面，一方面是所謂的粗略搜尋法，當選擇多個畫面時，年長者選擇同時觀看四段不同的視訊，如此一來能同時看不同地方或時間所發生的事情，找到遺忘物品時的可能地方，但由於人的眼睛同時

間觀看不同畫面，可能導致使用者無法專心關注某個視訊。另一方面則為精細搜尋法，只顯示一個畫面時，年長者可以非常專注於整個視訊的內容，可以很仔細的搜尋與回想遺忘物品的發生情境與地方。

4. 實驗結果

在實驗結果地方，在第一部分為實驗視訊錄製、第二部分說明門檻值設之研究。在實驗視訊錄製選擇了五樣物品並以最佳的解析度錄製且模擬老人應用系統的實際狀況繪製整體錄製視訊的流程圖，摒除外在因素所影響錄製的結果。在第三部分是視訊關鍵幅提取結果，第四部分則說明圖形化使用者互動介面的操作與呈現的結果。

4.1 實驗視訊錄製

選擇五樣年長者常遺忘的物品來做實驗，如手機、藥罐、書本、手錶、眼鏡，利用研究方法去提取關鍵幅。測試物品如圖 2。



圖 2. 測試物品

錄製視訊流程如圖 3。經由此流程錄製出視訊一至視訊五，作為後續實驗的測試用視訊。為了模擬年長者日常生活的習慣，選擇客廳、餐桌、玄關、浴室、飲水機旁和書房做為場景，分別錄製各場景的物件取放動作後，再把錄製的各場景視訊分別挑選所需片段後，將這些片段組合成一個年長者生活中有可能發生的情況，例如：某年長者三餐飯後必須服藥，早餐過後他拿著藥罐，在飲水機旁裝水及吃藥；午飯後，想到需吃藥時，便去飲水機旁倒水，並拿藥到客廳去吃藥；晚餐過後順利的吃完藥，就去就寢。隔天一早，忘了藥罐擺放在何處，此時便可利用此系統去尋找。



圖 3. 錄製視訊流程圖

另外，錄製視訊時，擺放物件的角度對於演算法的影響也需探討，由於我們利用熵值演算法取視訊關鍵幅，所以物件的擺放角度並無影響，但如果採用物件比對方式取關鍵幅的話，物件擺放的角度就極為重要，因此同一個物件，擺放的角度不同，演算法就可能判讀為另一物件。

4.2 門檻值設定研究

本段針對門檻值如何設定做詳細說明，視訊門檻值的設定影響因素，我們歸納出兩項：第一是拍攝時的環境因素，就是室外的環境與室內的環境會對門檻值有所影響，一般常見的視訊場景大部分以室外為主，例如：棒球場、公園、頂樓等等，但我們的研究是提供年長者居家生活品質，換句話說是設置一個快速回憶系統讓年長者找尋遺忘物品，因此場景基本上是室內環境。第二是所謂視訊的內容，也就是說視訊中動作不停的改變，門檻值需要設的較大，才不會造成大部分影像都成關鍵幅，相反的，如果視訊中的內容是變化比較小的，就如同我們系統所錄製的視訊，使用的對象是年長者，主要是模擬他們生活起居，因此視訊的內容較為穩定，不會不停改變，所以本文結合環境與視訊內容兩項因素，對我們的方法設定適合的門檻值。

4.3 關鍵幅提取結果

使用年長者常用的物件做為實驗測試物品，利用圖 3.錄製所得之測試視訊一至五做為測試視訊。首先，將該視訊分別讀入後，再利

用我們提出的方法，將視訊中的關鍵幅提取出來，結果詳如表 1。表格中的取出關鍵幅即利用演算法的提取結果，真正關鍵幅數則是我們利用肉眼判斷之結果，由此表可以清楚的看出我們的方法可以有效的將關鍵幅取出，該結果與真正關鍵幅幾乎一致；該結果同時顯示我們的方法可以有效的縮短瀏覽所需時間達 100 多倍。為呈現視覺的結果，於圖 4.為手機、圖 5.藥罐、圖 6.書本、圖 7.手錶和圖 8.眼鏡顯示各別視訊片段所取出的關鍵幅影像。

表 1. 視訊幅數與取出關鍵幅資料

視訊名稱	視訊一	視訊二	視訊三	視訊四	視訊五
視訊總幅數	1280	1130	1300	1250	1030
取出關鍵幅數	9	7	5	6	5
真正關鍵幅數	8	7	6	7	5



圖 4. 手機關鍵幅影像



圖 5. 藥罐關鍵幅影像



圖 6. 書本關鍵幅影像



圖 7. 手錶關鍵幅影像



圖 8. 眼鏡關鍵幅影像

4.4 圖形化使用者互動介面顯示

圖形化使用者介面的畫面，主要介面的首頁如圖 9.，畫面右側的選擇畫面觀看個數的區

塊，可以選擇要觀看的視訊個數，分別為兩個
 視訊 **2**，及四個視訊 **4** 的按鈕供選擇。當
 按下要選擇的視訊個數按鈕，隨即跳到所選擇
 個數的頁面；若要離開系統，則可按下方的紅
 色鍵 **離開系統**，以關閉此頁面。



圖 9. 首頁圖

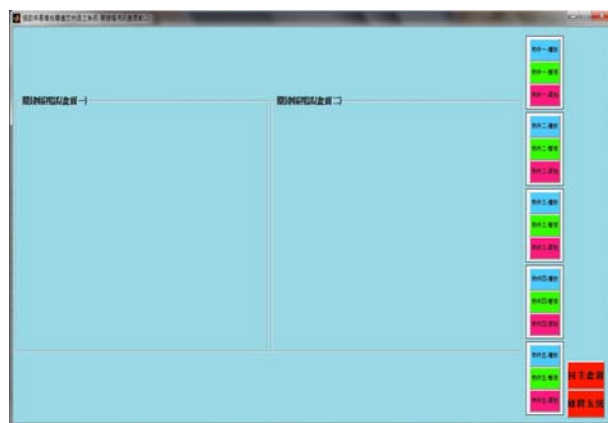


圖 10. 二個畫面顯示視窗

選擇想要的視訊個數，按下對應的按鈕。當按
 下二個畫面鈕，顯示視窗如圖 10。跳出為二
 個畫面的視窗，右側為選擇五種可能物件的按
 鈕，每組物件有三個鈕，分別為藍色的播放
 鈕、綠色的暫停鈕及桃紅色的原始鈕；當按
 下藍色播放鈕時，按鈕會從藍色變為黃色，而關
 鍵幅視訊(畫面一)會顯示隨機擷取某一物件
 的關鍵幅視訊；此時可再選擇另一組按鈕，按
 下播放鍵，此物件會顯示在關鍵幅視訊(畫面二)
 中，二個畫面播放視窗如圖 11.和圖 12.當按
 下綠色暫停鈕時，則播放鈕會由黃色變為原
 始的藍色如圖 12.和圖 13；視窗按鈕顏色隨
 視窗功能改變，可讓使用者更方便及容易的
 使用介面。在圖 12.詳細說明播放視窗按鈕
 之設定，

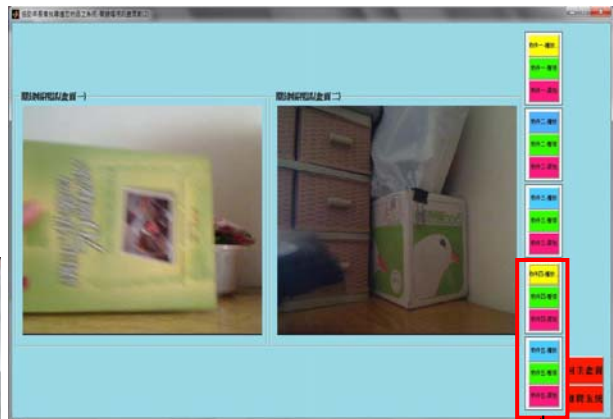


圖 11. 二個畫面播放視窗

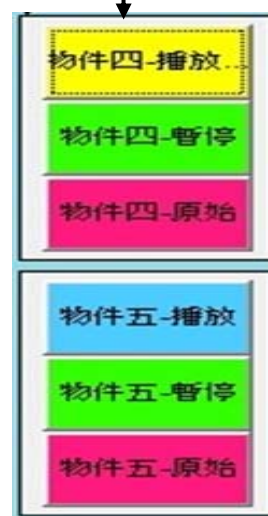


圖 12. 播放視窗按鈕

因為五種測試物件分別產生五段視訊，所以每
 一段視訊的播放都有一組按鈕，其按鈕包含播
 放、暫停和原始，功能分別是播放視訊，暫停
 播放視訊與索引至原始視訊片段。



圖 13. 二個畫面暫停視窗

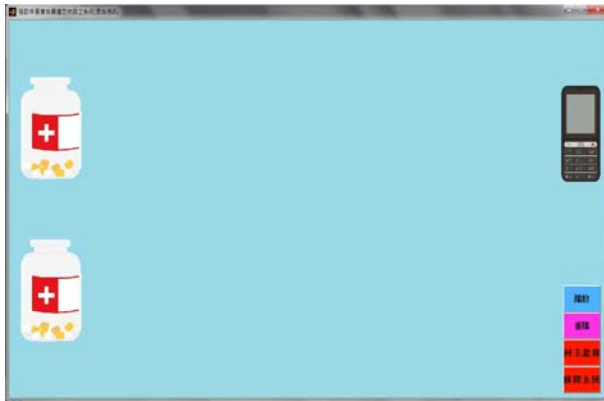


圖 14. 原始畫面顯示視窗

倘若按下桃紅色的原始鈕則顯示撥放原始影片的介面，如圖 14.原始畫面顯示視窗；畫面右下方分別有紫色重播鈕、藍色播放鈕、以及上述的回主畫面鈕和離開系統鈕。當按下藍色的播放鈕，按鈕顏色從藍色變成黃色，並顯示為播放中，而原始畫面框會輸出原始影片，如圖 15.原始畫面播放視窗；當原始影片播放完畢時，播放中的按鈕會變為灰色，並鎖住按鈕，若想再觀看原始畫面，可按下紫色的重播鈕，該按鈕會顯示為重播中，並且轉換為黃色，如圖 16.原始畫面重播視窗，若要關閉此原始畫面視窗，則同樣按下離開系統鈕即可。



圖 15.原始畫面播放視窗



圖 16. 原始畫面重播視窗

假如選擇視訊個數要改變，應先按下回主畫面鈕，跳回主頁面重整訊息，即回到圖 9.首頁圖，再進行選擇。當按下四個畫面鈕，則出現圖 17.四個畫面顯示視窗。同樣的右側會顯示您所選擇的各組物件的播放鈕、暫停鈕以及原始鈕。此時即可隨機選擇四個物件播放鈕，視窗會跳出隨機的物件畫面，可以個別觀看也可多個物品同時觀看，如圖 18.四個畫面同時播放視窗，亦可單一暫停個別視訊或同時暫停視訊，當按下暫停鈕則會如圖 19.四個畫面同時暫停視窗。若要結束此畫面，則同樣按下紅色的離開系統鈕即可，或可按下回主畫面鈕，回到圖 9. 首頁圖。



圖 17. 四個畫面顯示視窗



圖 18. 四個畫面同時播放視窗



圖 19. 四個畫面同時暫停視窗

5. 結論

由於高齡化時代已來臨，高年齡層的生活機能與社會福利已是不容忽視的問題。但現今真正能貼近生活，幫助他們的設計並不多，有鑑於此，本研究發展出一套視訊技術來協助年長者。透過一個友善的互動介面系統，先瀏覽取出的關鍵幅視訊，看是否可以使年長者快速回憶遺忘物品擺放之位置；若不能，再索引至原始視訊片段，觀看其完整的視訊內容，來達成快速提醒年長者遺忘物品的目的。簡言之，本文藉由提取視訊關鍵幅以濃縮視訊的長度，再利用索引的技術連結至原始視訊片段，最後利用友善的使用者互動介面來呈現，以有效提升年長者的生活品質。

參考文獻

- [1] A. Ferman, M. Tekalp, and R. Mehrotra, "Robust color histogram descriptors for video segment retrieval and identification," *IEEE Trans. Image Process.*, vol.11, no. 5, pp.-497-508, May 2002.
- [2] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [3] H. S. Chang, S. Sull, and S.-U. Lee, "Efficient video indexing scheme for content-based retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 8, pp. 1269-1279, Dec. 1999.
- [4] J.-N. Hwang, T.-D. Wu, and C.-W. Lin, "Dynamic frame-skipping in video transcoding," in *Proc. IEEE 2nd Workshop on Multimedia Signal Processing*, pp. 616-621, 1998.
- [5] J. Meng, Y. Juan, and S. F. Chang, "Scene change detection in a MPEG compressed video sequence," *Digital Video Compression: Algorithms and Technol.*, vol. SPIE-2419, pp. 14-25, Feb. 1995.
- [6] J. M. Gauch, S. Gauch, S. Bouix, and X. Zhu, "Real time video scene detection and classification," *Inform. Process. Manage.*, vol. 35, no. 3, pp.381-400, 1999.
- [7] M. A. Pentland, "Probabilistic visual learning for object detection," *Proceedings of the Fifth IEEE Conference on Computer Vision*, 1995, pp. 786-793, 1995.
- [8] M. A. Psarrou, "Key-frame extraction algorithm using entropy difference," *Proceedings of Multimedia Information Retrieval 04*, New York, USA, Oct. 2004.
- [9] P. S. Huang, C. P. Huang, C. H. Hsieh, B. J. Hwang, C. Y. Chiou and K. F. Hsiao, "A vision-based system to help senior citizens for memory recall of object-placing," *Proc. IEEE Conf. Instrumentation, Measurement, Computer, Communication and Control* pp. 357-360, 2011.
- [10] Rav-Acha, Y. Pritch, and S. Peleg, "Making a Long Video Short: Dynamic Video Synopsis," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 435-441, June 2006.
- [11] Y. Zhuang, Y. Rui, T. S. Huang, and S. Mehrotra, "Adaptive key frame extraction using unsupervised clustering," in *Proc. ICIP '98*, vol. I, pp. 866-870, 1998.