# 行政院國家科學委員會專題研究計畫 成果報告

## 智慧型人機互動界面之開發
## 研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 97-2221-E-212-035-
執 行 期 間 ：97年08月01日至98年09月30日
執 行 單 位 ：大葉大學電機工程學系

計畫主持人：黃登淵
共同主持人：陳木松
計畫參與人員：碩士班研究生-兼任助理人員：莊頌祥

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 98年09月22日

# 行政院國家科學委員會補助專題研究計畫 ■ 成 果 報 告 □期中進度報告

## 智慧型人機互動界面之開發

計畫類別：■ 個別型計畫　　□ 整合型計畫
計畫編號：
執行期間： 97 年 08 月 01 日至 98 年 07 月 31 日

計畫主持人：黃登淵
共同主持人：陳木松
計畫參與人員：莊頌祥

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
■出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：大葉大學電機工程學系

中　華　民　國 98　年　10　月　31　日

## 中文摘要

在近幾年當中，基於視覺影像與統計學習方法的人機介面(Human-Computer Interfaces；HCIs )，其相關研究越來越引人注目。由於手勢是一種最直接和最方便的輸入方式，使得它的應用範圍日益廣泛。常見的應用包括有取代傳統的鍵盤和滑鼠輸入、在虛擬實境中作為控制輸入器、或輔助聾啞人士進行交互溝通等。

本研究提出一基於 Gabor Filter + SVM 的方法來進行手勢識別。本研究之目的在於提出(1)適應性膚色模型選擇方法，以解決光線變化的問題；(2)手勢角度估算與校正方法，以解決手勢旋轉的問題；(3)手勢與手臂分割方法，以增加使用者穿著衣物(短袖或長袖)之彈性。透過以上問題的解決，本研究確實可以提昇手勢識別之強健性。

採用 Gabor Feature 做為手勢特徵描述的方法，其手勢識別率遠高於僅採用原始手勢影像之識別率，在所採用的 3 種分類器中，又以支持向量機(SVM)之手勢識別率最高。針對靜態手勢資料庫的識別試驗中，當應用 Gabor Filter 來表示手勢時，歐氏距離的識別率在特徵數量為 35 時可得到 93.0%，餘弦距離的識別率在特徵數量為 113 時可得到 94.0%，但支持向量機的識別率在特徵數量為 103 時卻可得到 96.1%之水準；經比較沒有經過 Gabor Filter 變換之原始影像，歐氏距離的識別率在特徵數量 25 時僅為 80.0%，餘弦距離的識別率在特徵數量 35 時亦為 80.0%，至於 SVM 分類器的識別率在特徵數量 29 時可得到 86.4%；由實驗結果得知採用 Gabor Feature＋SVM 可以獲得最好的手勢識別結果。

在動態手勢偵測方面，本研究所發展的即時手勢偵測與識別系統亦能有效地識別出各種旋轉角度下之正確手勢。本系統每幀影像處理時間約為250ms，大部分時間花在適應性膚色偵測上。另外，透過一系列動態視訊影像之檢測，在動態手勢識別時，分別對穿著長袖衣物和穿著短袖衣物進行實驗，當穿著短袖時，在270幀中成功辨識出252幀，SVM辨識率為93.3%；當穿著長袖時，在300幀中成功辨識出282幀，SVM辨識率為94%。

**關鍵詞**：手勢識別、Gabor Feature、主分量分析法、支持向量機

## 英文摘要

Human-Computer Interfaces (HCIs) based on visual image and statistical learning method have attracted more attention in the last decade. Because of its accessibility and convenience, hand gestures are widely adopted in a large number of HCI's applications. For instance, it can replace conventional keyboard and mouse as an input for computers, or as a controller in virtual reality or as an auxiliary communication way for the deaf-and-dumb community.

The effect of hand-pose orientation on recognition rate is greatly significant. In this study, we proposed a module called "hand-pose angle estimation and correction," which can effectively overcome the difficulties of hand gestures with various angles. In addition, this module can also correct the hand gestures to an acceptable small angle. In this work, we used Gabor feature as a hand gesture descriptor, and then adopted PCA for feature extraction to reduce the dimensionality of Gabor features. Furthermore, three classifiers including Euclidean distance, cosine distance, and SVM are employed for hand-gesture classifications.

Experimental results show that the recognition rates of Gabor-filtered images are higher than those of raw images. Among the three classifiers, the SVM classifier achieves the highest recognition rate for both with and without Gabor-filtered images. For static hand gestures on our own database, the result reveals that the combination of Gabor-filtered image with the SVM

method has the highest recognition rate of 96.1% with 103 feature vectors used. However, the recognition rate is only 86.4% for the same classifier but without using Gabor filters.

In addition, a dynamic gesture recognition system is presented for more real-life conditions. The method for segmenting the hand from the forearm is found to be effective. The recognition result is improved from 72.8% to 93.7% when the hand-pose correction module is applied; this indicates that using the responses of Gabor filters to estimate the hand-pose angle is effective.

**Keywords:** Hand gesture, Gabor feature, Principal component analysis (PCA), Support vector machine (SVM)

## 報告內容

### (一) 前言

在近幾年之中，基於影像與統計學習方法的人機介面相關研究越來越引人注目，例如：瞳孔辨識、人臉偵測與識別、手勢與手語識別等等。基於影像的優點，在於輸入時不需要任何的身體接觸，或是額外的輔助工具，例如磁卡或鑰匙，僅需利用數位相機或 CCD 所擷取的圖片來進行識別，對於注重隱私的現代人來說，這是相當重要的。基於統計學習方法的基本理論就是利用子空間轉換技術，將高維度的手勢影像空間投影至較低維度之影像空間，或轉換至頻域來進行分析，以加速訓練和提昇辨識的效果。手勢識別是一種最直接和最方便的輸入方式，它可應用的範圍相當的廣泛，例如：取代一般傳統的鍵盤和滑鼠來進行輸入、在虛擬實境中當作控制器來使用、或輔助聾啞人士做為溝通方式等。有鑑於此，開發更方便、更人性化的輸入介面，以迅速正確的進行手勢識別，就成為研究中相當重要的一項課題。

### (二) 研究目的

本研究提出一基於Gabor Filter + SVM的方法來進行手勢識別。本研究之目的在於提出(1)適應性膚色模型切換方法，以解決光線變化的問題；(2)手勢角度估算與校正方法，以解決手勢旋轉的問題；(3)手部與手臂分割方法，以允許使用者穿短袖衣服之彈性。透過以上問題的解決，可以提昇手勢識別的強健性。

### (三) 文獻探討

由於手部為一彈性體，故其外觀與形狀的變化幅度相當大；此外，由於手勢很容易受到光線、視角與手部遮蔽的影響，更增加「基於視覺手勢識別」研究上相當的難度。現有的主流技術包括有：主分量分析法(Principal Component Analysis; PCA)、彈性圖形匹配法(Elastic Graph Matching; EGM)及隱藏馬可夫模型(Hidden Markov Model; HMM)等方法。

Triesch與Malsburg[1,2]等人提出所謂的彈性圖形匹配法來針對複雜背景下的靜態手部姿勢進行分類與識別。在簡單背景下可以達成92.9%的識別率，在複雜背景下也有85.8%之辨識率。但由於試驗時所採用的訓練資料庫，其手勢影像視角變化均限制在一定的範圍內，

因此這個方法還是會受到不同視角的影響，因此本方法在應用上還是有其侷限性。

Chen 等人[3,4]為了解決猜拳遊戲(剪刀、石頭與布)中多角度的手勢辨識問題，採用三台攝影機自前方、左方與右方分別取得手勢影像，接著將所取得的影像資料用來訓練三個支持向量機分類器，待訓練完成後，再透過資訊融合的方式，以識別出猜拳遊戲中的手勢變化。最後實驗結果顯示：前方攝影機之辨識率為 73.3%，左方與右方攝影機則分別為 87.5%與 92.5%。此方法雖解決多重視角的問題，但由於背景單純，再加上識別的手勢僅有 3 種，使得整個問題簡化不少，但若應用在複雜背景下，或增加更多不同的手勢時，此方法之強健性仍有待考驗。

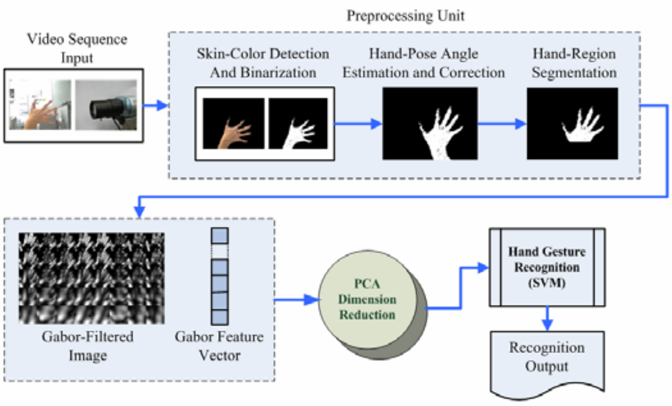
Qing Chen 等人[5,6]提出 Haar-Like Feature+SCFG (Stochastic Context-Free Grammar)組合而成的二階段式即時手勢辨識系統。首先利用 Haar-Like Feature+Adaboost Algorithm 來找出手勢位置與辨識手勢，接著利用第二階段的 Stochastic Context-Free Grammar 方式，利用其所產生的規則，找出輸入一連串手勢後所代表的意義，雖然此方法能夠快速的偵測出手勢的位置與其代表的意義，但是由於識別的手勢只有 4 種並且皆為單純的白色背景，而其容忍手勢旋轉的角度只能為±15°。

此外由於手勢的分類是透過學習過程來完成，所以用來訓練的影像數量是相當重要的。因此，Chen等人[7]採用2,000張影像序列，其中包含20個不同手勢與20個不同的個人，使用隱馬可夫模型可達成90.5%的辨識率；若再搭配傅利葉描述子(Fourier Descriptor)的總體不變性與運動特徵，辨識率可再往上提昇至93.%。

Amin等人[8]採用Gabor-PCA作為特徵擷取的方式，並利用Hough Transform的方式來作為手勢旋轉校正之前處理步驟，接著利用Fuzzy C-Mean來做為分類的方法，針對America Sign Language(ASL)內的26個字母來進行辨識，結果平均辨識率可達93.23%，但對於相似的字母其辨識率偏低。由於本方法手勢變化不大，同時也無法達到即時偵測與辨識的功能。

## (四) 研究方法

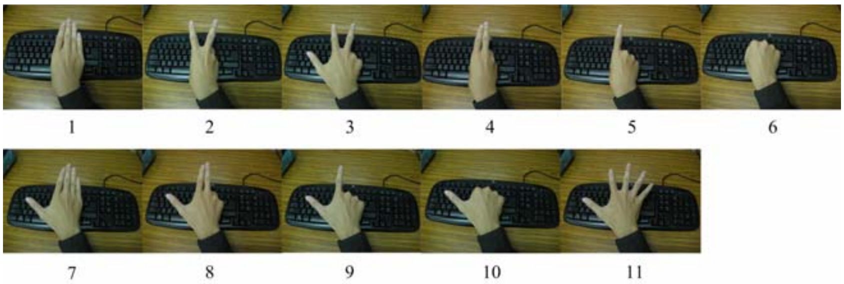
本研究主要是使用自行開發的前處理單元，其包含以下3個模組：(1)適應性膚色模型切換方法；(2)基於Gabor Feature的手勢角度估算與校正方法，與(3)手部切割方法。由於採用Gabor Feature來做為表示手勢的方法，所以必須結合主分量分析法來進行特徵降維，最後再利用支持向量機來進行手勢分類(見圖一)。本文將同時比較歐式距離、餘弦距離與SVM等分類器在手勢辨別上的優劣點。




圖一、本文所提手勢識別方法流程圖

### 4.1 自建手勢資料庫

不同的環境，例如背景的複雜度、光線的強弱、與攝影機的等級等，都是影響手勢影像品質的重要因素。本研究為避免上述因素之影響，決定自建手勢資料庫，手勢種類主要是參考文獻[9]。我們的手勢資料庫來自於10名使用者，並讓每位使用者比出11種不同的手勢(見圖二)，每一種手勢重複12次(見圖三)，接著透過CCD攝影機來擷取影像，之後再利用人工的方式將手勢區域擷取下來。因此，我們的手勢資料庫總共會有1,320張的手勢圖片，其中訓練與測試資料庫各為660張。

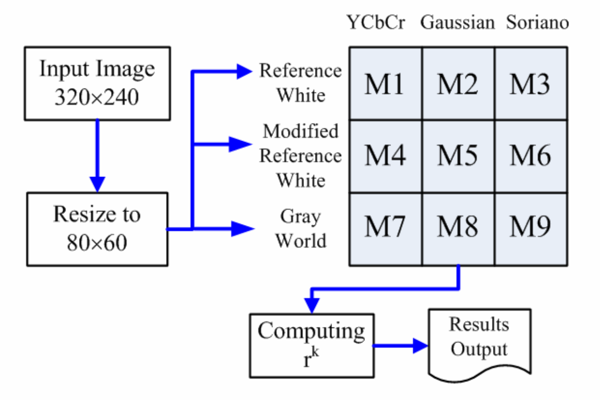



圖二、自建手勢資料庫內的 11 種手勢

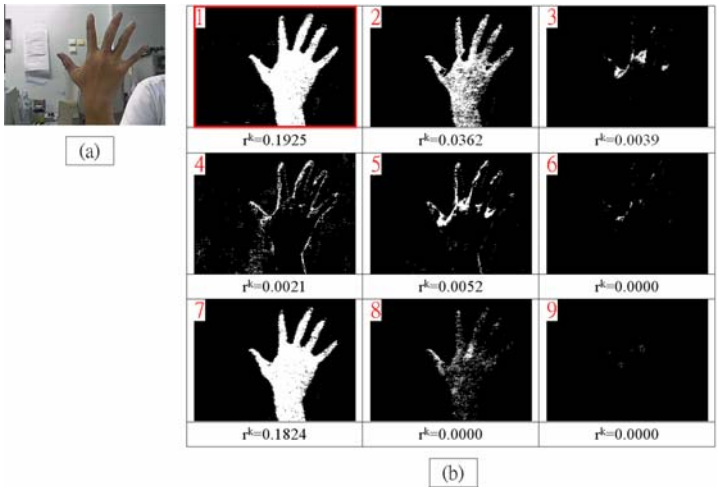
圖三、具有各種不同角度之手勢 1 圖片

### 4.2 自適應性膚色模型選擇

由於膚色分割常被用來做為手勢偵測之前處理方法，因為它的處理速度相當快。因此，成功的膚色分割對於即時應用扮演著相當重要的角色。本研究提出一新的方法(見圖四)-適應性膚色模型選擇。首先利用雙線性內插法將原始圖片(320×240像素)縮小為80×60像素，以提升前處理的速度，接著透過YCbCr色彩空間[10-12]、Soriano膚色模型[10]與Gaussian Model[12,13]等3種膚色分析方法，和搭配參考白方法[10]、修正參考白方法[14]與Gray World[15]等3種光線補償模式所組合而成的9種適應性膚色模型。最後利用「修正的四連通區域」與「膚色點」數之計算，進而提出「區域膚色點」之鑑別指標(若該連通區域內點數>100，即被視為所要保留之區塊)，來做為評估的準則。當膚色總點數除以總區塊數為最大時，該膚色模型即被視為最佳的模型。本指標能在不同的光線或複雜的背景中，自動挑選出最佳的膚色模型。而挑選出的膚色區域便是我們所想保留的手臂與手勢區域(見圖五)。




圖四、適應性膚色模型選擇流程圖




圖五、(a)原圖；(b)紅色框為經過適應性膚色模組選擇後之結果

### 4.3 基於Gabor Filter之手勢角度估算與校正方法

由於在一般的情形下，使用者作出的手勢，都會有不同程度的歪斜情況發生，若是直接對這些歪斜手勢進行識別，通常錯誤率會比較大。因此，為了解決手勢歪斜導致辨識率下降之問題，本研究提出一基於 Gabor Filter 之手勢角度估算與校正方法。其具體步驟如下(見圖六)：

(1) 進行影像二值化並擷取出手勢區域；緊接著，將分割出來的影像縮小至 20*20 像素，並與 Gabor Filter 進行迴旋積，其中 Gabor Filter 設定之參數如下：$\gamma = 0.785$，$\theta \in \{0^o, 90^o, 72^o, 45^o, 36^o, -72^o, -45^o, -36^o\}$，與 $\sigma = \{1,2,3\}$。

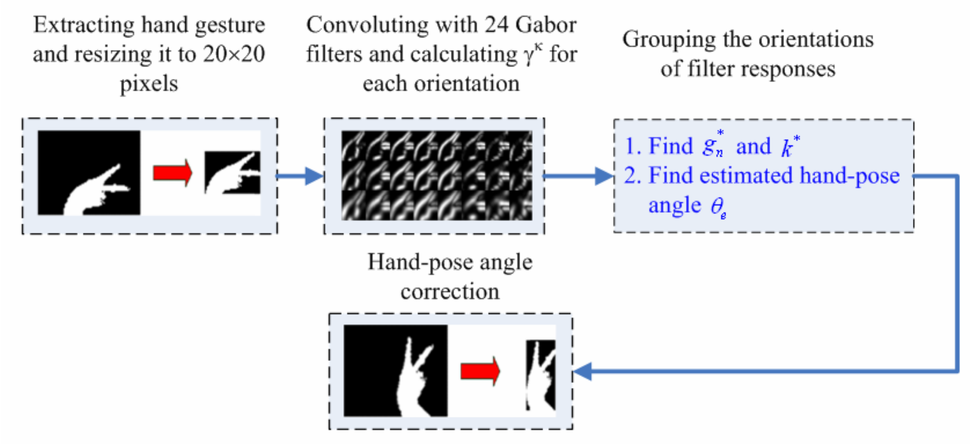
(2) 計算各個旋轉角度內所產生 Gabor Filters 的平均灰階值，亦即 $\gamma^k = \frac{1}{3wh}\sum_{j=1}^{3}\left(\sum_{x=1}^{w}\sum_{y=1}^{h} i(x,y)\right)$，其中 $k \in \{1, 2, \cdots, 8\}$，而 $w$ and $h$ 分別表示影像的寬與高。

(3) 將 $\theta = \{0°, 36°, 45°\}$；$\theta = \{36°, 45°, 72°\}$；$\theta = \{45°, 72°, 90°\}$；$\theta = \{0°, -36°, -45°\}$；

$\theta=\{-36°,-45°,-72°\}$；$\theta=\{-45°,-72°,-90°\}$各歸類為一組，全部共六組，並計算出各組的平均灰階值，亦即$\varsigma^{g_n}=\frac{1}{3}\sum_{k\in g_n}\gamma^k$，$n\in\{1,2,\cdots,6\}$。

(4) 選出平均灰階值最高的組別後，亦即$\varsigma_{\max}^{g_n^*}=\arg\max_{1\le n\le 6}\{\varsigma^{g_n}\}$。

(5) 再選出該組中平均灰階值最高的旋轉角度，亦即$\gamma_{\max}^{k^*}=\arg\max_{k\in g_n^*}\{\gamma^k\}$，因此可得到 Gabor Filter 最大響應的角度所在，該角度即為手勢所欲校正的角度。




圖六、手勢角度估算與校正方法流程圖

### 4.4 手部與手臂之分割

本文利用膚色分割的方式尋找出手部所在的位置。假設使用者為穿著長袖之狀態，則所擷取的範圍剛好是手勢區域。但若是使用者穿著短袖之狀態，便會擷取到手臂的部分。為了讓系統對於衣服(長袖或短袖)穿著具有更大的彈性，本文提出手部與手臂分割之方法。其具體實現步驟如下(見圖七)：

(1) 首先計算二值化影像每列之白點總數，並紀錄其最大值為$\alpha_n$。

(2) $\alpha_n$與水平軸頂點的距離記為$h^*$，並假設此圖片之寬為 w，高為 h。

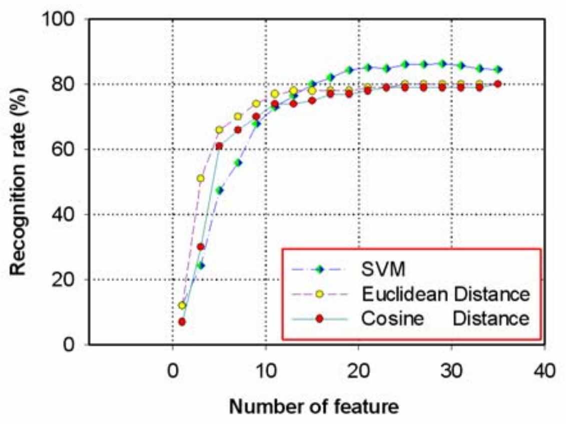
(3) 計算手部分割的大小$H$為：

$$H=\begin{cases} h^*+\frac{1}{4}\alpha_n & if\ h^*<\frac{4}{5}h \\ h & otherwise \end{cases}$$

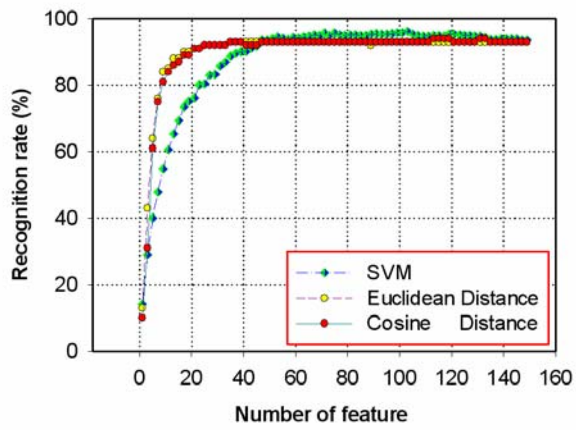

圖七、手部與手臂之分割示意圖

### (五) 結果與討論

首先針對靜態手勢資料庫進行本文所提方法之驗證。本實驗的訓練影像與測試影像分別有660張，其中包含各種旋轉角度之手勢影像，且訓練與測試影像不重複。圖八表示原始影像手勢識別率與所取特徵數之關係，由圖中可知，當所取特徵數高於15時，手勢識別率以SVM分類器最好，其中當特徵數等於29時，其識別率達最大為86.4%。反之，當利用Gabor filter來表示手勢特徵時，其手勢識別率明顯高於原始圖像(見圖八)，其中當特徵數等於103時，SVM分類器的識別率可達到96.1%之水準(見圖九)。由此可知，Gabor filter具有擷取手勢特徵之最大鑑別能量。




圖八、採用原始影像之手勢識別結果




圖九、採用Gabor-filtered Image之手勢識別結果

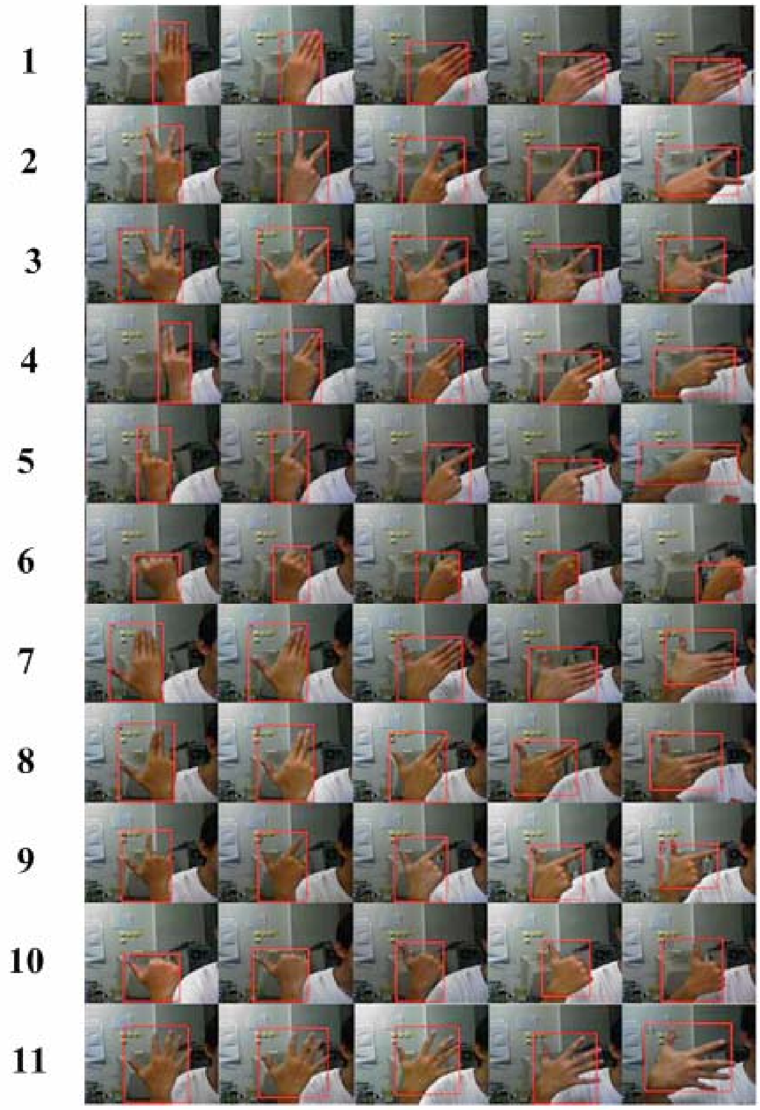
將以上結果總結於表一。由表中可知：不論採用何種分類器，利用Gabor filter來表示手勢特徵，其識別率皆高於原始手勢影像，最大者並可高出13%左右。在使用的3種分類器中，以SVM分類器的結果為最好。其中在採用Gabor filter的情況下，特徵數量為103時辨識率為96.1%；而在未採用Gabor filter的情況下，當採用特徵數量為29時，其辨識率為86.4%。

表一、不同分類器下 Raw image 與 Gabor-filtered image 之最高辨識率

| | Raw image (20×20 pixels) | | Gabor-filtered image (160×100 pixels) | |
|---|---|---|---|---|
| | 辨識率(%) | 特徵數量 | 辨識率(%) | 特徵數量 |
| Euclidean Distance | 80.0 | 25 | 93 | 35 |
| Cosine Distance | 80.0 | 35 | 94 | 113 |
| SVM (C=3, γ=0.9 for Raw image)<br>SVM (C=2, γ=0.125 for Gabor-filtered image) | **86.4** | **29** | **96.1** | **103** |

本文亦實現動態手勢識別系統，其結果如圖十所示。針對動態手勢識別，分別針對穿著長袖衣服與穿著短袖衣服進行試驗。在有使用手勢旋轉模組下：當穿著短袖時，在270幀中成功辨識出252幀，SVM分類器的辨識率為93.3%；當穿著長袖時，在300幀中成功辨識出282幀，SVM分類器的辨識率為94.0%，因此平均識別率為93.7%。當系統沒有採用手

勢旋轉模組時：當穿著短袖時，在260幀中成功辨識出189幀，SVM分類器的辨識率為72.7％；當穿著長袖時，在303幀中成功辨識出221幀，SVM分類器的辨識率為72.9％，因此平均識別率為72.8%。因此，由實驗結果可知採用手勢旋轉模組確實能有效的提升辨識結果。




圖十、11 種手勢在各種角度下之辨識結果

**(六) 結論與建議**

本研究成功地提出一基於Gabor filter + SVM手勢識別系統。本系統提出(1)適應性膚色選擇模組，以克服光線變化之影響；(2)手勢角度估算與校正模組，以克服手勢旋轉之影響；(3)手部與手臂切割模組，以克服穿著不同衣物對識別率之影響。同時，我們亦自建接近真實條件之手勢影像資料庫。針對本文所提方法，對於靜態手勢識別，若測試對象為原始影像時，SVM分類器之識別率為86.4%，但當採用Gabor filter來表示手勢特徵時，其識別率可以由86.4%提昇至96.1%。由此可知，Gabor filter具有擷取手勢特徵之最大鑑別能量。對於動態手勢識別，當不採用手勢旋轉模組時，SVM分類器之平均識別率為72.8%；但引入手勢旋轉模組後，其識別率可以由72.8%提昇至93.7%。由此可知，本文所提手勢角度估算與校正方法確實能有效地提昇手勢識別率。

# 參考文獻


[1] J. Triesch and C. von der Malsburg, “Robust classification of hand postures against complex backgrounds,” In: Proceedings of the IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 170-175, Killington, Vermont, USA, Oct. 1996.

[2] J. Triesch and C. Von Der Malsburg, “A system for person-independent hand posture recognition against complex backgrounds,” IEEE transactions on pattern analysis and machine intelligence, Vol. 23, No. 12, pp. 1449-1453, 2001.

[3] Y. T. Chen, and K. T. Tseng, “Multiple-angle hand gesture recognition by fusing svm classifiers,” In: Proceedings of IEEE conference on Automation Science and Engineering, pp. 527-530, Scottsdale, AZ, USA, Sep. 2007.

[4] Y. T. Chen, and K. T. Tseng, “Developing a multiple-angle hand gesture recognition system for human machine interactions,” In: Proceedings of the 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON), pp. 489-492, Taipei, Taiwan, Nov. 5-8, 2007,.

[5] Q. Chen, N. D. Georganas, and E. M. Petriu, “Real-time vision-based hand gesture recognition using haar-like features,” In: Proceedings of IEEE Instrumentation and Measurement Technology Conference Proceedings, pp. 1-6, Warsaw, Poland, May 1-3, 2007

[6] Q. Chen, N. D. Georganas, and E. M. Petriu, “Hand gesture recognition using Haar-like features and a stochastic context-free grammar,” IEEE Transaction on Instrumentation and Measurement, Vol. 57, No. 8, Aug 2008.

[7] F. S. Chen, C. M. Fu, and C. L. Huang, “Hand gesture recognition using a real-time tracking method and hidden markov models,” Image and Vision Computing, Vol. 21, No. 8, pp. 745-758, Aug. 2003.

[8] M. A. Amin, and H. Yan, “Sign language finger alphabet recognition from gabor-pca represention of hand gestures,” In: Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, August 19-22, 2007

[9] A. Caplier, L. Bonnaud, S. Malassiotis, and M. Strintzis, "Comparison of 2d and 3d analysis for automated cued speech gesture recognition," In: Proceedings of the 9th International Workshop on Speech and Computer (SPECOM '04), Saint-Petersburg, Russia, September 2004.

[10] R. L. Hsu, M. A. Mottaleb, and A. K. Jain, “Face detection in color image,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, pp. 696-706, 2002.

[11] C. Garcia, and G. Tziritas, “Face detection using quantized skin color regions merging and wavelet packet analysis,” IEEE Transactions on Multimedia, Vol. 1, No. 3, pp. 264-277, 1999.

[12] D. Chai, and A. Bouzerdoum, “A Bayesian approach to skin color classification in YCbCr color space,” In: Proceedings of TENCON, Vol. 2, pp. 421-424, 2000.

[13] C. Garcia, and G. Tziritas, “Face detection using quantized skin color regions merging and wavelet packet analysis,” IEEE Transactions on Multimedia, Vol. 1, No. 3, pp. 264-277,

1999.
[14] J. Y. Xu, “Face detection and recognition technology research in complex background,” In: Proceedings of M.S. thesis, Shandong university of technology, pp. 22-24, China, 2007.
[15] E. Y. Lam, “Combining gray world and retinex theory for automatic white balance in digital photography,” In: Proceedings of the Ninth International Symposium on Consumer Electronics, pp.134-139, Macau, 2005.

## 發表論文

1. **<u>Deng-Yuan Huang</u>**, Wu-Chih Hu, Sung-Hsiang Chang, and Mu-Song Chen, “Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination,” submitted to *Expert Systems with Applications* **(SCI)**.
2. **<u>Deng-Yuan Huang</u>**, Wu-Chih Hu, and Mao-Hsiang Hsu, "Adaptive skin color model switching for face tracking under varying illumination", accepted by and to appear in *the 4th International Conference on Innovative Computing, Information and Control (ICICIC2009)*, IEEE Computer Society Press, Kaohsiung, Taiwan, Dec. 7-9, 2009. **(EI)**
3. **<u>Deng-Yuan Huang</u>**, Wu-Chih Hu, and Sung-Hsiang Chang, "Vision-based hand gesture recognition using PCA+Gabor filters and SVM", *Proceedings of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP2009)*, IEEE Computer Society Press, Vol. 1, pp. 1-4, Kyoto, Japan, Sep. 12-14, 2009. **(EI)**

## 計畫成果自評

計畫成果自評部份，請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

本研究計畫原申請兩個子計畫，預計於兩年內完成，分別為「複雜背景下基於SVM之手勢識別」與「應用手勢識別於輔助手語翻譯系統之開發」，但因只核定通過第一個子計畫，因此將僅就該計畫之執行成果來進行自評。本計畫原目標設定在解決複雜背景下不同視角之手勢識別問題；針對此一目標，本研究提出一基於Gabor Filter + SVM之方法可有效提昇靜態與動態之手勢識別。本研究成果之重點在於提出(1)適應性膚色模型選擇方法，以解決光線變化的問題；(2)手勢角度估算與校正方法，以解決手勢旋轉的問題；(3)手勢與手臂分割方法，以增加使用者穿著任何衣物(短袖或長袖)之彈性。透過前述問題的解決，可以有效地提昇手勢識別之準確性與強健性。由實驗結果來評估當初原計畫所設定之目標，應已完全達到預期之目標。

本研究成果可應用於任何真實、無限制條件下之測試環境。本研究所提手勢識別方法同時具有光線變化強健性、手勢旋轉不變性，與使用者穿著任何衣物之彈性。本方法對於靜態手勢識別，可以達到96.1%識別率之水準。對於動態手勢識別，當加入手勢旋轉模組後，其識別率可以由72.8%提昇至93.7%。由此可知，本研究所提手勢角度估算與校正方法確實

能有效地提昇手勢識別率。根據以上結果，本研究方法確實極具學術與應用價值，其成果當然適合於國際研討會與國際期刊來進行發表。本研究總計發表2篇國際研討會論文，與產出一篇SCI等級的論文(已投稿至Expert Systems with Applications)(請參見論文發表列表)，因此本研發成果不論在質與量都有相當長足的進步。

附件二

# 可供推廣之研發成果資料表

□可申請專利　■ 可技術移轉　日期：98 年 10 月 31 日

| | |
|---|---|
| **國科會補助計畫** | 計畫名稱：智慧型人機互動界面之開發<br>計畫主持人：黃登淵<br>計畫編號：NSC 97-2221-E-212 -035　學門領域：圖形辨識 |
| **技術/創作名稱** | 光線變化下基於Gabor filter手勢旋轉校正之手勢偵測與識別技術 |
| **發明人/創作人** | 黃登淵 |
| **技術說明** | 本技術提出一基於Gabor Filter + SVM的方法來進行手勢識別。本技術之重點在於提出(1)適應性膚色模型切換方法，以解決光線變化的問題；(2)手勢角度估算與校正方法，以解決手勢旋轉的問題；(3)手勢與手臂分割方法，以允許使用者穿著短袖衣服之彈性。透過以上問題的解決，可以提昇手勢識別的強健性與準確性。 |
| | The author present a novel technology for hand gesture recognition based on Gabor filters and support vector machine (SVM) classifiers for environments with varying illumination. This technology (1) is robust against varying illumination, which is achieved using an adaptive skin-color model switching method; (2) is insensitive to hand-pose variations, which is achieved using a Gabor filter-based gesture angle estimation and correction method; (3) allows users to wear either a long- or short-sleeve shirt, which is achieved using a method that segments the hand from the forearm. |
| **可利用之產業**<br>**及**<br>**可開發之產品** | **可利用之產業：**<br>人機界面、視訊監控、治安保全、門禁監控、醫療診斷，與智慧家庭等。<br>**可開發之產品：**<br>虛擬實境、數位遊戲、手語翻譯等產品化。 |
| **技術特點** | 本技術特點具有(1)對光線變化下之手勢識別具強健性；(2)對手勢旋轉具有不變性；(3)對使用者的衣物穿著具有相當大的彈性。本方法對靜態與動態手勢識別率分別可達96.1%與93.7%之水準。 |
| **推廣及運用的價值** | 由於產業界對於人機界面之技術需求相當高，應用領域也不斷的擴大，因此本技術之推廣與運用價值均相當高。 |

1.每項研發成果請填寫一式二份，一份隨成果報告送繳本會，一份送　貴單位研發成果推廣單位（如技術移轉中心）。

2.本項研發成果若尚未申請專利，請勿揭露可申請專利之主要內容。

# Vision-based Hand Gesture Recognition Using PCA+Gabor Filters and SVM

Deng-Yuan Huang[1], Wu-Chih Hu[2], Sung-Hsiang Chang[3]
[1,3]Department of Electrical Engineering, Da-Yeh University
[1]kevin@mail.dyu.edu.tw, [3]r9603025@mail.dyu.edu.tw
[2]Department of Computer Science and Information Engineering, National Penghu University
wchu@npu.edu.tw

***Abstract*—In this paper we present a novel method for hand gesture recognition based on Gabor filters and support vector machine (SVM). Gabor filters are first convolved with images to acquire desirable hand gesture features. The principal components analysis (PCA) method is then used to reduce the dimensionality of the feature space. With the reduced Gabor features, SVM is trained and exploited to perform the hand gesture recognition tasks. To confirm the robustness of the proposed method, a dataset with large posed-angle (>45 deg.) of hand gestures is created. The experiment result shows that the recognition rate of 95.2% can be achieved when SVM is used. A real-time video system for hand gesture recognition is also presented with a processing rate of 0.2 s for every frame. This result proves the efficiency and superiority of the proposed Gabor-SVM method.**

***Keywords-gesture recognition; Gabor wavelet; SVM; PCA***

## I. Introduction

Hand gesture has become an important application in vision-based human-computer interfaces (HCIs) for the past decades because the traditional HCI input devices, such as mice and keyboard, cannot quickly respond to currently complicated interaction systems. For hearing impaired community, the development of automatic gesture translation based natural languages (e.g. the American Sign Language; ASL) is highly expected to improve their communication means among humans. For the recognition system of hand gestures based on images or videos, the posed-angle of gestures taking by a camera/webcam can usually be a critical factor in determining the effectiveness of the recognition systems.

Triesch and Malsburg [1] applied the method of elastic bunch-graph matching (EBGM) to the classification of hand postures for grayscale images. For EBGM method, the jets are described as vectors based on a 2D Gabor-wavelet transform. The results showed that their system can reach 86.2% recognition rates against complex background. This approach can achieve user-independent and scale-invariant recognition. However, this method is not view-independent and is computational inefficiency due to the matching process.

Chen and Tseng [2] proposed a multi-angle hand gesture recognition system for finger guessing games. To cope with various angles for hand gestures, they used three webcams set at front, left, and right directions of hand to capture gesture images. Then, three SVM classifiers are trained using the images acquired from the three cameras. After the training process, the constructed classifiers were fused by one voting and two plans of fusion to decide the gesture. The recognition rates of their system for the front, left, and right classifiers are 73.3%, 87.5%, and 92.5%, respectively. However, only 3 hand gestures were used in their work.

Amin and Yan [3] used PCA and Gabor filters to recognize the American Sign Language (ASL) finger alphabets from hand gesture images. The classification is then conducted with a method of fuzzy-c-mean clustering. The experimental results showed that the recognition rate of the ASL alphabets with average 93.23% accuracy can be achieved. However, the recognition rate of similar alphabets is relatively low in their approach.

In the past decade, the Gabor features have been successfully used in the fields of hand gesture and face recognitions [3,4]. However, Gabor features employed are of too high dimensionality to be used effectively. We proposed to deal with this problem by the PCA method to reduce the dimensionality of the feature space. The classification of hand gestures is then performed by the SVM method. Finally, a real-time video system on hand gesture recognition is presented.

The remainder of the paper is organized as follows: In Section 2, we describe how to construct the dataset of hand gestures and briefly outlines the methods of extracting the hand gesture features, followed by sketching the method of SVM. In Section 3, we present our experiment results. Finally, Section 4 gives a conclusion and some suggestions for future work.

## II. System Description

This section consists of the major components such as image capturing and preprocessing, feature extraction and classification of hand gestures. A brief description is provided in the subsequent sections.

### *A. Image Capturing and Preprocessing*

Some critical factors such as lighting condition, posed-angle and scale variability of hand gesture should be considered while collecting images for hand gesture recognition. The images of 11 hand gestures (see Fig. 1) for training and testing were collected with the same colored background by 10 signers, and each signer was requested to sign the same hand gesture 12 times; each time from a different angle and position. In order to verify the classification capability, the hand was cropped manually and resized to 20*20 pixels at the initial stage of the experiment. Fig. 2 shows that the pose of hand gestures with small angles

978-0-7695-3762-7/09 $26.00 © 2009 IEEE
DOI 10.1109/IIH-MSP.2009.96

(<45 deg.) and large angles (>45 deg.), which were taken six times separately to test the robustness of the proposed method. The dataset contains 120 (=10*12) images of each hand gesture and a total of 1320 (=10*12*11) images for 11 signs of hand gesture.

In a real time vision-based system, we first extracted the hands from a sequence of video images using the skin color information. Skin color may not be enough for tracking hands but it is often a fast convenient cue. We used the concept of "reference white" [5] as a method for lighting compensation and exploited the skin color model suggested by Soriano et al. [6] for converting a RGB color space into a normalized rgb color space. The segmented image of hand gesture was first resized to 20*20 pixels and then converted it into a grayscale image.

### B. *Feature Extraction of Hand Gestures*

Features of hand gesture are collected by the following three steps. We first convolved the hand gesture images with the Gabor filters. The PCA method was then used to reduce the dimensionality of the Gabor-coded images. Finally, the Gabor-coded images were concatenated by the rows to form a discriminating feature vector. The details of the two methods are described in the following sections.

*1) The Gabor filter:* Gabor filters (wavelets, kernels) can capture the most significant visual properties such as spatial locality, orientation selectivity, and spatial frequency characteristics. Considering the preferable characteristics, we chose the Gabor features to represent the hand gesture images.

Mathematically, a 2D isotropic Gabor filter is the product of a 2D Gaussian and a complex exponential function. The general expression can be expressed as

$$g_{\theta,\gamma,\sigma}(x,y)=\exp\left(-\frac{x^2+y^2}{\sigma^2}\right)\exp\left(\frac{j\pi}{\lambda}(x\cos\theta+y\sin\theta)\right) \quad (1)$$

The parameter θ represents the orientation, λ is the wavelength, and σ indicates scale at orthogonal direction. However, with this set of parameters the Gabor filter does not scale uniformly as the parameter σ changes. It is better to use a parameter γ=λ/σ to replace λ so that a change in σ corresponds to a true scale change in the Gabor filter. Also, it is convenient to apply a 90$^o$ counterclockwise rotation to (1), such that θ expresses the normal direction to the Gabor wavelet edges. Therefore, the Gabor filter can be alternatively defined as follows

$$g_{\theta,\gamma,\sigma}(x,y)=\exp\left(-\frac{x^2+y^2}{\sigma^2}\right)\exp\left(\frac{j\pi}{\gamma\sigma}(x\sin\theta-y\cos\theta)\right) \quad (2)$$

By selectively changing each of the parameters of the Gabor filter, one can tune the filter to a specific pattern arising in the image. Some examples of Gabor filter with different parameters (γ, θ, σ) are illustrated in Fig. 3.

By convolving a Gabor filter with image patterns, the similarity based on the Gabor response can be estimated. To emphasize three types of characteristics in images such as edge-oriented, texture-oriented, and a combination of both, one can change the parameters (σ, γ, θ) of the Gabor filter, where the variation of θ changes the sensitivity to edge and texture orientations, the variation of σ represents different "scales", and the variation of γ indicates the sensitivity to high/low frequencies.

A set of parameters of the Gabor filters used is γ=0.785, θ={0, π/2, 2π/5, π/4, π/5, 2π/11, π/7, π/8}, and σ={1, 2, 3, 4, 5}. Therefore, 40 Gabor responses from each image can be obtained. Each filter response is then converted into a pattern vector with 400 elements for a filter response of 20*20 pixels. By concatenating the 40 filter responses, the pattern vector has a dimensionality of 400*40=16,000. Fig. 4 shows the Gabor filter responses of hand gesture for sign "2".


Figure 1. Hand gestures in the dataset


Figure 2. Hand gesture images with small angle (<45 deg.) in top row and with large angle (>45 deg.) in bottom row

(a) (b) (c)

Figure 3. Examples of Gabor filters. Each example shows the real part of Gabor filter for different parameters. (a) γ={1/2, 3/2, 5/2, 7/2}; (b) θ={0, π/6, π/3, π/2}; (c) σ={4, 8, 12, 16}

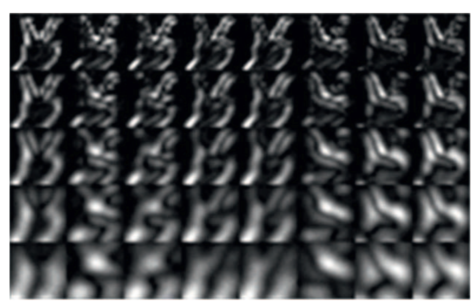


Figure 4. Gabor-coded image of hand-gesture "2"

*2) Principal component analysis (PCA):* The method of PCA [7] is a popular dimensionality reduction technique with the goal to find a set of orthonormal vectors in the data space, which can maximize the data's variance and map the data onto a lower dimensional subspace spanned by those vectors.

Consider a dataset with $M$ images $x_i \in \Re^N \ (i = 1, \cdots, M)$ belonging to C subjects, and $N$ is the number of pixels in the image. The total scatter matrix $S_T \in \Re^{N \times N}$ is defined as

$$S_T = \sum_{i=1}^{M} (x_i - \mu)(x_i - \mu)^T = AA^T \tag{3}$$

where μ is the global mean image of the training set, and $A = [x_1 - \mu \ \cdots \ x_M - \mu] \in \Re^{N \times M}$.

A direct computation of $S_T$ is impractical due to the huge size $N \times N$ of the matrix. Instead of direct finding the eigenvector $W_{PCA}$ of $S_T$, we solve the eigenvalue problem, $RV_{PCA} = V_{PCA}\Lambda$, to obtain the eigenvectors, $V_{PCA} \in \Re^{M \times P}$, and the eigenvalues, $\Lambda = diag[\lambda_1 \ \cdots \ \lambda_P] \in \Re^{P \times P}$, with decreasing order $\lambda_1 \geq \cdots \geq \lambda_P > 0$, where $\lambda_i$ is the nonzero eigenvalue of the matrix $R = A^T A \in \Re^{M \times M} \ (M \ll N)$. Then, the PCA subspace $W_{PCA}$ is formed by multiplying the matrix A with the eigenvectors $V_{PCA}$, that is, $W_{PCA} = AV_{PCA} \in \Re^{N \times P}$. Therefore, the feature vector **y** of an image **x** is acquired by projecting **x** into the coordinate system defined by the PCA subspace, that is

$$y = W_{PCA}^T (x - \mu) \in \Re^P \tag{4}$$

### C. *Classification of Hand Gestures*

In principle, one SVM classifier searches for an optimal hyperplane that maximizes the margins of their decision boundaries to ensure that their worst-case generalization errors are minimized, which is known as "structural risk minimization (SRM)."

To perform the classification between two classes, a nonlinear SVM classifier is applied by mapping the input data $(x_i, y_i)$ into a higher dimensional feature space using a nonlinear operator $\Phi(x)$, where $x_i \in \Re^d$ and $y_i \in \{+1, -1\}$. Therefore, the optimal hyperplane can be computed as a decision surface

$$f(x) = \mathrm{sgn}\left( \sum_i y_i \alpha_i K(x_i, x) + b \right) \tag{5}$$

where $\mathrm{sgn}(\cdot)$ represents the sign function, and $K(x_i, x) = \Phi(x_i)^T \Phi(x)$ is the predefined kernel function that satisfies Mercer's condition [8]. In this research, the radial basis function (RBF) is used and it is defined as follows

$$K(x_i, x) = \exp\left( -\gamma \|x_i - x\|^2 \right), \ \gamma > 0 \tag{6}$$

where γ=0.25. The coefficients $\alpha_i$ and $b$ in (5) can be determined by the following quadratic programming (QP) problem

$$\begin{aligned} &\max \left[ \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right] \\ &s.t. \ \sum_i \alpha_i y_i = 0 \\ &0 < \alpha_i < C, \ \forall i \end{aligned} \tag{7}$$

The parameter $C$ is a penalty that represents the tradeoff between minimizing the training set error and maximizing the margin, where $C = 8$ is determined empirically. Since the SVM is a binary classifier, it should be extended for an *m*-class problem in hand gesture recognition. We used the so called one against one approach, which is a pairwise method and needs to train $m(m-1)/2$ SVM classifiers. In addition, another two distance measures, i.e., Euclidean distance and Cosine similarity distance, are also computed to make a comparison with the SVM method.

## III. RESULTS AND DISCUSSION

The dataset of hand gesture images was classified as a training set and a testing set, for which data set has 6*11*10=660 images (3 small angles and 3 large angles data selected randomly from each of the 10 signers for 11 hand gestures). The image of hand gesture was first convolved with Gabor filters to form a Gabor-coded image. The data of 40 filter responses were concatenated by the rows to form a pattern vector with a dimensionality of 16,000, which is further reduced by the PCA method to construct a discriminating feature vector. Finally, the classification of hand gestures was performed by SVM (C=8 and γ=0.25), a Euclidean distance, and a cosine similarity distance, respectively.

Fig. 5 shows the recognition rates for the three methods when different number of features acquired from the PCA method is used. The maximum recognition rates using SVM, the Euclidean distance, and the cosine distance are 95.2%, 93%, and 93%, respectively, with corresponding numbers of features being 100, 50, and 50, respectively. This result confirms the outstanding performance of the proposed Gabor-SVM method when compared to the other two methods. The analysis of the confusion matrix performed by the SVM method with a number of features of 100 (see Table 1) verifies the results and reveals which gestures are sources of errors.

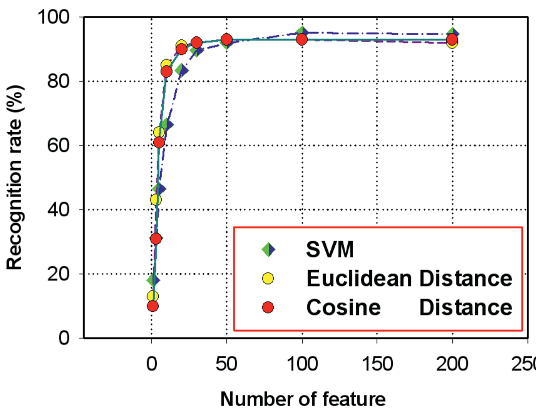



Figure 5. Results of recognition rate versus number of features used


Figure 6. Some detection results of hand gestures for the proposed system

For instance, gestures 4 and 5 are very similar and thus are partly confused with the SVM method. The rate of gesture 5 misclassified to gesture 4 is 8.3%. However, gesture 8 is confused with gesture 9 having a misclassification rate of 6.6%. This is not really surprising; indeed gestures 8 and 9 have been selected to test the limits of recognition.

A vision-based hand gesture recognition system was carried out on a Pentium PC with a 3.4 GHz processor and 4GB DDR II memory. A sequence of video images were captured by a webcam, and then processed by skin-color segmentation to extract the hand from the image. The cropped image is further resized to 20*20 pixels. Some detection results of hand gestures are shown in Fig. 6. The processing rate of the proposed system is about 0.2 second for every frame, which has readily achieved the requirement of a real-time system.

## IV. CONCLUSIN AND FUTURE WORK

We have proposed a novel Gabor-SVM method, which can achieve a recognition rate of 95.2% of hand gestures, and it is better than those of the other two methods, namely, Euclidean and cosine measures. The recognition results confirm the efficiency and superiority of the proposed Gabor-SVM method. Additionally, a hand gesture recognition system has been implemented with a processing rate of 0.2 second per frame. However, only 11 hand gestures are used with a limitation of wearing long sleeve clothes. In the future work, to increase the versatility of our hand gesture dataset, more different hand gestures need to be added and the limitation of wearing long sleeve clothes should be relaxed to accommodate the real environments.

## ACKNOWLEDGMENT

This research was fully supported by a grant from National Science Council, Taiwan, under contract NSC-97-2221-E-212-035.

## REFERENCES

[1] J. Triesch and C. von der Malsburg, "Robust classification of hand postures against complex backgrounds," In: Proc. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition, Killington, Vermont, USA, Oct. 1996, pp. 170–175.

[2] Y. T. Chen, and K. T. Tseng, "Multiple-angle hand gesture recognition by fusing SVM classifiers," In: IEEE conference on Automation Science and Engineering, Scottsdale, AZ, USA, Sep. 2007, pp. 527-530.

[3] M. A. Amin, and H. Yan, "Sign language finger alphabet recognition from Gabor-PCA representation of hand gestures," In: Proc. of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, August 2007, pp. 2218-2223.

[4] C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, pp. 572–581, May, 2004,.

[5] R. L. Hsu, A. M. Mohamed, and A. K. Jain, "Face detection in color images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, pp. 696-706, May, 2002,.

[6] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, "Skin detection in video under changing illumination conditions," In: Proc. of 15th International Conference on Pattern Recognition, Barcelona, Spain, vol. 1, 2000, pp. 839-842.

[7] M. Turk, and A. Pentland, "Eigenfaces for recognition," J. Cogn. Neurosci., vol. 3, pp. 71-86, January, 1991.

[8] V. N. Vapnik, Statistical learning theory, John Wiely and Sons, New York, 1998, pp. 423-424.

TABLE I. CONFUSION MATRIX FOR THE RECOGNITION RESULTS BY SVM WITH A NUMBER OF FEATURES OF 100.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 93.3 | 0 | 0 | 1.6 | 0 | 1.6 | 1.6 | 0 | 0 | 0 | 1.6 |
| 2 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 1.6 | 98.3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 1.6 | 0 | 0 | 96.7 | 0 | 0 | 1.6 | 0 | 0 | 0 | 0 |
| 5 | 0 | 1.6 | 0 | 8.3 | 88.3 | 0 | 0 | 0 | 1.6 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| 7 | 1.6 | 0 | 0 | 0 | 0 | 0 | 88.3 | 6.6 | 0 | 0 | 3.3 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 1.6 | 90.0 | 6.6 | 1.6 | 0 |
| 9 | 0 | 0 | 1.6 | 0 | 1.6 | 0 | 1.6 | 1.6 | 91.6 | 1.6 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |

# Adaptive Skin Color Model Switching for Face Tracking under Varying Illumination

Deng-Yuan Huang[1], Wu-Chih Hu[2], Mao-Hsiang Hsu[3]
[1,3]Department of Electrical Engineering, Da-Yeh University
[1]kevin@mail.dyu.edu.tw, [3]r9703030@mail.dyu.edu.tw
[2]Department of Computer Science and Information Engineering, National Penghu University
wchu@npu.edu.tw

***Abstract*—In this paper, an adaptive skin color model switching based on AdaBoost method for face tracking is proposed. Possible skin clusters under illumination varying scenes are detected by an optimal skin color model, which is adaptively selected by a well-defined quality measure. The possible facial candidates are further validated by AdaBoost to determine whether human faces exist in video sequences or not. The tracking sequences reveal that good and robust results are obtained from dim- to profile- to back-light scenarios. The performance of the proposed method can achieve an average tracking time of about 145.4 ms/frame and a detection rate of 94.4%.**

***Keywords-Skin color model; AdaBoost; Face tracking***

## I. INTRODUCTION

The most two important issues of face detection are face tracking and localization. Face detection are of importance due to a wide variety of applications such as public security, video surveillance, and access control. Face detection is often preceded by the extraction of skin-tone colors [1], since it is one of the most important cues of face features with invariance of the changes of face scales, poses, and facial expressions. However, the color-based approaches are quite difficult to robustly detect skin-tone color in the presence of complex background and varying illumination.

The compact skin clusters can be formed in YCbCr color space under a wide range of lighting variations [2]. However, the skin color models using YCbCr space frequently misclassifies non-skin pixels at low luminance as skin-tone pixels, and vice versa [1] due to its nonlinear dependence on luma. Since skin-tone pixels have a distinct shape in the normalized color space (r, g), Soriano et al. [3] use (r, g) color space to perform face detection with high capabilities under daylight, incandescent lamp, fluorescent light and a combination of these light sources. Under certain light conditions, the skin color distribution can be modeled by a Gaussian function in (r, g) color space [4]. However, since single skin color model cannot deal with similar skin-tone pixels in background and wide ranges of lighting changes, Stern et al. [5] proposed an adaptive color space switching method to tackle such difficulties and they claimed this method can achieve satisfied performance on face detection and tracking.

Much research for color constancy have been suggested but so far their performance has been inadequate. In this paper, we present a novel scheme based on adaptive switching for skin-color models (ASSM) with light compensation for face tracking in video sequences under unconstrained illumination. Extensive experiments show that switching between the skin color models results in better tracking performance when compared to using single skin color model throughout.

The remainder of this paper is organized as follows. In Section 2, the method of adaptive switching for skin color models is described. Section 3 presents the results of face tracking in video sequences under varying illumination. Conclusions and a future work are given in Section 4.

## II. ADAPTIVE SKIN COLOR MODEL SWITCHING METHOD

The proposed method, ASSM, is constructed by the possible combinations of three different skin color models, i.e., YCbCr model [1,2], Soriano's model [3], and Gaussian mixture model [4], with three types of lighting compensation, i.e., reference white [1], modified reference white [6] and gray world [7]. A quality measure for face tracking is presented to adaptively select an optimal skin color model from the combinations. The detected skin clusters are further validated by performing an AdaBoost method on each possible face candidate in video sequences.

### *A. Skin color models*

*1) YCbCr Skin color model:* YCbCr is a family of color spaces used as a part of the color image pipeline in response to increasing demands for digital approaches in handling video information, and has become a widely used model in digital video. In contrast to RGB, the YCbCr color space is luma-independent, leading to a better performance under varying lighting scenes. The corresponding skin cluster can be described as:

$$\begin{aligned}
&(Y,\ Cb,\ Cr)\ \textit{is classified as skin pixels if}:\\
&\qquad 60 \le Y \le 255\\
&\qquad 100 \le Cb \le 125\\
&\qquad 135 < Cr \le 170\\
&\textit{where}\ Y,\ Cb,\ Cr \in [0,\ 255]
\end{aligned} \tag{1}$$

*2) Soriano's skin color model:* The normalized RG color space (r, g) is first applied, and a pair of quadratic functions is used to define the upper and lower bounds of the skin locus [3]. To prevent grayish and whitish pixels from being labeled as skin, pixels are excluded from skin membership when falling within a circle with radius 0.02 around the white point (r=g=0.33). Here we slightly modified the Soriano's skin color model with some RGB constraints, $C_{RGB}$. Therefore, the skin cluster of value S can be determined with chromaticity (r, g) and original RGB space as (2) of which the dot ( $\cdot$ ) means the logical operator "and."

$$
\begin{aligned}
&S = \begin{cases} 1, & (g < g_u)\cdot(g > g_d)\cdot(R_W > 0.0004)\cdot C_{RGB} \\ 0, & otherwise \end{cases} \\
&where \\
&g_u = -1.3767r^2 + 1.0743r + 0.1452 \\
&g_d = -0.776r^2 + 0.5601r + 0.1766 \\
&R_W = (r-0.33)^2 + (g-0.33)^2 \\
&C_{RGB} = (R > 130)\cdot(B > 55)\cdot(G > B)\cdot((R-G) > 25)
\end{aligned} \tag{2}
$$

*3) Gaussian mixture skin color model:* Gaussian mixture models can be viewed as a form of generalized radial basis function (RBF) network and are used to describe the skin cluster in the RGB color space as:

$$
\begin{aligned}
&G(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
&where \\
&x \in [R,G],\ 51 \le R \le 102,\ 51 \le G \le 153
\end{aligned} \tag{3}
$$

where $\mu$ and $\sigma$ are mean as well as standard deviation estimated from the specified skin color regions of R and G channels, respectively. Let $G_{max} = \mu + 2\sigma$ and $G_{min} = \mu - 2\sigma$ be the upper and lower bounds of possible skin-tone pixels, respectively. The white point (R=G=84) is also excluded from the skin color membership. Thus, the corresponding skin cluster can be determined as:

$$
\begin{aligned}
&S(R,G) = \begin{cases} 1, & (G < G_{\max})\cdot(G > G_{\min})\cdot(R_W > 26) \\ 0, & otherwise \end{cases} \\
&where \\
&R_W = (R-84)^2 + (G-84)^2
\end{aligned} \tag{4}
$$

### *B. lighting compensation mathods*

*1) Reference white lighting compensation:* The concept of reference white was first presented by Hsu et al. [1]. In their method, the top 5% of the luma values in the image is regarded as the reference white if the number of these pixels is sufficiently large (>100). The R, G, and B components of a color image are then adjusted so that the average gray value of these reference-white pixels is linearly scaled to 255. Let $i \in [l_u,\ 255]$ be the top 5% gray levels and $f_i$ be the pixel number of gray level $i$ in the image. Thus, the modified RGB components can be estimated as:

$$
\begin{aligned}
&M_{top} = \sum_{i=l_u}^{255} i \cdot f_i \Big/ \sum_{i=l_u}^{255} f_i \\
&\chi_{new} = \chi_{old} / M_{top} \times 255,\ where\ \chi \in \{R,G,B\}
\end{aligned} \tag{5}
$$

*2) Modified Reference white:* This method was proposed by Xu [6], and it is a modified version of reference white [1]. In their method, the bottom 5% gray levels are also considered. Let $i \in [l_u,\ 255]$ and $i \in [0,\ l_d]$ be the top 5% and bottom 5% gray levels in the image, respectively. The modified (R, G, B) components can be calculated as:

$$
\begin{aligned}
&\chi_{new} = (\ln(\chi_{old}) - \ln(l_d))/(\ln(l_u) - \ln(l_d)) \times 255 \\
&where\ \chi \in \{R,G,B\}
\end{aligned} \tag{6}
$$

*3) Gray world lighting compensation:* Gray world [7] is one of the lighting compensation methods, which seeks to equalize the mean of the red (R), green (G), and blue (B) channels. The gray world assumption is based on the observation that for a typical scene, the average intensity of the red, green, and blue channels should be equal. Let *M* and *N* be the image height and width, respectively. First, the averages of RGB channels and gray levels, $\chi_{AVG}$ and $\mu_{AVG}$, are calculated, respectively, as follows:

$$
\begin{aligned}
&\chi_{AVG} = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N} I_\chi(x,y),\ where\ \chi \in [R,G,B] \\
&\mu_{AVG} = \frac{1}{3}(R_{AVG} + G_{AVG} + B_{AVG})
\end{aligned} \tag{7}
$$

Then, the scale ratios and modified pixels of original RGB channels, $A_\chi$ and $\hat{I}_\chi$, are estimated, respectively.

$$
A_\chi = \mu_{AVG} / \chi_{AVG},\ and\ \hat{I}_\chi(x,y) = A_\chi \cdot I_\chi(x,y) \tag{8}
$$

### *C. Face detection using AdaBoost algorithm*

Haar-like features (see Fig. 1) are widely used in face detection by AdaBoost algorithm adopting the concept of integral image that allows for efficient feature computation [8]. AdaBoost learning aims to select a small number of weak classifiers, which represent the local discriminative features of faces, and then to combine them into a strong classifier to decide whether an image is a face or a non-face. AdaBoost can deal with very large sets of weak classifiers due to its greedy characteristics. To significantly improve computational efficiency and also reduce false positive rate (FPR), a sequence of strong classifiers is concatenated as a so-called cascaded detector. More details of computing the integral image and implementing the AdaBoost algorithm can be found in [8].

### D. *Quality measure*

The skin clusters can be found by a 4-connected component labeling for skin-tone pixels greater than 50. Then the possible facial regions are fitted with a rectangle $W_r$ or an ellipse $W_e$ as shown in Fig. 2. To evaluate the quality of the segmented face regions for every possible combination of skin color models, a quality measure $r^k$ representing the $k^{th}$ skin color model is proposed and defined as:

$$r^k = \omega_1 r_1^k + \omega_2 r_2^k \tag{9}$$

where

$$r_1^k = \sum_{i=1}^{N_C} \left( \sum_{(x,y)\in W_r} p(x,y) \Big/ W_r \right) / N_C \tag{10}$$

and

$$r_2^k = \sum_{i=1}^{N_C} \left( (S_e^{W_e} + S_p^{W_e} + S_a^{W_e}) / 3 \right) / N_C \tag{11}$$

where $p(x, y)$ is a detected skin-tone pixel in $W_r$ or $W_e$, $\omega_1$ and $\omega_2$ are weights set to be 0.5, respectively, and $N_C$ is the total number of all 4-connected components with skin-tone pixels greater than 50. $S_e^{We}$, $S_p^{We}$ and $S_a^{We}$ represent the sensitivity (i.e., true positive rate = *TP*/(*TP*+*FN*)), the specificity (i.e., true negative rate = *TN*/(*TN*+*FP*)), and the spatial accuracy (=1-(*FP*+*FN*)/(*TP*+*FN*)), respectively. These values are estimated from the elliptical regions, $W_e$, where *TP*, *TN*, *FP* and *FN* represent the total number of true positive, true negative, false positive and false negative pixels, respectively.

## III. EXPERIMENTAL RESULTS

Experiments were performed on a computer with Intel(R) Core(TM)2 Quad Q8200 2.33 GHz processor and 3.25GB RAM. The algorithm was implemented in BCB (Borland C++ Builder) 6.0. Fig. 3 shows the flow chart of the proposed method for face tracking. The video image (320×240 pixels) is input and first resized to 80×60 pixels, and then the quality measures $r^k$ are computed by all possible combinations of skin color models. With maximum $r^k$, an optimal skin color model can be chosen. Afterwards, the possible facial candidates are further validated by the cascaded AdaBoost method. Fig. 4 shows the results of detected skin-tone pixels for all possible combinations, and an optimal combination of YCbCr + Modified reference white ( $r^k = 0.612$ ) is selected here.

Figure 5 shows the typical results of detected skin clusters of which some of the most important features of human face, i.e., eyes, may be lost due to severe lighting variations (see the red box with a size of W×H). Therefore, to keep as myriad face features as possible, the width of red box is expanded by 0.5W on both either side, and the height of red box is increased by 0.5H only on the top side. The detected possible face candidates (see the yellow box) are formed and then fed to the cascaded AdaBoost detector to further validate whether the possible candidate regions are a human face or not.

In order to evaluate the robustness of the proposed method, a tracking sequence with varying brightness of dim light, profile light and back light is performed as shown in Fig. 6. The tracking results of quality measure with adaptive skin color model switching in the video are shown in Fig.7. Note that only five models are used in this case. As revealed in Fig. 7, M1 and M7 are alternately used in the period of dim light, M7 dominates the period of profile light (scanning from left to right), and M5 (Gaussian model) is most preferred in the back light situation. Apparently, only single skin color model in this case cannot accommodate all the lighting variations from dim- to profile- to back light situations.

Finally, the performance of the proposed method in face tracking was evaluated by a sequence of video which is composed of 144 frames with a size of 320×240 pixels. There are totally 144 faces in the tested sequence (see Fig. 6). The detection rate, missing rate, and false alarm rate are $R_P$=94.4% (136/144), $R_M$=4.2% (6/144), and $R_{FA}$=1.4% (2/144), respectively. Furthermore, the average detection time for each frame is about 145.4 ms; this result also confirms the possibilities of real-time operation in face tracking for the proposed method.

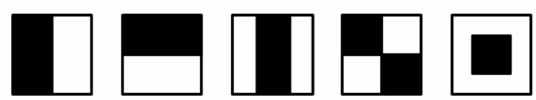


Figure 1. The set of Haar-like feature for AdaBoost

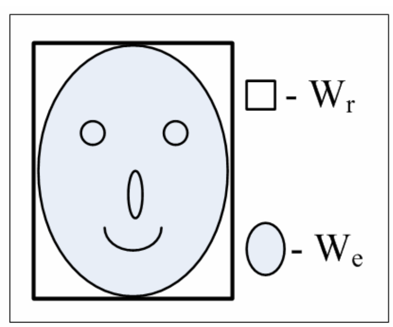



Figure 2. 4-connected component labeling for the estimation of quality measure

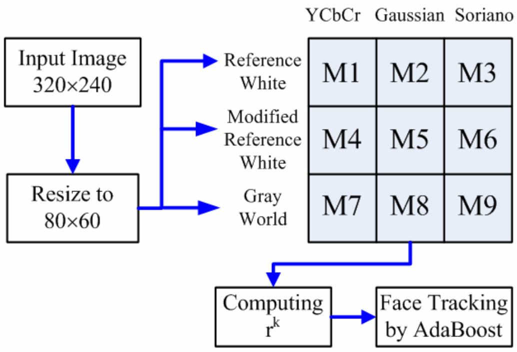
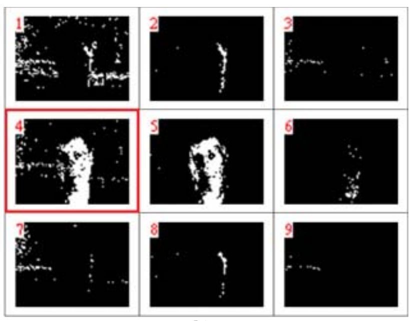


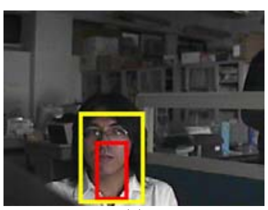

Figure 3. Flow chart for face tracking by adaptive skin color model switching method and AdaBoost

(a)

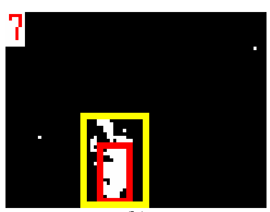


(b)

Figure 4. (a) Original image; (b) Skin-tone pixels detected by all possible combinations of skin color models



(a)



(b)

Figure 5. (a) Original image; (b) Results of detected skin cluster

(a) dim light (b) profile light (c) back light

Figure 6. The tracking sequence with great lighting variations

## IV. Conclusion and Future Work

In this paper, an adaptive skin color model switching with AdaBoost method is presented, and it can successfully deal with large lighting variations from dim- to profile- to back-light situations. The tracking time for the tested video sequence is about 145.4 ms/frame. In the proposed system, the view angles of faces are constrained within $\pm 30^o$. In a future work, we will focus on improving the tracking speed and on manipulating view angles greater than $30^o$.

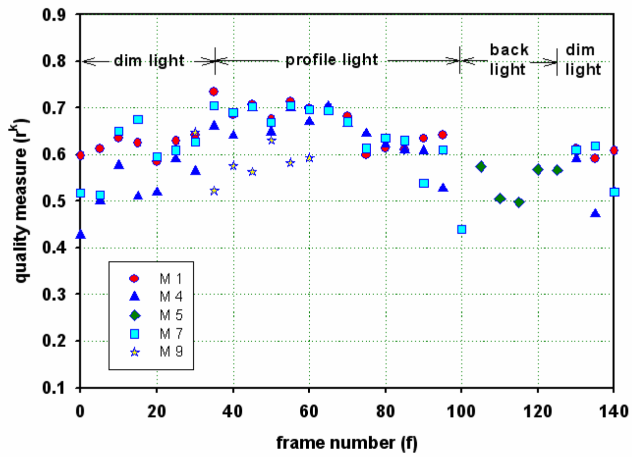



Figure 7. The variations of quality measure for face tracking with adaptive skin color model switching

## Acknowledgements

This research was fully supported by a grant from National Science Council, Taiwan, under contract NSC-97-2221-E-212-035.

## References

[1] R. L. Hsu, M. Abdel-Mottaleb, A. K. Jain, “Face detection in color image”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, 2002, pp. 696-706.

[2] D. Chai, and A. Bouzerdoum, “A Bayesian approach to skin color classification in YCbCr color space”, In: Proceedings of TENCON, Vol. 2, 2000, pp. 421-424.

[3] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, “Skin detection in video under changing illumination conditions”, In: Proceedings of the 15th International Conference on Pattern Recognition, Vol. 1, 2000, pp. 839-842.

[4] J. Yang and A. Waibel, “A real-time face tracker”, In: Proceedings of the IEEE Workshop on Applications of Computer Vision, Sarasota, Florida, USA, 1996, pp. 142-147.

[5] H. Stern, and B. Efros, “Adaptive color space switching for tracking under varying illumination”, Image and Vision Computing, Vol. 23, No. 3, 2005, pp. 353-364.

[6] J. Y. Xu, “Face detection and recognition technology research in complex background”, M.S. thesis, Shandong university of technology, China, 2007, pp. 22-24.

[7] E. Y. Lam, “Combining gray world and Retinex theory for automatic white balance in digital photography”, In: Proceedings of the Ninth International Symposium on Consumer Electronics, Macau, 2005, pp.134-139.

[8] P. Viola and M. J. Jones, “Rapid object detection using a boosted Cascade of simple features”, In: IEEE Proceedings of Computer Vision and Pattern Recognition, Vol.1, Hawaii, USA, 2001, pp. I511-I518.

# 行政院國家科學委員會補助國內專家學者出席國際學術會議報告

98 年 8 月 25 日

附件三

| | | | |
|---|---|---|---|
| 報告人姓名 | 黃登淵 | 服務機構及職稱 | 大葉大學電機系助理教授 |
| 會議 時間 地點 | 2009 年 8 月 12~14 日<br>中國瀋陽 | 本會核定補助文號 | NSC 97-2221-E-212-035- |
| 會議名稱 | (中文)<br>(英文)<br>The Ninth IEEE International Conference on Hybrid Intelligent Systems | | |
| 發表論文題目 | (中文)<br>(英文)<br>Video Object Segmentation by Integrating Motion Information and Gradient Compensation without Background Construction | | |

報告內容應包括下列各項：

一、參加會議經過

**98 年 08 月 11 日**

由於台灣到中國瀋陽並無班機直飛，因此必須透過小三通經由金門到廈門高崎國際機場，搭乘中國內陸航班往遼寧省會城市瀋陽。其間由於班機延誤加上轉機延遲等因素，以致抵達瀋陽桃仙國際機場時已接近傍晚時分，隨後即轉往本次會議地點瀋陽市中心洲際飯店(Intercontinental Hotel)附近之下塌旅館休息，會議第一天並無任何報告與論文發表。由於本日到達時間已晚，只能於隔天(8/12)辦理註冊手續。

**98 年 08 月 12 日**

本日首先至會場辦理註冊手續。這是HIS2009 會議第一天，首先由Program Committee Chair主持開幕儀式與致詞。隨即由Duke University教授 Paul P. Wang進行第一場Plenary Speech - Mathematics of Uncertainty for Intelligent Systems；與由Kyushu University副教授Hideyuki Takagi進行第二場Plenary Speech - IEC Prospector's Guide: Features of Some Interactive EC Frameworks。今天的議程相當緊湊，除了早上的二場Keynote speech之外，下午更安排了六個Session(每三個Session為同一時段)與四個Poster Session(每二個Poster Session為同一時段)，由於部分Session時間重疊，我選擇以下兩個比較感興趣的Sessions來參加，分別為：Session A02 – Agent-based user interface design and applications與Session A05 – Soft computing and their application to multimedia。

**98 年 08 月 13 日**

這是 HIS2009 會議第二天，大會首先安排台灣的逢甲大學講座教授張真誠進行第三場Plenary Speech - A Self-Reference Watermarking Scheme based on Wet Paper Coding；與由University of Glamorgan 教授 Peng Shi 進行第四場 Plenary Speech - Robust Filtering on Hybrid Dynamical Systems with Uncertainties。今天的議程也是相當的緊湊，除了早上安排的二場Keynote speech 之外，下午更安排了七個 Session 與五個 Poster Session，由於部分 Session 時間重疊，我今天選擇以下兩個 Session 來參加，分別為：Session B02 – Kernel learning, cluster analysis, and its applications 與 Session B05 – Innovative computing for image analysis (1)。

**98 年 08 月 14 日**

這是 HIS2009 會議的第三天，也是大會會議的最後一天。今天一整天都安排 Presentation Sessions and Poster Sessions，分別在早上安排九個 Sessions 與三個 Poster Sessions，下午則安排八個 Sessions 與四個 Poster Sessions 等。早上參加的 Sessions 分別為：Session C17 – Innovative computing for image analysis (2)，與 Session C08 – Information management & knowledge management (2)。下午參加的 Sessions 是：Session C12 – Advanced data processing technology，與 Session C15 – Soft computing for image and signal processing。同時早上的 Session C17 也是我們論文報告的 Session，本篇論文報告期間，與會人員發問相當踴躍、討論氣氛熱烈，由於時間限制，大家只得於會後繼續討論，彼此交換心得。

**98 年 08 月 15 日**

結束 HIS2009 會議行程，循原來小三通的方式回台灣。首先由瀋陽桃仙國際機場搭機至廈門高崎國際機場，然後再坐船至金門，再轉搭飛機回高雄小港國際機場。

二、與會心得

此次參加 HIS2009 會議，認識了許多世界各地同樣為 Intelligent computing 領域的研究人員，包括有馬來西亞、美國、日本、印度、英國等地的教授與博士生，當然也有接觸到中國當地的與會人士，皆與他們相談甚歡，同時也交換了許多在研究上的意見與想法，相信對於未來的研究會有更進一步的幫助。

除此之外，此次的最大收穫便是聽取了許多非常有趣的演講，其中讓我印象最為深刻的是Kyushu University副教授Hideyuki Takagi之keynote Speech - IEC Prospector's Guide: Features of Some Interactive EC Frameworks。本次Talk主要是講述如何利用植基於人類智識、經驗、偏好(preference)與直覺之Interactive Evolutionary Computation (IEC)方法來最佳化目標系統。一般說來，設計一個fitness function對大部分系統的最佳化是相當困難的，但IEC卻提供了一個解決的方法。因此，這個方法也啟示了研究者一個最佳化目標系統的一個方向，相當值得參考。

另一個收獲較大的是，本次有許多相關領域的台灣學者參加 HIS2009 會議，包括有雲科大張傳育教授、伍麗樵教授、虎尾科大郭文忠教授、澎湖科大胡武誌教授、嘉義大學柯建全教授、王智弘教授、盧天麒教授、王皓立教授、高雄大學張保榮教授、高應大王嘉男教授、

逢甲大學林志敏教授、龍華科大黃添財教授、大葉大學張顧耀教授等人，大家平時難得有機會可以聚在一起針對個人研究交換心得，透過這次的交流在研究領域的啟發與拓展真是獲益良多。

三、考察參觀活動(無是項活動者省略)

(略)

四、建議

主辦單位瀋陽師範大學相當用心，不論是會議議程之安排，或是其他參觀行程之排定，都相當完善。但此次研討會仍有一些美中不足之處，可能是瀋陽師大第一次承辦國際研討會，在辦理註冊手續時，由於承辦人員不熟悉整個流程，以致註冊手續花了大家相當長的時間才能夠辦理完畢，希望往後主辦單位能針對此點做一些改進。

五、攜回資料名稱及內容

下列為攜帶回國之資料，內容皆為此次研討會議所發表之論文。

Proceedings of 2009 Ninth International Conference on Hybrid Intelligent Systems, Volume 1.

HIS2009 論文光碟一片

六、其他

(略)

# 行政院國家科學委員會補助國內專家學者出席國際學術會議報告

98 年 9 月 21 日

附件三

| | | | |
|---|---|---|---|
| 報告人姓名 | 黃登淵 | 服務機構及職稱 | 大葉大學電機系助理教授 |
| 會議 時間<br>地點 | 2009 年 9 月 12~14 日<br>日本京都 | 本會核定<br>補助文號 | NSC 97-2221-E-212-035- |
| 會議名稱 | (中文)<br>(英文)<br>2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP2009) | | |
| 發表論文題目 | (中文)<br>(英文)<br>Vision-based Hand Gesture Recognition Using PCA+Gabor Filters and SVM | | |

報告內容應包括下列各項：

一、參加會議經過

**98 年 09 月 10 日**

本日搭乘日本航空班機 JL654，由桃園國際機場飛往大阪關西國際機場，本次航班(JL654)起飛時間為台北下午 13:45，到達時間為大阪下午 17:20。隨即入住距離大阪關西國際機場較近的神戶大倉飯店(Kobe Okura Hotel)，並將於次日到會場進行報到手續。

**98 年 09 月 11 日**

本日由神戶搭車至京都研討會場所在地 Mielparque Kyoto 進行報到，並領取會議相關資料、論文光碟等，由於本日無議程進行，因此以研讀相關資料及研討會論文為主。由於往後三日均有議程進行，因此這幾天均入住距離會場較近的京都新都 Hotel。

**98 年 09 月 12 日**

這是 IIH-MSP2009 會議第一天，首先由 Program Committee Chair 主持開幕儀式與致詞。隨即由 Department of Intelligence Science and Technology, Kyoto University 教授 Takashi Matsuyama 進行第一場 Keynote Speech - The state of the art of 3D video technology - Accurate 3D shape and motion reconstruction, high fidelity visualization, and efficient coding for 3D video。今天的議程相當緊湊，早上除了 Keynote speech 之外，更安排有 5 個 Session，其中本人選擇與自己研究領域最接近的 A05 session － Intelligent Surveillance and Pattern Recognition 來參加。下午則安排了 10 個 Session(每 5 個 Session 為同一時段)，由於部分 Session 時間重疊，我選擇以下兩個比較感興趣的 Sessions 來參加，分別為：Session A10 - Circuit Techniques for Multimedia Signal Processing 與 Session A13 - Advanced Multimedia Processing and Retrievals。

**98 年 09 月 13 日**

這是 IIH-MSP2009 會議第二天，大會首先安排 Graduate School of Information Management & Security, Korea University 教授 Hyoung Joong Kim 進行第二場 Keynote Speech － Data Compression by Data Hiding。今天的議程也是相當的緊湊，除了早上安排的 Keynote speech 之外，尚安排有 5 個 Session 來進行 oral presentation。同樣地，我也是選擇一個自己比較感興趣的 Session B02 - Intelligent Image and Signal Processing 來參加。至於下午則安排有 10 個 Session，由於部分 Session 時間重疊，我選擇以下兩個

Session 來參加，分別為：Session B10 - Image Processing and VLSI Implementation 與 Session B11 - Intelligent Watermarking Techniques, Image Authentication and Visual Cryptography (II)。

**98 年 09 月 14 日**

這是 IIH-MSP2009 會議的第三天，也是大會會議的最後一天。大會首先安排中國浙江大學教授潭建榮(Jianrong Tan)進行專題演講，講題為：Multimodal Information Fusion in the Virtual Environment and Its Applications in Product Design。今天一整天都安排 Presentation Sessions and Poster Sessions，分別在早上安排 4 個 Sessions，下午則安排有 4 個 Sessions 與 2 個 Poster Sessions 等。早上參加的 Sessions 分別為：Session C02 － Intelligent Video Processing。下午參加的 Sessions 有：Session C06 - Multi-dimensional Signal Processing, Modeling and Visualization，與 Session C07 - Multimedia Signal Processing for Plasma Diagnostics。同時早上的 Session C02 也是本人論文報告的 Session，本篇論文報告期間，與會人員發問相當踴躍、討論氣氛熱烈，由於時間限制，大家只得於會後再行討論，彼此交換研究心得，收獲頗多。

二、與會心得

此次參加 IIH-MSP2009 會議，認識了許多世界各地同樣為 Intelligent Information Hiding 與 Multimedia Signal Processing 等領域的研究人員。本次與會的學者遍佈相當廣泛，在亞洲方面包括有日本、台灣、南韓、中國、新加波與伊朗等，在歐洲方面則有德國、捷克、荷蘭、挪威、義大利與波蘭等，此外還有埃及與俄羅斯等國家的教授與博士生參與。在會議期間，當然也有接觸到中國當地的與會人士，彼此也交換了許多在研究上的意見與想法，相信對於未來的研究會有更進一步的幫助。

除此之外，此次的最大收穫便是聽取了許多非常深入且生動的演講，其中讓我印象最為深刻的是浙江大學潭建榮教授之專題演講 - Multimodal Information Fusion in the Virtual Environment and Its Applications in Product Design。本次Talk主要是講述如何利用3D模型來建構一個更真實的虛擬實境(virtual reality)。除此之外，潭教授亦講述如何利用這個技術來模擬一個汽車引擎的組裝過程，整個模擬過程相當真實。

另外一個較大的收獲是，本次有許多相關領域的台灣學者參加 IIHMSP2009 會議，包括有雲科大張傳育教授、伍麗樵教授、虎尾科大郭文忠教授、澎湖科大胡武誌教授

、高雄大學張保榮教授、高應大陳聰毅教授、王鴻猷教授，中山大學謝欽旭教授、逢甲大學林志敏教授、台南大學李建樹教授、彰師大王春清教授、崑山科大盧春林教授等人，大家平時難得有機會可以聚在一起針對個人研究交換心得，透過這次的交流在研究領域的啟發與拓展真是獲益良多。

參與本次國際研討會最大的收獲是在會議進行過程，除了可以聆聽來自世界各國最新的研究成果外，更可以利用休息的時間與相關學者專家進行聯誼交流。此外經由這次會議，認識了許多學者專家，對於國家及個人在國際能見度的提升確實有顯著的幫助。

三、考察參觀活動(無是項活動者省略)

(略)

四、建議

1. 國科會應多鼓勵年輕學者出席國際會議，給予較多的出國補助。
2. 若國內有多人同時參加一個會議。可組團前往，一方面節省旅費，一方面可相互照應。

五、攜回資料名稱及內容

下列為攜帶回國之資料，內容皆為此次研討會議所發表之論文。

IIHMSP2009 論文光碟一片與議程手冊一本

六、其他

(略)