行政院國家科學委員會專題研究計畫 成果報告

嬰幼兒特殊表情動作影像分析擷取
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 98-2221-E-216-029-
執 行 期 間 ：98 年 08 月 01 日至 99 年 10 月 31 日
執 行 單 位 ：中華大學資訊工程研究所

計 畫 主 持 人：黃雅軒

計畫參與人員：學士級-專任助理人員：黎竹芸
碩士班研究生-兼任助理人員：彭國達
碩士班研究生-兼任助理人員：陳禹仲
碩士班研究生-兼任助理人員：歐志鴻
碩士班研究生-兼任助理人員：王勻駿

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 ：本計畫涉及專利或其他智慧財產權，2 年後可公開查詢

中 華 民 國 100 年 01 月 12 日

1. 前言

父母希望能隨時紀錄小寶貝成長過程中點點滴滴，同時也希望能捕捉記錄小朋友除了一般正常攝影外，更多不同角度的可愛模樣與特殊情境，包括：不預期的、已知的，但具不同視角的各類喜怒哀樂珍貴又驚喜的真實紀錄。若攝影設備能在自然環境下，辨識嬰幼兒的一些特殊表情動作，並可自動取景攝影並挑選值得保留的相片，則能讓父母從此不會再有「來不及」捕捉小寶貝可愛模樣的失落感。本計劃應此需求，進行相關核心技術的研發。

2. 研究目的

本計畫將藉由研發電腦視覺相關的新技術，使得家中嬰幼兒生長過程的點點滴滴，包括不預期的、已知的但具不同視角的各類喜怒哀樂、珍貴又驚喜的真實紀錄等，均能被有效的捕捉和記錄。主要的研發項目包含 (1)嬰幼兒臉部表情影像資料庫建立，(2)嬰幼兒臉部影像偵測，(3)嬰幼兒臉部影像光線補償，(4)嬰幼兒臉部特徵點抽取，和(5)嬰幼兒臉部表情辨識。

3. 文獻探討

表情辨識不僅具有學術研究價值，亦具有高度的商業價值，其廣泛性的應用，在生活中亦可隨處可見，如應用在醫院裡，可以檢視病人的狀況，當病人痛苦的時候，表情辨識系統可以判斷患者的表情，並立即發出警訊，防止意外發生；應用在嬰幼兒的玩具上時，可以自動偵測嬰幼兒的表情，當嬰幼兒大笑或哭鬧時，可以自動拍照，或發出訊息通知父母有異常情況，讓父母親能注意他(她)們。表情辨識系統應用在日常生活中的例子，可以說不勝枚舉。從另一個觀點來看，人跟人之間溝通，表情也佔據一個很重要的因素，從對方的表情變化中，我們可以觀察出對方的心情狀況，並可視情況來做回應；在人機介面上，如果能讓電腦辨識出人類的表情變化，並且適時地做出回應，那將使現今一般人對電腦冷冰冰的印象改觀。為了能使電腦認識人類的表情變化，已有許多學者投入大量的心力在電腦視覺研究上，並致力於能讓電腦理解人類的表情。在近幾年來，表情辨識已經成為一個熱門的課題，因此有許多學者提出各種方法，來對表情進行辨識。表情辨識方法主要可以分成兩種， 其中一種是使用Facial Action CodingSystem(FACS)[1][2][3] ，另一種則是以特徵為基礎(Feature-based) [24][5][6][7][8][9][10]的辨識方法。

FACS是Ekman和Friesen於1978 年所提出用於人臉表情描述的編碼系統，在這套系統中，會依據人臉肌肉的分佈，以及一些肌肉群的運動狀況，定義出動作單元(Action Units)，每個動作單元表示臉部上特定區域的移動狀況，如眉毛上升、嘴角上揚等，共定義了44 種動作單元(如圖一所示)，透過動作單元的組合，來進行表情判斷。Tian[2]等人發展出一套自動臉部分析系統(Automatic Face Analysis，AFA)，能依照人臉上永久或暫時性的特徵，對人臉正面影像序列進行分析，辨識出每種單獨的動作單元。Donato[3]等人發現使用GaborWavelet來進行特徵擷取，再進行上半部和下半部人臉的FAUs分類，比傳統的幾何方法可以達到更好的效果。

除了基於動作單元的表情辨識方法之外，也有基於紋理特徵等方法的表情辨識研究。Bartlett 和Littlewort[4]等人將輸入的影像序列，偵測出正面的人臉位置，並經過Gabor Wavelet 擷取出紋理特徵，最後再使用一連串的SVM 分類器來分類出7 種不同的主要表情(包含自然、生氣、猶豫、恐懼、快樂、悲傷和驚訝)。Ma 和Khorasni[5]則使用離散餘弦轉換(Discrete Cosine Transform)對整張影像進行特徵偵測和抽取，並使用前饋式類神經網路(Feedforward NeuralNetworks)來進行辨識。Dubuisson 和Davoine[6]等人則先利用主成分分析法(Principal Component Analysis)和

LDA(Linear Discriminant Analysis)進行前處理將影像降低維度後，再進行辨識。有些基於2D 或3D(Model Template-Based)的表情辨識方法[7][8][9][10]，這些方法計算於3D 模型中，特徵點的幾何變化或是對應於2D 的紋理特徵變化，最後再經過辨識器，進行表情辨識。

參考文獻

1. P. Ekman and W.V Freisen, "The facial action coding system: a technique for the measurement of facial movement", San Francisco: Consulting Psychologists Press, 1978.
2. Y-L. Tian, T. Kanade, J.F. Cohn, "Recognition action units for facial expression analysis", IEEE Trans. Pattern Anal. Mach. Intell. 23(2) (2001) 87-115.
3. G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, T.J. Sejnowski, "Classifying facial actions", IEEE Trans.. Pattern Anal. Mach. Intell. 21(10)(1999) 974-985.
4. M.S. Bartlett, G. Littlewort,I. Fasel, and J.R. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction," in Proc. Conf. Computer Vision and Pattern Recognition Workshop, Madison, WI, Jun. 16-22, 2003, vol. 5, pp. 53-58.
5. L. Ma and K. Khorasani, "Facial expression recognition using constructive feedforward neural networks," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 34, no. 3, pp. 1588–1595, Jun. 2004.
6. S. Dubuisson, F. Davoine, and M. Masson, "A solution for facial expression representation and recognition," Signal Process.: Image Commun., vol. 17, no. 9, pp. 657–673, Oct. 2002.
7. I. A. Essa and A. P. Pentland, "Facial expression recognition using a dynamic model and motion energy," presented at the Int. Conf. Computer Vision, Cambrdige, MA, Jun. 20–23, 1995.
8. M. Pantic and L. J. M. Rothkrantz, "Expert system for automatic analysis of facial expressions," Image Vis. Comput., vol. 18, no. 11, pp. 881–905, Aug. 2000.
9. Irfan A. Essa, "Coding, analysis, interpretation, and recognition of facial expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp.757–763, Jul. 1997.
10. M. S. Bartlett, G. Littlewort, B. Braathen, T. J. Sejnowski, and J. R. Movellan, "An approach to automatic analysis of spontaneous facial expressions," presented at the 5th IEEE Int. Conf. Automatic Face and Gesture Recognition, Washington, DC, 2002.

4. 研究方法

由於現今公開的人臉資料庫中，並沒有專門針對嬰幼兒的部分，特別是嬰幼兒的表情影像，因此本計劃自行建立了一套以嬰幼兒為主的臉部表情影像資料庫。表情的種類分成六個部分，分別為自然表情(Natural)、哭(Cry)、笑(Smile)、生氣(Angry)、驚訝(Surprise)和討厭(Disgust)。建立用的攝影裝置採用微軟網路攝影機 VX-6000，配合使用自行編寫的應用程式，拍攝環境以室內為主，但不限於同一地點；被拍攝者與攝影機的距離約為 60 公分至 120 公分；影像存檔以 JPEG 格式為主，影像大小為 640*480。嬰幼兒臉部表情影像資料庫共有 47 人，其中 25 個男生、22 女生；每個人的表情影像張數不全相同。經過整理，共有 9259 張嬰幼兒表情影像，詳細數據如下表格一，而圖一顯示一些嬰幼兒臉部不同表情的樣本。

	人數	張數
自然表情(Natural)	44	3037
生氣(Angry)	21	1349
哭(Cry)	21	1216

笑(Smile)	30	1816
驚訝(Surprise)	18	1105
討厭(Disgust)	12	736

表格一、整理後的資料庫數據

圖一、嬰幼兒臉部表情樣本

接下來介紹本計畫所開發的主要技術成果。

(1) 嬰幼兒臉部偵測

嬰幼兒臉部偵測是嬰幼兒表情辨識的首要環節，其處理的問題是確認輸入影像中是否存在嬰幼兒臉部影像，如果存在，則對其進行定位。嬰幼兒臉部偵測的正確性將大幅影響之後嬰幼兒表情辨識的結果，故在此本計畫選擇使用 Adaboosting 學習演算法，特徵則使用海爾(Haar)特徵，利用其對於邊緣、線段較為敏感的特性，來提升嬰幼兒臉部偵測的正確性及可靠度。

在訓練嬰幼兒臉行偵測器時，以 3800 張的嬰幼兒臉部影像作為正向樣本與 2048 張不包含嬰幼兒臉部影像的圖片作為負向樣本來訓練 Adaboost。測試設備為個人電腦，CPU 是 Intel Core2 Q9400 2.66GHz，記憶體是 3GB，軟體平台為 Windows XP。第一組測試樣本共有 3227 張嬰幼兒臉部影像，實驗數據如表二所示。

表二、嬰幼兒臉部影像偵測實驗數據

File Name	Hits	Missed	Total Time(Second)
嬰幼兒臉部影像	3094	133	203.255

第二組測試影像，依照表情分類，每種表情各 300 張樣本，實驗數據如表三所示。

表三、各表情臉部影像偵測實驗數據

File Name	Hits	Missed	Total Time(Second)	Recognition rate
angry	232	68	18.752	77.3%
cry	247	53	18.042	82.3%
hate	267	33	18.734	89%
natural	288	12	18.172	96%
smile	274	26	19.081	91.3%
surprise	239	61	19.309	79.6%

(2) 嬰幼兒臉部影像光線補償

我們所提出光線補償的做法主要包含三個步驟：(1)同態濾波(Homomorphic Filtering)、(2)非等方向平滑化(Anisotropic Smoothing)和(3)統計資料正規化。此技術使得在不同光線環境中均能產生光線分布較均勻的嬰幼兒人臉影像，對後續的分析幫助很大。相關技術已發表在 the 17th International Conference on Multimedia Modeling (MMM-2011)，論文題目為 An Effective Illumination Compensation Method for Face Recognition。另外，此技術也已分別向中華民國和美國提出專利申請。

(3) 嬰幼兒臉部特徵點抽取

本計畫以動態形狀模型(Active Shape Models)來做臉部特徵點的定位，它包含了眼睛、眉毛、鼻子、嘴巴以及臉部輪廓等特徵點，總共包含 50 個特徵點點。我們也提出多種改良方法，特別是以加入物件偵測技術來先找出較具鑑別性的角點特徵點方式，可大幅提升定為準確度。相關技術已發表在 CVGIP2010，論文題目為 An Improved ASM-based Facial Feature Locating Method。圖二為一般的角點特徵點偵測結果。

圖二、一般的角點特徵點偵測結果

(4) 嬰幼兒臉部表情辨識

當前大部分的機器學習方法，皆是依靠樣本的學習來建立樣本模型，若輸入資料與樣本模型差異過大時，往往無法正確的辨識物件，因為此類方法欠缺對輸入資料的歸納性與根據過去的學習來幫助學習新物件的能力。所為輸入資料的歸納性是定義為對全新樣本歸類的能力，因此歸納性越佳的演算法，對未知的樣本，越能得到正確的分類結果。基於上述的原因，於去年度的計畫中，選擇了 HTM 演算法作為表情辨識的核心演算法，著眼於其優越的資料歸納性；HTM 能借由時間的資訊進行非監督式學習來歸納訓練資料，再以分群運算來產生時序相近的

特徵，建立可描述事物中具有不變性特徵的樣本模型；並且藉由階層式時序記憶學習網路，將過往訓練所得到的結果，在新的樣本加入學習時，一起做為學習網路的輸入資料進行歸納與學習，讓過去的學習經驗夠幫助網路來學習新的事物。

本計畫已將訓練用的連續表情資料透過 ASM 臉部特徵點抽取，將其具有表情變化的主要特徵區域〈即雙眼、鼻子與嘴唇〉轉換成感測向量來訓練 HTM 模型。而 ASM 特徵區域與感測向量的轉換可用下圖 A 來說明，根據 ASM 所偵測到的雙眼位置、鼻子與嘴唇位置來裁切出四個矩形區域，依序將左眼、右眼、鼻子與嘴唇矩形內的影像灰度值並串接起來，作為感測向量來訓練 HTM 模型。

5. 實驗結果

嬰幼兒臉部表情辨識實驗，採用兩個表情資料庫來進行效能的評估，第一組資料庫為自行拍攝的嬰幼兒表情資料庫，另一組資料庫則為成人的表情資料庫 Cohn-Kanade Facial Expression Database。由於嬰幼兒的表情不似成人般，具有豐富且細膩的變化，因此在經過評估後，嬰幼兒的表情主要分成三大類，第一類是自然表情(無表情)，第二類是高興表情，最後一類是驚訝表情。一開始，我們會將嬰幼兒資料庫分成三等份，並且採用 3-load 的驗證方式，來對資料庫進行評測，每次取一等份來當測試資料，其餘兩等份當訓練資料，如此重複三次，交互驗證，最後取其平均來當做整體的辨識率。實驗結果如圖三所示。整體平均辨識率為 91.5%，其中有些情形，如表情變化幅度太小和表情不明確等(圖四)，皆是主要造成辨識率下降的主因。辨識率計算公式如下：

$$\text{正確率}(\%) = \frac{\text{辨識正確張數}}{\text{總張數}} \text{x} 100\%$$

Output / Input	Happiness	Neutral	Surprise
Happiness	87.58%	7.07%	5.35%
Neutral	4.56%	90.11%	5.33%
Surprise	6.57%	9.90%	95.85%

圖三、三類表情的 Confusion-Matrix。

		
原始表情	高興	高興
辨識表情	驚訝	無表情

圖四、左邊的高興和驚訝表情相似，因此造成誤判，右邊則為表情不明顯。

另一實驗資料庫我們採用 Cohn-Kanade 臉部表情資料庫，Cohn-Kanade 表情資料庫現今多

被研究表情辨識的學者，用來當作效能評估的表情資料庫，其中，資料庫裡共有 97 個人，485 段影像序列 ，剔除掉人眼無法直接辨識表情的序列後，最後我們採用 93 人 325 段影像序列來當作演算法的效能的評估。和嬰幼兒資料庫一樣，一開始會將資料庫分成五個部分，接下來採用 5-load 的做法，一次拿其中的一部分當測試資料，其餘當訓練資料，重複五次，最後取其平均的結果來當作最後的效能。實驗結果如圖四十八所示。Cohn-Kanade 資料庫平均辨識率為 85%，其中厭惡(Disgust)、高興(Happiness)和驚訝(Surprise)表情最容易辨識成功，因其表情變化較明顯。

Output / Input	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	71.4%	1.7%	5.1%	5.1%	13.7%	2.9%
Disgust	5.5%	91.0%	2.0%	0.0%	1.5%	0.0%
Fear	6.5%	6.5%	60.0%	11.0%	14.0%	2.0%
Happiness	2.8%	1.5%	2.4%	92.0%	1.3%	0.0%
Sadness	9.3%	0.0%	5.6%	2.3%	79.1%	3.7%
Surprise	0.0%	0.0%	1.4%	0.0%	3.0%	95.6%

圖四十八、Cohn-Kanade 資料庫之六類表情 Confusion-Matrix。

6. 結果與討論

本計畫發展完成的內容包含有(1)嬰幼兒臉部表情影像資料庫建立，(2)嬰幼兒臉部影像偵測，(3)嬰幼兒臉部影像光線補償，(4)嬰幼兒臉部特徵點抽取，(5)嬰幼兒臉部表情辨識，與(6)實驗結果。嬰幼兒臉部表情影像資料庫建立，在計畫初定之時，即開始進行建立與收集，時至今日，已完成相當數量的資料可供各項相關技術發展使用；嬰幼兒臉部影像偵測已達到可接受的正確率；嬰幼兒臉部影像光線補償已完成演算法架構的設計，且已將此模組整合至嬰幼兒臉部表情辨識系統的應用之中；嬰幼兒臉部特徵點抽取的發展進度，雖然其中遇到多次的困難，但現今已一一克服，已順利達成計畫目標。

本計畫的成果已發表四篇論文以及二件專利申請，

- “An Effective Illumination Compensation Method for Face Recognition,” MMM, 2011. (EI Index)
- “A Novel ASM-Based Two-Stage Facial Landmark Detection Method,” Pacific-Rim Conference on Multimedia, 2010. (EI Index)
- “Facial landmark detection by combining object detection and active shape model,” Third International Symposium on Electronic Commerce and Security, 2010. (EI Index)
- “Facial Expression Recognition Based on Fusing Weighted Local Directional Pattern and Local Binary Pattern,” CVGIP, 2010.
- 「影像紋理信號的萃取方法、影像識別方法與影像識別系統」和「Method and system for image extraction and identification」已分別向中華民國(申請案號：09114876)和美國(申請案號：12/835263)提出專利申請。
-

下列為二篇論文(“An Effective Illumination Compensation Method for Face Recognition” 和 ”A Novel ASM-Based Two-Stage Facial Landmark Detection Method”)的內容。

A Novel ASM-Based Two-Stage Facial Landmark Detection Method

Ting-Chia Hsu, Yea-Shuan Huang, and Fang-Hsuan Cheng

Computer Science & Information Engineering Department,
Chung-Hua University, Hsinchu, Taiwan

Abstract. The active shape model (ASM) has been successfully applied to locate facial landmarks. However, in some exaggerated facial expressions, such as surprise, laugh and provoked eyebrows, it is prone to make mistaken detection. To overcome this difficulty, we propose a two-stage facial landmark detection algorithm. In the first stage, we focus on detecting the individual salient corner-type facial landmarks by applying a commonly-used Adaboosting-based algorithm, and then further apply a global ASM to refine the positions of these landmarks iteratively. In the second stage, the individual detection results of the corner-type facial landmarks serve as the initial positions of active shape model which can be further iteratively refined by an ASM algorithm. Experimental results demonstrate that the proposed method can achieve very good performance in locating facial landmarks and it consistently and considerably outperforms the traditional ASM method.

Keywords: Active Shape Model, Facial Landmark Location.

Introduction

Facial feature extraction is a very popular research field in the recent years which is essential to various facial image analyses such as face recognition, facial expression recognition and facial animation. In general, based on different kinds of information extraction, the technology of facial feature extraction can be divided into two categories. First, local method, which is to detect local face components such as eye pupils, eye corners, mouth corners, etc. Secondly, global method, which makes use of the whole geometric structure of face components to locate the interested facial landmarks. In local method, because the feature models of facial landmarks are mutually independent, the detection result is easy to be affected by the variation of lighting and poses. In global method, because it uses a set of feature landmarks to form a global facial structure model, it usually has more ability to endure the detection error of individual landmark. Therefore, the global method generally obtains better performance in locating facial landmarks. At present, three kinds of the most commonly-used methods are deformable templates (DT) [1], active shape models (ASM) [2][3][4] and active appearance models (AAM) [5]. Both ASM and AAM are provided by Cootes, they iteratively decrease an energy function to obtain the optimized facial landmark locations.

G. Qiu et al. (Eds.): PCM 2010, Part II, LNCS 6298, pp. 526–537, 2010.

In recent years, ASM has been successfully applied to medical image analysis, such as computed tomography (CT), and it also can be applied to locating facial feature landmarks. However, the accuracy of the facial feature localization is still a problem because face images are much complex than medical images. Therefore, researchers keep on proposing new methods to improve its performance, such as Haar-wavelet ASM [6], SVMBASM [7] and ASM based on GA [8]. In general, these new methods have better accuracy than the original ASM, but they all are still prone to make mistaken detection in exaggerated facial expressions.

In this paper, we present a novel two-stage algorithm to improve the performance of facial landmark detection. The traditional method of ASM uses the average facial shape template to initialize the positions of facial landmarks, and it iteratively finds the best landmark positions only along the normal direction of edge contours. This process may contain two kinds of drawbacks. First, the average facial shape template may deviate considerably from the genuine landmark positions, therefore the landmarks are not able to be found correctly. Secondly, the genuine landmark position may not be located on the normal direction of edge contour, which will accordingly produce unsatisfactory landmark positions. Furthermore, when people have made exaggerated facial expressions, the traditional ASM often performs poorly because the shapes of exaggerated facial expressions usually are very different from the average facial shape. However, through analyzing the structure of human face compositions, we can understand the shape variation of human face mainly depends on the positions of the left/right eye inner and outer corners, the left/right inner and outer eyebrow corners and the left/right mouth corners. If these corner positions can be found correctly at the first stage, it will be able to set more approximate initial positions for the facial landmarks. Accordingly, better landmark locations can be found through ASM iteration and the accuracy of landmark detection can be much improved. From the above discussion, an improved landmark detection method is proposed which detects the corner-type landmark first, uses the detected corner-type landmarks to initialize the facial feature positions in the second stage, and then applies ASM to obtain the final landmark positions.

This paper is organized as follows. Section 2 introduces the classical ASM method and Section 3 describes the proposed two-stage ASM. Experimental results are given in Section 4, and finally, conclusions are drawn in Section 5.

2 Review of the Active Shape Model (ASM)

ASM is one of statistical models, which contains a global shape model and a lot of local feature models. Section 2.1 decides the shape model; Section 2.2 describes the local feature models and Section 2.3 describes the ASM algorithm.

2.1 The Shape Model

Suppose there are n facial feature points and each one is located at obvious face contour. The positions of these n points are arranged into a shape vector X, that is

$$X = [\, x_1, y_1, x_2, y_2, \ldots, x_k, y_k, \ldots, x_n, y_n \,]^T \tag{1}$$

where x_k and y_k are the horizontal coordinate and the vertical coordinate of the kth feature point respectively.

Using the PCA operation, the eigenvectors of the covariance matrix corresponding to main shape variations can be generated. Then, a shape model can be represented as:

$$x = \bar{X} + Pb \tag{2}$$

where $\bar{X}$ is the mean shape model, $P = [\Phi_1 \Phi_2 \ldots \Phi_t]$ is the eigenvectors corresponding to the t largest eigenvalues, and b is the shape parameter which is the projection coefficient that X projects onto P. Usually, b_i is constrained within the range of $\pm 3\sqrt{\lambda_i}$, so that a constructed face shape will not degenerate too much.

2.2 The Feature Model

In general, we suppose a landmark is located on a strong edge. According to the normal direction of a landmark, we can get m pixels on both sides of this landmark. So, for each landmark, there are in total 2m+1 gray-level values which form a gray-level profile $\mathrm{g_i} = [\mathrm{g_{i0}}, \mathrm{g_{i1}}, \ldots, \mathrm{g_{i(2m)}}]$, where i is the landmark index. In order to capture the frequency information, the first derivative of profile $\mathrm{dg_i}$ is calculated as

$$dg_i = [g_{i1} - g_{i0}, g_{i2} - g_{i1}, \ldots, g_{i(2m)} - g_{i(2m-1)}]. \tag{3}$$

In order to lessen the influence of image illumination and contrast, dg_i is normalized as

$$y_i = \frac{dg_i}{\sum_{k=0}^{2m-1} |dg_{ik}|}, \text{ where } dg_{ik} = g_{i(k+1)} - g_{ik}. \tag{4}$$

The feature vector y_i is called "grayscale profile".

2.3 ASM Algorithm

The ASM searching algorithm uses an iteration process to find the best landmarks which can be summarized as follows:

1. Initialize the shape parameters b to zero (the mean shape).
2. Generate the shape model point using the $x = \bar{X} + Pb$.
3. Find the best landmark z by using the feature model.
4. Calculate the parameters b′ as $b' = P^T(z - \bar{X})$.
5. Restrict parameter b′ to be within $\pm 3\sqrt{\lambda_i}$.

If $|b' - b|$ is less than a threshold value, then the matching process is completed; else $b = b'$, and return to step 2.

3 The Proposed Method

The traditional ASM only uses the grayscale profile as its feature model. However, the grayscale profile is inherently a one-dimensional feature which in general is too simple to represent the distinct information of a landmark point. Basically, there are

two other drawbacks of the traditional ASM. The first is that it selects the target points only from the candidates along the normal direction of edge contour. If the target point is not located in the candidate points, it will cause the found target point incorrect. The second is it uses a fixed search range for different landmark points. But, different landmarks in fact may require different search ranges because they have different variation extents.

In general, the facial feature landmarks can be attributed into two types: corner-type landmarks and edge-type landmarks. The corner-type landmarks (such as the left/right eye inner/outer corners) have very unique 2-D shapes looked like corners, but the edge-type landmarks (such as the landmarks of eyelid or mouth lip) have non-unique 1-D shapes shown as a line. Fig. 1 shows some examples of corner-type landmarks and edge-type landmarks. Obviously, the corner-type landmarks are much easier to detect than the edge-type landmarks. Therefore, in this paper we propose a novel two-stage facial landmark detection algorithm. The first stage is to locate the corner-type landmarks, and the second stage is to locate the whole facial landmarks by using the locations of the detected corner-type landmarks in the first stage as the initial positions of ASM. In this study, we define a total of 10 corner-type landmarks which are the left/right eye inner and outer corners, the left/right eyebrow inner and outer corners, and the left/right mouth corners. Another difference of our method to the traditional ASM is to define variable search ranges for different edge-type landmarks according to their variation degrees. That is if from the training data the positions of an edge-type landmark differ a lot, the search range of this landmark will be accordingly large. On the contrary, if the positions of an edge-type landmark are very stable, the corresponding search range will be relatively small. The proposed method will be introduced in the following.

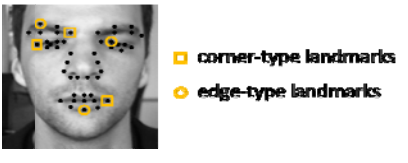


Fig. 1. Examples of the corner-type landmarks and the edge-type landmarks, where '◦' denotes corner-type landmarks and '□' denotes edge-type landmarks

3.1 The First Stage

Adaboosting algorithms have been extensively used for object detection and they often obtain outstanding detection performance. Therefore, for each corner-type landmark we used the Adaboosting algorithm [9] to construct a detector in the first stage. Samples of the 10 corner-type landmarks are shown in Fig. 2 in which the black spots indicate the corner representative positions and they are not necessary to be located at the center of the image blocks. In order to improve the issue on search range, we defined different search ranges for different corner-type landmarks in Fig. 3.

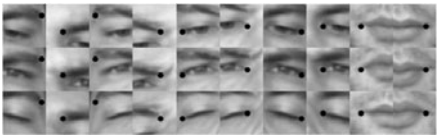


Fig. 2. Image samples of corner-type landmarks

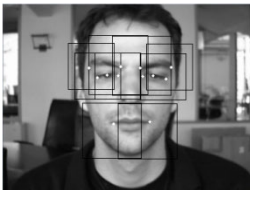


Fig. 3. The two-dimensional search ranges of 10 corner-type landmarks

With a constructed Adaboost-based detector, it may obtain several candidates of one landmark in the defined search range, and accordingly it is necessary to select the correct one among them. Because different corner-type landmarks are located at different facial geometric compositions (i.e. eyebrow, eye and mouth), their candidate selections should be designed according to their own affecting external factors (such as hair and glasses). With the above understanding, we categorized the 10 facial landmarks into three groups (eyebrow, eye and mouth) based on their functions and their geometric positions. Let e(x,y) be the edge strength of pixel (x,y) and s(x,y) be the detection score of a specific Adaboost-based corner-type landmark detector. Then, each group has its own candidate selection design as described in below.

3.1.1 Candidate Selection of Eyebrow Corners

Conceptually, an eyebrow corner should have a strong horizontal edge strength and a weak vertical edge strength. But, because the eyebrow may be covered by hair, just using the edge strength cannot get good candidate selection result. Instead, a HOG (Histogram of Oriented Gradients) [12] feature is also used to select the eyebrow corners. Therefore, in order to select the correct candidate, three factors are taken into consideration as

$$F_{eyebrow}(x,y) = \alpha \log s(x,y) + \beta \log e(x,y) + \gamma \log(\frac{1}{h(x,y)}) \quad (5)$$

where α, β and γ are three weight parameters, s is the detection score of the Adaboost-based eyebrow detector, e is the edge strength and h is the Mahalanobis distance of the HOG features between the corresponding eyebrow model and the eyebrow candidate at pixel (x,y). Among the eyebrow corner candidates, the one having the largest $F_{eyebrow}$ is the selected candidate.

3.1.2 Candidate Selection of Eye Corners

Because an eye corner and its near pupil present a rather stable distance, this property can be used to select the eye corners. Since our face detection algorithm can detect

not only face positions but also both pupil positions, both eye corners accordingly can be roughly estimated from the detected pupil positions. Among the eye corner candidates, the one closest to its estimated eye corner is selected. Fig. 4 displays an example of eye corner selection.

Fig. 4. An example of eye corner selection, where '●' denotes an eye corner candidate, and '▲' denotes the estimated eye corner

3.1.3 Candidate Selection of Mouth Corner

Because the mouth corner candidates usually are located either at the correct mouth corners or at the facial wrinkle corners with medium-large edge strengths, it will be ineffective if the edge strength is used to select the correct mouth corner candidate. However, the two kinds of candidates have very different variances. In general, a true mouth corner has a larger variance than a wrinkle corner does. With this understanding, $F_{mouth}(x, y)$ is designed to reflect the possibility that a candidate is truly a mouth corner as $F_{mouth}(x, y) = \mu \, log \, s(x, y) + \omega \, log \, v(x, y)$, where μ and ω are weight parameters, v is a variance function. Among the mouth corner candidates, the one having the largest F_{mouth} is selected to be the correct one.

Sometimes, especially when one opens his/her mouth widely or compresses his lip mightily; the largest F_{mouth} may correspond to a wrong candidate. In fact, when a mouth is widely opened, it is difficult to detect by an adaboost-based detector because the current mouth image deviates significantly from the normal mouth appearance, and sometimes even all the detected candidates do not contain the correct mouth corner. Similarly, when one compresses his/her lip mightily, the variance of a facial wrinkle corner may be larger than that of the correct mouth corner. Therefore, we further proposed a method to improve the correctness of mouth corner selection.

3.1.4 Further Improvement of Mouth Corner Selection

If the angle between the line passing two eye pupils and the line passing two mouth corner candidates is larger than a threshold, it indicates the current mouth direction is inconsistent to the current eye direction and this constitutes an abnormal face composition. Therefore, it will be very useful for us to make a certain modification so that a wrongly selected candidate can be updated to a correct one. Our experiments showed when encountering an abnormal face composition, most probably one mouth corner candidate (called 'candidate A') is correctly selected and the other one (called 'candidate B') is incorrectly selected. Therefore, we try to predict the correct position of candidate B by using the correct candidate A. Experiments showed the two eye pupils are easier to detect than the two mouth corners and they have higher detection accuracy. So we can remedy the wrongly detection mouth corners from the detected eye pupils. First, from the two selected mouth corner candidates, we need to decide which one is correct and which one is incorrect. To serve this end, we simply select

the candidate with the larger detection score as the correct one and the other one as the incorrect one. The selected correct candidate is called the "base point". From the two detected pupils, a middle separating line can be constructed which has the same distance to the two pupils. Then, from the "base point" and the middle separating line, an "anchor point" located at the other side of the middle separating line can be found. The base point and the anchor point have the same distance to the middle separating line. Then, a segment can be defined by taking the anchor point as its center, having 1/3 length of the distance between two pupils, and being along the line direction parallel to the two pupils. Within the segment, the most appropriately predicted candidate is obtained by the following design. For each candidate C, two sub-blocks can be defined, one is in the left side of C and the other is in the right side of C. For a true mouth corner candidate, one of its sub-blocks contains a large portion of lip pixels which is called "lip-attributed sub-block" (LASB), and one of its sub-blocks contains a large portion of skin pixels which is called "skin-attributed sub-block" (SASB). Basically, the most appropriately predicted candidate C satisfies two conditions. First, the intensity of the corresponding LASB is smaller than that of the corresponding SASB. Second, the intensity variance of the corresponding LASB is larger than that of the corresponding SASB. However, for each pixel candidate, it is not necessary to explicitly decide which sub-block is the LASB and which is the SASB. Instead, this can easily be decided by simply considering the physical composition of the current processing candidate. If the candidate under consideration corresponds to a left mouth corner, the left sub-block is the SASB and the right sub-block is the LASB. On the contrary, if the candidate under consideration corresponds to a right mouth corner, the left sub-block is the LASB and the right sub-block is the SASB. Therefore, for a true mouth corner candidate, it must follow

$$\begin{cases} Var(\text{LASB}) > Var(\text{SASB}) \\ Avg(\text{LASB}) < Avg(\text{SASB}) \end{cases} \tag{6}$$

Here, Var and Avg denote the intensity variance and the average intensity of a sub-block, respectively. If there are more than one candidate meet the above conditions, the one having the largest sum of Var(LASB) and Var(SASB) is selected to be the most appropriately predicted candidate.

In Fig. 5, the square point and the circle point denote the selected candidates of the left mouth corner and the right mouth corner, respectively. The triangle point denotes the found anchor point and the line segment passing the triangle point is the search range of which the most appropriately predicted candidate of the right mouth corner is decided.

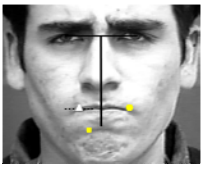


Fig. 5. An illustrative graph for further improving mouth corner selection

3.2 The Second Stage

This second stage is to detect the whole facial landmarks by using the detected locations of the corner-type landmarks from the first stage as the initial positions of the second-stage ASM model. Beside the initialized landmark positions of eyes, eyebrows and mouth, the nose landmark positions are also initialized according to a new composition of eye corners and mouth corners. The average position (S_x, S_y) of the 4 corner-type landmarks (including the left-eye inner corner, the right-eye inner corner, the left mouth corner and the right mouth corner) taken from the average face shape is computed as a reference point. The new nose landmark positions can be estimated by the following three steps:

Step 1. Compute the new average position (C_x, C_y) of the 4 corner-type landmarks obtained from the first stage;

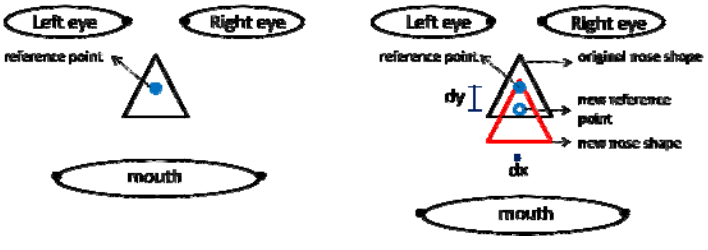
Step 2. Compute the displacement (d_x, d_y) between the reference point and the new reference point, i.e.

$$(d_x, d_y) = (C_x\text{-}S_x\,,\ C_y\text{-}S_y)$$

Step 3. Shift each nose landmark position $\bar{X}_i$ by (d_x, d_y), i.e

$$\bar{X}_i = \bar{X}_i + (dx, dy), i \in \text{landmarks of nose}$$

Fig. 6 shows the initialization method of nose landmarks, and the blue solid circle is the reference point, the blue hollow circle is the new reference point of the currently being processed face, the black triangle denotes the original nose shape, the red triangle denotes the re-initialized nose shape after the displacement of d_x and d_y.



(a) (b)

Fig. 6. Initialization of nose landmarks (a) average shape;(b)result of first stage

As for defining the appropriate search range of each edge-type landmark, the standard deviation of its position is adopted. First, the feature landmarks are divided into six groups, as shown in Fig.7. Group 1 denotes the eyebrow-related landmarks; Group 2 denotes the eye-related landmarks; Group 3 denotes both side nose-both-side-related landmarks; Group 4 denotes the bottom nose-bottom-related landmarks; Group 5 denotes the upper-lip-related landmarks; and Group 6 denotes the lower-lip-related landmarks. On purpose, all the landmarks in the same group use a same search range SR_k which is defined as:

$$sd(j) = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(x_{ij} - \bar{x}_j)^2} \tag{7}$$

$$SR_k = \max_{j \in group_k} sd(j) \tag{8}$$

where k is the group index, j is the landmark index, $\bar{x}_j$ is the average position of the jth landmark, and $sd(j)$ is the standard deviation of the jth landmark position.

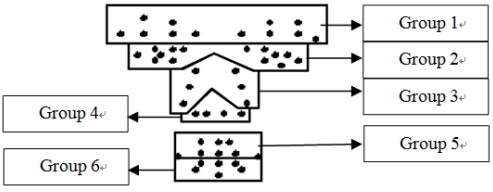
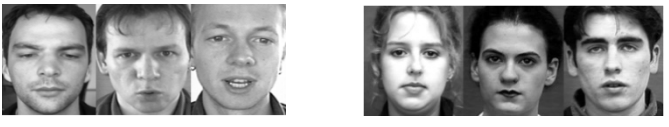


Fig. 7. A illustrative diagram of the six-groups facial landmarks

4 Experimental Results

We use the well known BioID face database and a part of the Cohn Kanade database to train the ASM shape model and the 10 corner-type Adaboost-based detectors. In total, there are 3016 BioID face images (include mirrored images) and 6588 Cohn Kanade face images used in the training stage. Because the Cohn Kanade database in total contains 9005 face images, the remaining 2417 face image is used to test the performance of landmark localization. Fig. 8 shows some samples of both databases. The 50 landmark points are manually labeled for all the images.



(a) (b)

Fig. 8. (a) Examples of the BIOID database, and (b) Examples of the Cohn Kanade database

The hit rate of each corner-types landmark is calculated and is listed in Table 1. In this paper, the hit rate of each landmark is defined as

$$hit\ rate(\%) = \frac{\sum_{i=0}^{N} f(i)}{N} * 100\% \tag{9}$$

$$f(i) = \begin{cases} 1 & \text{, if } |M_i - D_i| < 0.3 * M^w \text{ and } \dfrac{M^w}{1.5} < D_i^w < 1.5 * M^w; \\ 0 & \text{, otherwise} \end{cases} \tag{10}$$

where N is the total number of test images, M_i is the manually marked position of this landmark of the i-th image, D_i is the representative position of the detected landmark

block of the i-th image, D_i^w is the width of the detected landmark block of the i-th image, and M^w is the width of the manually marked landmark block.

Table 1. The hit rates of different corner landmarks. The index 1/2 is the left outer/inter eyebrow corner, the index 3/4 is the right outer/inter eyebrow corner, the index 5/6 is the right outer/inter eye corner, the index 7/8 is the left outer/inter eye corner, the index 9/10 is the right/left mouth corner.

Index	1	2	3	4	5	6	7	8	9	10
Hit rate(%)	97.4	94.9	93.9	99.3	96.3	96.6	98.8	98.4	94.7	98.2

For evaluating the accuracy of landmark localization, the error rate E is defined as

$$E_j = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{pt_{ij} - ans\_pt_{ij}}{dist_i} * 100\%\right) \tag{11}$$

where pt_{ij} is the detected position of the j-th landmark of the i-th image, ans_pt_{ij} is the manually marked position of the j-th landmark of the i-th image, and $dist_i$ is the distance between the two pupils of the i-th image.

The overall performance of the proposed two-stage ASM method and the traditional ASM are compared by showing their individual errors of each landmark in Fig. 9.

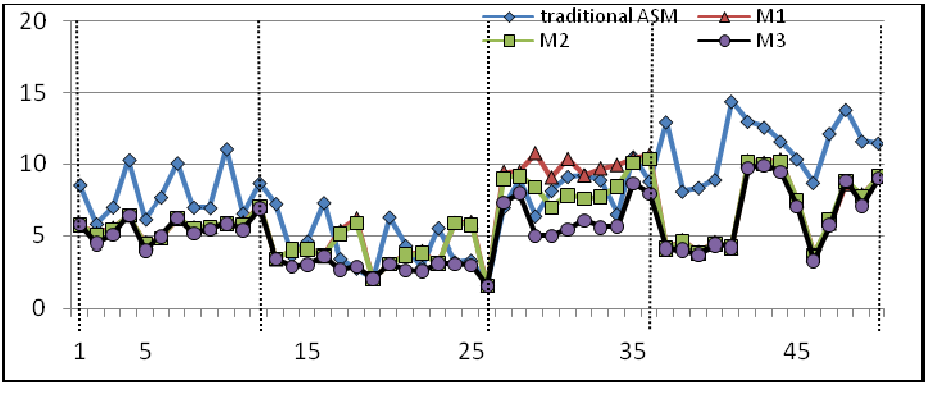


Fig. 9. The error rate of each corner landmark. The horizontal axis is the corner index in which No.1~No.12 correspond to the eyebrow-related landmarks, No.13~No.27 correspond to the eye-related landmarks, No.28~No.36 correspond to nose-related landmarks, and No.37~No.50 correspond to mouth-related landmarks. The vertical axis indicates the error rate.

There are several improvements proposed in this paper. In order to verify their effectiveness, several experiments are conducted by using different combination of the proposed methods. M1 denotes using the first stage to locate the corner-type landmarks with the Adaboost-based detectors and the traditional ASM to locate the edge-type landmarks without initializing the nose shape, M2 denotes using the first

stage 1 to locate the corner-type landmarks and the improved ASM to locate the edge-type landmarks with initializing the nose shape, and M3 denotes using the first stage to locate the corner-type landmarks and the improved ASM to locate the edge-type landmarks by using both nose shape re-initialization and different landmark-related search ranges.

Form Fig. 9 we can see the error of M1 in the nose part is larger than traditional method. Because when we initialized the eye, eyebrow and mouth without initialized the nose, sometimes it would undermine the overall facial structure. Therefore, it will cause large errors. But in M2, because we using the eye corner and mouth to initialize the nose position, the error rate in nose can be reduced.

Obviously, M3 performs much better than the traditional ASM. This reveals that both the Adaboost-based corner-type landmark detectors and the variable rectangular search ranges are very useful in detecting the corner-type landmarks of eyebrow, eye and mouth, such as the 1th, 4th, 7th, 10th, 13th, 16th, 20th, 23th, 37th and 41th landmarks in Fig. 9. When a human face has made an exaggerated facial expression, due to the two-stage ASM design, most landmarks can still be detected correctly. By using different search ranges for different landmarks can also improve the landmark localization accuracy. Although none of nose-related landmarks belongs to the corner-type landmark, they can still using the eye corner and mouth corner to improvement. Fig. 10 shows the detected positions of 50 landmarks by using the traditional and the proposed two-stage ASM methods, individually, the first row is the processed results of the traditional ASM, and the second row is the processed results of the proposed method M3. Obviously, the proposed method M3 gets much better results than the traditional ASM.

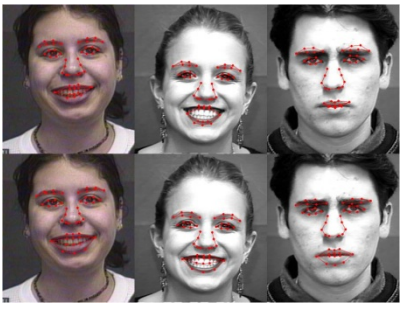


Fig. 10. Some results on the Cohn Kanade database. The top row shows the detected landmarks by the traditional ASM method and the bottom row shows the detected results by the proposed ASM method.

5 Conclusion

In this paper, we have proposed a two-stage ASM method to improve the facial landmark detection. The first stage uses an Adaboosting algorithm to locate 10 corner-type landmarks, which are attributed into three classes (i.e., eyebow, eye and mouth) and each class has its own candidate selection method from the detected candidates. The second stage is to detect the whole facial landmarks by using the

locations of the detected corner-type landmarks in the first stage as the initial positions of ASM, and different facial landmarks correspond to different search ranges based on their variation extents. From the experimental results, it demonstrates clearly that the proposed method outperforms the traditional ASM algorithm, especially in corner-type landmarks. In the future work, we will try to design a 2D feature model instead of the tradition 1D feature model for the edge-type landmarks. Expectedly, it can further improve the accuracy of localizing facial landmarks.

Acknowledgments. This research is supported by Taiwan NSC under contract NSC 98-2221-E-216-029.

References

1. Zhang, B., Ruan, Q.: Facial feature extraction using improved deformable templates. In: The 8th International Conference on Signal Process., vol. 4 (2006)
2. Coots, T.F., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
3. Zhou, D., Petrovska-Delacr'etaz, D., Dorizzi, B.: Automatic Landmark Location with a Combined Active Shape Model. In: IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems (2009)
4. Pu, B., Liang, S., Xie, Y., Yi, Z., Heng, P.-A.: Video Facial Feature Tracking with Enhanced ASM and Predicted Meanshift. In: Second International Conference on Computer Modeling and Simulation (2010)
5. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: Proc. European Conference on Computer Vision (1998)
6. Zuo, F., de With, P.H.N.: Fast facial feature extraction using a deformable shape model with Haar-wavelet based local texture attributes. In: Proceedings of IEEE Conference on ICIP (2004)
7. Du, C., Wu, Q., Yang, J., Wu, Z.: SVM based ASM for facial landmarks location. In: 8th IEEE International Conference on Computer and Information Technology, CIT 2008 (2008)
8. Wan, K.-W., Lam, K.-M., Ng, K.-C.: An accurate active shape model for facial feature extraction. Pattern Recognition Letters 26(15) (November 2005)
9. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2001)
10. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005) (2005)

An Effective Illumination Compensation Method for Face Recognition

Yea-Shuan Huang, Chu-Yung Li

CSIE Department, Chung-Hua University, Hsinchu, Taiwan

707, Sec.2, WuFu Rd., Hsinchu, Taiwan 300, R.O.C.

e-mail: yeashuan@chu.edu.tw

Abstract. Face recognition is very useful in many applications, such as safety and surveillance, intelligent robot, and computer login. The reliability and accuracy of such systems will be influenced by the variation of background illumination. Therefore, how to accomplish an effective illumination compensation method for human face image is a key technology for face recognition. Our study uses several computer vision techniques to develop an illumination compensation algorithm to processing the single channel (such as grey level or illumination intensity) face image. The proposed method mainly consists of four processing modules: (1) Homomorphic Filtering, (2) Ratio Image Generation, and (3) Anisotropic Smoothing. Experiments have shown that by applying the proposed method the human face images can be further recognized by conventional classifiers with high recognition accuracy.

Keywords—Face Recognize, Illumination Compensation, Anisotropic Smoothing, Homomorphic Filtering.

1. Introduction

In recent years, digital video signal processing is very popular because digital audio and video technology have made a lot of progress, the price of large data storage is lower and the cost of the optical photographic equipments also decreases. Most importantly, artificial intelligence and computer vision technology are getting mature. So intelligent video processing systems gain much attention to the public, especially it has become a very important role in the safety monitoring field. In this field, the accuracy of face recognition is an essential goal to pursue, so we address this issue here, and hope to be able to develop a high accuracy of face recognition.

For face recognition, there are several problems which will affect the recognition accuracy. Among them, ambient lighting variation is a very crucial problem because it will affect the system performance considerably. Currently, most face recognition methods assume that human face images are taken under uniform illumination, but in fact the background illumination is usually non-uniform and even unstable. Therefore, the face images of the same person often have very different appearances which make face recognition very difficult. Furthermore, slanted illumination probably produces different shadows on face images which may reduce the recognition rate greatly. So

this research focuses on this topic and proposes an illumination compensation method to improve the recognition accuracy under different background illumination.

There are many approaches have been proposed already, such as Retinex [1], Illumination Cone [2], Quotient Image [3], Self-Quotient Image [4], Intrinsic Illumination Subspace [5] , Columnwise linear Transformation [6], Logarithmic Total Variation model [7] , Discrete Cosine Transform [8] algorithm and Gradient faces [9] method. Retinex is an algorithm to simulate human vision which main concept is the perception of the human eye will be affected by the object reflectance spectra and the surrounding lighting source. Therefore, in order to get the ideal image it computes each pixel's albedo by subtracting the intensity of this pixel and those of its surrounding eight pixels, which results in the original Retinex algorithm, also called Single Scale Retinex, SSR. In recent years, several algorithms based on this concept but using more neighboring pixels also were proposed and they proclaimed to produce better performance than Retinex, just like Multi-Scale Retinex, MSR [10] and Multi-Scale Retinex with Color Restoration, MSRCR [10]; Illumination Cone constructs a specific three-dimensional facial model for each person, then various illuminated two-dimensional images of one person can be constructed from his own three-dimensional facial model. All of Quotient Image, Self-Quotient Image and Intrinsic Illumination Subspace adopt an image preprocessing. Quotient Image (QI) has to input at least three images under different background illumination in order to remove the information of lighting source. Self-Quotient Image (SQI) is derived from Quotient Image and it needs only one input image to perform lighting compensation. Therefore, it is easily applied to all kinds of recognition systems. Being similar to QI and SQI, Intrinsic Illumination Subspace first uses a Gaussian Smoothing Kernel to obtain the smoothed image, and then it reconstructs an image with the basis of the intrinsic illumination subspace. Columnwise linear Transformation assumes that by accumulating each column of each human face image the intensity distributions of different persons are very similar. So, the average intensity distribution A nontrivial is computed from all the training face images first, and the intensity distribution B of the current processed face image is also computed, then by transforming B to A, a compensated face image can be derived. The Logarithmic Total Variation (LTV) model is derived from the TV-L^1 model [11] and the TV-L^1 model is particularly suited for separating "large-scale" (like skin area) and "small-scale" (like eyes, mouth and nose) facial components. So the LTV model is also retain the same property. The Discrete Cosine Transform (DCT) algorithm transforms the input image from spatial domain to frequency domain first. Finally, Gradient faces method use Gaussian kernel function to transform the input image to gradient domain and get the Gradient faces to recognition.

However, these methods still have their shortcomings and deficiencies. For example, both Illumination Cone and Quotient Image require several face images of different lighting directions in order to train their database; all of Retinex, Self-Quotient Image, Intrinsic Illumination Subspace , Columnwise linear Transformation, LTV model, DCT algorithm and Gradient faces method cannot tolerate the face angle deviation and certain coverings (such as sunglasses) on faces. For the above reasons, our approach references the previous approaches to propose a novel illumination compensation method. The proposed method is based on "combination" and "complementarity" two key ideas to combine three distinct illumination compensation methods. It can efficiently eliminate the effect of background lighting change, so a subsequent recognition system can accurately identify human face images under different background illumination.

This paper is arranged into 4 sections. Section 2 describes the concept and the processing steps of the proposed compensation algorithm; Section 3 describes the testing database and experimental results; finally, conclusion is drawn in Section 4.

2. The Proposed Illumination Compensation Method

In order to eliminate the effect of background lighting, we assume that (x, y) is the coordinate of an image pixel $\boldsymbol{P}$, $f(x, y)$ is the gray value of P. So based on a Lambertian model [12], $f(x, y)$ can be expressed by the multiplication of two functions [2, 3, 12], which is

$$f(x, y) = i(x, y)r(x, y). \quad (\mathbf{1})$$

In this function, $i(x, y)$ is the illuminance of $\boldsymbol{P}$ and $r(x, y)$ is the reflectance of P. In general, the illumination values of neighboring pixels are similar to each other, so $i(x, y)$ can be regarded as one kind of low-frequency signal in an image. However, the reflectance will show the contrast arrangement of different composite materials (such as skin, eyebrows, eyes and lips, etc.) of this image. Therefore, $r(x, y)$ can be regarded as a high-frequency signal which closely corresponds to texture information of face.

Based on this understanding, our research uses the digital filtering approach to reduce the low frequency signal, and emphasize the high frequency signals of a face image at the same time. We expect to decrease the influence of background lighting on facial analysis and recognition. So the facial texture features can be strengthened to achieve the better face recognition accuracy.

The proposed illumination compensation method consists of (1) Homomorphic Filtering, (2) Ratio Image Generation, and (3) Anisotropic Smoothing, which are shown in Fig.1.

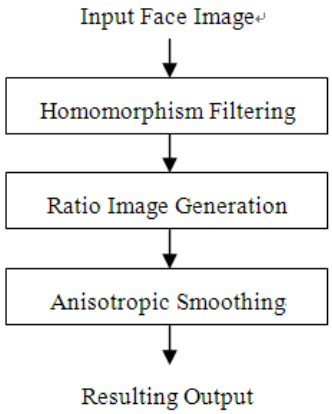


Fig. 1. The processing diagram of the proposed method.

2.1 Homomorphic Filtering

In reality, face images are influenced to many conditions and factors (such as lighting and face angle), so an original image may contain lot of noises. Therefore, we use a homomorphic filtering to adjust the image intensity by strengthening the high-frequency signal and decreasing the low-frequency signal.

First, we adopt the logarithm operation to separate the illumination and reflection coefficient from image, that is,

$$\begin{aligned} Z(x,y) &= ln\, f(x,y) \\ &= ln\, i(x,y) + ln\, r(x,y) \end{aligned} \quad (2)$$

Next, we adopt the Fourier Transform to compute the left and right sides of the above equation,

$$\begin{aligned} F\{Z(x,y)\} &= F\{ln\, i(x,y)\} + F\{ln\, r(x,y)\} \\ Z(u,v) &= F_i(u,v) + F_r(u,v) \end{aligned}$$

where $Z(u,v)$, $F_i(u,v)$ and $F_r(u,v)$ are the Fourier Transform results of $Z(x,y)$, $ln\, i(x,y)$ and $ln\, r(x,y)$ respectively. Then, we use a low-frequency filtering function H (u, v) to multiply the above formula and get

$$\begin{aligned} S(u,v) &= H(u,v)Z(u,v) \\ &= H(u,v)F_i(u,v) + H(u,v)F_r(u,v) \end{aligned} \quad (3)$$

Furthermore, we use an inverse Fourier Transform to get

$$\begin{aligned} SS(x,y) &= F^{-1}\{S(u,v)\} \\ &= F^{-1}\{H(u,v)F_i(u,v)\} + F^{-1}\{H(u,v)F_r(u,v)\} \\ &= i'(x,y) + r'(x,y) \end{aligned} \quad (4)$$

where

$$\begin{aligned} i'(x,y) &= F^{-1}\{H(u,v)F_i(u,v)\} \\ r'(x,y) &= F^{-1}\{H(u,v)F_r(u,v)\} \end{aligned}$$

Finally, we apply the exponential operation to the above formula and obtain

$$\begin{aligned} g(x,y) &= e^{SS(x,y)} \\ &= e^{\{i'(x,y)+r'(x,y)\}} \\ &= e^{i'(x,y)} e^{r'(x,y)} \\ &= i_0(x,y) r_0(x,y) \end{aligned} \quad (5)$$

After performing all of the above steps, $g(x,y)$ is the final filtered image. Because of $H(u,v)$ is a low-frequency filtering function, it will significantly reduce the intensity of low-frequency signal. So $g(x,y)$ can not only effectively preserve the high-frequency texture information, but also reduce the impact of illumination variation.

In general, $H(u,v)$ can be designed as

$$H(u,v) = (r_H - r_L)[1 - e^{-C[D^2(u,v)/{D_0}^2]}] + r_L \quad (6)$$

where $r_H > 1$, $r_L < 1$, and D_0 is a cut-off frequency. The constant c is a parameter to control the increasing degree of the exponential function. Figure 2 shows an illustrating graph of H (u, v).

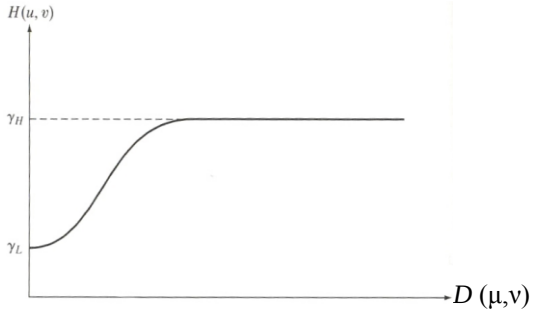


Fig. 2. A low-frequency filtering function $H(u, v)$.

The low-frequency signal not only includes the illumination information but also includes the texture information of human face image. So the r_L should be set to a small value but not zero, if we want not destroy the texture information of face image. Because of above reason, in order to remove the illumination information, we proposed the second steps: Ratio Image Generation.

2.2 Ratio Image Generation

We have used the homomorphic filter to reduce the influence of illumination, but we cannot eliminate all low-frequency signals because the low-frequency signal may also contain some facial features which are useful to recognition. So instead of setting r_L = 0 to completely eliminate the low-frequency signal, r_L is set to be 0.5. Consequently, the filtered image still contains part of illumination information. For further reducing the illumination information, a second operation called "Ratio Image Generation" is proposed to eliminate the low-frequency signal. From the experiment, it clearly shows that using both of Homomorphic Filtering and Ratio Image Generation outperform than using Homomorphic Filtering only.

Since $g(x, y)$ denotes the value of a filtered image pixel, based on a Lambertian model [8], it can also be formulated as

$$g(x, y) = r(x, y)i(x, y) \tag{7}$$

where $\mathrm{r(x, y)}$ is the albedo, and $i(x, y)s(x)$ is the illumination value of pixel (x, y). As described before, $\mathrm{r(x, y)}$ denotes the texture information of the image and $\mathrm{i(x, y)}$ denotes the low-frequency information. Let $\mathrm{W(x, y)}$ be a smoothed image information by convoluting $\mathrm{g(x, y)}$ with a Gaussian function G. That is

$$\mathrm{W(x, y) = g(x, y) * G.} \tag{8}$$

Basically, the lighting factor can be implicitly attributed to W. Because both $\mathrm{i(x, y)}$ and $\mathrm{W(x, y)}$ correspond to the low frequency signal of an image at pixel (x, y), we can use $i(x, y) \approx c \times W(x, y)$ to present the approximate relationship between both low-

frequency data and *c* is a constant value. If $g(x,y)$ is divided by $W(x,y)$, a new image $N(x,y)$ can be constructed which inherently reveals the high frequency attribute $r(x,y)$. That is

$$N(x,y) = \frac{g(x,y)}{W(x,y)} = \frac{r(x,y)\, i(x,y)}{W(x,y)} \approx cr(x,y) \tag{9}$$

where *N* can effectively reflect the intrinsic information of an image, which is called the ratio image.

2.3 Anisotropic Smoothing

While a ratio image *N* can effectively reflect the high-frequency signal of image, but it is very sensitive to noise. Therefore, we use an anisotropic smoothing operation to reduce the interference of noise. However, the general smoothing algorithms will not only reduce noise, but also undermine the image texture characteristics because they belong to high frequency signal. In order to reduce the noise effect and avoid the degeneration of normal texture information, we purposely design an anisotropic smoothing algorithm to produce the smoothed image. Here, some variables about the anisotropic smoothing operation are defined as below:

$N_{x,y}$ is the image value of pixel (*x*, *y*) in a ratio image

$$\begin{aligned} \Delta_E &= N_{x+1,y} - N_{x,y} \\ \Delta_W &= N_{x-1,y} - N_{x,y} \\ \Delta_S &= N_{x,y+1} - N_{x,y} \\ \Delta_N &= N_{x,y-1} - N_{x,y} \end{aligned} \tag{10}$$

Δ_E, Δ_W, Δ_S and Δ_N represent respectively the 4-directional image differences between pixel (*x*, *y*) and its adjacent image pixels. During the smoothing operation, a large degree of smoothing will be executed on the uniform parts of image, but a much small degree of smoothing will be executed on the boundary of image. Consequently, the smoothed image will preserve its boundary information effectively. To serve this purpose, a weighting function based on image difference is designed as

$$w_k = exp^{-\frac{\Delta_k \cdot \Delta_k}{\delta}} \quad \text{for } k \in \{E, W, S, N\} \tag{11}$$

where δ is the bandwidth parameter to control the change rate of the exponential function. Then, the smoothed image are computed by

$$g_{x,y}^{t} = g_{x,y}^{t-1} + \lambda\left(w_E g_{x+1,y} + w_W g_{x-1,y} + w_S g_{x,y+1} + w_N g_{x,y-1}\right) \tag{12}$$

where $g_{x,y}^{t}$ is the image value of pixel (*x*, *y*) after *t* times smoothing operations. Finally, in order to obtain more consistently filtered face images, a histogram equalization operation is applied to the anisotropic smoothed image.

3. Experimental Results

In order to estimate the performance of the proposed method, the present study uses two famous face databases (Banca [13] and Yale database B [14]) to evaluate the recognition rate. The Banca database contains human frontal face images grabbed from several sections to reflect different variation factors. Among all sections, the section 1, 2, 3 and 4 of the "controlled" classification are used in our experiment. In each section, there are 10 images for each person, and in total there are 52 persons (26 males and 26 females), therefore it consists of 2,080 images in total. For performance comparison, we adopted three pattern matching methods (RAW, CMSM [15] and GDA [16]) to evaluate the recognition accuracy. RAW refers to the nearest-neighbor classification based on the image value in the Euclidean distance metric. CMSM (Constrained Mutual Subspace Method) constructs a class subspace for each person and makes the relation between class subspaces by projecting them onto a generalized difference subspace so that the canonical angles between subspaces are enlarged to approach to the orthogonal relation. GDA (Generalized Discriminant Analysis) adopts kernel function operator to make it easy to extend and generalize the classical Linear Discriminant Analysis to a non-linear one. Because CMSM needs to construct a mutual subspace, the images of 12 persons are selected to serve this end. Therefore, the face images of the rest 40 persons are used to test the recognition performance in this experiment. By randomly separating the 40 persons, different enrollment and unenrollment sets are constructed. An enrollment set contains the face images of the persons which have enrolled themselves to the recognition system and an unenrollment set contains the face images of the persons which have not enrolled to the system. During each random separation, there are 35 persons are selected in the enrollment set and 5 persons are in the unenrollment set. With this design, hundreds of experiments can be easily performed. Among the four sections, only the first section is used for serving the training purpose, and the other three sections are for testing. As for the Yale database B, it contains 5760 single light source images of 10 subjects each was taken pictures under 576 viewing conditions (9 poses x 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured. Hence, the total number of images is in fact 5760+90=5850. But we only test 1 pose (pose 0) of them; it means we only use 640 images to test the recognition rate. Then these 64 images are further separated into 6 sections (about 10 images per section), and only the first section is used for serving the training purpose, and the other five sections are for testing. Among 10 peoples, 5 of them are selected for enrollment, and the other 5 are for unenrollment.

The specific settings of parameters in our experiments are $r_H = 1.6$, $r_L = 0.5$, $D_0 = 15$, and $c = 10$. For CMSM, the base number is set 1000, and for GDA, the kernel sigma is set 4400 and its feature dimension is 200. Figure 3 shows some images examples of which the first row is the original images, the second row is the images after applying the homomorphic filter, the third row is the ratio images, and the fourth row is the images operated by the anisotropic smoothing algorithm which indeed are the output images of our illumination compensation method.

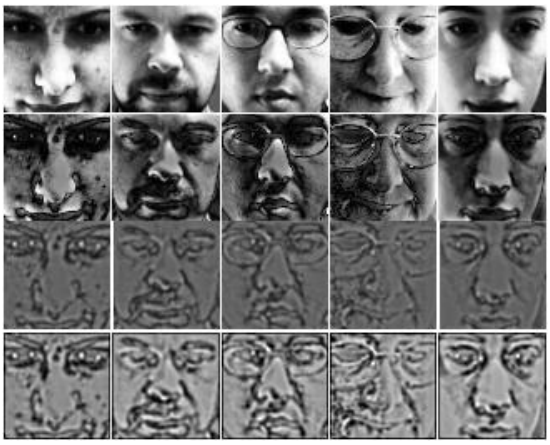


Fig. 3. Image examples of different processing steps, from the first row to the fifth row: input images, hormomorphic filtered images, ratio images, and anisotropic smoothed images.

Table 1 lists the recognition results of three different pattern matching methods, and FAR, FRR, and RR denote false accept rate, false rejection rate, and recognition rate individually. From this table, it shows all the recognition rates of the three recognition methods are larger than 90%, and the recognition rate of CMSM even is up to 95%. Thus, this experiment demonstrates that the compensated image by using the proposed approach can be further recognized by general recognition methods.

Table 1. The recognition results of three different pattern matching methods on the compensated Banca face database.

	CMSM	RAW	GDA
FAR	4.6%	6.7%	6.1%
FRR	4.8%	7.3%	6.5%
RR	95.1%	92.6%	93.4%

In addition, this study also compared the recognition rates with eight other illumination compensation methods: (1) Original, means we used original image to processing image without illumination compensation, (2) HE, means Histogram equalization method, (3) Retinex, (4) DCT means the Discrete Cosine Transform algorithm, (5) RA, means that we used ratio image generation + anisotropic smoothing, (6) HA is means homomorphic filtering + anisotropic smoothing, (7) LTV means Logarithmic Total Variation model, and (8) Gradient faces method. Besides, the recognition result of the original images is listed as a reference. Table 2 shows the experimental results to compare our algorithm with other compensated algorithms. Obviously, our method outperforms the other methods.

Table 2. Recognition results of different illumination compensation algorithms adopt Banca database.

Illumination Compensation Method	CMSM	RAW	GDA
Original	88.2%	57.6%	60.3%
Histogram equalization	88.5%	60.1%	64.3%
Retinex	81.5%	65.0%	75.4%
DCT	88.3%	85.1%	82.0%
RA	89.1%	88.3%	84.1%
HA	91.8%	85.0%	81.7%
LTV	92.3%	90.0%	90.6%
Gradient faces	92.5%	87.4%	90.7%
The proposed method	95.1%	92.6%	93.4%

Because the Banca database does not contain images with significant illumination variation, we purposely used a few human face images from the Yale Face database [16] to demonstrate the effectiveness of our illumination compensation method. Visually, from Figure 4, the original images in the first row show different appearances, but the final output images in the fourth row in fact appear quite similar to each other. Table 3 lists the recognition rates of our proposed method and the other illumination compensation algorithms with the Yale database B.

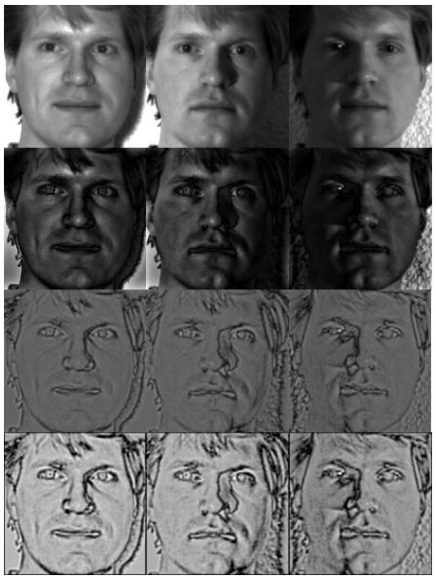


Fig. 4. Image examples from Yale faces database. The first column to the third column is respectively “central-light source image”, “left-light source image”, and “right-light source image”.

Table 3. Recognition results of different illumination compensation algorithms and different databases.

Illumination Compensation Method	CMSM	RAW	GDA
Original	90.0%	82.6%	92.2%
Histogram equalization	96.1%	91.6%	97.9%
Retinex	95.8%	87.8%	97.8%
DCT	94.0%	88.0%	100.0%
RA	93.9%	83.7%	95.9%
HA	92.1%	87.6%	97.8%
LTV	86.2%	93.0%	98.2%
Gradient faces	94.1%	93.8%	100.0%
The proposed method	97.8%	95.6%	100.0%

In Table 3, We can find the recognition rate of Yale database B can be up to 100%. It's because the Yale database B contains variation only in illumination and keeps other conditions (ex. background, pose, expression and accessory) the same. However, the recognition rate of Banca database is lower (at most 95.1%) because it basically contains more variation and more peoples than Yale database B. So we can say the recognition of Banca database is more difficult than that of Yale database B. From the above experiments, it obviously shows that our purposed method consistently performs best than the other commonly used illumination compensation methods for the Banca, and Yale B face databases.

4. Conclusion

In this paper, we propose a set of illumination compensation technique use for human face recognition. The proposed technique uses digital filtering to reduce the low-frequency signal and strengthen the high-frequency signal to reserve the facial texture information. And the proposed technique also can reduce the effect of background lighting change to increase the accuracy of face image recognition. Experiments have shown that the proposed method can achieve very promising recognition accuracy for the Banca database and Yale B faces database of each recognition method. It confirms the proposed algorithm is indeed more feasible and applicable. Actually, the proposed method is a general lighting compensation method which is not only limited in recognizing human faces. In the future, we will try to apply this method to other applications (such as OCR and Video surveillance).

Reference

1. Ching-Liang Lu, Yuan-Kai Wang and Kuo-Chin Fan, "Face Recognition under Illumination and Facial Expression Variation", CVGIP, 2009.
2. P.N. Belhumeur and D.J. Kriegman, "What is the set of images of an object under all possible lighting conditions?", In Proceedings, IEEE International Conference on Computer Vision and Pattern Recognition, pp. 52 - 58, 1997.
3. T.R. Raviv and A. Shashua, "The quotient image: Class based re-rendering and recognition with varying illuminations", In Proceedings, IEEE International Conference on Computer Vision and Pattern Recognition, pp. 566-571, 1999.
4. H.T. Wang, S.Z. Li, Y.S. Wang, "Face Recognition under Varying Lighting Conditions Using Self Quotient Image", In Proceedings, IEEE International Conference on Automatic Face and Gesture Recognition, pp.819-824, May 2004.
5. C.P. Chen and C.S. Chen, "Lighting Normalization with Generic Intrinsic Illumination Subspace for Face Recognition", IEEE Conference on Computer Vision and Pattern Recognition, October 2005.
6. Moonhwi Lee, and Cheong Hee Park, "An Efficient Image Normalization Method for Face Recognition Under Varying Illuminations", IEEE Transactions on Pattern Analysis and Machine Intelligence, 711 - 720, 1997.
7. W. Chen, M. J. Er, and S. Wu, "Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain", IEEE Transactions on Systems, Man, and Cybernetics, vol. 36, no. 2, April 2006.
8. T. Chen, W. Yin, X. Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 9, pp. 1519-1524, September 2006.
9. Taiping Zhang, Yuan Yan Tang, and Zhaowei Shang "Face Recognition Under Varying Illumination Using Gradientfaces", IEEE Transaction on image processing, vol. 18, no. 11, November 2009.
10. Yun-Ju Li, "Color Image Enhancement Using Hybrid Retinex Algorithm", Master Thesis of Department of Graphic Communications and Technology, 2004.
11. T. F. Chen, and S. Esedoglu, "Aspects of Total Variation Regularizes L^1 Function Approximation", CAM Report 04-07, Univ. of California, Los Angeles, February 2004.
12. Michael Oren and Shree K. Nayar, "Generalization of the Lambertian model and implications for machine vision," *International Journal of Computer Vision*, Volume 14 , Issue 3,pp. 227-251,1995
13. The BANCA Database – English part; http://banca.ee.surrey.ac.uk/.
14. The Yale faces database - http://cvc.yale.edu/projects/yalefaces/yalefaces.html
15. Giraudon, C., "Optimum antenna processing: a modular approach", Proc. NATO Advanced Study Inst. on Signal. Processing and Underwaler Acoustics, pp. 401-410, 1977.
16. G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach", Neural Comput., vol. 12, pp. 2385-2404, 2000.

國科會補助專題研究計畫項下出席國際學術會議心得報告

日期: 98 年 9 月 16 日

計畫編號	NSC 98-2221-E-216-029		
計畫名稱	嬰幼兒特殊表情動作影像分析擷取		
出國人員姓名	黃雅軒	服務機構及職稱	中華大學資工系　副教授
會議時間	98 年 9 月 12 日至 98 年 9 月 14 日	會議地點	日本　京都
會議名稱	International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP2009)		
發表論文題目	Face Recognition Based on Complementary Matching of Single Image and Sequential Images		

一、參加會議經過

- 此次會議有 25 個國家投稿，總共投稿篇數為 410，每篇論文由兩位評審來審稿，最後有 326 篇被接受為 Oral 和 poster 論文。
- 此會議為 multiple tracks，同一時間有 5 個不同主題的 sections 同時舉行，事先要做功課，才能有效的知道要參加的 section。其實，有時候感興趣的論文，可能在同一時段橫跨不同的 sections，則需有效的利用時間，才能得到最好的學習效果。
- 會議中有許多的時間可以彼此討論，有很好經驗交談和學習的機會，甚至談及互訪和未來合作的可能。例如，在第二天晚宴時，大阪市立大學名譽教授 Hiromitsu Hama 談到 12 月份將來台灣參加研討會，他希望有機會拜訪其他大學，我們則表示歡迎他來本校訪問，或許也可邀請他來演講。這個議題將會繼續透過 email 來規劃。
- 會議第二天當我發表論文時，有許多聽眾參加，包含本會議榮譽共同主席 Prof. Joshiaki Shirai 和議程委員會共同主席 Dr. Hitoshi Sakano。當我發表結束時，Dr. Sakano 問了二個問題，當會議結束後，我前去致意，才知道於 2002 年就與 Dr. Sakano 於 ICPR 會議見過面，當時他曾在人臉辨識技術上給我建議。想不到 7 年後我們能在本會議中重逢，雙方都很驚喜，那時會議榮譽共同主席 Prof. Joshiaki Shirai 也加入我們的談話，大家交換了一些研究上的心得，也建立起關係。
- 此會議安排三個 Invited talks，分別是

- ✧ (9/12) The state of the art of 3D video technologies – accurate 3D shape and motion reconstruction, high fidelity visualization, and efficient coding for 3D video, by Prof. Takashi Matsuyama, Kyoto university;
- ✧ (9/13) Data compression by data hiding, by Prof. Hyoung Joong Kim, Korea university;
- ✧ (9/14) Multimodal information fusion in the virtual environment and its applications in produce design, by Prof. Jianrong Tan, Zhejiang university..

- 此會議總共包含 39 sections，其中我參加的 sections 有
  - ✧ Multimedia Signal Processing for Intelligent Applications
  - ✧ Intelligent Surveillance and Pattern Recognition
  - ✧ Advances in Biometrics(I)
  - ✧ Advances in Biometrics(II)
  - ✧ Intelligent Image and Signal Processing
  - ✧ Behavior Analysis and Abnormal Event Detection
  - ✧ Statistical Image Processing and Application
  - ✧ Application of Intelligent Computing to Signal and Image Processing

二、與會心得

- International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP2009)會議包含廣泛的研究議題，都是電腦視覺領域近年來重要的研究領域，藉由與其他學者的交談，可以擴展研究者視野，刺激老師進行產學合作計畫的動機，頗適合學校老師的參與。
- 由於大阪市立大學名譽教授 Hiromitsu Hama 有意願來台參訪，今後將繼續聯絡，促成此事，或許有助於學校在國際化和國際合作等事務的推廣有幫助。

三、建議

- 本校老師多參與國際性會議，除了介紹研究成果，增加學校的知名度以外，也能快速擴展視野，建立合作管道，對未來研究和教學有很大的幫助。
- 鼓勵學校的研究生參與這種研究與應用結合的國際性會議，讓他們更了解研究的實用價值，以激發學習和研究的熱誠。

四、攜回資料名稱及內容

會議議程一本和光碟片一片

附件三

國科會補助專題研究計畫項下出席國際學術會議心得報告

日期: 98 年 9 月 29 日

計畫編號	NSC 98-2221-E-216-029		
計畫名稱	嬰幼兒特殊表情動作影像分析擷取		
出國人員姓名	黃雅軒	服務機構及職稱	中華大學資工系　副教授
會議時間	98 年 9 月 20 日至 98 年 9 月 25 日	會議地點	中國　上海
會議名稱	Pacific Rim Conference on Multimedia (PCM)		
發表論文題目	A novel ASM-based two-stage facial landmark detection method		

報告內容應包括下列各項：

一、參加會議經過

- 此次會議共有 4 天，第一天有 2 個 Tutorial Talk：
  - ✧ Histogram in tersection kernel learning for multimedia applications
  - ✧ MPEG activities for 3D video coding

  而第二天到第四天均為論文發表，其中每天都安排一個 Keynote Speech，題目分別是

  1. (9/22) A New Machine Learning Framework with Application to Image Annotation, Prof. Zhi-Hua ZHOU
  2. (9/23) The Evolution of Image Search, Prof. Kyros Kutulakos, Prof. Yong RUI;
  3. (5/24) Learning to Rank: Pushing the Frontier of Web Search, Prof. Tie-Yan LIU.
- 此會議總共包含 15 sections，大部分時段都有 poster section，口頭報告時段的主題包含四大類，分別為

  1. Multimedia Analysis and Retrieval
  2. Multimedia Systems and Applications
  3. Multimedia Compression, Communication and Optimization
  4. Multimedia Security and Right Management

- 我的論文發表於第二天下午第一時段(13:30~15:30)，題目為"A novel ASM-based two-stage facial landmark detection method"，由於上一篇論文作者沒到場發表，故我有充分時間講解論文，在

場聽眾事後表示本論文發表非常清楚和完整，充分達到研究心得交流目的。下二圖為 PCM 會議和論文發表時的照片

圖一　PCM 會場與看板

圖一　PCM 論文發表

- 此會議參加人數比預期的少，故可以有較多的時間彼此討論。由於會議專注於多媒體(特別是影像和電腦視覺)的技術和應用，而所有與會人員都是從事於此領域研究，所以有很好經驗交談和學習的機會。

二、與會心得

- 此會議共有 261 篇論文投稿，最後只選出 75 篇口頭報告和 56 篇海報論文，可見此會議是一個具有較高水準的會議。
- 然而，個人對主辦單位感到失望，整體會議並沒有經過精心的規畫和佈置，不但沒有見到醒目的會議旗幟，第二天開幕時也只見到少數著名學者出席，似乎主辦單位(復旦大學)沒有投入足夠的人力和邀請來舉辦此會議。
- 有很多大陸學生參與此會議，感覺他們很有勇氣，積極的發問，有強烈的企圖心，相信他們在此會議中會有很好的學習。
- 幾乎看不到台灣學生的參加，比起大陸學生的積極參與，個人對台灣年輕一代的研究感到憂慮，若不努力，恐怕會落後大陸越來越多。

三、建議

- 本校老師多參與國際性會議，除了介紹研究成果，增加學校的知名度以外，也能快速擴展視野，建立合作管道，對未來研究和教學有很大的幫助。
- 鼓勵學校的研究生參與這種研究與應用結合的國際性會議，讓他們更了解研究的實用價值，以激發學習和研究的熱誠。

四、攜回資料名稱及內容

會議論文集上、下二冊

國科會補助計畫衍生研發成果推廣資料表

日期:2010/12/31

國科會補助計畫	計畫名稱: 嬰幼兒特殊表情動作影像分析擷取
	計畫主持人: 黃雅軒
	計畫編號: 98-2221-E-216-029- 學門領域: 圖形辨識

無研發成果推廣資料

98 年度專題研究計畫研究成果彙整表

計畫主持人：黃雅軒	計畫編號：98-2221-E-216-029-
計畫名稱：嬰幼兒特殊表情動作影像分析擷取	

成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	1	0	100%		挣 ’ Facial Expression Recognition Based on Fusing Weighted Local Directional Pattern and Local Binary Pattern,’ CVGIP, 2010.
		專書	0	0	100%		
	專利	申請中件數	1	0	100%	件	「影像識別方法與影像識別系統」申請中華民國專利(申請案號：09114876)
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	3	3	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	1	1	100%		
國外	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		

		研討會論文	3	0	100%		1.’An Effective Illumination Compensation Method for Face Recognition,’ MMM, 2011. (EI Index) 2.’A Novel ASM-Based Two-Stage Facial Landmark Detection Method,’ PCM, 2010. (EI Index) 3.’Facial landmark detection by combining object detection and active shape model,’ ISECS, 2010. (EI Index)
		專書	0	0	100%	章/本	
	專利	申請中件數	1	0	100%	件	「Method and system for image extraction and identification」申請美國專利(申請案號：12/835263)
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)		無					

	成果項目	量化	名稱或內容性質簡述
科	測驗工具(含質性與量性)	0	

教處計畫加填項目	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

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（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、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。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：■已發表 □未發表之文稿 □撰寫中 □無

專利：□已獲得 ■申請中 □無

技轉：□已技轉 □洽談中 ■無

其他：(以 100 字為限)

我們所提出光線補償和臉部特徵點偵測技術已在國際會議中發表。另外，光線補償技術也已分別向中華民國(申請案號：09114876)和美國(申請案號：12/835263)提出專利申請。

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）(以 500 字為限)

本計畫藉由研發電腦視覺相關的新技術，使得家中嬰幼兒生長過程的點點滴滴，包括不預期的、已知的但具不同視角的各類喜怒哀樂、珍貴又驚喜的真實紀錄等，均能被有效的捕捉和記錄。主要的研發項目包含 (1)嬰幼兒臉部表情影像資料庫建立，(2)嬰幼兒臉部影像偵測，(3)嬰幼兒臉部影像光線補償，(4)嬰幼兒臉部特徵點抽取，和(5)嬰幼兒臉部表情辨識。本計畫在智權方面，已發表了四篇論文和二件專利申請，也正與廠商洽談技術移轉合約。