

# 多人臉影像視覺辨識技術

吳明芳\* 李振興\* 王炳聰\* 詹慧珊\*\* 黃建邦\*\*\*

\*崑山科技大學電機工程系 副教授

\*\*國立台南護專化妝品應用科 副教授

\*\*\*崑山科技大學電機工程系 研究生

## 摘要

隨著科技的發展，舉凡門禁管制系統、金融管理、犯罪偵查與電腦認證等等，都是需要借助一套強而有力的身份確認系統。目前有關人臉辨識或偵測的研究已有許多，但多為單人辨識技術或多人人臉偵測方法。至於多人人臉的辨識技術仍有許多發展的空間，所以本研究目的在發展結合單人人臉辨識技術及多人人臉偵測方法的多人人臉辨識系統，此類系統特別有利於犯罪的偵查。

做法上，本研究實際應用臉部特徵點處理的技術於多人人臉場合的影像中，發展以臉部特徵為主的多人人臉身份確認系統。此系統利用各種不同的基本影像處理方法自動找出特徵點的座標後，計算出其所對應的特徵向量組合。在比對方面，本研究利用歐幾里德距離和漢明距離之差異值計算法來比對特徵向量間的相似程度，並考慮人臉偏轉情形進行修正，以求較高的辨識率。完成理論的推演後，本論文最後以實際單人及多人合照影像進行實驗，以驗證本研究所提之身份確認系統及改善辨識率方法的實際功效。

**關鍵詞：**視覺技術、特徵、多人臉、辨識

## 壹、緒論

所謂”影像”乃泛指所有實際存在含有某種消息的信號，如照相機拍得的相片，此相片含有人、事、物等影像，而數位影像處理乃是以數位方式處理這些信號並進而分析此信號，以得到所需的資訊[1]。影像視覺技術是近幾年來被廣泛討論以及研究的一門領域，因為人們可以由一張影像，利用許多特徵性質的確認，則可辨識出整個影像以及所需要的影像資訊。使用彩色影像處理的兩個主要因素[2]。第一、彩色是一種有效的描述子，常可簡化目標辨識和從景物中抽取。第二、人類可以分辨幾千種彩色的色度和強度。

人臉辨識是近幾年來興起的一項研究熱潮，它是利用個人臉部特性來取得身份證明的一種技術，即使用臉部特徵點做為識別碼以供記錄，並用來辨識身份。使用人臉辨識方法作為辨認的技術可廣泛地應用在電腦登錄認證、門禁控制、犯罪偵查、以及各種通



行的控制。雖然臉部辨識在各個領域都有非常具潛力的應用，但是由電腦來實現人臉辨識仍然是一件不易的工作。其主要原因在於實際場合中，臉部取像時人臉並不受強制性的固定，這使得取像後的人臉大小、角度、光線及背景等均存在極大的變異性，因而造成了臉部特徵的不穩定性因素，增加了辨識上的困難。所以本研究使用人臉大小正規化，與人臉偏轉角度修正等方法來提高系統的辨識率。目前一般研究多以單人人臉辨識或多人人臉追蹤為主，然而在許多應用場合中，受檢測影像常具多人臉孔（例如車站出入口之多人影像的通緝犯偵測），要從多人人臉的影像中辨識各個人臉身份的技術仍有很大研究空間。本研究為了能實現多人人臉的辨識目的，以無仰視角效應的黃種人臉正面影像為對象，將單人人臉辨識及多人人臉追蹤兩項技術結合為一，以發展多人人臉辨識系統，可更有利於多種場合下的使用。

本研究主要採用人臉特徵以作為辨識，過去二十幾年中，有許多研究人員投入這個領域，也獲得相當豐碩的成果。在以特徵向量為基礎的人臉辨識，早期是利用樣本匹配(template matching)[3-4]；Turk 與 Pentland [5]則利用特徵臉(eigenface)成功的辨識出人臉；Pentland et al. [6-7] 利用主要分量分析法(principal component analysis, PCA)，使用較低維度的特徵向量去描述人臉，且設計了一個距離函數(distance from feature space, DFFS)，作為原始輸入影像與重建後影像間差距的評估。在以幾何(geometry-base)為基礎的人臉辨識上，部份研究者利用人臉上的五官，例如：眼睛、鼻子、嘴巴、下巴及臉形等特徵間的幾何關係，或彼此相對位置實現人臉辨識的工作[8-10]，而最早使用人臉的幾何特徵做辨識的是在 1973 年的 Kanade[11]。此外，1983 年後有些研究者運用類神經網路[12-14]的技術在影像中找尋人臉特徵，文獻[12]的做法是一個典型的例子，該作者以一種連結的模組來表示臉部，其方法是使用金字塔的結構來表示影像的資料。金字塔中的每一層都是用網路來表示其內容，包括一個輸入、一個隱藏、及一個輸出層。訓練的資料集包括 6 個手繪不同表情的臉部。通常單純使用 RGB 色彩空間進行偵測易受光線影響，所以有所多研究將其轉換至其他對光線依賴性較低的色彩空間(如 CIE、HUV)等[15-16]。特徵偵測方面也有許多使用了不同方法的研究，如：運用特徵與周圍顏色的差異性，以進行特徵搜尋[17-19]；以形狀作比對的特徵物體偵測[20]；運用紅外線照射人眼瞳孔的反射現象，來做眼睛特徵判別[21]等。目前人臉追蹤或辨識的應用被廣泛研究，許多大學與研究機構，也都有相關的論文與技術報告的發表，如：將人臉追蹤應用於監控系統的研究[22]；使用進化演繹法、橢圓形狀來追蹤人臉並使用人臉特徵配合類神經網路以辨識臉部表情的研究[23]；應用小波理論於人臉辨識的研究[24]；多人人臉追蹤方面的論文[25]等等。

綜觀上述學術研究可以發現，目前人臉辨識在技術上的缺點，主要有資料庫佔用過多空間，以及臉部易受光線或陰影等因素影響特徵擷取的正确性。因此，本研究將 RGB 色彩空間轉換成不受光線強弱影響的 HSV 色彩空間，依照膚色色度分佈集中的特性，來偵測出影像中人臉的位置，以減少受光線或陰影所造成的影響。在特徵擷取方面，只需基本的影像處理計算，即能求出代表整張臉部的特徵向量。當系統進行辨識時，僅需比對特徵向量間的差異值，即可判斷出身份，因此比對的資料庫不用儲存影像檔，改善了資料庫空間佔用過多的問題，同時簡單的計算量也增加了比對的速度。



## 貳、人臉區域偵測

為了從多人人臉影像中，去尋找所有可能的人臉候選區域(Face Candidate Region)，本研究運用膚色偵測、影像處理和幾何規則判定等方法，以讓人臉特徵擷取的程序能成功順利的進行，同時也會對所找到的人臉候選區域進行角度轉正的動作，以便後續演算法的進行。整體流程運作如圖 2.1 所示，亦即輸入人臉影像後，系統依流程圖的各處理程序自動找出可能的人臉候選區域，並經由人臉幾何規則的判定程序，找出真正的人臉區域。

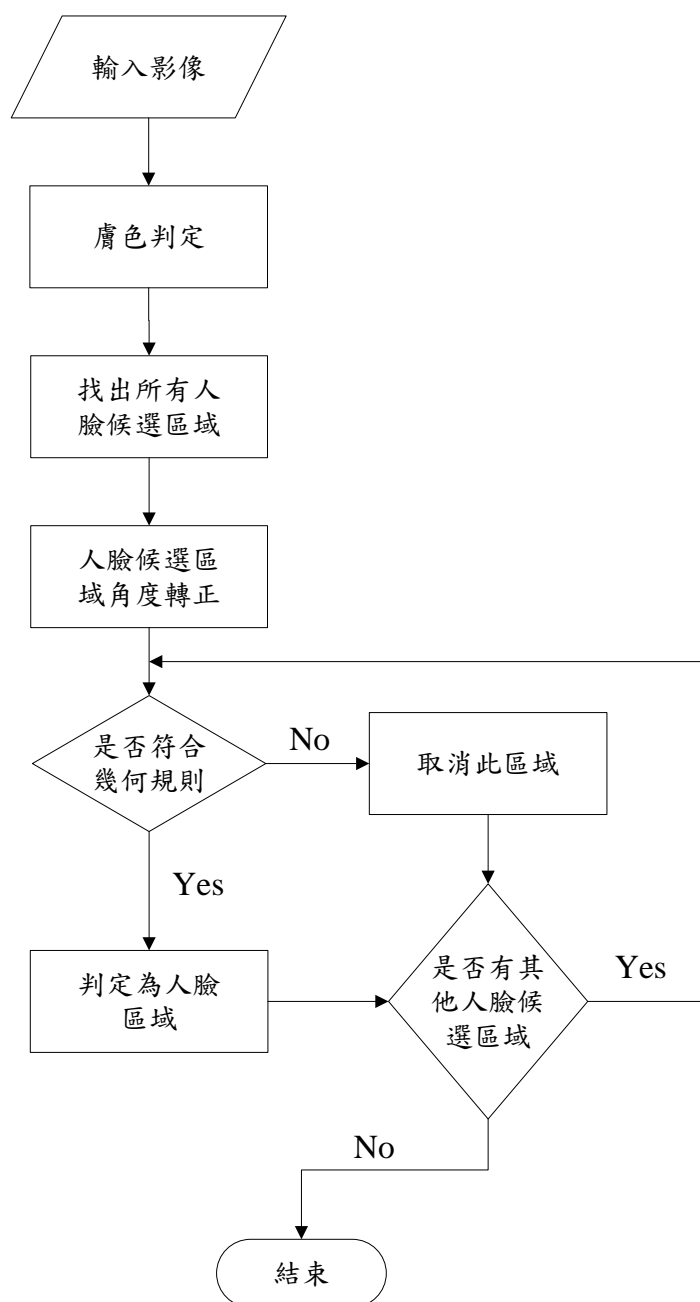


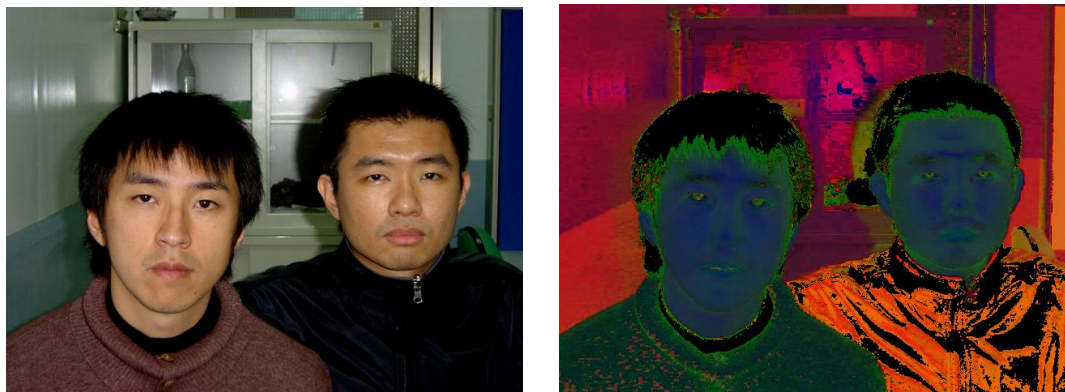
圖 2.1 臉部區域判別流程



在膚色判定的流程中，若給予如圖 2-2(a)所示之輸入影像，其中包含了許多複雜的背景區域和人臉區域，雖然人類的皮膚顏色和背景顏色通常有一定程度的差異，但如果使用 RGB 色彩空間，那麼將會不易從影像中將皮膚顏色與背景顏色分離，因為 RGB 色彩空間對於影像光線的敏感度太大，所以不適合用來描述膚色的分佈範圍。因此本論文以黃種人的膚色特質為基礎，選擇使用可以將亮度和色度分離的 HSV 色彩空間，將輸入影像的原始 RGB 色彩空間轉換為 HSV 色彩空間後，便可以比較容易將皮膚顏色與背景顏色分離出來。兩者的轉換公式如下所示[27]：

$$\begin{aligned}
 H1 &= \cos^{-1} \left\{ \frac{0.5[(R-G)+(R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right\} \\
 H &= \begin{cases} H1 & \text{if } B \leq G \\ 360^\circ - H1 & \text{if } B > G \end{cases} \\
 S &= \frac{\text{Max}(R,G,B) - \text{Min}(R,G,B)}{\text{Max}(R,G,B)} ; V = \frac{\text{Max}(R,G,B)}{255}
 \end{aligned} \tag{2.1}$$

其中色相 H(Hue)是作為偵測人臉的主要依據，主要因為 H 較不容易受到光強弱的影響。彩度 S (Saturation)代表顏色中的飽和度，亦即顏色中滲入白色的程度，其值介於 0~1 之間；而 V(Value)代表顏色的明暗度，也是介於 0~1 之間。圖 2.2(a)經色彩空間轉換後的影像如圖 2.2(b)所示。



(a) RGB 色彩空間

(b) HSV 色彩空間

圖 2.2 RGB 色彩空間到 HSV 色彩空間的轉換

將 RGB 色彩空間轉換至 HSV 色彩空間後，接下来的工作就是要將皮膚顏色與背景顏色分離。膚色在 HSV 色彩空間中分佈的情形(S 和 V 座標值為原值放大 100 倍)如圖 2.3 所示，我們可以發現到膚色色分佈都集中在某一區域範圍內。因此可以利用數學式來描述出屬於膚色的範圍，以求得影像中屬於膚色的部份。



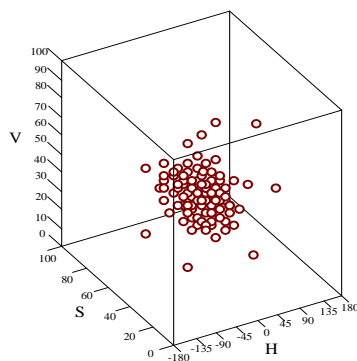
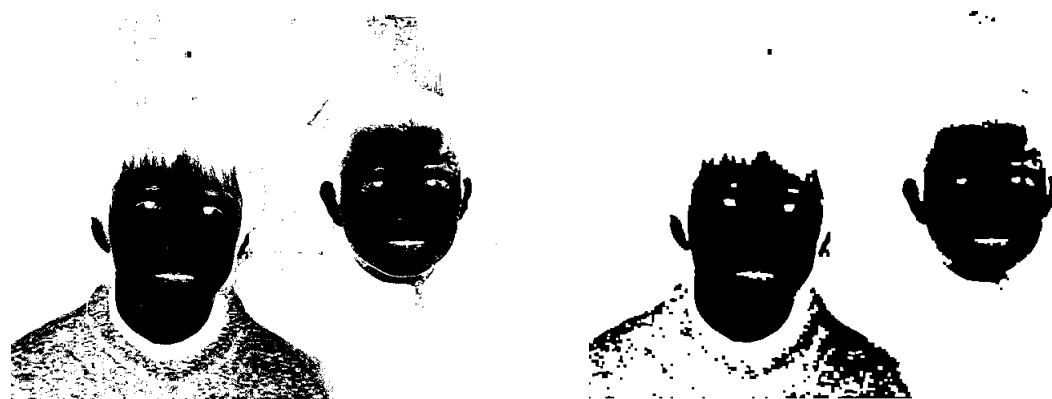


圖 2.3 膚色在 HSV 色彩空間分布情形

要將膚色部分與背景分離可使用臨界值法將影像做二值化處理。其作法是用一些方程式的組合來定義出膚色在 HSV 色彩空間的分佈。下列即為本研究用於定義膚色範圍的方程式組合[27]：

$$\begin{aligned}
 &S \geq 10, V \geq 25, S \leq -H - 0.1V + 110, H \leq -0.4V + 75, \\
 &S \leq \begin{cases} 0.08(100-V)H + 0.5V & \text{if } H \geq 0 \\ 0.5H + 35 & \text{if } H < 0 \end{cases} \quad (2.2)
 \end{aligned}$$

經由這些方程式組合可以找出影像中所有符合條件的像素點，這些像素點即是歸屬於膚色的部份，而我們並將這些符合的像素點二值化值設為 0，其他不符合的則設為 1。圖 2.4(a)所示即是圖 2.2(b)經此程序處理後的結果，黑色為膚色部份(值為 0)，白色為非膚色部分(值為 1)。



(a) 膚色在 HSV 色彩空間二值化的結果 (b) 膚色二值化的影像經開閉運算後的結果

圖 2.4 膚色二值化及開閉運算

由圖 2.4(a)可以看出二值化後的影像包含了人臉區域與非人臉區域，此時我們會先經由開閉運算對影像加以處理，以消除雜訊並使嘴巴部份適合於之後作角度的計算。經過開閉運算的結果如圖 2.4(b)所示，可以發現到一些人臉區域外的小雜訊被消除了，影像也變得較為平滑。對於開閉運算的處理，本研究是先作閉合運算再作斷開運算，其目



的是為了要使嘴巴部分的輪廓明顯，並讓與人臉區域藕斷絲連的其他非人臉區域分離。

由於在特徵擷取中，本文是利用相對位置來處理，因此為了使相對位置的尋找較為容易，必須先對有偏轉角度的人臉做一個轉正的處理。在做人臉角度轉正的動作之前，必須先使用到連通成份標示(connected-component labeling)[2]，把影像中屬於同一物件之相連像素找出，以便獲得物體面積的資訊，並藉此來標定出影像中的各個獨立之物件。其方法是利用 8 鄰接方式將鄰近的點連接在一起，讓所有連接的點都會被標示成一個物體。在本研究中偵測膚色及五官的過程，都將會利用連通成分的標示來分辨所要擷取的物體。

為了方便統計出每一個獨立區塊所佔像素的大小，本文將文獻[2]所述之連通成分標示的演算法則加以修正，修正後的法則可以更有次序的表示出每一獨立區塊的大小與座標，以方便區塊大小的比對。以圖 2.5 為例，圖 2.5(a)代表數個獨立區塊的集合，圖 2.5(c)定義出標示連通成分用之 3x3 遮罩結構元素，並以此遮罩將影像由上而下由左而右掃描。從掃描到第一個區塊的第一點開始，利用圖 2.5(c)找出所有獨立區塊所佔的像素點個數，結果如圖 2.5(b)所示。本研究修正後之連通成分標示的演算法則敘述如下：

- (1) 以掃描到第一個區塊的第一點為出發點，使用如圖2.5(c)所示的結構元素，依照A、B、C、D、E、F、G、H順時鐘方向的順序尋找下一點，找到後依序編入流水號記錄其座標，並用0取代中心點。直到A~H皆為0時進入步驟(2)。
- (2) 當A=B=C=D=E=F=G=H=0時，依照流水號順序往後遞減以尋找新的中心點，直到A、B、C、D、E、F、G、H中有一點為1為止，再重複步驟(1)。
- (3) 當流水號往後遞減至1時，代表整個獨立區塊已經搜尋完畢，此時影像重新由上而下由左而右掃描尋找下一個區塊的第一點並標示為一個新的物體，再重複步驟(1)。
- (4) 當影像由上而下由左而右掃描，在影像內找不到任何1的點時，代表所有區塊搜尋完畢，並結束標示程序。

搜尋完整張影像後，再依照先前所記錄的座標，逐一以 1 標示出所有座標點。如圖 2.5(b)所示，最後找出整張影像共有 2 個區塊，其像素點大小分別為 13 點和 15 點。

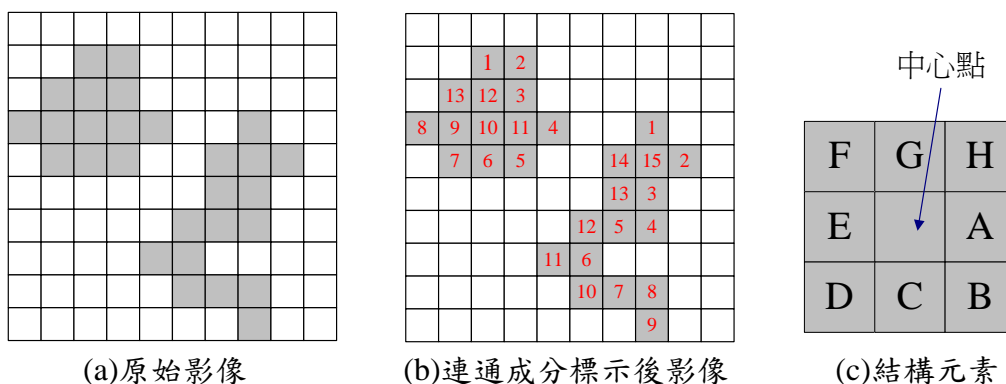


圖 2.5 連通成分標示

在連通成分標示的過程中，本研究同時設定了區塊像素數目的範圍，只擷取合適的



區域，因為太小的區域即使是屬於真正的人臉區域，特徵也是難以辨識的。一方面也藉此再次去除其他細小的背景區域或是雜訊點。圖 2.6(a)所示為圖 2.4(b)經過連通成份標示後的結果，共擷取了三個膚色區域，但此時仍然無法確定這些區域是否皆為人臉區域，因為它極有可能只是符合了膚色判斷的條件而已，很有可能它是手臂部份或跟膚色相近的衣服、背景物品等等。所以經過連通成份標示後的區域還必需經過人臉高寬比例的檢測，來排除非人臉膚色的區域。在一般正常情況之下，人臉區域(不論包不包含頸部)的高度必大於其寬度[25]，依此原則以及實際經驗，本研究的高(Height)及寬(Width)比例判斷式定為 $0.8 \leq \frac{Height}{Width} \leq 3$ 。同時，圖 2.6(a)經人臉高寬比例檢測後可擷取到如圖 2.6(b)所示的兩個人臉區域。圖 2.6(c)所示則為人臉區域套上原影像人臉膚色的結果。

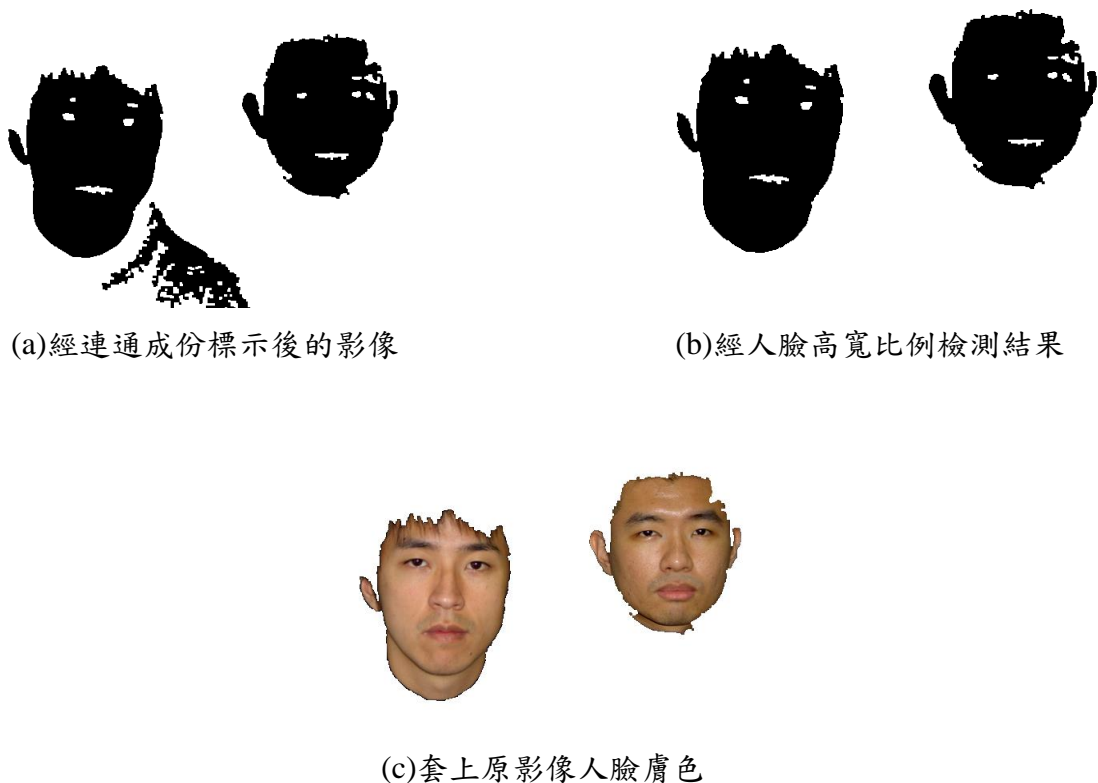


圖 2.6 連通成份標示及人臉高寬比例檢測處理

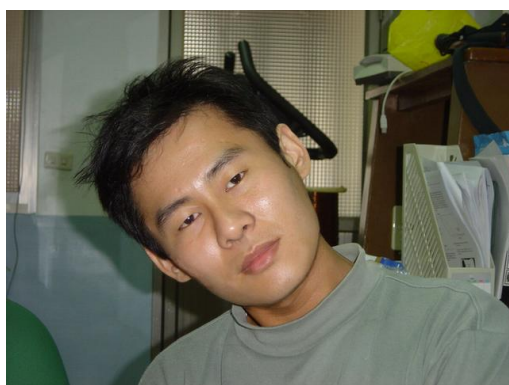
完成連通成份標示及人臉高寬比例檢測處理之後，為了準確擷取人臉特徵，接下來就是要對可能人臉區域做角度轉正的動作了。要做角度轉正，首先必須知道所要旋轉的角度值為何？針對此點，本研究經過多張不同的二值化影像發現到，嘴巴部分在人臉區域中的所有封閉缺洞(白色)中所佔區域是最大的，加上嘴巴兩端點也就是嘴角的連線，基本上會隨著人臉偏轉而呈現相同的角度轉動。因此本文利用此一性質來求取轉正人臉所應旋轉的角度 $\theta$ 。雖然得到角度 $\theta$ 之後，即可進行影像旋轉，但此時必須使用座標轉換的方法，讓原影像的座標像素轉換至新影像的座標上。假設原像素與旋轉中心的相對



位置為 $(x, y)$ ，旋轉後的像素位置與旋轉中心相對位置為 $(x', y')$ ，則新舊對應像素點的關係如下：

$$\begin{aligned} x' &= x \cos(-\theta) + y \sin(-\theta) \\ y' &= -x \sin(-\theta) + y \cos(-\theta) \end{aligned} \tag{2.3}$$

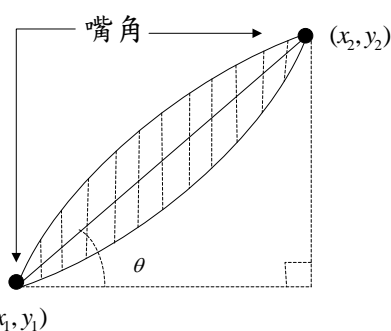
經過旋轉後座標位置 $(x', y')$ 的像素值即用原座標位置 $(x, y)$ 的像素值來代替，但由於像素的位置為整數值，而經上述的旋轉方程式後， $x', y'$ 不為整數，有一些像素位置會變成沒有像素值。因此，本文使用雙線性內插法[1]來解決此問題。圖 2.7(a)所示為人臉偏轉角度較大的一個例子，經二值化處理後的人臉候選區域如圖 2.7(b)所示，人臉區域擷取嘴巴部份並將其放大表示後如圖 2.7(c)所示，由兩嘴角座標 $(x_1, y_1)$ 及 $(x_2, y_2)$ 求得偏轉角度 $\theta = \tan^{-1} \frac{y_2 - y_1}{x_2 - x_1}$ 後，將人臉轉正的結果如圖 2.7(d)所示。



(a) 原始影像



(b) 人臉候選區域



(c) 以嘴巴區域求轉正角度



(d) 轉正後影像

圖 2.7 人臉區域轉正過程

### 參、人臉特徵擷取

由一張影像中取得臉部正確區域後，接下來就是要在該區域內找出臉部所代表的特





徵。由於光源分佈不均的因素，使得臉部每個區域受光程度有所不同，灰階值也會有明顯的差異。假如對整張影像以傳統的二值化方法，都設同一個臨界值作處理，並不能得到很好的效果。因此我們必須將臉部劃分成多個小區域個別處理。依此方式，本研究共統計了 96 張人臉影像，以人臉偵測後所得的影像寬度  $d$  為依據，計算出每一張影像中五官的相對距離，再取全部的平均值。如圖 3.1 所示，其全部相對距離均以寬度  $d$  的倍數表示之。

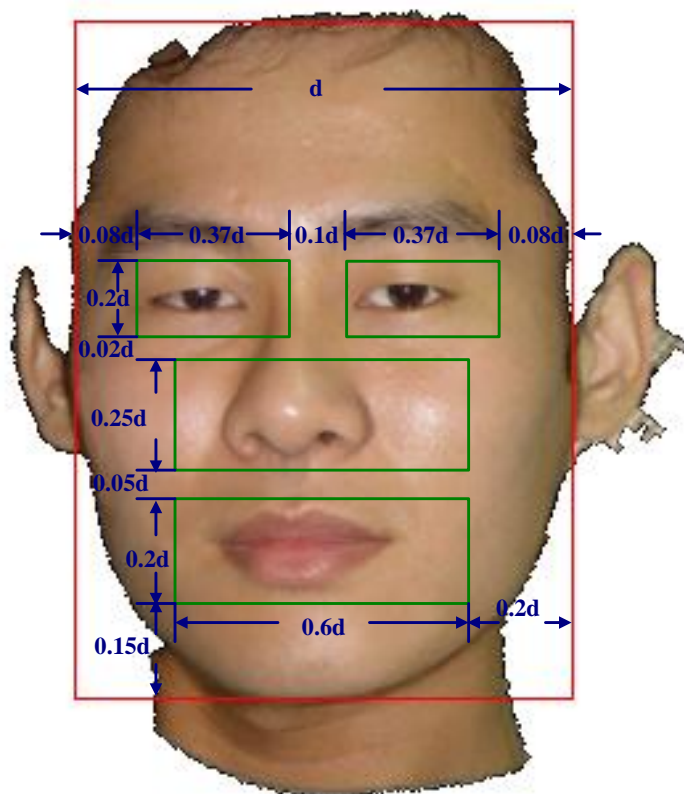


圖 3.1 五官相對距離

為了解決前述臉部因光源強度差異所造成的二值化問題，本文採用適應性臨界值 (Adaptive thresholding)[28]，對於待處理之區域影像取直方統計圖(histogram)。若期望二值化後的影像黑色區域面積佔全部影像的 10%，則由直方統計圖的最黑部分(灰階值為 0)開始算起，累加到個數為此區域的總像素個數的第 10% 個，則將這個像素所在的像素值設為臨界值。若區域內像素值小於臨界值，則設為 0；否則設為 1。一般而言，對於不同的區域，其定義所佔黑色區域的比例也不同，且通常區域範圍越大黑色區域的比例就會越小。這種方法不僅可以應用於整個臉部影像，也可以只針對其中某個特徵區域作處理。另外，由於臨界值是以二值化後黑色所佔的比例為依據，這使得在光線強度的變化下所得到的效果都會相同，解決了外在光線所造成臨界值不易選擇的問題。

接下來的步驟即是擷取嘴巴的特徵影像，處理流程如圖 3.2 所示。在進入圖 3.2 處理流程前，須先找出嘴巴的區域。根據圖 3.1 所示，嘴巴約在人臉由下往上  $0.15d \sim 0.35d$  及



由左往右0.2d~0.8d 的區域內，擷取後之嘴巴區域如圖3.2(a)所示。因為輸入的影像為RGB 的彩色影像，為了處理上的方便，必須經由下列方程式[1]將彩色影像轉換成灰階影像：

$$I_{gray} = 0.299 \times I_r + 0.587 \times I_g + 0.114 \times I_b \quad (3.1)$$

其中 $I_{gray}$ 是影像的灰階值， $I_r$ 、 $I_g$ 和 $I_b$ 分別是影像在R, G 和B 的值。經過灰階處理後的嘴巴影像如圖3.2(b)所示。

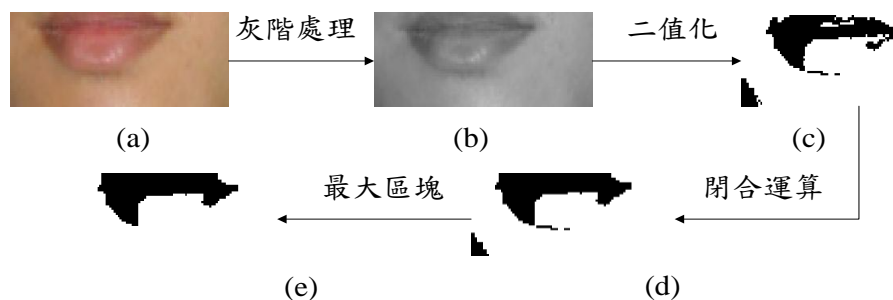


圖 3.2 嘴巴特徵影像擷取流程

接著依照適應性臨界值法，將嘴巴依照灰階值的比例作二值化，在此將比例值設為20%，經過二值化處理後的嘴巴影像如圖3.2(c)所示。由於二值化後影像會產生許多小雜點，必須再經過形態學的閉合運算，將很接近的黑色部位連接起來，並消除孤立的小點。經過閉合運算後的嘴巴影像如圖3.2(d)所示。最後運用連通成分標示法，找出圖中所有的區域，並經過比對後找出最大區域，此區域即代表嘴巴的特徵影像。經過最大區塊搜尋的嘴巴特徵影像如圖3.2(e)所示。

接著擷取鼻子的特徵影像。唯在擷取鼻子特徵影像前，和嘴巴一樣須先找出鼻子的區域。根據圖3.1所示，鼻子約在人臉由下往上0.4d~0.65d 及由左往右0.2d~0.8d 的區域內。鼻子的影像處理過程和嘴巴一樣，擷取到的鼻子影像先經過灰階處理，再做二值化運算(比例值設為12%)以及閉合運算最後找出代表鼻子的特徵影像，所有處理流程如圖3.3所示。

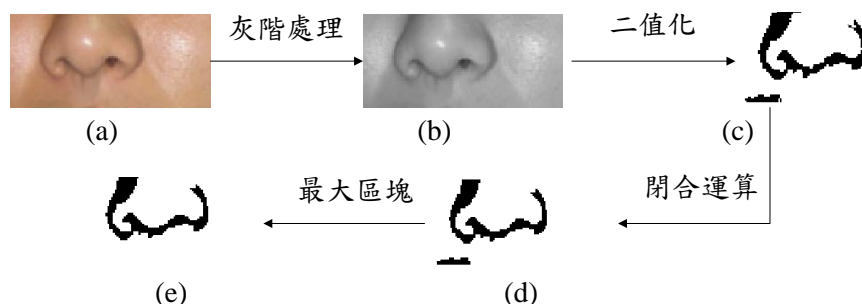


圖 3.3 鼻子特徵影像擷取流程

最後擷取眼睛的特徵影像，處理流程如圖3.4所示。眼睛可分為左眼及右眼兩個區域。兩眼的上下距離一樣，約在人臉由下往上0.67d~0.87d 的範圍。左右距離則分別落在由左往右0.08d~0.45d 及0.55d~0.92d 的區域內。眼睛區域擷取後的影像如圖3.4(a)所示。接著運用邊緣偵測觀念，以Sobel運算子[2]找出眼睛的邊緣。經過邊緣偵測處理後的眼晴影像如圖3.4(b)所示。此時影像還是RGB的彩色模式，須再利用(3.1)式將彩色影



像轉換成灰階影像。經過灰階處理後的眼睛影像如圖3.4(c)所示。因為本研究所期望的眼睛特徵影像是以黑色的方式呈現，所以必須將處理好的灰階影像(圖3.4(c))再經反白處理，使原本黑色的背景轉成白色，而白色的眼睛影像轉成黑色。經過反白處理後的眼睛影像如圖3.4(d)所示。接下來的幾個步驟和嘴巴的處理程序一樣，分別為二值化運算(比例值設為15%)、閉合運算以及搜尋最大區塊，最後找出代表眼睛的特徵影像。處理流程如圖3.4(e~g)所示。

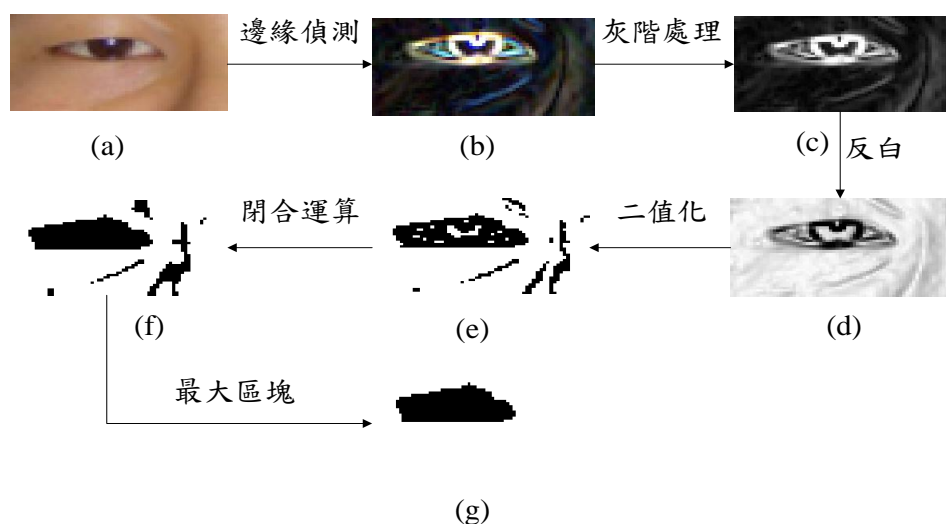


圖 3.4 眼睛特徵影像擷取流程

影像經過基本處理過程後，隨即針對所獲得的特徵影像施以特徵點的擷取程序。在整張臉部中有許多可用的特徵，但並不是每個特徵都可以很容易的被擷取出來。因此，著重於處理八個較明顯且具代表性的特徵點，分別為兩個嘴角、鼻子特徵影像的左右兩端點以及眼睛的四個眼角。為了能在特徵影像中找出上述八個特徵點的座標，首先利用影像投影方法[26]，將二值化的臉部特徵影像做水平方向的投影，其結果如圖3.5所示。

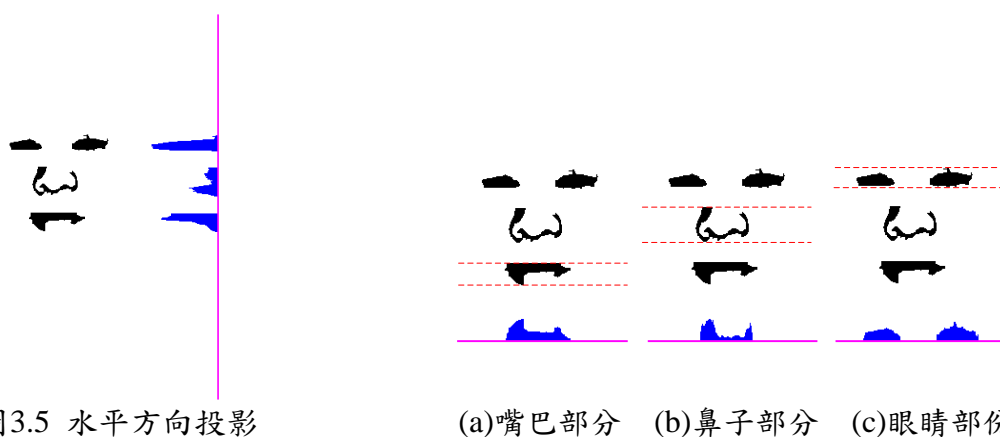


圖3.5 水平方向投影

(a)嘴巴部分 (b)鼻子部分 (c)眼睛部份

圖3.6 垂直方向投影

由圖3.5中可以很清楚的發現，特徵影像經過水平投影後，嘴巴、鼻子和眼睛的水平方向，都會產生一段很明顯的峰值，與周圍的值有明顯的落差。所以，本研究利用水平投影的方式紀錄了三個峰值區域的縱軸分佈範圍。緊接著對應於所紀錄的三個峰值區域的縱向範圍分別施以垂直方向的投影。這樣一來便可以得到各特徵影像在橫座標的分佈



範圍，投影結果如圖3.6所示。

由圖3.6(a)中可以看出嘴巴區域經過垂直方向投影後，會產生一段峰值區域，而這段峰值區域的起始點和終止點，就是嘴巴兩個嘴角的橫座標。同理，由圖3.6(b)中可以得到鼻子兩端點的橫座標，以及圖3.6(c)中得到眼睛四個眼角的橫座標。得到了八個特徵點的橫座標之後，縱座標的值就可以依橫座標的位置在相對應的縱軸範圍內找到。如圖3.7所示，A~H為搜尋到的八個特徵點。另外，若再定義嘴巴的中心點I(A和B的平均值)、鼻子的中心點J(C和D的平均值)以及兩眼的中心點K(E~H的平均值)，則臉部總共可得11個特徵點。

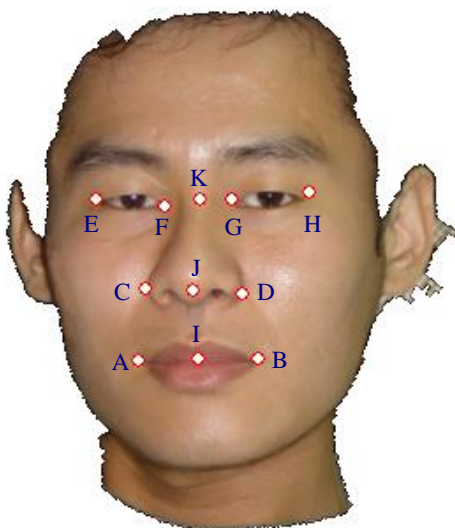


圖3.7 特徵點位置

因為進行特徵比對程序時，特徵向量間的差異值是比對的主要依據，所以本文從所得到的11個特徵點中，考慮23個具代表性的特徵向量，如表3.1所列。

表3.1 臉部特徵向量

1	$\overrightarrow{AB}$	9	$\overrightarrow{HB}$	17	$\overrightarrow{HD}$
2	$\overrightarrow{CD}$	10	$\overrightarrow{EB}$	18	$\overrightarrow{ED}$
3	$\overrightarrow{EF}$	11	$\overrightarrow{HA}$	19	$\overrightarrow{HC}$
4	$\overrightarrow{GH}$	12	$\overrightarrow{FA}$	20	$\overrightarrow{FC}$
5	$\overrightarrow{FG}$	13	$\overrightarrow{GB}$	21	$\overrightarrow{GD}$
6	$\overrightarrow{KJ}$	14	$\overrightarrow{FB}$	22	$\overrightarrow{FD}$
7	$\overrightarrow{KI}$	15	$\overrightarrow{GA}$	23	$\overrightarrow{GC}$
8	$\overrightarrow{EA}$	16	$\overrightarrow{EC}$		

由於計算出的向量，會隨著取像的遠近而改變其大小，所以必須將每個向量經正規化處理。如圖3.8所示，正規化程序首先計算出A~H八個特徵點的中心點N1、嘴巴的中心點N2以及兩眼的中心點N3和N4。其次計算出N2, N3和N4這三點與中心點N1間的平均距離。只要將表3.1中的每個向量除上這個平均距離，即完成了正規化的動作。



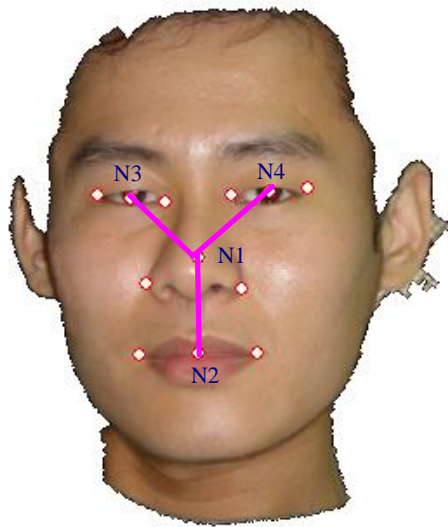


圖3.8 正規化參考點

#### 肆、實驗結果

本實驗所使用的硬體設備包含acer travelmate 290筆記型電腦一部、以及Sony p92數位像機一台。作業系統為Windows Xp，辨識系統中所有的演算法則是用Visual C++ 6.0版[29]來完成介面視窗化的實現。在實驗測試方面，本研究拍攝了16個人的臉部影像以供辨識率測試，並從16人中任意選擇拍攝多人合照影像。為了建立資料庫及供應比對，每個人拍攝6張臉部影像，共96張640×480(像素×像素)的單人彩色影像，其中的4人影像如圖4.1所示；並有20張兩人合照影像及20張三人合照影像，如圖4.2所示。而每張影像是分別在不同時間及不同背景下所拍攝而得，可作為資料庫內及供應比對的所有影像。

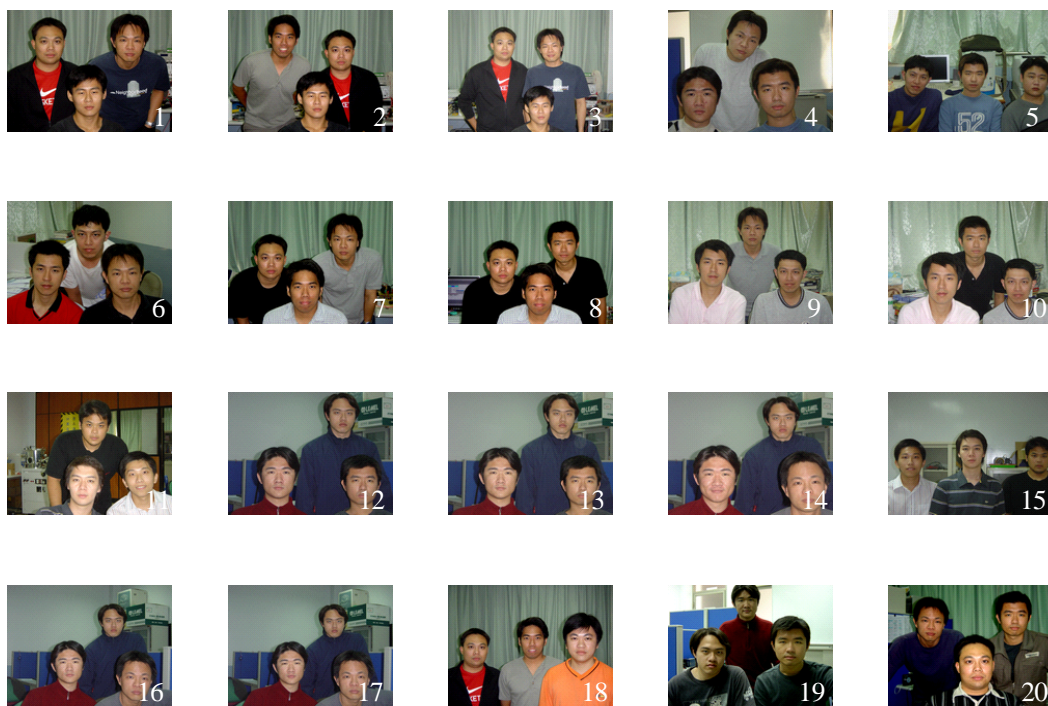


圖 4.1 單人臉部影像





(a) 二人合照影像



(b) 三人合照影像

圖 4.2 多人臉部影像

為了順利進行特徵比對工作，本研究使用了歐幾里德距離與漢明距離[30]兩種計算差異值的方法，分別敘述如下：

一、 歐幾里德距離

歐幾里德距離在距離測量方面是最常用來計算差異值的方法。此種相似程度的運



算主要是直接計算兩向量之間的差異，計算後所得到的值越小代表兩向量差異量越小，計算公式如下所示[30]：

$$d_E = \sqrt{\sum_{i=1}^L (K_i - H_i)^2} \quad (4.1)$$

其中， $d_E$  稱為歐幾里德距離， $k_i$  為輸入的特徵向量的第  $i$  個成分， $H_i$  為資料庫儲存的特徵向量的第  $i$  個成分， $L$  則為特徵向量的維度。由於每個特徵的數值等位不同，因此在計算歐幾里德距離前，必須針對每一個特徵給予不同的權值(Weighting)。同時考慮第  $j$  個特徵向量以及加入權值後的歐幾里德距離公式如下所示

$$d_{Ej} = \sqrt{\sum_{i=1}^{L_j} W_{ij} (K_{ij} - H_{ij})^2}, \quad j = 1, 2, \dots, M \quad (4.2)$$

其中， $d_{Ej}$  為第  $j$  個特徵向量加入權值後的歐幾里德距離， $L_j$  為第  $j$  個特徵向量的維度， $W_{ij}$  為第  $j$  個特徵向量的第  $i$  個成分的權重， $K_{ij}$  為輸入的第  $j$  個特徵向量的第  $i$  個成分， $H_{ij}$  為資料庫儲存的第  $j$  個特徵向量的第  $i$  個成分， $M$  則為特徵向量的總數。所以考慮所有特徵向量後， $d_{Ej}$  的總和可以表示成

$$d_{ET} = \sum_{j=1}^M d_{Ej} \quad (4.3)$$

其中  $d_{ET}$  為考慮所有特徵向量及加入權值後的歐幾里德距離總數。

## 二、漢明距離

漢明距離與歐幾里德距離之定義不同，它並非直接計算兩向量之間的差異值，而是計算兩向量在預設的誤差範圍內有幾個成分不一樣。因此當兩個向量越相近，其分量之差異值落在容許的誤差範圍內的數目就越多，所相對應的漢明距離就越小，計算公式如下所示[30]：

$$d_H = \sum_{i=1}^L \#\{|K_i - H_i| > \Delta_i\} \quad (4.4)$$

其中， $d_H$  稱為漢明距離， $\#$  為計算個數的符號， $K_i$  為輸入的特徵向量的第  $i$  個成分， $H_i$  為資料庫儲存的特徵向量的第  $i$  個成分， $\Delta_i$  為特徵向量的第  $i$  個成分的誤差容許值， $L$  則為特徵向量的維度。對於每個特徵向量，只要差異量大於個別的誤差容許值，就可以得到個數“1”，因此  $d_H$  的最大值為  $L$ 。當  $d_H=0$  時，則表示所有特徵向量的差異值都落在容許的誤差範圍內。

當每個特徵的數值等位不同時，在計算漢明距離前，亦須對每一個特徵給予不同的權值。所以，利用這種方法計算出的數值越高，代表兩身份之間的相似度越低。同時考慮第  $j$  個特徵向量以及加入權值後的漢明距離公式如下所示：

$$d_{Hj} = \sum_{i=1}^{L_j} \#\{W_{ij} |K_{ij} - H_{ij}| > \Delta_{ij}\}, \quad j = 1, 2, \dots, M \quad (4.5)$$

其中， $d_{Hj}$  為第  $j$  個特徵向量加入權值後的漢明距離， $L_j$  為第  $j$  個特徵向量的維度， $W_{ij}$  為第  $j$  個特徵向量的第  $i$  個成分的權重， $K_{ij}$  為輸入的第  $j$  個特徵向量的第  $i$  個成分， $H_{ij}$  為資料庫儲存的第  $j$  個特徵向量的第  $i$  個成分， $\Delta_{ij}$  為第  $j$  個特徵向量的第  $i$  個成分的誤差容許值， $M$  則為特徵向量的總數。所以考慮所有特徵向量後， $d_{Hj}$  的總和可以表示成：

$$d_{HT} = \sum_{j=1}^M d_{Hj} \quad (4.6)$$



其中， $d_{HT}$ 為考慮所有特徵向量及加入權值後的漢明距離總數。

特徵值經前述差異值方法的處理後，隨即針對本研究所擷取到的臉部影像進行特徵比對的工作。在資料庫建檔和比對數量方面，本研究首先將16個人所擷取到的第1組單人照影像，經過特徵計算之後儲存在資料庫等待比對，其餘單人照的2~6組影像則當作輸入以供辨識之用。單人照的比對總數共(16人×5組)80筆，二人合照的比對數為(2人×20組)40筆，而三人合照的比對總數為(3人×20組)60筆。

資料庫建檔1組的所有比對結果辨識率整理如表4.1所示。由表中顯示使用歐幾里德距離及漢明距離兩種差異值計算方法的單人臉部辨識率分別為92.5%和88.8%；二人合照之臉部辨識率分別為75.0%和72.5%；三人合照辨識率分別為66.7%和61.2%。

表4.1 資料庫建檔1組所有比對結果辨識率

歐幾里德距離				
照片人數	比對總數	成功	失敗	辨識率
單人	80	74	6	92.5%
二人	40	30	10	75.0%
三人	60	40	20	66.7%
漢明距離				
照片人數	比對總數	成功	失敗	辨識率
單人	80	71	9	88.8%
二人	40	29	11	72.5%
三人	60	37	23	61.2%

為了使人臉特徵辨識系統能再達到更好的辨識率，本研究將資料庫建檔數量增為2組。所以每個人所擷取到的第1~2組影像，經過特徵計算之後儲存在資料庫等待比對。其餘的3~6組影像則當作輸入辨識用，故單人照的比對總數共16×4=64筆。而兩人及三人合照影像的比對總數則跟之前一樣分別為40筆與60筆。資料庫建檔2組時的所有人數組合的比對結果辨識率整理如表4.2所示。由表中可以看出，如果和資料庫建檔1組的辨識率作比較，辨識成功率有些許的提高，證明了增加資料庫建檔組數以提升辨識率的可行性。

表4.2 資料庫建檔2組所有比對結果辨識率

歐幾里德距離				
照片人數	比對總數	成功	失敗	辨識率
單人	64	60	4	93.8%
二人	40	31	9	77.5%
三人	60	41	19	68.3%
漢明距離				
照片人數	比對總數	成功	失敗	辨識率
單人	64	57	7	89.1%
二人	40	30	10	75.0%
三人	60	39	21	65.0%





此外，考慮到人臉的水平偏轉可能造成人臉辨識上的困難，進而影響到辨識成功率，所以本研究也針對此點可能的影響因素進行修正。此步驟中假設人臉以 y 軸旋轉，如圖 4.3 所示，本文以所得到的特徵點座標，計算出頭部以 y 軸旋轉的角度，再求出各個特徵點的轉正角度，然後代入轉正公式便可得到新的特徵點座標。為了導出轉正公式，本文考量人臉在正面取像的情況下，兩眼是等寬的，而鼻子和嘴巴兩端點也應該是對稱於中心軸(人臉寬度一半的座標軸線)，所以如有偏轉則兩端點就不會對稱，與中心軸的距離也就不會一樣，因此本文虛擬一個如圖 4.3(a)所示的 3D 空間座標點，並由圖 4.3(b)所示的眼睛部份水平投影關係來求其各點的幾何關係。

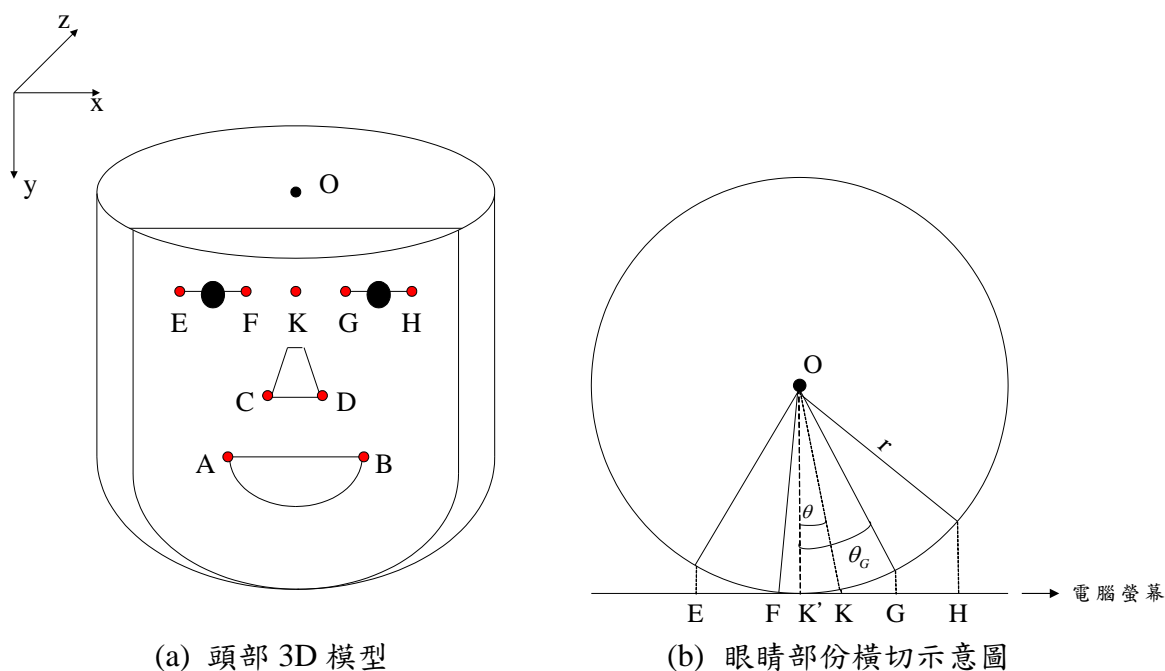


圖 4.3 人臉水平偏轉修正

圖 4.3(b)的 E、F、G、H 及 K 等各點，皆對應於圖 4.3(a)，代表眼睛的特徵點，而 O 點為中心軸上的座標點，K' 是人臉的中心點，r 表人臉寬度的一半， $\theta$  為 O 點至兩眼中點 K 與 K' 的夾角， $\theta_G$  為 O 點至特徵點 G 與 K' 的夾角。因為人臉是以 y 軸旋轉，所以每個特徵點的 y 座標值皆固定不變，因此我們只須要求出每個特徵點轉正後新的 x 座標值。當人臉轉正後，也就是 K 點移動到 K' 的位置。對於特徵點 G，轉正後與中心點 K' 的角度應該是  $\theta_G - \theta$ ，而  $r \times \sin(\theta_G - \theta)$  即為轉正後 G 與 K' 的距離長度，因為 K' 的座標值是不變的，所以可求得特徵點 G 的新座標值，其他的特徵點也是依照此種方式計算求得新的座標值。以特徵點 G 為例的轉正計算式如下：

$$\begin{aligned}
 Z_K &= \sqrt{r^2 - (X_K - X_{K'})^2} & ; & \quad \theta = \tan^{-1}\left(\frac{X_K - X_{K'}}{Z_K}\right) \\
 Z_G &= \sqrt{r^2 - (X_G - X_{K'})^2} & ; & \quad \theta_G = \tan^{-1}\left(\frac{X_G - X_{K'}}{Z_G}\right) \\
 X_{G'} &= r \times \sin(\theta_G - \theta) + X_{K'}
 \end{aligned} \tag{4.7}$$



其中  $X_G$  即為轉正後特徵點 G 的新 X 座標值。在圖 4.3 中，只要 K' 落於 F 及 G 之間，人臉水平偏轉修正皆可以得到不錯的效果，本研究依實際經驗所能處理的最大頭部水平偏轉角度約為 20 度。當每個特徵點都經過計算而得到新的座標值後，本研究就把這些新的數值代入差異值計算式中，來重新進行辨識計算工作。轉正處理後資料庫建檔 1 組所有比對結果的辨識率如表 4.3 所示。

表4.3 轉正處理後資料庫建檔1組所有比對結果辨識率

歐幾里德距離				
照片人數	比對總數	成功	失敗	辨識率
單人	80	76	4	95.0%
二人	40	31	9	77.5%
三人	60	40	20	66.7%
漢明距離				
照片人數	比對總數	成功	失敗	辨識率
單人	80	73	7	91.3%
二人	40	30	10	75.0%
三人	60	38	22	63.3%

由表 4.3 跟表 4.1 比較，可以發現在單人及二人辨識方面的效果比較顯著，可能是由於單人及二人人臉部分的影像較大，特徵點的尋找較為準確，因而辨識效果會比較好；而三人照片則因為臉部影像較小，特徵點與中心點的角度也會比較小，所以轉正處理後的辨識率改變量就較不顯著。本研究在轉正處理後再次以資料庫建檔 2 組來辨識，以提升辨識率。轉正處理後資料庫建檔 2 組時所有人數組合的比對結果辨識率整理如表 4.4 所示。由表 4.4 與表 4.3 比較，可以發現資料庫多增加 1 組，辨識率稍有提升。並且由表 4.4 與表 4.2 比較，同樣是資料庫建檔 2 組，經過轉正處理後的單人及二人辨識率比未經過轉正處理的辨識率高，由此可以證明轉正處理與資料庫建檔 2 組的結合，對於提升辨識率有其功效。

表4.4 轉正處理後資料庫建檔2組所有比對結果辨識率

歐幾里德距離				
照片人數	比對總數	成功	失敗	辨識率
單人	64	62	2	96.9%
二人	40	32	8	80.0%
三人	60	41	19	68.3%
漢明距離				
照片人數	比對總數	成功	失敗	辨識率
單人	64	59	5	92.2%
二人	40	31	9	77.5%
三人	60	39	21	65.0%



## 伍、結論

本研究設計出一個多人人臉身份辨識系統，且成功地將此系統的演算法用視窗介面的方式來完成。在系統中，由於所有的特徵向量代表人臉影像的資訊，所以利用歐幾里德距離及漢明距離之差異值計算法進行特徵向量比對，這樣的資料量雖少，亦能維持不錯的辨識率，所以其優點是電腦儲存的資料量不會佔太大空間，並可使處理速度較快。由實驗結果可以得知，多人人臉身份辨識成功率比單人人臉身份辨識成功率低，究其原因應該是人臉區域較小時，特徵點的判別比較不易，且於室內多人照時照相機光照分散，明暗度較不一致，也是影響辨識結果的原因之一。由實驗數據顯示，當資料庫多增加一組樣本時，更能提高系統的辨識率，而經過轉正處理後的身份辨識率也能再度提高其辨識率。結合以上兩者的轉正處理後資料庫建檔 2 組辨識率更有不錯的提升效果，證明了本研究所提方法的正確性及可行性。對於辨識率的提升，未來可以使用色彩矯正法來改善影像受光線影響的缺點，或利用類神經網路與小波理論來增強辨識效果。再者由於本研究只考慮 11 個特徵點來取特徵向量，忽略掉臉上一些不容易取得的特徵，所以未來可以發展出以更多特徵點擷取技術的人臉辨識系統。

## 參考文獻

- [1] 連國珍，“數位影像處理”，儒林圖書有限公司，2008 年 1 月五版一刷。
- [2] R. C. Gonzalez and R.E. Woods, “Digital Image Processing 3/E,” Prentice-Hall, New Jersey (2007)
- [3] J. H. Huang, “A Potential-based approach for shape matching and recognition,” Pattern Recognition, Vol. 29, No. 3, 1996, pp. 463-470.
- [4] R. Brunelli and T. Poggio, “Face recognition: Features versus templates,” IEEE Trans. Pattern Analysis and Machin Intelligence, 1993, pp. 1042-1052.
- [5] M. Turk and A. Pentland “Eigenfaces for recognition,” J. Cognitive Neuroscience, Vol.3, No.1, 1991, pp.71-86.
- [6] K. K. Sung and T. Poggio, “Example-based learning for view-based human face detection,” In Proc. Image Understanding Workshop, Monterey, Calif., Nov. 1994, pp.843-850.
- [7] E. A. Pentland, “View-based and modular eigenspaces for face recognition,” Tech. Rep 245. M. I. T. Media Lab Vismod, 1994.
- [8] S. H. Jeng, H. Y. Mark Liao, C. C. Han, M. Y. Chern and Y. T. Liu, “An efficient approach for facial feature detection using geometrical face model,” Pattern Recognition, Vol.31, No.3, 1998, pp.273-282.
- [9] X. I. Jia and M. S. Nixon, “Extend the feature vector for automatic face recognition,” IEEE Trans. Pattern Analysis and Machin Intelligence, Vol. 17, No. 12, Dec. , 1995, pp. 1167-1176.
- [10] L. Xiaobo and R. Nicholas, “Face contour extraction from frontview images,” Pattern Recognition, Vol.28, No.8, 1995, pp. 1179-1667.
- [11] T. Kanade, “Picture processing by computer complex and recognition of human faces,” Tech. Rep, Kyoto Univ., Dept. Inform Sci., 1973
- [12] R. Brunelli and T. Poggio, “Face recognition: Features versus templates,” IEEE Trans. Patt. Anal. and Mach. Intell., Vol.15, 1993, pp.1042-1052.
- [13] A. Rahardja, A. Sowmya and W. Wilson, “A neural network approach to component versus



- holistic recognition of facialexpressions in images, ” in SPIE Proc.:Intell. Tobots and Computer Vision X: Algorithms and Techn., Vol.1607, 1991, pp.62-70.
- [14] A.F.James and M.S.David, “Neural Networks Algorithms, Applications and Programming Techniques ,”Addison- Wesley Publishing Company, 1991, pp.1-43.
- [15] Shinjiro Kawato, Jun Ohya, “Automatic Skin-color Distribution Extraction for Face Detection and Tracking”, Proc.Int. Conf. On Signal Processing, Vol. II, 2000, pp. 1415-1418.
- [16] J.L.Crowley, K.Schwerdt, “Robust Tracking and Compression for Vedio Communication” In IEEE Computer Society Int. Conf. On Computer Vision, Workshop on Facie and Gesture Recognition, Corfu, Greece, September 1999.
- [17] Xiaozheng Zhang and Russ M.Mersereau, “Lip Feature Extraction Towards an Automatic Speechreading System”, In Proc. International Conference on Image Processing, Vancouver, Sept. 2000.
- [18] Rein-Lien Hsu; Abdel-Mottaleb, M.; Jain, A.K. “Face Detection in Color Images”IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, Issue: 5, pp.696-706, May 2002.
- [19] Yao, Hongxun and Wen Gao, “Face Detection and Location Based on Skin Chrominance and Lip Chrominance Transformation From Colr Images”, Pattern Recognition, vol.34, number8, 2001, pp. 1555-1564.
- [20] Wark,T.;Thambiratnam, D.;Sridharan, S. “Person Authentication Using Lip Information”, TENCON'97. IEEE Region 10 Annual Conference.Speech and Image Technologies for Computing and Telecommunications, 1997, Vol.1,pp.153-156.
- [21] Morimoto, C.H.; Flickner, M. “Real-time Multiple Face Detection Using Active Illumination”, Fourth IEEE International Conference on Automatic Face and Gesture Recognition. Proceedings. , 2000, pp.8-13 ,
- [22] 黃敏峰, “人臉追蹤法應用於監控系統之研究”, 碩士論文, 國立成功大學電機工程學系, 2003。
- [23] 吳明衛, “自動化臉部表情分析系統”, 碩士論文, 國立成功大學資訊工程學系, 2003。
- [24] 吳佳珍, “以鑑別性小波參數為主之人臉辨識系統”, 碩士論文, 國立成功大學資訊工程學系, 2001。
- [25] 黃一庭, “規則導向之多人人臉即時偵測”, 碩士論文, 國立東華大學資訊工程學系, 2002。
- [26] R. Jain, R. Kasturi and B. G. Schunck, “Machine Vision,”1995,pp. 35-36.
- [27] 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, Sept.1999, pp. 264-277 .
- [28] J. H. Lai, P. C. Yuen, Wen-Sheng Chen, S. Lao and M. Kawade, “Robust facial feature point detection under nonlinear illuminations, ”IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001,pp.168-174.
- [29] C. Petzold, “Programming Windows,” fifth Edition, SoftChina Corporation, 2000.
- [30] O. D. Richard, E. H. Peter and G. S. David, “Pattern Classification, ” John Wiley & Sons,2001.



# The Technique of Image-Vision Recognition of Multiple Faces

\*Ming-Fang Wu   \*Jen-Hsing Li   \*Ping-Tsung Wang  
\*\*Hui-Shan Chan   \*\*\*Chien-Pang Huang

\*Associate professor of Department of Electrical Engineering, Kun Shan University.

\*\*Associate professor of Department of Applied Cosmetology, National Tainan Institute of Nursing.

\*\*\*Postgraduate of Department of Electrical Engineering, Kun Shan University.

## ABSTRACT

Along the development of modern technology, the entrance control system, financial management, criminal detection, and computer certification etc., which all require a set of strong identity verification system. Although many studies have been proposed for identity verification or detection of face, but most current study only concentrate on the identity verification technology of single face or the detecting method of multiple faces. There are still a lot of development-space of identity verification technology of multiple faces. So the purpose of this thesis is to develop a system for the recognition of multiple faces by combining both the single face recognizing technology and the multiple faces detection method. And this system is especially favorable to the investigation of the crime.

On the method, this study use the processing technology of face features to deal with the images with multiple faces, and to develop a set of identity verification system that relying mainly on face features. In this system, the basic image processing techniques are used to find out the coordinates of feature points automatically, and then to calculate their combination of corresponding eigenvectors. On the comparison aspect, this study has made the use of difference value calculation methods including Euclidean distance and Hamming distance to compare and to check the degree of similarity among the eigenvector to achieve the goal of identification. This study also considers the correction of face's angle to upgrade the identification rate. After the completion of the theoretical inference, this study, in addition to use practical single and multiple people images for experiments to test and verify the practical efficiency of the multi-face identity-verification system proposed in this thesis.

**Keywords: vision technique, feature, multiple faces, recognition.**

