

行政院國家科學委員會專題研究計畫成果報告

雙眼彩色臉孔辨識之研究

A COLOR FACE RECOGNITION SYSTEM USING BINOCULAR IMAGES

計畫編號：NSC 88-2213-E-009-063

執行期限：87年8月1日至88年7月31日

主持人：陳玲慧 國立交通大學 資訊科學研究所

一、中文摘要

在研究計畫中我們提出一種利用立體視覺模型所做的彩色臉孔辨識方法。首先，將兩個攝影機並排以達成如人眼般的視覺模型，並對攝影機前的人取兩張影像。藉著由各個影像中找到特徵點並將兩張影像的相同特徵點做對應，可以算出特徵點的三度空間位置。辨識所需的主要特徵值都是從這些位置所導出的，另外還加上由區域邊緣點所導出的不變量。因為使用三度空間位置，可以避免做大小的調整，也可以避免因為投影所造成的誤差。本系統的訓練過程中含有特徵選取功能，可以選擇分離度較大的特徵用於辨識。本系統的辨識方法是估計不明人物是資料庫中某人的機率。除此之外，為了容易研究與使用，我們為本系統發展了一個方便的PC上的使用者介面。

關鍵詞：臉孔辨識、三度空間、特徵點

Abstract

In this project, a color face recognition system using stereo vision model is proposed. Two cameras are calibrated to be parallel and an image pair is captured from a person. 3D positions of feature points on face are first found, the feature vector including distances and ratio values derived from feature points and invariant moments calculated from edge points are then extracted. In the training process, a feature selection algorithm is provided to choose features with higher disjunction degree. The recognition process estimates the a posteriori probability of the unknown belonging to each person. A user friendly interface is developed on PC for convenient research and use.

Keywords: face recognition, 3D, feature point

二、緣由與目的

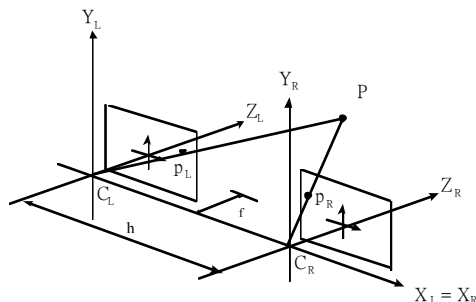
在以往的研究中 [1-9]，大部份都是由一張臉孔的影像中找尋特徵。但是因為只使用一張影像，所以並不能得到深度的資訊（如鼻子高度等）。若是使用了正面與側面兩張影像 [10-12]，則要面對攝影機組之架設與兩張影像大小不一等問題。而若是運用視訊中的影像序列 [13-14]，則往往會需要

人工的輔助。況且，因為投影幾何的原理，在影像中會有投影誤差（Projection error），這些誤差往往會影響我們所擷取的特徵。此外，隨著使用者介面的進步，一個好的臉孔辨識系統應該有親切的使用介面與方便的操作步驟。因此，本研究計畫的幾項重要目的如下：

1. 發展一可以取得三度空間資訊的臉孔辨識系統。
 2. 設計簡易的攝影機組架設方法。
 3. 定義臉孔的特徵點並尋找快速的特擷取方式。
 4. 製作方便且親切的使用者介面與操作步驟。
- 三、研究報告的內容

在本研究計畫中，建立了一個臉孔辨識的系統，整個系統包含影像取得、特徵擷取、訓練與辨識等步驟。首先由兩組攝影機取得兩張臉孔影像，然後經過特徵擷取後得到特徵值。若這些特徵值是用於訓練本系統，則將其存入資料庫，並經由特徵選取過程選擇適用於辨識之特徵；若取得的影像是作為辨識之用，則由資料庫取出資料進行辨識。

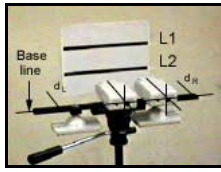
本計畫提出的新臉孔辨識方法，是使用兩架平行的攝影機模擬人的眼睛，建立立體的視覺模型 [15]（如圖一），並由左右兩張臉孔的影像中取得經由我們所定義的許多臉孔特徵點的三度空間位置。



圖一 立體視覺模型

在架設的過程中，因為要達到攝影機組的平行關係，所以必須要有一個攝影機定位方法。為了達成此目的，我們將利用平行線在影像中具有消失點的性質，對各個攝影機單獨做調整。首先將兩條水平平行線印在紙上並將紙貼在牆上，再將攝影機架

立於兩平行線之前，並利用水平儀讓兩條平行線與攝影機架均保持絕對水平(如圖二)。接下來，測量攝影機架兩端與牆的距離並使其相等。如此一來攝影機架便平行於牆上的兩條水平平行線。



圖二 攝影機定位

在所架設的攝影機的影像中，可以發現兩條平行線影像可能並不平行，這是由投影幾何的消失點原理所造成。此原理可簡述如下：“如果空間中一組平行線與影像平面不平行，則在影像中的投影將是不平行的直線，而這些直線在影像中的交點稱為這組平行線的消失點”。因為牆上的兩平行線是水平的，所以消失點的垂直座標為 0，假設消失點的水平座標為 u_v ，且攝影機常數為 f ，則攝影機要旋轉的角度為 $\theta = \tan^{-1} \frac{f}{u_v}$ 。

在攝影機定位問題解決後，接著便是要從所得的兩張影像中同時找出特徵點的位置。首先，對於每一水平位置，可以用下面這個函數來估計此位置上的水平“不”對稱性，而不對稱性最小的地方就是臉孔的水平中心。

$$HDS(u) = \sum_{v=v_i}^{v_t} \sum_{d_u=1}^{W_u/2} \sum_{c=r,g,b} |f_c(u-d_u, v) - f_c(u+d_u, v)|$$

where

- v_t = top of the searching window ,
- v_b = bottom of the searching window ,
- W_u = width of the searching window , and
- f_r, f_g, f_b = red, green, and blue values of the image.

左右眼的水平位置也是利用不對稱函數來尋找，因為眼睛是臉中水平對稱性次高的地方。這三個水平位置尋找的結果如圖三。



圖三 臉孔中心與兩眼的水平位置

下一步驟是尋找眉毛與眼睛的高度。在此是對眼睛水平位置附近的點做水平投影，但是在投影時是以紅色為基準，並且計算另外兩色的色差。如下式所示：

$$SICR(v) = \frac{|e^{v/nc} - e^{v/ey}||^2}{\sum_{du=-|e^{v/nc} - e^{v/ey}||/2}^{|e^{v/nc} - e^{v/ey}||/2} IC R(e^{v/ey} + du, v), \quad (3)$$

where

$$e^{v/ey} = \begin{cases} l^{v/ey}, & \text{in the left image.} \\ r^{v/ey}, & \text{in the right image.} \end{cases}$$

and

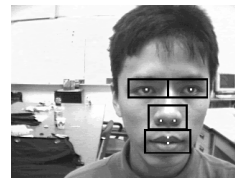
$$ICR(x,y) = \alpha \cdot f_r(x,y) + \beta \cdot [|f_r(x,y) - f_g(x,y)| + |f_r(x,y) - f_b(x,y)|]$$

因為眉毛與眼睛附近的投影量變化較劇烈，所以我們對水平投影作二階導數，如此一來則在眼睛與眉毛的位置上形成兩個高峰。如圖四



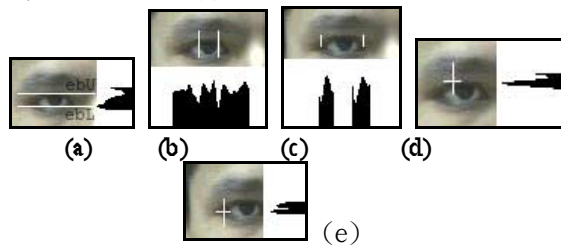
圖四 眉毛與眼睛的高度

接下來便運用類似方法尋找鼻孔與嘴的高度，但在投影時是直接針對顏色的強度。有了上述的粗略位置之後，便能夠藉以找出特徵點的精確位置。眼球與鼻孔中心是尋找 5 乘 5 均值最低的像素，而嘴巴中心線是取垂直投影量最高的水平位置，並在此位置上取山谷值最高的點為中心、再往上找邊緣值最高的點為嘴唇上緣中心。所有的六個特徵點如下圖所示。



圖五 人臉上的六個特徵點

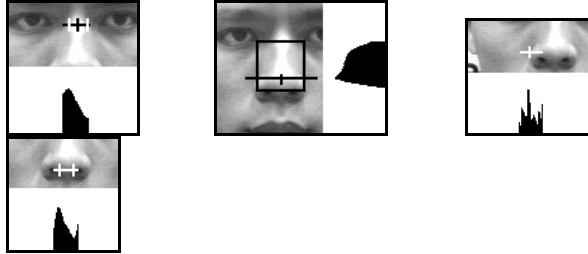
找到六個特徵點後，我們分別為眼睛找四個特徵點、為鼻子找六個特徵點、為嘴巴找四個特徵點。眼睛的工作分別如下(a)眼球的上、下界用水平投影與一階微分得到。(b)眼球的左、右界用垂直投影與二階微分得到。(c)眼睛的左、右界用垂直投影與一階微分得到。(d)眼球左右緣上利用像素的邊緣值尋找與上眼瞼的交點。最後是利用像素的山谷值在眼睛左右端尋找眼角位置。



圖六 尋找眼睛的特徵點

鼻子的特徵點另有 6 個。首先是在眼球高度利

用像素的水平邊緣值找鼻樑的左右邊緣，並得到此高度的鼻樑中心。其次是利用水平投影在鼻孔中間尋找鼻頭的高度。接著是用像素的水平山谷值尋找兩個外側的邊緣點，此兩點的高度為鼻頭與鼻孔的中間。最後是用像素的水平邊緣值尋找兩個鼻孔內側的特徵點，此兩點的高度與鼻孔相同。這些步驟均顯示於圖七。



第一步 第二步 第三步 第四步

圖七 尋找鼻子的特徵點

嘴巴的特徵點另有 4 個。尋找的步驟如圖八所示。首先在嘴巴區域中的每一個水平位置上尋找紅色頻道的最大垂直山谷值，接著對這些最大的垂直山谷值取水平的一階導數，則在左邊與右邊嘴角會分別造成極大與極小值。如此則找到兩嘴角位置。然後，從嘴角往內尋找每一水平位置上紅色頻道垂直邊緣值最大的像素。在嘴巴的兩側，這些像素中位置最高的點就是嘴唇上緣的端點。



第一步 第二步 第三步

圖八 尋找嘴巴的特徵點

待兩張影像的所有特徵點都找到之後，便要做特徵點位置的定位。定位的方法，是對每一個特徵點，先固定其在左影像中的位置，並在右影像中該特徵點位置周邊範圍內尋找最相似的點；然後再固定右影像中所得的點，在左影像原特徵點位置周邊範圍內找一個最相似點。經過定位後，由兩張影像的各特徵點位置便可以算出特徵點在三度空間中的位置。

找到 24 個特徵點的三度空間位置之後，接著便是從中取得 50 個特徵值。這些特徵值共有 5 型，分別為 X-型、Y-型、Z-型、距離型與比例型，這些型態都是取兩個特徵點的相對關係，定義如下：

$$XDIS(i, j) \equiv FP_i.x - FP_j.x$$

$$YDIS(i, j) \equiv FP_i.y - FP_j.y$$

$$ZDIS(i, j) \equiv FP_i.z - FP_j.z$$

$$DIS(i, j) \equiv \sqrt{\begin{pmatrix} (FP_i.x - FP_j.x)^2 + \\ (FP_i.y - FP_j.y)^2 + \\ (FP_i.z - FP_j.z)^2 \end{pmatrix}}$$

$$XYR(i, j, k, L) \equiv (FP_i.x - FP_j.x) / (FP_k.y - FP_l.y)$$

另一方面，利用眼睛與嘴唇的特徵點，可在各張影像中尋找到眼睛與嘴唇上緣的邊緣像素。利用這些邊緣點可將影像定義成一個二值函數並藉以算出中心動量 (central moment) 與七個不變量 [16]：

綜上所述，每一組臉孔影像 (含左右兩張影像) 共可得到 64 個特徵值。這些特徵值可用於訓練或辨識。

統的訓練的過程是將所得到的特徵值存入資料庫，並且對資料庫中所有的特徵值進行特徵選取的動作。特徵選取的目的是要選出分離性較大的特徵，以便在辨識的時候能夠有正確的結果。對於每一特徵，分離性的估算是先考慮該特徵在某兩人之常態分佈下的離異性，再對所有人口配對所得的離異性做平均。如下式：

$$Disjoin(m, n, i) = \frac{|\mu_{m,i} - \mu_{n,i}|}{\sigma_{m,i} + \sigma_{n,i} + \alpha}$$

$$DISJOIN(i) = \frac{1}{N_p} \sum_{\substack{m, n \\ m < n}} Disjoin(m, n, i)$$

上式中 i 代表某一特徵， m, n 分別代表不同的人， μ 與 σ 為常態分佈的平均值與標準差，而 N_p 則是人口配對數。經過特徵選取後，分離性較大的特徵將用於往後的辨識過程。

本系統的辨識方法分為兩個階段。第一階段是一個投票演算法，每個特徵值會根據其與資料庫中各人該特徵平均值選擇距離最近者，亦即每一特徵值會投下某人一票。待所有特徵值都選擇完畢後，選出得票最高的三個人當作可能的辨識結果。

如果得票最高者的票數比其他兩人的票數和還高，則確定得票最高者為辨識結果。否則，則進入辨識的第二階段。在第二階段是取受測者的每一特徵值來計算其屬於資料庫中某人的該特徵值分佈之機率，並將所有特徵值的機率加總當作分數。對於第一階段中所得的三個候選人，分數最高者則為辨識的結果。如下式：

$$Score(p, i) = \begin{cases} 0, & f_i \text{ is not selected.} \\ \frac{1}{\sqrt{2\sigma_{p,i}}} \exp\left[-\frac{(f_i - \mu_{p,i})^2}{2\sigma_{p,i}^2}\right], & f_i \text{ is selected} \end{cases}$$

$$SCORE(p) = \sum_{i=0}^6 Score(p, i)$$

然而，如果贏家的分數太低或是前兩名的得分太接近，則視為不可信任的結果。

四、實驗結果

本系統的實驗分為兩個部分進行。第一部份是測試經過訓練後系統的辨識效果，這一部份共有十位受測者（六男四女）。實驗結果如表一所示，在此實驗中可發現本系統在經過訓練後能夠有很好的辨識效果。實驗中不能正確辨識的原因都是因為特徵點的尋找不正確所致，而不是因為定義的特徵不佳。

表一 辨識結果

Name	Train	Test	CR (%)	ER (%)	RR (%)
H.S. Tang	6	6	83.3	0	16.6
C.T. Yeh	6	4	100	0	0
Eastun	8	4	75	25	0
S.J. Lin	3	2	100	0	0
Y.K. Lee	6	4	75	0	25
Eve	3	2	100	0	0
M.S. Liaw	4	2	100	0	0
Jean	3	2	100	0	0
Guest 1	4	3	100	0	0
Guest 2	2	2	100	0	0
Total	45	31	90.3	3.2	6.5

其中 Train、Test、CR、ER、RR 分別為訓練影像、測試影像、測試影像、錯誤率拒絕率。

第二部份的實驗是針對沒有存在於系統中的人。此部份共用兩人來做實驗，結果均判定為拒絕進入。表二是辨識過程中第二階段兩人的得分。從第一個實驗中我們發現經過訓練的人都能有 80 以上的得分。因此當遇到未經過訓練的人時，本系統確實能夠判定其為不認識的人。

表二 兩位闖入者的得分

Name	Average highest score
Intruder 1	45.21

Intruder 2	51.43
------------	-------

五、重要結論

本研究計畫提出一種新的臉孔辨識方法。主要是利用模擬雙眼的立體視覺模型取得臉孔特徵點的三度空間資訊。如此可以不受臉的前後左右移動之影響，且可以消除因為影像的投影而產生的誤差。

為了達成立體視覺模型，在本計畫中使用了兩架平行的攝影機，並且提出一個利用攝影機取得的影像做攝影機架設的方法。架設完成後的攝影機組可以方便地移動。

在特徵點的選擇上，本計畫所定義的特徵點都是臉上較特別的位置，目的是使得特徵擷取過程能夠快速且正確。當然，從這些特徵點出發，仍有許多特徵點可以尋找，甚至也可以選擇其他類型的特徵以增加系統的可適性。這些都是往後可研究的方向。

本研究計畫所用的訓練與辨識方法具有快速的特性，利用特徵選擇的機制，可以就資料庫中人們的特徵值分佈選擇適合用於辨識的特徵。實驗結果顯示錯誤的辨識往往是因為尋找特徵點時發生錯誤，如何避免各種狀況所造成的錯誤也是未來的研究主題。

六、參考文獻

- [1] Ashok Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey", *Pattern Recognition*, Vol. 25, No. 1, pp. 65-77, 1992.
- [2] Rama Chellappa, C. L. Wilson, and Saad Sirohey, "Human and machine recognition of faces: A survey", *Proceedings of the IEEE*, Vol. 83, No. 5, pp. 705-740, 1995.
- [3] Craw, H. Ellis, and J. R. Lishman, "Automatic extraction of face-features", *Pattern Recognition Letters*, Vol. 5, No. 2, pp. 183-187, 1987.
- [4] Venu Govindaraju, Sargur. N. Srihari, and David B. Sher, "A computational model for face location", *IEEE*, 1990.
- [5] Alan L. Yuille, Peter W. Hallinan, and David S. Cohen, "Feature extraction from faces using deformable templates", *International Journal of Computer Vision*, Vol. 8, No. 2, pp. 99-111, 1992.
- [6] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, 1990.
- [7] Matthew Turk and Alex Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991.

- [8] T. Aibara, K. Ohue, and Y. Matsuoka, "Human face recognition of P-type Fourier descriptors", *SPIE Proceedings*, Vol. 1606:Visual Commun. And Image Process., 1991, pp. 198-203.
- [9] Ebihara K, Ohya J, and Kishino F, "Real-time facial expression detection based on frequency domain transform", *Proceedings of the SPIE - The international society for optical engineering*, Vol. 2727, pt. 2, pp. 916-926, 1996.
- [10] David J. Beymer, "Face recognition under varying pose", *IEEE*, 1994.
- [11] W. H. Tsai and P. J. Chen, "Human face recognition using both front and side views", *National Chiao Tung University, Master thesis*, 1992.
- [12] Takaaki Akimoto, Yasuhito Suenaga, and Richard S. Wallace, "Automatic creation of 3D facial models", *IEEE Computer Graphics & Applications*, pp. 16-22, Sep. 1993.
- [13] Irfan A. Essa and Alex Pentland, "A vision system for observing and extracting facial action parameters", *IEEE Computer Vision & Pattern Recognition*, pp. 76-83, 1994.
- [14] Gozde Bozdagi, A. Murat Tekalp, and Levent Onural, "3-D motion estimation and wireframe adaptation including photometric effects for model-based coding of facial image sequences", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 4, No. 3, pp. 246-256, 1994.
- [15] Robert M. Haralick and Linda G. Shapiro, "Computer and robot vision", pp. 47, 1993.
- [16] Rafael C. Gonzalez and Richard E. Woods, "Digital image processing", *Addison Wesley*, pp. 514, 1992.