行政院國家科學委員會專題研究計畫　期中進度報告

子計畫三：利用人臉表情及唇形動態資訊進行身分確認之研究(2/3)

計畫類別：整合型計畫
計畫編號：NSC93-2213-E-002-037-
執行期間：93 年 08 月 01 日至 94 年 07 月 31 日
執行單位：國立臺灣大學資訊工程學系暨研究所

計畫主持人：洪一平

計畫參與人員：江岳軒、柯政宏、楊惠菁

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 94 年 6 月 1 日

行政院國家科學委員會補助專題研究計畫期中精簡報告

結合音訊與視訊之多模組身分確認之研究 － 子計畫三: 利用人臉表情及唇形動態資訊進行身分確認之研究(2/3)

計畫類別：□個別型計畫　　■整合型計畫

計畫編號：NSC -93-2213-E-002-037

執行期間：93 年 8 月 1 日至 94 年 7 月 31 日

計畫主持人：洪一平

本成果報告包括以下應繳交之附件：

☐ 赴國外出差或研習心得報告一份

☐ 赴大陸地區出差或研習心得報告一份

☐ 出席國際學術會議心得報告及發表之論文各一份

☐ 國際合作研究計畫國外研究報告書一份

執行單位：國立台灣大學資訊工程學系暨研究所

中　華　民　國　94 年 5 月 30 日

行政院國家科學委員會專題研究計畫期中精簡報告

結合音訊與視訊之多模組身分確認之研究 — 子計畫三: 利用人臉表情及唇形動態資訊進行身分確認之研究(2/3)

Biometric Person Authentication Using Dynamic Information of Facial Expression and Lip Tracking

計畫編號：NSC-93-2213-E-002-037
執行期限：93 年 8 月 1 日至 94 年 7 月 31 日
主持人：洪一平
計畫參與人員：江岳軒、柯政宏、楊惠菁

一、中文摘要

本子計畫的目標在於結合人臉表情與唇形追蹤等動態資訊進行身分確認。在傳統利用靜態資訊的方法中，由於影像中人臉的大小、方位，以及取像時的背景、光線均存在極大的變異性，因此使得人臉辨識的困難度變得比較高。在本子計畫中，我們嘗試利用使用者的臉部動態表情中所蘊含的資訊來進行身分確認。在動態唇形身分確認中，每一個使用者分別用一個外觀高維曲面(appearance manifold)來表示，而每一個外觀高維曲面則由一群姿勢高維曲面(pose manifold)來表示。為了建立這複雜且非線性的外觀高維曲面，對於一個使用者的連續影像，首先我們利用 K-means 分群演算法將其分成數群影像集合，並利用主成份分析法(PCA)來求得一個近似的主成份平面。為了將唇形在時間軸上的變化列入考量，我們計算唇形影像在其所屬的姿勢高維曲面彼此之間的轉變機率。而為了解決取像時，光線變異性對系統的影響，在此計畫中我們採用對光源環境作彈性分類(soft classification of lighting conditions)加上局部線性鑑別分析(locally linear discriminant analysis)的機制解決光源的問題。此方法首先會求出最佳的光源環境類別，再將訓練影像經過彈性光源分類，最後配合彈性分類的結果進行局部線性鑑別分析。

關鍵詞：身分確認、生物測定學、人臉表情、唇形追蹤、外觀高維曲面、姿勢高維曲面、光源環境的彈性分類、局部線性鑑別分析。

Abstract

The goal of this project is to combine dynamic information of human face, such as facial expression and lip tracking, for person authentication. In the traditional methods that only utilize static information, the large variations in face size, face pose, lighting and background increase the difficultly of face verification. Therefore, in this research project, we will use the dynamic information contained in face expressions for person authentication. To investigate the dynamic lip information extracted from the image sequences, each person is represented by an appearance manifold, which consists of a collection of pose manifolds. To construct this complex nonlinear appearance manifold for each person, we apply the K-means algorithm to cluster the image sequences of the talking lip for each person. We represent each cluster as a plane, which is computed by principal component analysis (PCA). To take the dynamic information of the talking lip into account, the transition probability between the pose manifolds is computed from the image sequences. On the other

hand, to solve the lighting variation problem for face recognition, we adopt soft classification of lighting conditions (SCLC) with locally linear discriminant analysis (LLDA) in this sub-project. The basic idea of SCLC+LLDA is to find the optimal lighting condition classes which best describe the lighting variation, and then apply a soft lighting classification to each training image. With the soft classification result, a locally linear transformation would be applied to find the global optimal subspace for face recognition.

Keywords: Person Authentication, Biometrics, Facial Expression, Appearance Manifold, Pose Manifold, Soft Classification of Lighting Conditions, Locally Linear Discriminant Analysis.

二、緣由與目的

隨著科技的進步，自動身分確認已成為一個愈來愈重要的問題。基於憑證(token-based)或基於知識(knowledge-based)的方法已經愈來愈不敷安全及便利的需求。因此，利用生物特徵的身分確認系統在近年來是一個非常熱門的課題。在各種生物特徵中，人臉是最明顯的外露特徵。就人類視覺而言，在相當遠的距離時我們便可藉由人臉來分辨出對方的身分，因此人臉辨識與確認一直是電腦視覺領域持續關注探討的問題。

所謂的「人臉身分確認」(face authentication)，與「人臉識別」(face recognition)並不是完全相同的工作，前者是要針對處理對象所宣稱的身分做出確認的動作，後者則是要判斷處理對象的身分是誰。這兩種工作在決策方式與評估方法不盡相同。相關的研究大致可以根據人臉特徵的資訊來源分成兩大類，一類使用人臉的靜態資訊，這一類的方法會希望所處理的人臉盡量不要有表情變化；另一類則是利用人臉的動態資訊，這一類的方法希望所處理的人臉最好能有一些表情或唇形變化。

在本年度的計畫中，我們的研究方向分成兩個部分，一個是利用動態唇形的資訊來做身分確認；另一個是利用光源環境的彈性分類和局部線性鑑別分析來解決人臉影像拍攝時，光線變異性對辨識系統的影響。我們分別在第三和第四小節介紹我們對這兩個部分的研究方法。

三、使用外觀高維曲面來做動態唇形身分確認

過去做動態唇形影像身分確認的方法中，Broun 等人[1][2]利用色調和飽和度來取出唇形影像，並利用影像切割的方法來取出嘴唇部分的長、寬等資訊，最後利用分類器來做身分確認。Luettin 等人[3]利用主成份分析法來取得唇形影像的投影，此外同時採用唇形輪廓來當分類器的輸入。但是這些方法在處理唇形於時間軸上的動態變化都沒有彈性，因此正確率在先天上便會有所限制。在這裡，我們參考人臉辨認領域裡常用的方法。Lee 等人[4]提出利用外觀高維曲面來對影片中動態的人臉做辨識。在此計畫中，我們利用類似的概念，來對動態的唇形影像做身分確認。

在動態唇形身分確認上，我們先建立唇形影像資料庫，對每一個使用者，我們錄製數段使用者說出特定語彙時的唇形變化影像 (如《圖一》)。

《圖一》使用者的唇形影像：某一使用者唸 “image processing and pattern recognition”時的某五張唇形變化影像。

1. 研究方法

為了考量不同的人說出同一句特定語彙在時間軸上所造成的不同特徵，每個人會擁有一個能描述自己唇形影像變化的外觀高維曲面 (appearance manifold)。當得到屬於每個人的外觀高維曲面後，對某張

測試唇形影像 I，我們根據和 I 最接近的外觀高維曲面來判斷 I 的身分。唇形影像 I 的身份 k^* 可以利用下面式子決定：

$$k^* = \arg\min_k d_H(I, M_k).$$

其中 d_H 代表唇形影像 I 和外觀高維曲面 M_k 的最小距離。

每一個外觀高維曲面是由 m 個姿勢高維曲面 (pose manifold) $\{C^1,...,C^m\}$ 組成。對於每一個使用者，這方法使用下面三個步驟來得到他的外觀高維曲面：首先，我們收集有此使用者數段的連續唇形變化影像，並利用 K-means 分群演算來將這些連續的唇形變化影像分成 m 個互斥子集合 $\{S_1,...,S_m\}$。之後，對每一個互斥子集合 S_k，我們對其使用主成分分析法，來得到近似於姿勢高維曲面 C^k 的主成分平面 L_k。在得到所有姿勢高維曲面的近似後，我們計算姿勢高維曲面之間的轉變機率。此轉變機率可用來描述唇形影像在時間軸上的變化。這方法定義轉變機率 $p(C^j \mid C^i)$ 為：

$$p(C^j \mid C^i) = \frac{1}{\Lambda}\sum_{q=2}^{l}\delta(I_{q-1} \in S_i)\delta(I_q \in S_j),$$

其中當 $I_q \in S_j$ 時，$\delta(I_q \in S_j) = 1$，否則為 0。在此，Λ 的目的是為了確保對任一個姿勢高維曲面 C^i 而言，$\sum_{j=1}^{m} p(C^j \mid C^i) = 1$。

接著，定義 C^{ki} 為外觀高維曲面 M_k 中的第 i 個姿勢高維曲面；並定義 L_{ki} 為利用主成份分析法得到的用於近似姿勢高維曲面 C^{ki} 的主成份平面。給定一張唇形影像 I，它與外觀高維曲面 M_k 之間的距離為與姿勢高維曲面的距離期望值:

$$d_H(I, M_k) = \sum_{i=1}^{m} p(C^{ki} \mid I) d_H(I, C^{ki})$$

此外，我們用 I 與主成分平面 L_{ki} 之間的距離來近似 I 與姿勢高維曲面 C^{ki} 之間的距離:

$$d_H(I, L_{ki}) \approx d_H(I, C^{ki}).$$

同時，為了使 $p(C^{ki} \mid I)$ 能加入時間先後的資訊，在每一個時間點 t，都將之前在時間點為 *0* 到 *t-1* 的唇形影像都列入考慮，即 $p(C^{ki} \mid I_t, I_{0:t-1})$。我們可以將此考慮時間先後因素的事後機率寫成遞迴型式:

$$p(C^{ki} \mid I_{t,} I_{0:t-1}) = \alpha * p(I_t \mid C_t^{ki}) * \Delta$$

其中

$$\Delta = \sum_{j=1}^{m} p(C_t^{ki} \mid C_{t-1}^{kj}) p(C_{t-1}^{kj} \mid I_{t-1} I_{0:t-2})$$

當 *t=0* 時，我們令 $\Delta = 0$。機率可能函數為：

$$P(I \mid C^{ki}) = \frac{1}{\Lambda_{ki}} \exp\left(\frac{-1}{2 * \sigma^2} \hat{d}_{ki}^2\right)。$$

由上述方法，依事後機率:

$$p(k \mid I) = \frac{1}{\Lambda} \exp\left(\frac{-1}{\sigma^2} d_H^2(I, M_k)\right)$$

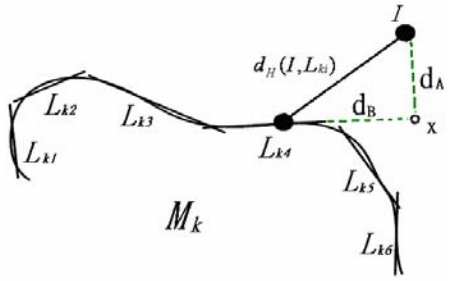
即可將此待辨識的唇形影像 I 辨識為

$$k^* = \arg\max_k p(k \mid I)。$$

2. 待解的問題

在實作的過程中，此一方法目前仍有下面三個問題尚待解決。

a. 距離定義的選擇問題：



《圖二》外觀高維曲面示意圖

令 x 為唇形影像 I 投影到主成分平面後的點 (如《圖二》所示)。計算 $d_H(I, L^{ki})$ 時，計算方法如下：

$$d_H(I, L^{ki}) = \alpha * d_A + (1-\alpha) * d_B$$

其中 d_A 即為 I 與 x 之間的歐氏距離，d_B 為 x 與主成分平面中點的距離。目前我們設

定參數α為0，意即未將d_A列入考慮，且d_B的計算方法是使用歐氏距離。未來希望可以使用瑪氏距離(Mahalanobis distance)，將資料分布的特性加入計算，期望能更精準地估計出I到姿勢高維曲面的距離。

b. 取樣頻率問題：

當唇形變化影像的取樣頻率不夠高時，外觀高維曲面沒辦法維持其連續的特性。實驗結果通常沒辦法得到滿意的辨識率。在未來的一年中，我們將設法利用影像形變(image morphing)的方法來解決此一問題。

c. 對位問題：

在拍攝唇形影像時，人臉的位置難免有所移動，因此唇形的位置也會跟著移動。我們目前並沒有去處理此一現象所帶來的影像對位問題。當唇形影像沒有經過正確的對位，會導致影像之間上下或是左右差距數個像素，這問題會影響到之後辨識的結果。在未來的一年中，我們將設法解決此一問題。

四、光源環境的彈性分類與局部線性鑑別分析(SCLC+LLDA)

光源對影像的影響是人臉辨識中最困難的問題之一。如之前所提及的，此方法首先會求出最佳的光源環境類別，再將訓練影像經過彈性光源分類(soft classification of lighting conditions)，最後配合彈性分類的結果進行局部線性鑑別分析[6] (locally linear discriminant analysis)。

1. 研究方法

給定一組N維的影像$Z=\{z_1,z_2,...,z_J\}$，每一個影像都屬於L個光源環境類別$\{Z_1,Z_2,...,Z_L\}$中的一個。其中任兩個光源環境類別的距離定義如下：

$$D(Z_i,Z_j)=1-\frac{m_i^T m_j}{\|m_j\|\|m_j\|}$$

其中m_i代表第i個光源環境類別的平均影像。

接著，我們可以定義最佳的一組光源環境類別$G_K=\{G_{K1},G_{K2},...,G_{KK}\}$，它可以將$\overline{G}_K$和$G_K$間的距離縮到最小：

$$G_{K\ \ OPT}=\arg\min_{G_K}\sum_{Z_j\notin G_K}D(G_K,Z_j)$$

$$D(G_K,Z_j)=\min_{Z_i\in G_K}D(Z_i,Z_j)$$

然而尋找最佳的一組光源環境類別是屬於NP-Complete的問題。以下的機制可以求得一個近似解：

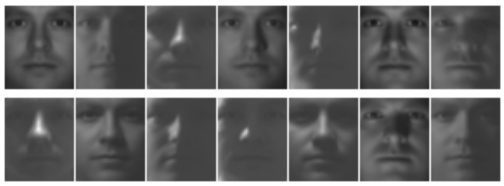
1. 以空集合初始化G_K

$$G_K=\{\phi\}$$

2. 用一個索引值k從1執行到K，在每一個回合都選擇最佳的G_{Kk}

$$G_{Kk}=\arg\max_{Z_i\notin G_K}\sum_{Z_j\notin G_K}D(G_K\cup Z_i,Z_j)$$

《圖三》是我們實驗結果中$K=14$最佳的光源環境類別，而《圖四》是光源環境關係程度圖。



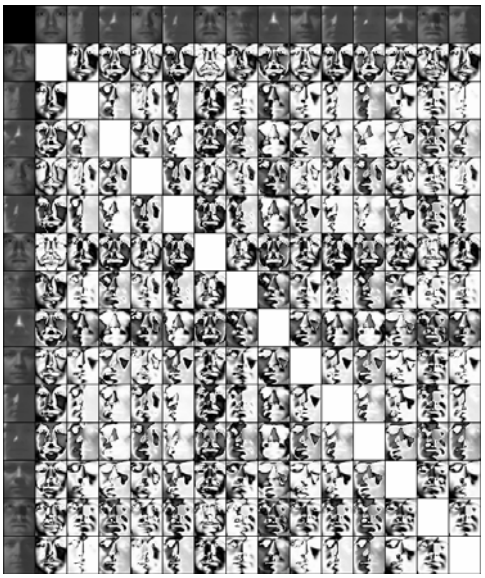
《圖三》在K=14中最佳的一組光源環境類別：此圖顯示了最佳光源環境類別的平均影像。

給定另一組N維的人臉訓練影像$X=\{x_1,x_2,...,x_M\}$、每一張都屬於C個人臉類別$\{X_1,X_2,...,X_C\}$中的一個，而對光源環境的彈性分類結果$v_i=\{v_{i1},v_{i2},...,v_{iK}\}$是一個$K$維的向量，它的定義如下：

$$v_{ik}=\begin{cases}\dfrac{1}{N_j}\cdot\dfrac{x_i^T m}{|x_i||m_k|} & \dfrac{x_i^T m}{|x_i||m_k|}\geq 0\\ 0 & otherwise\end{cases}$$

其中N_k是一個正規化的常數：

$$N_k=\sum_{k=1}^{K}v_{ik}=1$$



《圖四》光源環境類別的光源環境關係程度圖：圖中的第一行和第一列是索引，上面放的是所對應的光源環境類別的平均影像，而剩下的影像則是光源環境關係影像。影像越白的圖表示所對應的兩個類別關係程度高，反之則低。由上圖可知所得的最佳光源環境類別彼此不相像，此正交的關係更適合於表示所有光源的影響。

有了光源環境的彈性分類結果，我們就可以將局部線性鑑別分析的轉換向量 $U_k=\{u_{k1},u_{k2},...,u_{kN}\}$ 列式如下：

$$y_i=\sum_{k=1}^{K}v_{ik}U_k^T(x_i-\mu_k)$$

而第 k 個光源環境的平均影像 μ_k 定義如下：

$$\mu_k=(\sum_{i=1}^{M}v_{ik}x_i)/(\sum_{i=1}^{M}v_{ik})$$

所有轉換過的向量的總體平均 $\tilde{m}$ 為：

$$\tilde{m}=\frac{1}{M}\sum_{i=1}^{M}y_i=\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{K}v_{ik}U_k^T(x_i-\mu_k)$$

而含有 M_c 個資料的第 c 個類別的平均定義為：

$$\tilde{m}_c=\frac{1}{M_c}\sum_{x\in X_c}y=\sum_{k=1}^{K}U_k^Tm_{ck}$$

其中 m_{ck} 為第 c 個類別中屬於第 k 個光源環境類別的平均：

$$m_{ck}=\frac{1}{M_c}\sum_{x_i\in X_c}v_{ik}(x_i-\mu_k)$$

我們重新定義轉換後的類別間分散矩陣及類別中的分散矩陣：

$$\tilde{S}_B=\sum_{c=1}^{C}M_c(\tilde{m}_c-\tilde{m})(\tilde{m}_c-\tilde{m})^T$$

$$\tilde{S}_W=\sum_{c=1}^{C}\sum_{x\in X_c}M_c(y-\tilde{m})(y-\tilde{m})^T$$

而所求的轉換向量則是將以下的式子最大化的解：

$$J=(1-\alpha)\left|\tilde{S}_B\right|-\alpha\cdot\left|\tilde{S}_W\right|$$

將 J 對 u_{kn} 偏微後可得到：

$$\frac{\partial J}{\partial u_{kn}}=\left(2(1-\alpha)B_k-2W_k\right)u_{kn}+\sum_{i=1,i\neq k}^{K}\left(2(1-\alpha)B_{ki}-2W_{ki}\right)u_{in}$$

其中

$$B_k=\sum_{c=1}^{C}M_cm_{ck}m_{ck}^T\ 、\ B_{ij}=\sum_{c=1}^{C}M_cm_{ci}m_{cj}^T$$

$$W_k=\sum_{c=1}^{C}\sum_{x\in X_c}(p(k\mid x)(x-\mu_k)-m_{ck})(p(k\mid x)(x-\mu_k)-m_{ck})^T$$

$$W_{ij}=\sum_{c=1}^{C}\sum_{x\in X_c}(p(i\mid x)(x-\mu_i)-m_{ci})(p(j\mid x)(x-\mu_j)-m_{cj})^T$$

運用微分的結果，我們可以用以下的步驟得到一組最佳的 u_{kn}，從 $n=1$ 執行到 N 、$k=1$ 到 K：

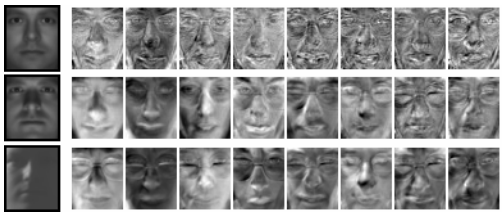
1. 隨機初始 K 個單位向量 u_{kn}。
2. 計算 $\frac{\partial J}{\partial u_{kn}}$，並以適當的更新速率 η 更新 u_{kn}：

   $$\Delta u_{kn}\leftarrow\eta\frac{\partial J}{\partial u_{kn}}$$
3. 保持每一組解的正交性：

   $$u_{kn}\leftarrow u_{kn}-\sum_{i=1}^{n}\left(u_{kn}^Tu_{ki}\right)u_{ki}$$
4. 將所求的向量標準化：

   $$u_{kn}\leftarrow u_{kn}/\|u_{kn}\|$$

《圖五》則是我們求出的幾組 u_{kn}



《圖五》幾組所求的u_{kn}：每一列最左方的圖是所對應的類別的影像，剩下的圖則是其前八個u_{kn}。

2. 實驗結果

我們使用 BANCA [7]人臉資料庫測試我們的實驗結果。BANCA 資料庫中共有 1560 張人臉影像、26 個人。每人分 12 次取像、一次取 5 張，12 次取像中共有三種光源環境。以下的實驗結果顯示我們的方法所獲得的辨識率可高達 93.3%。

《表一》BANCA 人臉資料庫的測試結果

Error Rate (%)	FAR = FRR			Min(FAR+FRR)		
	FA	FR	TE	FA	FR	TE
PCA+LDA	21.0	20.5	20.8	21.2	15.5	18.4
SCLC+LLDA	14.0	12.6	13.4	**7.9**	8.7	8.3
SCLC+LLDA with best LCCs	**9.5**	**8.7**	**9.1**	9.3	**4.1**	**6.7**

五、計畫成果自評

在本年度的計畫中，我們的研究主要是想利用人說話時嘴唇的動態資訊來進行身份確認的工作。目前我們所採用的方法是，針對待辨識的每一個人，將一連串的唇形動態序列影像建成一個外觀高維曲面，並使用主成分分析法來近似構成外觀高維曲面的姿勢高維曲面，以節省資料庫的存放空間。雖然目前仍有一些待解問題，但此方法能彈性的抓取時間軸上的動態資訊以達到高辨識率，是很值得繼續深入研究的方向。下一年度我們將著重在解決目前發現的待解問題，並擴大唇形資料庫。而在解決光線變異性的問題上，我們使用的局部線性方法同時擁有了線性方法中的高效能和非線性方法中的高正確率。但是目前的光源類別只考慮了單一光源的情況，而多個光源的影像仍得靠對光源的彈性分類來達成。下一年度我們會嘗試在光源類別上增加非單一光源的資料，以及嘗試其他求 SCLC+LLDA 轉換向量的解法。

六、參考文獻

[1] Xiaozheng Zhang and C. C. Broun, “Using Lip Features for Multimodal Speaker Verification”, *In A Speaker Odyssey - The Speaker Recognition Workshop, Crete, Greece, June 2001*.

[2] C. C. Broun, X. Zhang, R. M. Mersereau, M. Clements, “Automatic Speechreading with Application to Speaker Verification”, *In Proc. ICASSP, Orlando, May 2002.*

[3] Juergen Luettin, Neil A. Thacker, Steve W. Beet, “Speaker Identification by Lipreading”, in Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP'96), 1996.

[4] K. C. Lee, J. Ho, M. H. Yang, D. Kriegman, “Video-Based Face Recognition Using Probabilistic Appearance Manifolds”, *Proceedings of the 2003 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, pp. 313-320, vol. 1, Madison, June, 2003.

[5] A.S. Georghiades and D.J. Kriegman, “Illumination Cone Models for Face Recognition under Variable Lighting and Pose”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, 2001, pp 643-660.

[6] Y.-K. Kim, J. Kittler, “Locally Linear Discriminant Analysis for Multimodally Distributed Classes for Face Recognition with a Single Model Image”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, no. 3, March 2005.

[7] Bailly-Bailliere, S. Bengio, and K. Messer *et al.* “The BANCA Database and Evaluation Protocol”, *International Conference on AVBPA*, 2003, pp. 625-623.