

使用機率圖形模式建構影片中人臉辨識之資訊技術架構

A Framework of Probabilistic Graphical Model for Video-Based Face Recognition

江政杰

德明財經科技大學資訊科技系

Department of Information Technology, Takming University of Science and Technology.

Email: pluto@takming.edu.tw

摘要

在本論文中，我們設計一套建構於機率圖形模式下的資訊系統架構，來處理影片中人臉辨識的問題。我們所建構的機率模型將影片中的人臉辨識問題劃分為兩大部分：可能性評估與遞移性評估。關於可能性評估這部份，我們以傳統的單張影像中人臉辨識來處理。我們採用 two-dimensional linear discriminant analysis (2DLDA) 來判斷系統所觀察到的人臉影像資訊，以此為可能性評估的標準。同時，對於遞移性評估的計算，代表著由前次估計出的系統狀態結果轉移到現在狀態的可能性；在我們的設計中，這部份包含著人與人的識別轉變與人臉姿勢轉變的可能性。我們同時展示一些實驗，顯示我們設計的架構能夠在影片中人臉辨識的問題得到良好的效果。

一、緒論

人臉辨識(face recognition)技術在電腦視覺與圖形識別的資訊科技中，一直是一個熱門且重要的研究主題；在很多的多媒體應用系統上，人臉辨識也常常是一個關鍵的技術。人臉辨識的主要目的，在於辨識出一張影像上的人臉畫面是屬於哪一個人，當我們知道影像內出現者的身分後，例如可以在安全系統上自動辨識人員的身分，在後續的多媒體系統設計上有很大的發揮空間。

傳統上人臉辨識的處理會被視為一個監督式學習(supervised learning)的問題—根據事前準備每個人的影像所形成的訓練資料，使用不同的機器學習(machine learning)機制學習出人臉辨識的分類器，而對於未知的人臉影像以此分類器來判斷此人臉的身分。過去的研究中，各種不同的分類器設計已經被廣泛的討論過，例如 eigenface [12]、PCA 與 LDA [2]、或者是 SVM [3] 都是很常見的方法。此外，在 [11] 與 [16] 這兩篇論文中，對於影像上的人臉辨識問題與各種研究方法，都有很詳細的討論與研究。

近來，研究人員將焦點移至影片中的人臉辨識問題，也就是希望判斷在一個影片中出現人臉的身分。由於一段影片可以視為很多的單張影像連續播放所形成，因此理論上來說，對於影片中的人臉辨識問題，可以簡化成對這些連續影像的人臉辨識問題，然後把這些辨識結果整合起來。然而這樣的思

考太過於單純了，因為在影片內連續的影像所包含的時間資訊(temporal information)並沒有被考慮到。舉例來說，影片中一個人的臉部姿勢會一直改變，這些姿勢變動對單張影像的人臉辨識會很困難，於是我們如果對於整個影片的人臉辨識中可以加入姿勢變動的時間資訊，對於整體的辨識成功會很有幫助。最簡單的整合影片內各張影像的人臉辨識結果是使用投票的方式 [9]；更詳細分析連續影像在特徵向量空間中所形成的 manifold，可以針對每個人學習獨自的 manifold 而達到辨識的效果，因此這個觀點可以把人臉辨識的問題，轉換成在高維空間上 manifold 的比對 [1][6][13]。此外，有些研究人員針對影片中連續出現的人臉資訊，建構對應的 3D 模型，因此將人臉辨識的問題轉換成 3D 模型的比對與辨識的問題 [4][7][10]。

本研究主要的目標在處理影片中人臉辨識的問題，我們假設影片拍攝在一個固定單純的環境之下，例如在一個固定教室的場景內，所欲辨識的學生人數是固定，且事前可以獲得學生的照片等訓練資訊。我們假設有 K 個人員在整個系統內，影片中這些人可能會出現各種不同的臉部姿勢與角度；我們的研究目標在於建立一個機率模型來辨識哪個人出現在影片中。

在本論文中，我們的基本想法是以類似視覺追蹤(visual tracking)的方式來處理人臉辨識的問題，也就是在影片中隨著時間的演進，對 K 個可能出現候選者的挑選過程進行追蹤，而這個追蹤的歷程是根據觀察每張影片畫面所呈現的視覺特徵(visual feature)而定；這樣的想法啟發我們採用狀態空間模型(state space model)來建構人臉辨識的機率模型。狀態空間模型已經廣泛應用在控制時間資訊相關的動態系統上。我們所建構的機率模型將影片中的人臉辨識問題劃分為兩大部分：可能性評估與遞移性評估。在可能性評估的工作上，就如同傳統單張影像上的人臉辨識問題，判定在目前觀察到的影片畫面上出現者的身分，這部分我們採用 2DLDA 方法來實現單張影像的人臉辨識問題，而決定可能性評估的結果。在遞移性評估方面，主要是判定每個時間點的狀態估計會轉移到其他狀態的可能性，也就是說當可能性估計出現錯誤時，可以藉由遞移性估計的部份，提供一個轉移到正確狀態的機會。我們希望藉由整合可能性評估與遞移性評估兩個部份，在我們的模型中能同時包含傳統的

人臉辨識工作與影片中人臉姿勢的時間資訊。

本論文的組織架構如下。在第二節內我們會簡單說明狀態空間模型的運作基礎，第三節將介紹我們建構的機率圖形模型如何處理影片中的人臉辨識問題。然後我們會在第四節內說明如何使用2DLDA來計算可能性評估的值，而遞移性評估的運作方法會在第五節內介紹。第六節裡我們展示數個實驗的內容，以顯示本研究所提出方法的效能。最後，我們會在第七節提出本論文的結論，並討論未來可能繼續的研究主題與方向。

二、狀態空間模型

狀態空間模型的主要基礎是以貝氏網路(Bayesian Network)來分析一個動態系統，所謂的動態系統是指在時間的演進之下，根據觀察到的資訊來估計系統目前的狀態[5][8]。一個狀態空間模型在時間 t 時包含了兩類的節點：(i)表示系統狀態的 x_t ，(ii)表示觀察到的資訊 z_t 。圖1中顯示了一個典型狀態空間模型的圖形模式。

為了後續能簡潔的以數學模式來表示我們方法，在時間 t 的情形下，我們將使用 $X_t=\{x_1, \dots, x_t\}$ 代表在系統的狀態，而 $Z_t=\{z_1, \dots, z_t\}$ 則代表著觀察到的資訊。在使用狀態空間模型時一般都有兩個基本的假設，這些假設可以由貝氏網路的d-separation特性[8]所推演而得。第一個假設稱之為第一階的馬可夫特性(the first-order Markov property)，也就是：

$$p(x_t | X_{t-1}) = p(x_t | x_{t-1}) \quad (1)$$

而第二個假設是模型中的觀察資訊彼此是互相獨立的，也就是：

$$p(z_t | X_t, Z_{t-1}) = p(z_t | x_t) \quad (2)$$

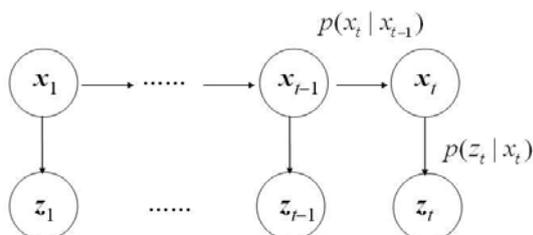


圖 1 狀態空間模型的機率圖形結構

三、人臉識別處理的架構

這一節我們將針對影片中人臉辨識的問題，以貝氏網路為基礎來建構狀態空間模型。假設系統中有 K 個人員，因此狀態向量(state vector) x_t 指向在時間 t 系統所辨識出的人員；同樣的，觀察資訊 z_t 代表著在這個時間系統所取得的人臉資訊。於是，觀察資訊集合 $Z_t=\{z_1, \dots, z_t\}$ 蒐集了影片在這個時間內的人臉影像，而 $X_t=\{x_1, \dots, x_t\}$ 代表著對這些觀察資訊的人臉辨識結果。

然而，以圖1顯示的基本狀態空間模型來處理人臉辨識時，將會因影片中人臉姿勢的改變而無法

得到準確的結果，因此我們必須將人臉姿勢的因素考慮進來。假設一個人的臉部姿勢可以分成 R 種，我們以 $H=\{h_1, \dots, h_R\}$ 表示。在我們建構的機率模型中，我們將在基本的基本狀態空間模型加入人臉姿勢的節點當成系統的事前資訊，加入後的模型如圖2所示。

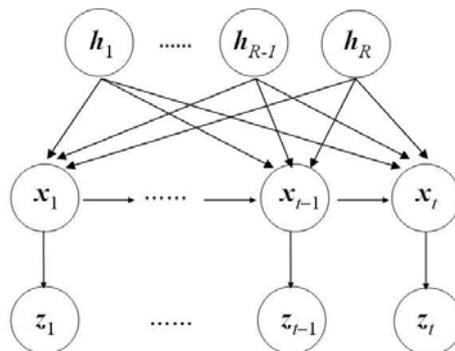


圖 2 我們針對影片中人臉辨識問題所建立的機率圖形模式

定理：

假設在圖2的圖形機率模型內，人臉姿勢為 $H=\{h_1, \dots, h_R\}$ 且系統的觀察資訊在時間 t 是 $Z_t=\{z_1, \dots, z_t\}$ ，則系統狀態 x_t 的事後機率可以由下列式子所計算

$$p(x_t | Z_t, H) \propto p(z_t | x_t) \int p(x_t | x_{t-1}, H) p(x_{t-1} | Z_{t-1}, H) dx_{t-1} \quad (3)$$

證明：

根據式(1)與(2)，同時套用貝氏網路的d-separation [8]來分析圖2的模型，我們可以得到下面四個條件獨立的特性

$$p(x_t | X_{t-1}, H) = p(x_t | x_{t-1}, H),$$

$$p(z_t | x_t, Z_{t-1}, H) = p(z_t | x_t),$$

$$p(x_t | X_{t-1}, Z_{t-1}) = p(x_t | X_{t-1}),$$

與 $p(H | X_t, Z_t) = p(H | X_t)$ 。因此

$$\begin{aligned} p(x_t | Z_t, H) &\propto \int p(X_t, Z_t, H) dX_{t-1} \\ &= \int p(H | X_t, Z_t) p(X_t, Z_t) dX_{t-1} \\ &= \int p(H | X_t) p(X_t, Z_t) dX_{t-1} \\ &= \int p(H | X_t) p(z_t | x_t) p(x_t | X_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\ &= \int \frac{p(H, X_t)}{p(X_t)} p(z_t | x_t) p(x_t | X_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\ &= \int \frac{p(H, x_t | X_{t-1})}{p(x_t | X_{t-1})} p(z_t | x_t) p(x_t | X_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\ &= p(z_t | x_t) \int p(x_t | X_{t-1}, H) p(H | X_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\ &= p(z_t | x_t) \int p(x_t | X_{t-1}, H) p(H | X_{t-1}, Z_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\ &= p(z_t | x_t) \int p(x_t | X_{t-1}, H) p(H, X_{t-1}, Z_{t-1}) dX_{t-1} \\ &\propto p(z_t | x_t) \int p(x_t | x_{t-1}, H) p(x_{t-1} | Z_{t-1}, H) dx_{t-1} \end{aligned}$$

而得證。

因此，要計算狀態 x_t 的事後機率可以由三個項次所決定：(i) $p(z_t|x_t)$ 代表對現在的觀察資訊的可能性評估、(ii) $p(x_t|x_{t-1}, H)$ 代表基於人臉姿勢的遞移性評估資訊、(iii) $p(x_{t-1}|z_{t-1}, H)$ 是前一次疊代計算的遞迴結果。關於可能性評估與遞移性評估，我們分別在第四與第五節詳細說明。整個系統在一開始需要一個系統狀態的初始值，我們同樣的使用 2DLDA 來判別影片中剛開始的人臉資訊。

四、使用 2DLDA 實現單張影樣人臉辨識來計算可能性評估

可能性評估，也就是在式(3)內的 $p(z_t|x_t)$ ，代表目前對特定系統狀態(也就是某個人)符合目前觀察資訊的可能性。可能性評估可以由目前的觀察資訊與每個人的人臉訓練資訊之間的相似度所估計。因此，可能性評估的計算，可以視為在影片中對各張影像進行傳統的單張影像人臉辨識工作。

本研究中採用 2DLDA (two-dimensional linear discriminant analysis) [15]來完成單張影像人臉辨識的工作。2DLDA 執行 IMLDA (uncorrelated image matrix-based linear discriminant analysis) [14] 兩次，一是水平一是垂直方向如圖 3 所示。理論上，2DLDA 會挑選出影像內在水平與垂直方向上最具分辨性的特徵資訊。

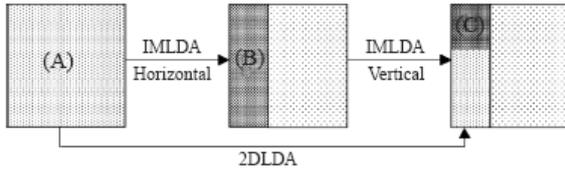


圖 3 2DLDA 運作的示意圖

五、遞移性評估

遞移性評估，也就是在式(3)內的 $p(x_t|x_{t-1}, H)$ ，代表在系統中評估錯誤時會有改變的可能性。根據以下的推導：

$$\begin{aligned} & p(x_t | x_{t-1}, H) \\ &= \frac{p(H | x_t, x_{t-1})p(x_t | x_{t-1})p(x_{t-1})}{p(H | x_{t-1})p(x_{t-1})} \quad (4) \\ &= p(x_t | x_{t-1}) \frac{p(H | x_t, x_{t-1})}{p(H | x_{t-1})} \end{aligned}$$

我們可以將遞移性評估分成兩個部份：一個是不考慮人臉姿勢時的遞移性資訊 $p(x_t|x_{t-1})$ ，一是在前後次疊代計算中人臉姿勢的關係。

關於第一項的 $p(x_t|x_{t-1})$ ，我們由每個人的人臉影像資訊在特徵空間上的相似狀況來學習此項的遞移性評估，為了與可能性評估的計算一致，這裡我們仍然將人臉資訊投影到 2DLDA 的向量空間上，據以計算每個人之間的相似度，然後用高斯分佈加以正規化。

關於第二項的 $p(H|x_t, x_{t-1})/p(H|x_{t-1})$ ，這一向很難以數學方式加以簡化成一個單純的算式，因此我們採用近似的估計方式：視為在連續時間 t 與 $t-1$

之間人臉姿勢轉換的可能性。舉例來說，在正臉(旋轉角度 0°)時轉到一個較小角度的姿勢(例如 30°)的機會，會比轉到一個較大角度的姿勢(例如 60°)的機會大。以此方法計算這一項的遞移性評估有兩件事情要完成。首先得判斷每一次人臉所呈現的姿勢，這一點我們仍然以 2DLDA 建立人臉姿勢的分類器。其次，我們必須定義好每個人臉姿勢角度之間的轉移機會，我們統計了許多影片中人臉滋事變動的順序，並整理出表 1 的轉換機率。請注意表 1 所呈現的機率是非對稱性的，例如在右側臉時轉到正臉時的機會，當然會比反過來的高。

表 1 人臉不同姿勢之間的轉換機率

$t-1 \backslash t$	Front	Right	Left	Up	Down	Right up	Left up
Front	0.2778	0.1667	0.1667	0.1667	0.1667	0.028	0.0278
Right	0.24	0.4	0.04	0.08	0.08	0.04	0.12
Left	0.24	0.04	0.4	0.08	0.08	0.12	0.04
Up	0.2222	0.1111	0.1111	0.3704	0.037	0.074	0.0741
Down	0.24	0.12	0.12	0.04	0.4	0.04	0.04
Right up	0.25	0.0417	0.125	0.0833	0.0417	0.417	0.0417
Left up	0.25	0.125	0.0417	0.0833	0.0417	0.042	0.0417

六、實驗結果

本實驗採用 Honda/UCSD Video Database [6] [17]為實驗的人臉資料影片，在此資料庫中，我們任選 10 個人的影片，圖 4 列舉這 10 個人的照片與名字，影片中每個人都會一定頭部造成臉部姿勢的改變。對於每個人的影片，我們取一半當訓練資料，另一半為測試資料。

第一個實驗是評估對影片內人臉辨識的正確率，表 2 列出由式(3)所完成辨識的每個人平均辨識正確率；同時我們採用 2DLDA 與 PCA 來完成人臉辨識與姿勢辨識以供比較。例如，“PCA+2DLDA”代表在可能性評估時使用 PCA 方法對單張影像辨識人臉，而在計算遞移性評估時使用 2DLDA 辨識人臉姿勢。此外，我們依照第四與第五節的描述進行但不計算 $p(x_t|x_{t-1})$ ，而紀錄成”(2DLDA+2DLDA) without transition”以顯示我們設計的遞移性評估所達成的效果。表 2 清楚的顯示我們的方法，也就是第一欄的”(2DLDA+2DLDA) with transition”，可以比其他的作法有更多的效能改進。

此外，我們嘗試討論本架構在包含可能性評估與遞移性評估後，隨著時間演進的收斂狀況。圖 5 包含一個範例，顯示在時間 8、14、23、與 28 的人臉辨識結果。我們試驗的人是”James”，在圖 5 橫軸顯示的人員代號是 4，這個例子是 James 由正臉轉向左臉。剛開始，觀察影像被辨識成另一個人”yokoyama”，但是經過時間的演進，透過不斷的可能性評估的修正與遞移性評估的輔助，最終他在時間 28 時就被正確辨識出。為了簡化圖表的展現，

圖 5 只同時顯示計算過程中三個人的機率變化，雖然我們的實驗實際上一次計算十個人的辨識。



圖 4 實驗資料使用的人員影片資訊

表 2 人臉不同姿勢之間的轉換機率

Name	2DLDA+ 2DLDA+ with transition	2DLDA+ 2DLDA+ without transition	PCA+ 2DLDA+ with transition	PCA+ PCA+ with transition
Behzad	0.91906	0.8225	0.86684	0.84334
Danny	0.82289	0.7847	0.68119	0.65938
Fuji	0.65862	0.6621	0.65862	0.67931
James	0.88179	0.77	0.96166	0.96166
Jeff	0.73801	0.7168	0.64425	0.60708
Joey	0.66567	0.5701	0.43582	0.49851
Ming	0.76607	0.7404	0.46787	0.49356
Rakesh	0.98201	0.9537	0.94087	0.94858
Wei	0.77236	0.7175	0.68293	0.66666
Yokoyama	0.85953	0.7559	0.69565	0.66889
AVG	0.806601	0.74937	0.70357	0.70269

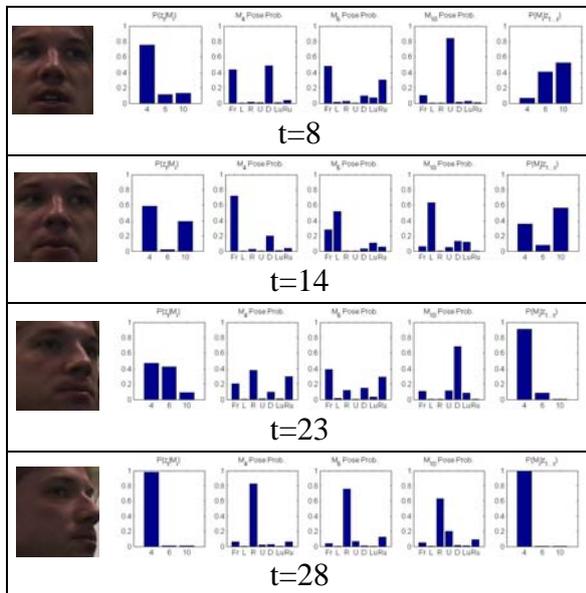


圖 5 本論文架構下對影片中人臉辨識之範例，顯示由時間 8 到 28 的辨識過程中，辨識結果由最初的錯誤到最後收斂於正確。

七、結論與未來展望

這篇論文中，我們提出一個機率圖形模型的架

構，以處理影片中人臉辨識的問題。我們的架構包含了可能性評估與遞移性評估這兩個主要的部份，在此架構之下，不但可以處理傳統的單張影像內人臉辨識的問題，也將影片中人臉姿勢的變動資訊納入，進而改善辨識的準確度。在單張影像內人臉辨識的問題上，我們採用 2DLDA 來建立人臉辨識的分類器。同時我們也設計一套處理遞移性評估的方式，可以將人臉姿勢的異動來輔助整個人臉辨識的進行。

關於本研究未來的展望，首先我們將使用更多人臉的影片資訊當實驗的資料，讓我們的方法能夠有更多的實驗數據當支撐。我們正在研究對本論文中的機率圖形模型是否有改善的空間，以便更精準的描述影片中人臉辨識的問題，達到更準確的辨識結果。此外，關於人臉姿勢的辨識與轉換問題，也是很值得繼續深入著墨之處。

參考文獻

- [1] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell, "Face Recognition with Image Sets Using Manifold Density Divergence," in Proceedings of CVPR, 2005.
- [2] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Trans. on PAMI, Vol. 19, No. 1, pp. 711-720, 1997.
- [3] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach," in Proceedings of ICCV, 2001.
- [4] Y.-X. Hu, D.L. Jiang, S.-C. Yan, L. Zhang, and H.-J. Zhang, "Automatic 3D reconstruction for face recognition," in Proceedings of International Conference on Automatic Face and Gesture Recognition, 2004.
- [5] Z. Ghahramani, "An Introduction to Hidden Markov Models and Bayesian Networks", International Journal of Pattern Recognition and Artificial Intelligence, 15(1): 9-42, 2001.
- [6] K.C. Lee and J. Ho and M.H. Yang and D. Kriegman, "Visual Tracking and Recognition Using Probabilistic Appearance Manifolds," Computer Vision and Image Understanding, Vol. 99, No. 3, pp. 303-331, 2005.
- [7] X.-G. Lu, A.K. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," IEEE Trans. on PAMI, Vol. 28, No. 1, pp. 31-43, 2006.
- [8] K. P. Murphy, "Dynamic Bayesian Networks: Representation, Inference and Learning", U. C. Berkeley, PhD. Thesis, 2002.
- [9] G. Shakhnarovich, J. W. Fisher, and T. Darrell, "Face recognition from long-term observations," in Proceedings of ECCV, pp. 851-865, 2002.
- [10] F. B. ter Haar and R. C. Veltkamp, "3D Face Model Fitting for Recognition," in Proceedings of ECCV, 2008.
- [11] A. S. Tolba, A.H. El-Baz, and A.A. El-Harby,

- “Face Recognition: A Literature Review”,
International Journal of Signal Processing, Vol.
2, No. 1, pp. 88-103, 2005.
- [12] M. Turk and A. Pentland, “Eigenfaces for
recognition”, Journal of Cognitive
Neuroscience, Vol. 3, pp. 72-86, 1991.
- [13] R.-P. Wang, S.-G. Shan, X.-L. Chen, and W.
Gao, “Manifold-Manifold Distance with
application to face recognition based on image
set,” in Proceedings of CVPR, 2008.
- [14] J. Yang, J.-Y. Yang, A.F. Frangi, and D. Zhang,
“Uncorrelated projection discriminant analysis
and its application to face image feature
extraction,” International Journal of Pattern
Recognition and Artificial Intelligence, Vol. 17,
No. 8, pp. 1325–1347, 2003.
- [15] J. Yang, D. Zhang X. Yong, and J. Yang,
“Two-dimensional Discriminant Transform for
Face Recognition,” Pattern Recognition, Vol.
38, No. 7, July 2005.
- [16] W. Zhao, R. Chellappa, P.J. Phillips, and A.
Rosenfeld, “Face recognition: a literature
survey”, ACM Computing Surveys, 35(4):
399-458, 2003.
- [17] The Honda/UCSD Video Database,
[http://vision.ucsd.edu/~leekc/HondaUCSDVide
oDatabase/HondaUCSD.html](http://vision.ucsd.edu/~leekc/HondaUCSDVideoDatabase/HondaUCSD.html).