

國立臺灣師範大學
資訊工程研究所碩士論文

指導教授：李忠謀 博士

利用機率圖模型於影片上之人臉辨識研究

Video-Based Face Recognition Using A

Probabilistic Graphical Model

研究生：詹依佳 撰

中華民國 九十八 年 六 月

摘要

利用機率圖模型於影片上之人臉辨識研究

詹依佳

針對影片上的人臉辨識問題，本論文提出一個機率圖模型來解決並將其公式化。首先，我們將此問題分成兩個部份來探討，分別為相似度之計算與遞移機率，其中相似度之計算可被視為傳統的單張影像之人臉辨識的結果，在此篇論文中，我們採用二維線性鑑別分析法(2DLDA)摘取特徵，再藉由高斯分佈來估算相似度。而遞移機率則是計算先前時間點的狀態轉移到此時間點的狀態之機率，我們可將遞移機率分成兩個部份來估算，其一為人與人的遞移機率，另一個則為姿勢轉換的遞移機率，希望藉由相鄰影像的時間關係修正錯誤的辨識結果與提升準確率。在本論文的實驗中，我們使用在國際上常採用的 Honda/UCSD 資料庫以及本實驗室自行建立的VIPlab資料庫。實驗證明本研究提出之方法可適用於不同的資料庫，且實驗結果也有90%以上的正確率。

關鍵字：人臉辨識、機率圖模型、二維線性鑑別分析法、高斯分佈

ABSTRACT

Video-Based Face Recognition Using A Probabilistic Graphical Model

by

Yi-Chia Chan

We present a probabilistic graphical model to formulate and deal with video-based face recognition. Our formulation divides the problem into two parts: one for likelihood measure and the other for transition measure. The likelihood measure can be regarded as a traditional task of face recognition within a single image, i.e., to estimate how similar to a specified person this observing face image is. In our work, two-dimensional linear discriminant analysis (2DLDA) is employed for feature extraction, and then we use a Gaussian distribution to assess the likelihood measure. The transition measure is estimated via two terms, person transition and pose transition. The transition terms could fix some incorrect recognition results because of considering the information between adjacent frames. In the face recognition experiments, we adopt two datasets, Honda/UCSD dataset and VIPlab dataset. Finally, it is demonstrated that our proposed approach is robust in different datasets and produces good recognition accuracy which is more than 90%.

Keywords: probabilistic graphical model, face recognition, two-dimensional linear discriminant analysis (2DLDA), Gaussian distribution

誌謝

看著自己的論文完成，真的有說不出的喜悅，這一切都是因為有你們！

在這兩年裡，首先要感謝李忠謀老師的指導，對於我的研究給予絕對的自主權，使我能依照自己的想法做研究，並且適時的提供意見，讓我不會失了方向；感謝政杰學長，在每個星期二早上九點，不管颶風下雨，都到實驗室和我們 meeting，和我一起慢慢建構出論文的雛型，以及每次我在研究上碰壁時，傾聽我的疑問，和我一起尋求解決的方法；感謝同 team 中的強者家維和凱民，因為你們，我才不敢偷懶，才不斷督促自己要追上你們的進度，也感謝你們常聽我抱怨，幫我解決一些研究或程式上的問題，還有感謝靖雅、定翔、建斌和百璋，這兩年，在課業上，我們一起熬夜寫作業，一起討論考試的重點，一起把論文拼出來，在娛樂上，一起過聖誕節，慶祝每個人的生日，以及一起上山下海的遊玩，這些有苦有歡笑的回憶，都還歷歷在目，我會一直牢牢記得這些和如此特別你們的回憶！

最後感謝我的家人，你們始終支持我，不管是物資或是精神上，都給予我最好的，使我能無後顧之憂的完成學業！

研究所的生涯要告一段落了，雖然開心自己完成了學業，但也充滿不捨，我會永遠記得這兩年的一切，再一次感謝大家，僅以本文獻給心愛的你們！



目錄

目錄.....	I
附圖目錄.....	III
附表目錄.....	V
第一章 緒論	1
1-1 研究動機.....	1
1-2 研究目的.....	2
1-3 研究範圍與限制.....	2
1-4 論文架構.....	3
第二章 文獻探討	4
2-1 單張影像上之人臉辨識.....	4
2-2 影片上之人臉辨識.....	8
2-2.1 未加入相鄰影像之資訊.....	9
2-2.2 加入相鄰影像之資訊.....	10
第三章 理論基礎	14
3-1 狀態空間模型(State Space Model).....	14
3-2 二維線性鑑別分析法(2DLDA).....	15
第四章 機率圖模型之人臉辨識	20
4-1 機率圖模型與公式推導.....	20
4-2 相似度之計算.....	23
4-3 遞移機率.....	24
4-3.1 人與人的遞移機率.....	25
4-3.2 姿勢轉換的遞移機率.....	26
4-4 流程與方法總結.....	27
第五章 實驗結果與分析	31
5-1 Honda/UCSD 資料庫.....	31
5-2 本研究使用 Honda/UCSD 資料庫之實驗.....	32
5-2.1 訓練流程.....	33
5-2.2 辨識結果分析.....	36

5-2.3 辨識率估算	38
5-2.4 機率分佈圖之探討	39
5-3 單張影像上人臉辨識技術之比較	40
5-4 使用 VIPlab 資料庫之辨識結果	41
第六章 結論與未來展望	44
參考文獻	45

附圖目錄

圖 2-1 人臉辨識技術之步驟	5
圖 2-2 支持向量機(SVM)之示意圖	6
圖 2-3 PCA 和 ICA 所建立的特徵臉影像.....	6
圖 2-4 資料在 PCA 及 ICA 上的分布情況	7
圖 2-5 上排為 2DPCA 重建後的結果；下排為 PCA 重建後的結果.....	8
圖 2-6 個人人臉表示法(identity surfaces)	10
圖 2-7 一串有序的人臉影像之隱藏式馬可夫模型	11
圖 2-8 M_k 這個人在低維空間的表示法	12
圖 2-9 一串有序影像之辨識情況	12
圖 2-10 一般的人臉機率模型和個人的人臉機率模型之轉換	13
圖 3-1 狀態空間模型之機率圖模型	14
圖 3-2 2DLDA 示意圖	16
圖 4-1 傳統的狀態空間模型	21
圖 4-2 PGFR 機率圖模型	21
圖 4-3 人與人之遞移機率之計算方式	26
圖 4-4 2DLDA 的人臉姿勢之分類	27
圖 4-5 本研究之訓練步驟	29
圖 4-6 本研究之人臉辨識演算法	30

圖 5-1 在 Honda/UCSD 資料庫中，原始影像的一些例子.....	32
圖 5-2 在 Honda/UCSD 測試用之資料庫中，截取出人臉影像的一些例子.....	33
圖 5-3 使用 2DLDA 將高維資料降至各維度之比較.....	35
圖 5-4 未加入遞移機率與本研究方法在各維度上之比較.....	38
圖 5-5 在四個時間點的機率分部情況.....	40
圖 5-6 在 VIPlab 資料庫中，原始影像的一些例子.....	42

附表目錄

表 5-1 影像解析度之大小對辨識率的影響	34
表 5-2 前處理方法之比較	34
表 5-3 使用 2DLDA 將高維資料降至低維後的辨識率	35
表 5-4 從訓練用之人臉影像所計算出的不同人之姿勢遞移機率表	36
表 5-5 人臉姿勢變化的遞移機率表	36
表 5-6 單張影像之人臉辨識與加入人和人的遞移機率之比較	37
表 5-7 單張影像之人臉辨識與加入兩種遞移機率之比較	38
表 5-8 不同的 window size 所測出的辨識率	39
表 5-9 本研究方法和三種單張影像之人臉辨識技術之比較	41
表 5-10 使用本研究方法於 VIPlab 資料庫與 Honda/UCSD 資料庫上之辨識率	43

第一章 緒論

1-1 研究動機

在現今的社會中，大家愈來愈重視自身的隱私和安全，因此身份認證已經佔個人生活中不可或缺的一環，更由於此，你我每天都會隨身攜帶著各式各樣的身份證明文件，舉凡身份證、駕駛執照、健康保險卡、金融卡、學生證、公司識別證等。上述這些傳統識別方法均無法有效的防止偽照、盜用等情況，降低了使用證件的安全性與可靠性。傳統識別方法只靠卡片加上密碼，或單純的以密碼來加以維護，一但遺失卡片，或忘記密碼，或被人竊取，都將會造成使用者極大的損失與不便。因此，近年來學術界與科技界都熱衷於尋求更好的方式來替代傳統識別方法，使得身份識別能具有更強大的監控能力與安全性，並提升使用上的便利性與可靠性。

由於科技進步與電腦運算速度增快，因此利用生物特徵做為識別的方法為近十年來的主要趨勢。目前較為大家所熟知的辨識技術包括人臉辨識、指紋辨識、語音辨識與虹膜辨識等方法，皆可透過電腦等硬體設備自動的來運作，使用者不需要再牢記一堆擾人的密碼，大大提升了身份識別的方便性。在這些方法中，人臉辨識無需運用特殊設備，無需繁雜的設定，無需接觸到機器，只需透過簡單的攝影機裝置，即可得到辨識時所需的資料，比其他方法更具便利性。人臉辨識若結合視訊監控系統，應用範圍相當廣泛，舉凡:場所的人員進出管制及身份確認等，因此人臉辨識的研究在生物認證技術上又為最活絡的一項技術。

1-2 研究目的

傳統的人臉辨識是以單張影像為辨識基礎，在單張影像中，過大的環境變化，會產生連電腦甚至是人眼都無法判斷的情況，因此單張影像的辨識結果通常是不可靠且容易出錯的。近年來，開始有較多人投入以影片為基礎的人臉辨識研究，即不單單只靠一張影像，武斷的判斷結果，而是藉由多張影像來決定最終的辨識結果，因此，除了能有效的提升辨識率之外，也能提升此辨識結果的信任度。

在此篇論文中，我們將影片上的單張影像辨識與時間資訊做結合，希望當二維影像降維後，所得到的特徵向量能更加簡潔且更具鑑別力，並且能有效的利用這一串資訊的時間關係，進而達到最好的辨識效果。

1-3 研究範圍與限制

給定一串有序影像，首要步驟需先使用人臉偵測之技術決定人臉是否在影像中，並定位出人臉位置。本研究假設，每段有序影像都已先行完成人臉截取與追蹤，因此只針對人臉辨識的方面進行討論。

此篇論文中，我們將針對一串有序之人臉影像做研究，希望所測試的人臉影像能趨近於真實世界的情況，是不受限制的。在此種狀況下，我們需面對以下的问题：

- 姿勢(Pose): 在一串有序影像中，人臉可能在某些時間點有姿勢的轉變，舉凡左臉 90°、左臉 45°、正臉、右臉 45°、右臉 90°、向上、向下，這七種姿勢，本研究希望在處理這七種姿勢的情況下，都能達

到準確的辨識結果。

- 臉部表情(Facial Expression): 在一串有序影像中，人臉可能在某些時間點的
 臉部特徵產生變化，如:閉眼睛，大笑，驚訝等，會直接影響到
 臉部整體形狀的改變，因而造成辨識上的困難，本研究希望當
 碰到上述問題時，仍能得到準確的辨識結果。
- 遮蔽(Occlusion):在一串有序影像中，人臉可能在某些時間點被其他物體遮
 蓋，如:頭髮、眼鏡、手等，本研究希望即使人臉有部份被遮蓋，
 仍可以得到準確的辨識結果。

1-4 論文架構

本論文共分成六個章節，第一章為緒論，第二章為研究之文獻探討，在單張影像和一串有序的影像上，對於前人方法加以討論，第三章及第四章則詳述本研究所提出之方法與步驟，第五章為實驗之結果與討論，而第六章為本研究之結論與未來展望。

第二章 文獻探討

有許多的學者曾致力於人臉辨識上的研究，我們大致可以將人臉辨識的研究方向分為二類[1][33][28]:

1. 單張影像上之人臉辨識

- 以單張影像進行人臉辨識

2. 影片上之人臉辨識

- 未利用相鄰影像之資訊:

(1) 給定一串有序影像，藉由單張影像之人臉辨識技術加上投票機制來

求得辨識結果。

(2) 在一串有序影像上建造二維或三維的人臉結構來進行人臉辨識。

- 利用相鄰影像之資訊:

(1) 在單張影像的辨識技術中加入一串有序影像所隱含之重要資訊，建

立屬於此人的機率模型來求得辨識結果。

下列將對於上述各類的研究之優缺點進行初步之探討。

2-1 單張影像上之人臉辨識

首先，在討論單張影像之人臉辨識前，必須先建立一個觀念，「辨識時，若以原始的人臉影像直接當作測試值或訓練值，是不可行的」，在原始高維空間中做運算，不僅計算量會很大並且容易受到雜訊的干擾，因此在人臉辨識前，會先

將人臉影像以保留最大特徵的方式來降維，而後在低維的子空間中再進行比對，

如圖 2-1[1]。

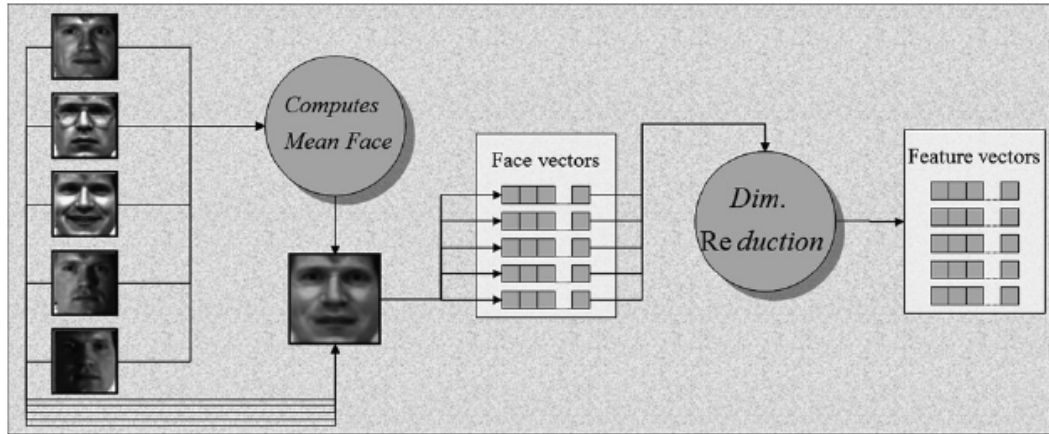


圖 2-1 人臉辨識技術之步驟[1]

在單張影像之人臉辨識中，大多數的方法都是將整張的人臉影像當作輸入，待完成降維的動作後，再行辨識，因此降維方法衍然成為辨識率或高或低之一大關鍵。而不同的人臉辨識技術常會伴隨著不同的降維方法，因此我們探討人臉辨識之主要技術如下：

- 特徵臉(EigenFaces)[29][2]:先對輸入之人臉影像以主要成份分析法(Principal Component Analysis, PCA)[24]之技術把變異較大的資訊保留下來，而後求得屬於此影像的特徵臉。再以最近距離法(Nearest Neighbor classifier, NN)來比對，此為人臉辨識運用最為廣泛的方法。
- 費雪臉(FisherFaces,FLD)[2][18]:採用線性鑑別分析法(Linear Discriminant Analysis, LDA)[24]之技術，把最具有鑑別力的資訊保留下來。在線性鑑別分析法(LDA)所求得之子空間中，能滿足以下特點:降維後，可拉開不同人的人臉影像之彼此距離並且拉近同一個人的人臉影像之彼此距離。

- 支持向量機(Support Vector Machine, SVM)[26][10]:運用支持向量機(SVM)之技術來完成不同人臉之分類。支持向量機(SVM)，簡單來說就是當資料降維後還是沒辦法分開時，能找到一組超平面(Hyperplane)，把降維後的資料再升維，讓不同的集合得以分開。在解決非線性以及高維的識別問題中表現出許多特有的優勢，如圖 2-2。

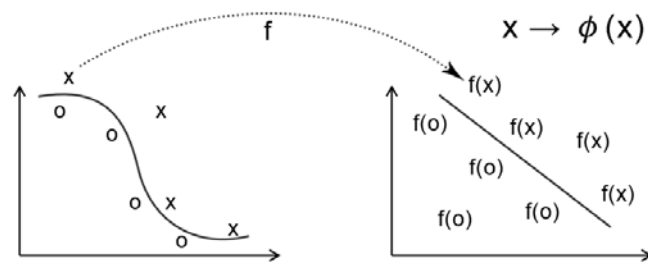


圖 2-2 支持向量機(SVM)之示意圖

- 獨立成份分析法(Independent Component Analysis, ICA)[3][6]:獨立成份分析法(ICA)可被視為主要成份分析法(PCA)的延伸。在應用上主要成份分析法(PCA)強調資料維度的降低，而獨立成份分析法可能降低維度、維度不變、或增加維度，因此獨立成份分析法(ICA)比主要成份分析法(PCA)更能廣泛應用於現實中。如圖2-3[6]、圖2-4[3]。



圖 2-3 PCA 和 ICA 所建立的特徵臉影像[6]

上排為取前 8 大特徵根所建立的特徵臉影像
下排為用 ICA 取得的 8 個之特徵向量

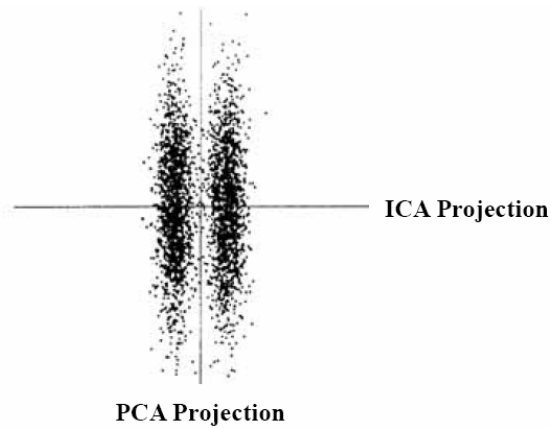


圖 2-4 資料在 PCA 及 ICA 上的分布情況[3]

圖中資料分佈明顯分成兩群，最大變異方向為垂直的軸(PCA Projection)，而獨立成份分析法則找出能將資料分成兩群的方向也就是橫軸(ICA Projection)

■保留局部關係之投影(Locality Preserving Projections, LPP)[11]:主要成份分析法

(PCA)及線性鑑別分析法(LDA)是保留全域結構(global structure)，不會考慮資料與資料之間的相關性。保留局部關係之投影(LPP)則考慮了不同資料之間的相關性，較能有效的描述多樣化(manifold)的結構，如某個人在某一時刻由各個角度同時拍下的照片，用保留局部關係之投影(LPP)較能描述此人臉之結構。

以上所提出的人臉辨識之技術皆是把二維人臉影像轉成一維後再做運算，如此便會失去一些原始空間中點與點之間相關性的資訊。在近年來的研究中，如：二維主要成份分析法(2DPCA)[30]、二維線性鑑別分析法(2DLDA)[31]、二維保留局部關係之投影(2DLPP)[5]皆是直接對二維的影像做降維的動作，即是不需要把原始影像轉成一維的向量方可做運算。在[30]、[31]、及[5]中，皆有實驗證明二維的辨識效果比一維的較佳，如圖 2-5[30]，而另一方面也提及了，當直接對二維影像做處理時，無論空間跟時間之複雜度都比一維影像來的降低許多。

在[5]的實驗中，將上述三種二維的方法之辨識率與訓練速度做比較。其中三種二維方法皆違反了一維方法的一種常理現象，「當降維後保留越多的維度時，此三種二維方法的辨識率反而都會降低」，因此當使用二維的方法做降維時，保留的維度不能太高。再者，當降維到較低維度時，比較這三種二維的方法，二維線性鑑別分析法(2DLDA)之辨識率會高於二維主要成份分析法(2DPCA)，更值得提及的是，雖然二維線性鑑別分析法(2DLDA)之辨識率比二維保留局部關係之投影(2DLPP)低一些，但是在訓練之速度上，二維線性鑑別分析法(2DLDA)卻比二維保留局部關係之投影(2DLPP)來的快得多，因此，上述的三種二維的方法，又以二維線性鑑別分析法(2DLDA)兼具準確度與速度，能擁有最好的效能。

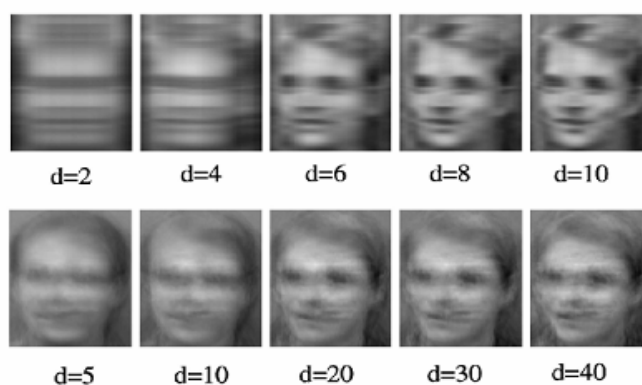


圖 2-5 上排為 2DPCA 重建後的結果；下排為 PCA 重建後的結果[30]

2-2 影片上之人臉辨識

在現今科技發達的時代，攝影機的價位已不再昂貴，因此目前的人臉辨識之研究已有越來越多以影片為主，希望能在整合整串有序影像之結果後，來達到即時且辨識率高的辨識結果。

在影片上之人臉辨識，雖然會擁有比單張影像更多的資訊來提高準確率，卻也擁有更多的不確定因素之挑戰需克服。典型的一串有序影像上之人臉辨識方法分成兩個階段:(1)人臉追蹤(2)人臉辨識。當給定一串有序影像時，第一個步驟即是對人臉的位置做追蹤，方可定位，以截取出每張影像中屬於人臉位置之區域，接著再將這些截取出的人臉影像交給第二個步驟做辨識，本研究主要探討的是第二個步驟之方法，因此對於人臉追蹤的部份就不多加討論。

一串有序影像中，相鄰影像間的關係性可用來提升辨識率，以下即針對在影片上做人臉辨識時，是否加入相鄰影像間的關係性之資訊，來做分類及探討。

2-2.1 未加入相鄰影像之資訊

由單張影像之人臉辨識變成一串有序影像之人臉辨識時，剛開始，尚未加入每張影像之間的關係性，即是每張影像之辨識結果是互相獨立的[9][27]，因此我們可以用以下三種方式來求得此串有序的人臉影像之最終辨識結果。當每張影像各自做完單張影像之人臉辨識後，(一) mean rule:將此串影像的所有辨識率取平均，辨識率最大者，即是最終之辨識結果，(二) majority rule:由於每張影像之辨識結果是獨立的，因此可用投票的方式來求得最終之辨識結果，(三) product rule:把所有影像之單張影像的辨識率相乘，值越大者即是最終的辨識結果，上敘不管採用何種方式，在最終之辨識率上，還是會受到單張影像之辨識率的影響。另外，還有許多的研究是利用多角度之人臉影像建造出屬於此人的二維或三維之人臉結構模型[12][16][34][7]，此方式雖然可以有效的改善光線以及角度的問題，但

是，相對來說，建造人臉模型也需要更複雜的程序及演算法，此方式也沒有善加利用有序影像中的時間資訊，因此章節 2-2.2 探討了如何在加入時間的資訊後，在訓練與辨識之過程簡單化的基礎下，提升辨識率之研究。

2-2.2 加入相鄰影像之資訊

一串有序影像相對於單張影像有許多優勢，其中包括了不同的視角、相鄰影像的資訊等，這些資訊都能提升人臉辨識的準確度。起初，在一串有序影像上做人臉辨識時，是將每張人臉影像利用時間資訊串連起來，Li 等人在[19]中所提出之理論，是將每個人的各種姿勢建構成一個屬於個人的人臉(identity surface)，利用截取核心特徵(kernel feature)之演算法來建立一條軌道(trajectory)，使得無論何種角度之人臉都能遵循這條軌道變形成正面之人臉，如圖2-6[19]。然而，此方法除了利用帶有權重的時間性投票機制來解出模型參數外，他並沒有完全去利用相鄰畫面的連貫性資訊。

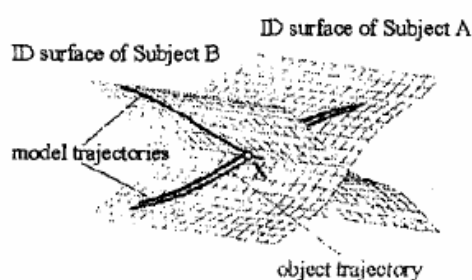


圖 2-6 個人人臉表示法(identity surfaces)[19]

Zhou 和 Chellappa 在[35]中提出，藉由個人可變量(identity variable)與運動向量(motion vector)的加入，可同步解決人臉追蹤與辨識之問題，也將狀態空間模型(state space model)之架構運用於有序影像之分類上。而Liu 和 Chen 在[20]

中提出使用隱藏式馬可夫模型(Hidden Markov Models, HMM)來執行人臉辨識。

一般的馬可夫模型(Markov Model)只能描述狀態(state)與狀態(state)之間的關係，而在隱藏式馬可夫模型(HMM)中，實際改變的狀態(hidden state)是觀測不到的，即是有序影像中人臉姿勢(pose)變化的狀態是無法觀測到的，因此我們只能透過觀測到的觀測值(observation)估計出觀測值和隱藏狀態(hidden state)之間的轉移關係，而能計算出觀測值(observation)所對應之機率密度函數，如圖2-7[20]。

在[20]中，使用主成份分析法(PCA)計算出人臉影像之多個姿勢(pose)中具有強大辨識力之特徵，但是並不適用於人臉影像有多種姿勢(pose)的情況，因而在[17]中，採用線性鑑別分析法(LDA)，可以比主成份分析法(PCA)較容易取得更具鑑別性之特徵，來將同個人不同姿勢區分的更開，而後再使用隱藏馬可夫模型(HMM)來做一串有序影像之人臉辨識。

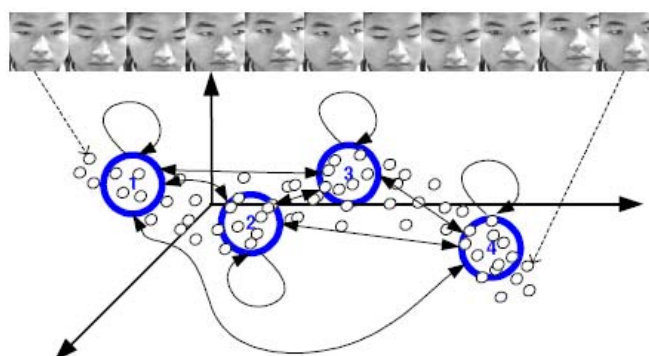


圖 2-7 一串有序的人臉影像之隱藏式馬可夫模型[20]

再者， Lee 等人又在[21]和[22]中提出，將每個人之不同角度的影像降維後，建立一個在低維子空間中屬於自己的人臉外貌(appearance manifold)，如圖2-8[21]所示。其中，他將每個人的每個姿勢(pose)運用主要成份分析法(PCA)，

計算出在低維空間中具代表性的PCA-plane，而後將每個人之不同姿勢的PCA-plane收集起來，即能代表此人在子空間中的人臉外貌(appearance manifold)。接著在辨識時，融合時間之資訊，去找尋在子空間中和此串有序影像最近之人臉外貌(appearance manifold)，如圖2-9[21]，即能得到辨識結果。

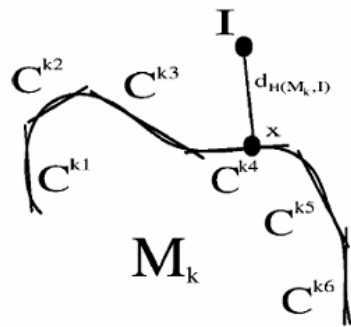


圖 2-8 M_k 這個人在低維空間的表示法[21]

C^{ki} 表示此人第 i 個姿態的 PCA-plane

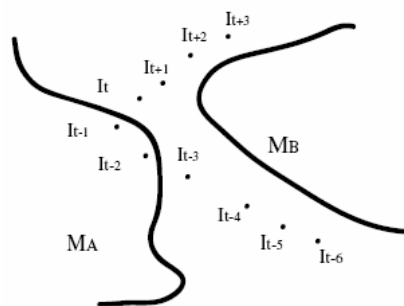


圖 2-9 一串有序影像之辨識情況[21]

$I_{t-6} \dots I_{t+3}$ 是一串動態之人臉影像。雖然在 I_{t-1}, I_{t-2} 發生錯誤之辨識結果，但是受到前後之辨識結果影響，整串影像之最終判斷仍會正確

在[21]中，雖然可以即時的(on-line)做人臉辨識，但是屬於個人的人臉機率模型(appearance manifold)還是必需在事前(off-line)先訓練完成，因此 Lee 等人又在[23]中提出，希望藉由[4]、[13]和[14]的方法，使得個人的人臉機率模型

(appearance manifold)也可以即時的訓練，也就是在辨識的過程中，我們先建立一個一般的人臉機率模型(a general appearance manifold)，而後我們將辨識成功的影像去重新訓練這個一般的人臉機率模型，在重新訓練之次數足夠的情況下，我們即可訓練出屬此於此人的人臉機率模型(a specific appearance manifold)，如圖 2-10[23]。如此，便能改善因為事前訓練，而造成訓練資料跟測試資料差異甚大的情況，但是也必須使用更複雜的數學模型來建立之。

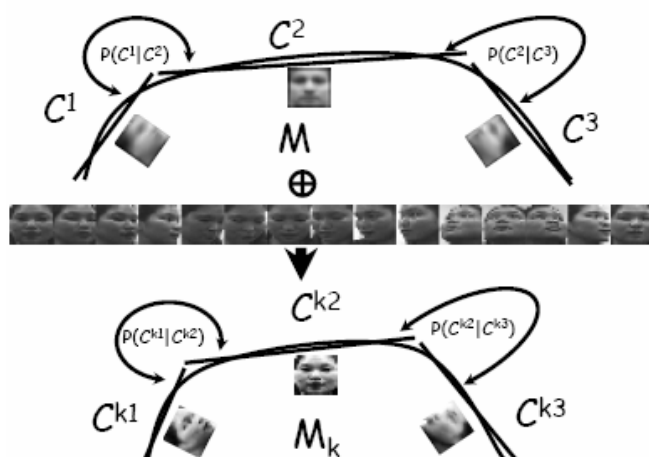


圖 2-10 一般的人臉機率模型和個人的人臉機率模型之轉換[23]

第三章 理論基礎

在介紹本研究之方法前，先提及兩個重要的機率模型與圖形辨識之方法，分別為「狀態空間模型(State Space Model)」[8][25]和「二維線性鑑別分析法(two-dimensional Linear Discriminant Analysis, 2DLDA)」[31]，此兩者在本研究中均佔非常重要的角色，因此在此章中會詳細介紹狀態空間模型與二維線性鑑別分析法。

3-1 狀態空間模型(State Space Model)

狀態空間模型(State Space Model)是以貝式網路(Bayesian network)[25]的架構來建造出的動態機率模型，並且隨著時間的變化來估計此模型的狀態。一般而言，狀態空間模型(State Space Model)，於時間點 t 包含兩種形式之節點，(i) x_t 表示在時間點 t ，此模型的狀態，(ii) z_t 表示在時間點 t 時，所觀測到的觀測值，如圖 3-1，即是狀態空間模型(State Space Model)的機率圖模型：

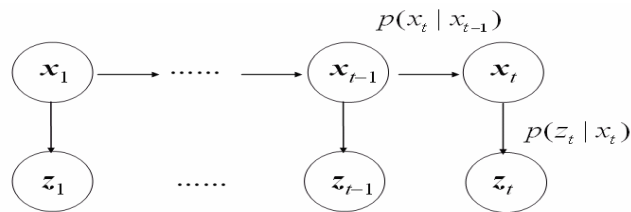


圖 3-1 狀態空間模型之機率圖模型

接著，我們用兩個符號來表示圖 3-1 中兩個節點隨著時間變化所得到的資訊，“ $X_t = x_{1, \dots, t}$ ”表示時間點 1 到時間點 t 中，各個時間點所得到的狀態之集合，

“ $Z_t = z_{1,\dots,t}$ ”表示時間點 1 到時間點 t 中，所觀測到的一串有序觀測值之集合。

參照圖 3-1，我們可以對此狀態空間模型(State Space Model)定出兩個假設：

(i) 第一階馬可夫特性(The first-order Markov property)：每個時間點之

狀態，只受到前一個時間點所產生之狀態影響。

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

(ii) 互相獨立：每個時間點之觀測值，只受到此時間點之狀態影響，並且和先

前時間點的觀測值是互相獨立的。

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

由以上兩個假設，我們即可對此狀態空間模型(State Space Model)利用貝式網路(Bayesian network)中的 d -分割性質(d-separation property)[25]加以應用，第四章會詳細說明如何將此模型運用在人臉辨識之研究上。

3-2 二維線性鑑別分析法(2DLDA)

若有一張 $M \times N$ 維的影像，傳統的線性鑑別分析法(Linear Discriminant Analysis, LDA)[2][24]只能對一維的向量做降維的動作，因此會把此影像的維度轉換成 $1 \times (M \times N)$ 維來執行，接著利用將組內變異變小及組間變異變大之概念，提高不同人之間的差異性(Between-Class Variation)，並同時降低同一人不同影像間的差異性(Within-Class Variation)，在計算上，會產生 $(M \times N) \times (M \times N)$ 的超高維矩陣，若對如此高維度的矩陣求得特徵根(eigenvector)和特徵向量(eigenvector)

時，會增加空間與時間之複雜度，因此線性鑑別分析法(LDA)並不適合用於高解析度的影像上。而二維線性鑑別分析法(2DLDA)[31]，則是直接在二維影像上完成降維的動作，不需先轉換成一維向量即可完成降維，如此一來，即會保留原始影像中空間的相關性，也不會有矩陣維度過大的情況，而缺點就是比線性鑑別分析法(LDA)需要使用更多的係數來代表一張圖。以下逐步說明二維線性鑑別分析法(2DLDA)[31]之運算方式。

二維線性鑑別分析法(2DLDA)，簡單來說，即是對影像實施了兩個階段的IMLDA(uncorrelated image matrix-based Linear Discriminant Analysis)[32]: 第一次，對水平方向的資訊執行一次IMLDA降維，使得原始影像的列被壓縮；第二次，將原始影像的列已經被壓縮過後的圖，對鉛直方向的資訊再做一次IMLDA降維，如圖3-2[31]。原則上，二維的線性鑑別分析法(2DLDA)即是對原始的影像從兩個方向去選取最具有代表性之特徵向量。

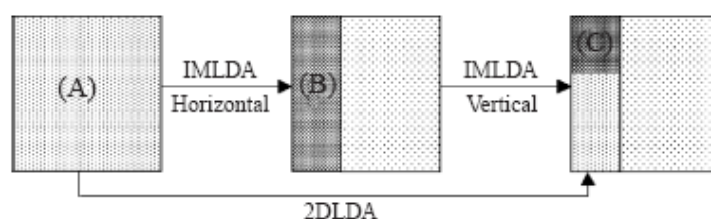


圖 3-2 2DLDA 示意圖[31]

假設在資料庫中共有 c 個人的影像，分別代表 c 個組別(class)；每個人都有 $M_i, i = 1 \dots c$ 張訓練用之人臉影像，每張人臉影像的大小為 $m \times n$ ； M 表示資料庫中共有多少張的人臉影像。因此我們可以計算出具有代表性的散佈矩陣，如下：

不同組間的散佈矩陣(the image between-class scatter matrix):

$$G_b = \frac{1}{M} \sum_{i=1}^c M_i (\bar{A}_i - \bar{A})^T (\bar{A}_i - \bar{A}) \quad (3)$$

(\bar{A}_i :表示第 i 個組別的平均; \bar{A} :表示資料庫中所有人臉影像的平均)

同組間的散佈矩陣(the image within-class scatter matrix):

$$G_w = \frac{1}{M} \sum_{i=1}^c \sum_{j=1}^{M_i} (A_j^{(i)} - \bar{A}^{(i)})^T (A_j^{(i)} - \bar{A}^{(i)}) \quad (4)$$

($A_j^{(i)}$:表示第 i 個組別的第 j 張圖; $\bar{A}^{(i)}$:表示第 i 個組別的平均)

我們都希望所有訓練用之人臉影像，經過二維線性鑑別分析法(2DLDA)降維後，能使不同組(不同人)的資料距離越遠越好， G_b 值越大越好，同組(同個人)的資料距離越近越好， G_w 值越小越好，經過 Lagrange method 的運算，我們可以將 Eq.(3)、Eq.(4)結合如下:

$$J(u) = \frac{u^T G_b u}{u^T G_w u} \quad (5)$$

由Eq.(5)，求得一個向量 u^* ，使得Rayleigh quotient function $J(u)$ 達到最大值， u^* 即被稱作是影像之最佳投影方向(optimal projection direction)。一般來說，對於多個組別的分類，單一個投影方向之辨識能力是不足夠的，因此我們需要一群具有強大辨識能力的鑑別向量 u_1, u_2, \dots, u_q (discriminant vectors)來表示，使得Eq.(5)達到最大值。另一方面，二維線性鑑別分析法(2DLDA)需使影像投影到各特徵向量的資訊彼此是無相關性的，因此必需滿足Eq.(6)。 G_t 代表是全域散佈矩陣， A 代表是原始影像， Y_i 代表是原始影像投影到 u_i 特徵向量的特徵值， Y_j 代表是原始影像投影到 u_j 特徵向量的特徵值[32]，我們希望 Y_i 和 Y_j 的相關性能越低越好，推

導如下:

$$\begin{aligned}
Cov(Y_i, Y_j) &= E\{(Y_i - EY_i)^T (Y_j - EY_j)\} \\
&= E\{[Au_i - E(Au_i)]^T [Au_j - E(Au_j)]\} \\
&= u_i^T \{E[(A - EA)^T (A - EA)]\} u_j \\
&= u_i^T G_i u_j = 0, i \neq j, i, j = 1, 2, \dots, q \tag{6}
\end{aligned}$$

依照上敘的限制以及Eq.(3)至Eq.(6)的計算，我們可以得到 G_b 和 G_w 所對應之廣義特徵向量(generalized eigenvectors)之解集，而後，截取前 q 大之廣義特徵值(generalized eigenvalues)所對應之廣義特徵向量(generalized eigenvectors)，

如: $G_b u_j = \lambda_j G_w u_j$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q$. , 所對應之 u_1, u_2, \dots, u_q 即為所求。

$U = (u_1, u_2, \dots, u_q)$ 即可以當做是對人臉影像執行特徵摘取之投影矩陣。

$$B = AU, \text{ where } U = (u_1, u_2, \dots, u_q) \tag{7}$$

- A 是一張 $m \times n$ 之人臉影像

- U 是 $n \times q$ 之投影矩陣

- B 是原始的人臉影像之列向量經過壓縮後呈 $m \times q$ 之影像

當我們對原始的人臉影像完成了第一階段的 IMLDA 後，原始的人臉影像 A

經過 Eq.(7) 得到特徵矩陣 B ，接下來，第二階段的 IMLDA，即是對特徵矩陣 B^T

去做處理:

不同組間的散佈矩陣(between-class scatter matrix):

$$H_b = \frac{1}{M} \sum_{i=1}^c M_i (\bar{B}_i - \bar{B})(\bar{B}_i - \bar{B})^T \tag{8}$$

(\bar{B}_i :表示第 i 個組別的平均； \bar{B} :表示資料庫中所有人臉影像的平均)

同組間的散佈矩陣(within-class scatter matrix):

$$H_w = \frac{1}{M} \sum_{i=1}^c \sum_{j=1}^{M_i} (B_j^{(i)} - \bar{B}^{(i)})(B_j^{(i)} - \bar{B}^{(i)})^T \quad (9)$$

($B_j^{(i)}$:表示第 i 個組別的第 j 張圖； $\bar{B}^{(i)}$:表示第 i 個組別的平均)

(其中 $B_j^{(i)} = A_j^{(i)}U, \bar{B}^{(i)} = \bar{A}^{(i)}U, \bar{B} = \bar{A}U$)

依循第一階段的 IMLDA 之 Eq.(5)、Eq.(6)之限制，我們也可以得到 H_b 和 H_w 所對應之廣義特徵向量(generalized eigenvectors)之解，而後，截取前 p 大之廣義特徵值(generalized eigenvalues)， $\mu_1, \mu_2, \dots, \mu_p$ ，所對應之廣義特徵向量(generalized eigenvectors)， $V = (v_1, v_2, \dots, v_p)$ ，即可對第一階段壓縮過後的特徵影像 B^T 執行第二階段的降維，如下:

$$C^T = B^T V \quad (10)$$

- B^T 是第一階段壓縮後的 $q \times m$ 之人臉影像

- V 是 $m \times p$ 之投影矩陣

- C^T 是原始的人臉影像經過列以及行之壓縮後 $q \times p$ 之矩陣

$$C = V^T B = V^T A U \quad (11)$$

二維的線性鑑別分析法(2DLDA)會使得原始影像 A ($m \times n$ 矩陣)保留重要的特徵向量，降維至 C ($p \times q$ 矩陣)，另一方面 p 、 q 的大小通常遠小於 m 、 n 之值，因此二維的線性鑑別分析法(2DLDA)除了可以保留影像中空間上的相關性，還可將資料做大量的壓縮。

第四章 機率圖模型之人臉辨識

第四章中，我們先對本研究所提出之機率圖模型做說明，使得我們可以公式化的描述在影片上的人臉辨識問題，而後，再將本研究方法所碰到的問題分成三個子問題來討論，最後，則總結此研究之事前訓練步驟以及整體架構之演算法。

4-1 機率圖模型與公式推導

傳統的狀態空間模型(state space model)之人臉辨識的主要架構如圖 4-1 所示。假定在此人臉辨識的架構中共有 K 個人，而 x_t (state vector)代表在時間點 t ，此人臉辨識的結果，將 x_1, \dots, x_t 集合起來， $X = \{x_1, \dots, x_t\}$ 代表在時間點 1 到時間點 t 所辨識出來的結果之集合； z_t (observation)代表在時間點 t ，於觀測到的資料中所截取出的人臉影像，將 z_1, \dots, z_t 的影像集合起來， $Z = \{z_1, \dots, z_t\}$ 代表在時間點 1 到時間點 t 從有序影像中所截取出來的人臉影像之集合。

當有序影像所截取出來的人臉影像，有姿勢(pose)的改變時，傳統的狀態空間模型(state space model)需要使用大量的狀態(state)才足以描述每個人不同姿勢(pose)的情況。例如：一個系統內有 20 個人，每個人有 7 種姿勢，則需要 140 個狀態才能夠含蓋所有可能發生的情況，否則無法得到正確的辨識結果。如此一來，傳統的狀態空間模型在碰到有姿勢變化的有序影像時，便會很複雜。因此，本研究主要的目的，就是建造一個機率圖模型 (a Probabilistic Graphical model for video-based Face Recognition, PGFR)，如圖 4-2，依照事前資訊，得知人臉的姿

勢(pose)可能有哪些變化，來建立新的節點， $H = \{h_1, \dots, h_R\}$ ，並將此新節點 H 加入傳統的狀態空間模型中。當我們在決定此觀測點的辨識結果時，同時也必需參考過去各種姿勢(pose)的狀態，如此一來，若遇到有大角度的姿勢變化之有序影像時，本研究所提出的 PGFR 模型只需要少量的狀態(state)，即能表示發生的情況。

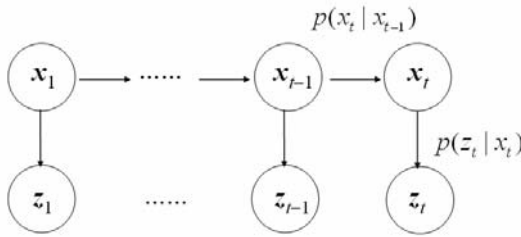


圖 4-1 傳統的狀態空間模型

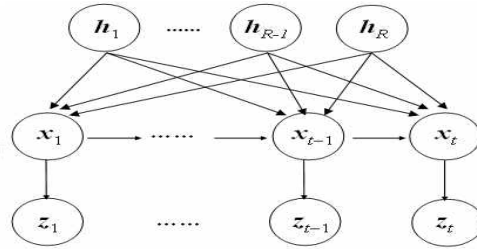


圖 4-2 PGFR 機率圖模型

定理一：

在時間點 t ，狀態 x_t 的事後機率 (posterior probability)，可由以下定理求得：

$$p(x_t | Z_t, H) \propto p(z_t | x_t) \int_{X_{t-1}} p(x_t | x_{t-1}, H) p(x_{t-1} | Z_{t-1}, H) dX_{t-1} \quad (12)$$

($H = \{h_1, \dots, h_R\}$: 有序影像中姿勢變化的資訊，共有 R 種姿勢)

($Z_t = \{z_1, \dots, z_t\}$: 一串有序影像中，每張影像所截取到的臉部區域之集合)

證明 Eq.(12):

依照圖 4-2 的貝式網路 (Bayesian network)，結合 Eq.(1) 和 Eq.(2)，以及運用

d -分割性質 (d-separation property) [25]，列出以下四種獨立情況：

$$- p(x_t | X_{t-1}, H) = p(x_t | x_{t-1}, H) \quad (13)$$

$$- p(z_t | x_t, Z_{t-1}, H) = p(z_t | x_t) \quad (14)$$

$$- p(x_t | X_{t-1}, Z_{t-1}) = p(x_t | X_{t-1}) \quad (15)$$

$$- p(H | X_t, Z_t) = p(H | X_t) \quad (16)$$

由以上的情況，推導如下：

$$\begin{aligned}
p(x_t | Z_t, H) &\propto \int_{X_{t-1}} p(X_t, Z_t, H) dX_{t-1} \\
&= \int_{X_{t-1}} p(H | X_t, Z_t) p(X_t, Z_t) dX_{t-1} = \int_{X_{t-1}} p(H | X_t) p(X_t, Z_t) dX_{t-1} \\
&= \int_{X_{t-1}} p(H | X_t) p(z_t | x_t) p(x_t | X_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\
&= \int_{X_{t-1}} \frac{p(H, X_t)}{p(X_t)} p(z_t | x_t) p(x_t | X_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\
&= \int_{X_{t-1}} \frac{p(H, x_t | X_{t-1})}{p(x | X_{t-1})} p(z_t | x_t) p(x_t | X_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\
&= p(z_t | x_t) \int_{X_{t-1}} p(x_t | X_{t-1}, H) p(H | X_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\
&= p(z_t | x_t) \int_{X_{t-1}} p(x_t | X_{t-1}, H) p(H | X_{t-1}, Z_{t-1}) p(X_{t-1}, Z_{t-1}) dX_{t-1} \\
&= p(z_t | x_t) \int_{X_{t-1}} p(x_t | X_{t-1}, H) p(H, X_{t-1}, Z_{t-1}) dX_{t-1} \\
&\propto p(z_t | x_t) \int_{X_{t-1}} p(x_t | x_{t-1}, H) p(x_{t-1} | Z_{t-1}, H) dX_{t-1} \quad (17)
\end{aligned}$$

由Eq.(17)得知，在時間點 t ，有三個因素會影響狀態 x_t 之辨識結果，(i) $p(z_t | x_t)$ ：表示在時間點 t ，得到的人臉影像 z_t ，和資料庫中每個人之相似度，這部份在章節4-2會詳細說明。(ii) $p(x_t | x_{t-1}, H)$ ：代表待測的人臉影像中，時間點 t 的辨識結果和前一個時間點 $t-1$ 的辨識結果之相似程度和姿勢變化的機率值，這部份在章節4-3會詳細說明。(iii) $p(x_{t-1} | Z_{t-1}, H)$ ：每個時間點，都依循此遞迴演算法，將先前時間點的辨識結果累計起來，這部份在章節4-4會詳細說明。此外，在最初的時間點， $t=1$ ，我們直接應用二維的線性鑑別分析法(2DLDA)，計算出 z_1 和每個人

的相似度，因而求得在時間點1時的辨識結果。

另一方面，在進行人臉辨識前，必須先滿足兩個假設：第一點，在有序影像中，藉由實行人臉偵測方法，假定每張影像都能截取出完整的人臉位置；第二點，假定每張影像所截取出的人臉，都可被歸類成某個人的某個姿勢變化，換言之，即是在所有訓練或測試的人臉影像裡，每張影像都可被歸屬到 R 個姿勢中的其中一個，在本研究中，我們採用 K -means 分群法，粗略的將人臉影像分成 R 個子群，並且手動的檢查這些人臉影像之分類是否有錯誤。

本小節先以公式化的機率形式來描述我們的人臉辨識方法(PGFR)，接著將此方法分成三個子問題，在以下各小節內，會詳細敘述三個子問題的解決方法。

4-2 相似度之計算

在 Eq.(12)中， $p(z_t | x_t)$ ，即是估計在時間點 t ，此張人臉影像 z_t 和每個在資料庫中之人之相似程度。然而，雖然我們取得的是一串有序影像，但在此我們不考慮時間的因素，只考慮空間上的分佈情況，因此我們可以把它視作單張影像之人臉辨識。在本研究裡，我們採用二維線性鑑別分析法(2DLDA)[31]來做特徵提取，而後實行單張影像之人臉辨識。

給定 K 個人訓練用之人臉影像，我們利用二維線性鑑別分析法(2DLDA)，將所有訓練用之人臉影像($m \times n$ 像素大小)降維至 $p \times q$ 的子空間內，並且能夠有效的將這些訓練用之人臉影像分割成 K 群(K 個人)。在本研究中，原始的人臉影

像大小經過二維線性鑑別分析法(2DLDA)降維後，每張影像都降維到 $b \times b$ 維，而後，當需要一維的計算時，我們可用 b^2 維的平面來表示。接下來，我們以 $m_k (k=1, \dots, K)$ 來表示第 k 個人之訓練用之人臉影像經過二維線性鑑別分析法(2DLDA)投影過後之平均值； z_t 表示在時間點 t ，所截取出之人臉影像；而 \bar{z}_t 代表 z_t 經過二維線性鑑別分析法(2DLDA)投影到低維子空間後之結果。完成上述的流程後，即能計算出，在時間點 t 待測之人臉影像， z_t ，和每個人在子空間上之距離。我們可利用高斯分佈將此距離正規化，而 z_t 所對應 x_t 的機率密度函數如下：

$$p(z_t | x_t = k) = (2\pi)^{-\frac{d^2}{2}} |C_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\bar{z}_t - m_k)C_k^{-1}(\bar{z}_t - m_k)'\right), k=1 \dots K \quad (18)$$

(C_k 表示 k 這個人訓練用之人臉影像的分散矩陣，covariance matrix)

(m_k 表示 k 這個人訓練用之人臉影像投影到子空間後的平均值)

經由章節 4-2 的計算，我們可以將 $p(z_t | x_t)$ 視作是「在時間點 t 時，待測的人臉影像和資料庫中的每個人個別的相似程度」。

4-3 遞移機率

在 Eq.(12) 中， $p(x_t | x_{t-1}, H)$ 即是代表在時間點 t 和上一個時間點 $t-1$ 之間的遞移關係之機率值。假定資料庫中有 K 個人的人臉影像，我們必須保證這 K 個人都至少有一段有序的人臉影像在資料庫中，如此，我們才能對連續時間點之間的遞移機率做事前的估算。當測試的人臉影像產生錯誤的判斷結果時，遞移機率可以修正判斷結果，將很離譜的錯誤去除掉或者是提高一些有可能的狀態之機率

值。在計算遞移機率時，我們將 $p(x_t | x_{t-1}, H)$ 轉換如下：

$$\begin{aligned}
 & p(x_t | x_{t-1}, H) \\
 &= \frac{p(H | x_t, x_{t-1})p(x_t | x_{t-1})p(x_{t-1})}{p(H | x_{t-1})p(x_{t-1})} \\
 &= \frac{p(H | x_t, x_{t-1})}{p(H | x_{t-1})} p(x_t | x_{t-1}) \tag{19}
 \end{aligned}$$

由 Eq.(19)得知此遞移機率可以分解成兩個部份來討論，(i)在沒有任何姿勢資訊的情況下，計算出 $p(x_t | x_{t-1})$ ，也就是在連續時間點的辨識結果中，人和人之間的遞移機率(ii) $\frac{p(H | x_t, x_{t-1})}{p(H | x_{t-1})}$ 即是對於兩個相鄰的時間點 t 及時間點 $t-1$ ，計算出姿勢辨識結果之比值。

由以上的論述，我們必須從每個時間點的人臉影像獲得兩個重要的資訊，第一：此時間點之人臉影像和資料庫中每個人的相似程度，第二：對於資料庫中的每個人而言，此時間點之人臉影像屬於此人的哪一個姿勢。總體來說，在每個時間點會改變的因素為「人」以及「姿勢」，如此，我們即可對兩個連續的時間點 t 和時間點 $t-1$ 之間的狀態，來觀察遞移機率。以下，即針對兩個遞移機率的運算做詳細的說明。

4-3.1 人與人的遞移機率

由 Eq.(19)，先討論如何計算 $p(x_t | x_{t-1})$ 之值。在連續時間點之辨識結果中，人和人之間的遞移機率，可在正式進行人臉辨識前，先行做估算，並且假設「此遞移機率值是不會隨著時間而改變的」，換言之，我們只需對資料庫中訓練用的人臉影像，經由章節 4-2，投影到低維的子空間後，即可計算出不同人之間的相

似度，計算方式如下：

$$d(i, j) = \frac{1}{|I_i|} \left(\sum_{r \in I_i} (r - m_j)(r - m_j)^t \right)^{1/2} \quad (20)$$

$d(i, j)$ 是計算 i 這個人和 j 的距離關係，如下圖 4-3， m_j 表示 j 這個人之訓練用之人臉影像經過章節 4-2 做特徵摘取後所求得之平均值； r 表示屬於 i 這個人的影像； I_i 表示 i 這個人訓練用之人臉影像做特徵摘取後的集合；Eq. (21) 利用高斯分佈再將此距離正規化成 i 和 j 的相似度之機率值：

$$\text{sim}(i, j) = \frac{1}{\Lambda} \exp\left(\frac{-1}{\sigma^2} d(i, j)\right) \quad (21)$$

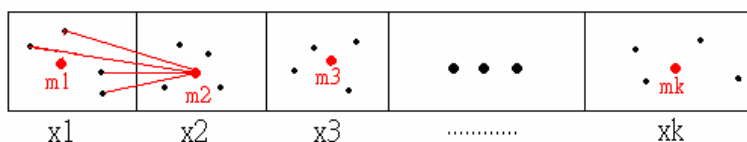


圖 4-3 人與人之遞移機率之計算方式

4-3.2 姿勢轉換的遞移機率

由 Eq.(19)，接著要討論 $p(H | x_t, x_{t-1}) / p(H | x_{t-1})$ 之值該如何計算。對於 $p(H | x_t, x_{t-1}) / p(H | x_{t-1})$ ，我們無法將此推導出一個封閉的形式求解，因而利用計算兩個相鄰時間點 t 及時間點 $t-1$ 中，人臉影像的姿勢之遞移機率來求近似解。

首先，對於資料庫中的每個人而言，我們先將各自訓練用之人臉影像，採用 K -means 分群法分成七類 $H = \{h_1, \dots, h_7\}$ ，左臉 90° 、左臉 45° 、正臉、右臉 45° 、右臉 90° 、向上、向下，這七種姿勢(pose)，接著，再利用章節 3-2 所提及的二維線性鑑別分析法(2DLDA)，以姿勢為組別，將每個人在資料庫中的人臉影像降維後分成七個類別，如圖 4-4。

再來，我們必須事前估計人臉姿勢變化之遞移機率，Eq.(22)採用統計的方式

計算出姿勢(pose)與姿勢(pose)間的遞移機率：

$$p(h_j | h_i) = \frac{1}{\delta_i} \sum_{t=2}^i \alpha(I_{t-1} \in h_i) \alpha(I_t \in h_j) \quad (22)$$

($\alpha(I_t \in h_j) = 1$ ，表示人臉影像 I_t 屬於 h_j 這個姿勢)

($\alpha(I_t \in h_j) = 0$ ，表示人臉影像 I_t 不屬於 h_j 這個姿勢)

(δ_i 是正規化常數，目的為確保 $\sum_{j=1}^n p(h_j | h_i) = 1$)

並且假設「不管是哪一個人，不同姿勢間之遞移機率都會是相似的」，舉例說明：

對每個人而言，在時間點 $t-1$ 時，所辨識出的人臉姿勢為左臉 45° ，下一個時間

點 t 時，所辨識出的人臉姿勢為正臉的機率比右臉 45° 的機率高。因此由 Eq.(22)

可將每個人的姿勢遞移機率計算出來而後再加總取平均，即是代表此資料庫所有

人的遞移機率。

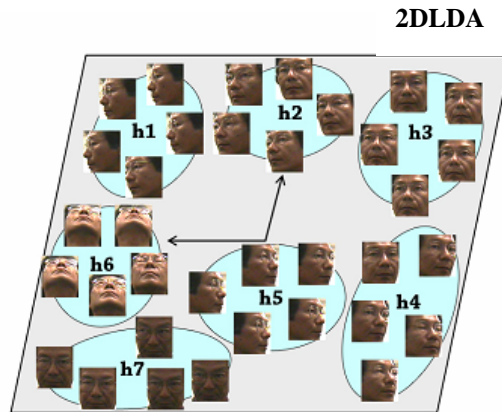


圖 4-4 2DLDA 的人臉姿勢之分類

4-4 流程與方法總結

原先，在時間點 t 之待測人臉影像， z_t ，我們希望可以估計出 $p(x_t | z_t, H)$ ，

但是由於我們處理的是一串有序的人臉影像，因此我們可以將時間的資訊加入其中，“ $p(x_t | z_t, H)$ ”可被視為聯合條件機率(joint conditional probability)“ $p(x_t | Z_t, H)$ ”，其中 Z_t 表示從開始 $t=1$ 到時間點 t 的一串有序之人臉影像，回顧 Eq.(12)，我們得到以下遞迴公式：

$$p(x_t | Z_t, H) \propto p(z_t | x_t) \int_{X_{t-1}} p(x_t | x_{t-1}, H) p(x_{t-1} | Z_{t-1}, H) dX_{t-1}$$

再由章節 4-3 的推論，可將 Eq.(12)修正如下：

$$\begin{aligned} p(x_t | Z_t, H) & \propto p(z_t | x_t) \int_{X_{t-1}} p(x_t | x_{t-1}, H) p(x_{t-1} | Z_{t-1}, H) dX_{t-1} \\ & \propto p(z_t | x_t) \int_{X_{t-1}} \frac{p(H | x_t, x_{t-1})}{p(H | x_{t-1})} p(x_t | x_{t-1}) p(x_{t-1} | Z_{t-1}, H) dX_{t-1} \end{aligned} \quad (23)$$

$p(x_t | x_{t-1})$ 和 $\frac{p(H | x_t, x_{t-1})}{p(H | x_{t-1})}$ 即隱含了時間的資訊，對於一串有序影像之人臉辨識可提升不少的準確率，並且避開了單張影像之人臉辨識中，只受一個因素影響辨識結果的缺點。

在本方法中，我們可將整體流程分成「訓練」以及「辨識」兩個階段。首先，我們必須先訓練資料庫中的人臉影像，其步驟如下：

- 步驟一：對資料庫中所有人臉的影像，以人分群，共分成 K 群(K 個人)
，再用二維線性鑑別分析法(2DLDA)來做訓練，得到一組 2DLDA 之子空間
- 步驟二：將步驟一投影後的人臉影像，由 Eq. (20)、Eq. (21)計算出人和人之間的相似度

圖 4-5 本研究之訓練步驟

- 步驟三: 用 K -means 分群法，以姿勢分群，對第一個人的訓練用之人臉影像分成 R 群，再對此資料用二維線性鑑別分析法(2DLDA)來做訓練，得到一組 2DLDA 之子空間
- 步驟四: 對第二個人的訓練用之人臉影像再重覆一次步驟三，依此類推，假設資料庫中有 K 個人，藉由步驟三，我們即可得到 K 組以姿勢分群的 2DLDA 子空間之資訊
- 步驟五: 由 Eq. (22) 計算出姿勢遞移機率

圖 4-5 本研究之訓練步驟

事前訓練完成後，當輸入一串有序的觀測值來做測試時，所需之辨識步驟如下:

- 輸入: 一串有序的觀測值 $\{z_1, \dots, z_{end}\}$ 及事前訓練所得之 2DLDA 子空間、人與人的相似機率、姿勢遞移機率
- (初始化): 輸入之觀測值為 z_1 時
- 步驟一: 由圖 4-5 之步驟一所求得的 2DLDA 子空間，計算出此觀測值 z_1 和每個人的相似度， $p(z_1 | x_1 = k)$ ， $k = 1 \dots K$
- $p(x_1 | Z_1, H) \leftarrow p(z_1 | x_1)$
- 步驟二: 觀測值 z_1 對應至圖 4-5 之步驟三及步驟四所求得之 2DLDA 子空間，計算出此觀測值 z_1 屬於 $x_1 = k$ ($k = 1 \dots K$) 的何種姿勢
- $t = t + 1$
- $t = 2$
- (重覆之步驟): 此時輸入之觀測值為 z_t
- 步驟三: 由圖 4-5 之步驟一所求得的 2DLDA 子空間，計算出此觀測值 z_t 和每個人的相似度， $p(z_t | x_t = k)$ ， $k = 1 \dots K$

圖 4-6 本研究之人臉辨識演算法

■ 步驟四：由圖 4-5 之步驟三及步驟四所求得之 2DLDA 子空間，計算出此觀

測值 z_t 屬於 $x_t = k$ ($k = 1 \cdots K$) 的何種姿勢

■ 步驟五：計算 $p(z_t | x_t) \int_{X_{t-1}} p(x_t | x_{t-1}, H) p(x_{t-1} | Z_{t-1}, H) dX_{t-1}$

， $x_t = 1 \cdots K$

■ 步驟六： $t = t + 1$

回到(重覆之步驟)再依序輸入觀測值

■ 步驟七：直到觀測值輸入完為止，即 $t = end$

而 x_{end} 可求得：

$$k = \arg \max_{x_{end}} P(x_{end} | Z_{end}, H)$$

$$\propto p(z_{end} | x_{end}) \int_{X_{end-1}} p(x_{end} | x_{end-1}, H) p(x_{end-1} | Z_{end-1}, H) dX_{end-1}$$

， $x_{end} = 1 \cdots K$ 之 k 即為此段影片之辨識結果

■ 輸出： k

此段影片之辨識結果

圖 4-6 本研究之人臉辨識演算法

第五章 實驗結果與分析

人臉辨識系統，包括人臉的追蹤、截取與辨識。本實驗是在人臉辨識的領域上做分析與探討，因此必須在假設影片中的人臉都有好的「追蹤」與「截取」的情形下，而後再將追蹤與截取得到的人臉影像來進行本研究方法之實驗。

本章之目的在於藉由實驗值的估算，來驗證本研究所提出的新方法(a Probabilistic Graphical Model)可在一串有序影像上之人臉辨識有所加強與改進。另一方面，並和現有的技術比較，以歸結出本研究之方法的優勢之處與適用之環境。

5-1 Honda/UCSD 資料庫

Honda/UCSD 資料庫中[36]有 20 人，每個人皆有兩段以上不同的影片，共有 52 段影片。每段影片皆是在室內的環境下由 SONY EVI D30 所錄製，每秒有 15 張影像，而影像之解析度為 640×480 ，每段影片的長度至少 20 秒以上。

Honda/UCSD 資料庫之特點在於每段影片人的頭部皆有重大的 2-D 和 3-D 之旋轉，如圖 5-1(a)、圖 5-1(b)和圖 5-1(c)所示。在每段影片中的前 15 秒內，人的頭部通常都會有穩定的速度和規律的轉動，而剩餘的時間，人的頭部除了有較大的轉動外，也會有表情的改變，如圖 5-1(d)，與部份遮蔽的情況發生，如圖 5-1(e)。

Honda/UCSD 資料庫包含了許多真實世界的情況，是目前在影片上之人臉辨識上最普遍且最具代表性的資料庫。



(a) Harsh 之人臉影像



(b) Hide 之人臉影像



(c) James 之人臉影像



(d) Chia 之人臉影像



(e) Jeff 之人臉影像

圖 5-1 在 Honda/UCSD 資料庫中，原始影像的一些例子

5-2 本研究使用 Honda/UCSD 資料庫之實驗

我們選取 Honda/UCSD 資料庫中 40 段影片，每個人皆有兩段不同的影片，一段當訓練用的資料，另一段則是當測試用之影片，如圖 5-2。

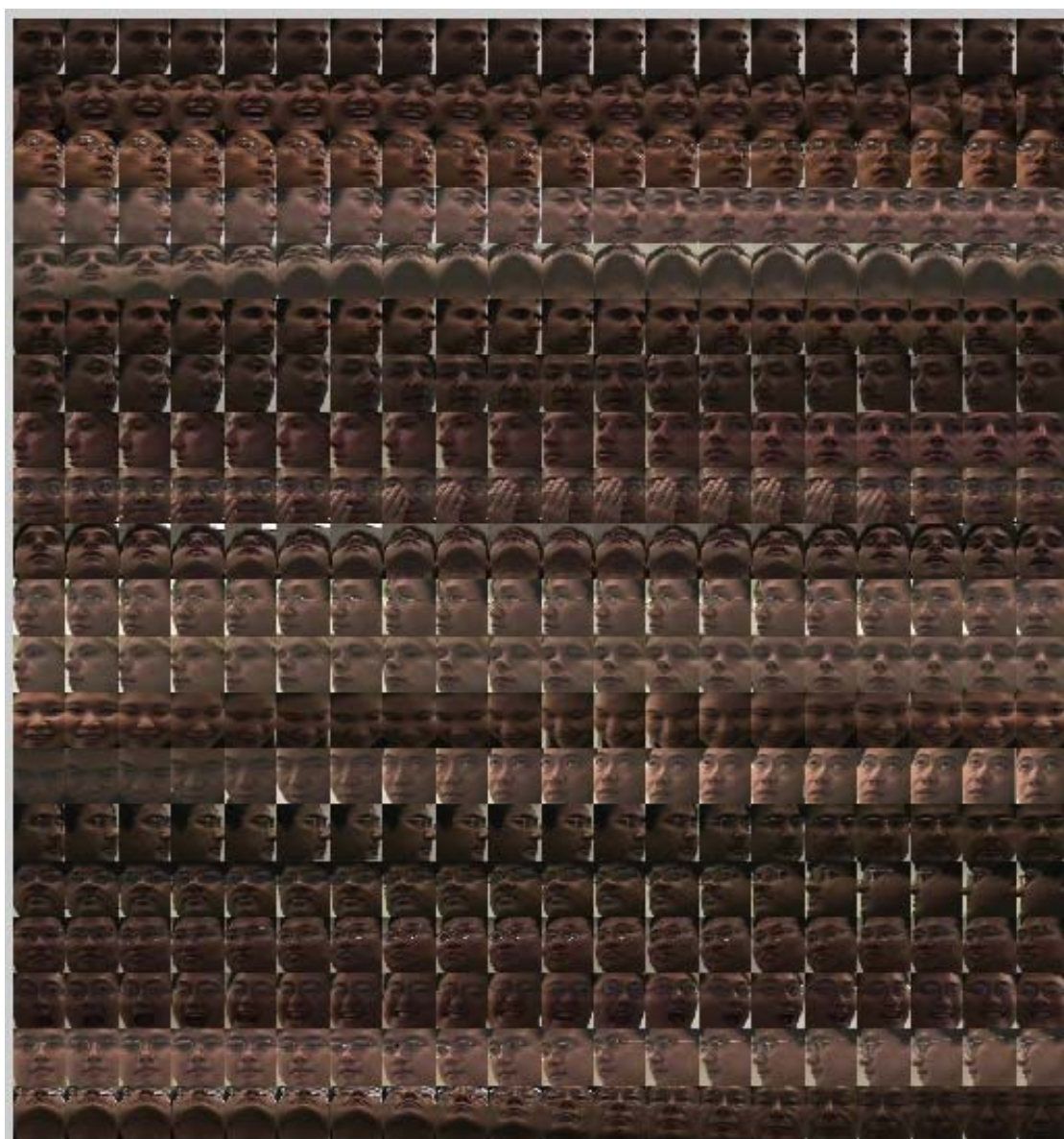


圖 5-2 在 Honda/UCSD 測試用之資料庫中，截取出人臉影像的一些例子

5-2.1 訓練流程

在我們選取 Honda/UCSD 資料庫中 40 段影片，每個人皆有兩段不同的影片且每段影片的長度皆大於 20 秒。每段影片經由追蹤與截取所得到的人臉資料都必需縮放成一定的大小，如圖 5-2。從實驗結果得知，如表 5-1，若影像解析度太高，容易受到雜訊的干擾，若解析度太低，又會使得影像失真過於嚴重，因此

縮放成 100x90 之像素大小能擁有最好的辨識結果。

表 5-1 影像解析度之大小對辨識率的影響

Resolution	80x72	100x90	150x135	200x180
Accuracy	79.05%	80.17%	78.84%	78.62%

另一方面，為了將光線不同所造成的影響降至最低，所有縮放至相同大小的影像在訓練前，必須先經過前處理。我們做過以下兩種嘗試，如表 5-2，分別為直方圖等化(hist equalization)和正規化(normalization: $I_{new} = \frac{I - \text{mean}(I)}{\text{std}(I)}$, I : 單張人臉影像)，由實驗結果發現直方圖等化的效果較好，因此在接下來的實驗中，所有影像的前處理皆使用直方圖等化。

表 5-2 前處理方法之比較

method	normalization	histogram equalization
Accuracy	69.54%	80.17%

當所有訓練用之人臉影像縮放至相同大小並且做完直方圖等化後，接下來依照章節 4-2 至章節 4-4 之步驟開始做訓練。第一，將所有訓練用之人臉影像(每個人皆有一串影像做為訓練用，共有 20 段人臉影像)，以人分類，並使用二維線性鑑別分析法(2DLDA)將資料從高維空間(100x90 的像素大小)降至低維的子空間中。表 5-3 中辨識率的計算，是在沒有任何時間資訊的情況下，採用傳統的單張人臉影像辨識之方法，辨識率為此串影像辨識正確的影像數與所有影像數之比值。由圖 5-3 中可得知，當影像降維至 5x5 維時，擁有最高的辨識率，一但保留過多的資訊(例如:9x9 維)，非但沒辦法提升辨識率，反而使得辨識率下降許多，因此選擇適當的維度降維，對於整體辨識率有極大的影響。另外，在章節 4-3.1

中提到，相鄰兩張影像，我們必須計算出人與人的相似機率，即是當資料被降至低維後，採用 Eq.(20)和 Eq.(21) 所估算出來的 20x20 的遞移矩陣 ($T, 0 \leq T_{ij} \leq 1$)。

表 5-3 使用 2DLDA 將高維資料降至低維後的辨識率

dim \ accuracy	3x3	4x4	5x5	6x6	7x7	8x8	9x9
2DLDA	63.80%	74.23%	80.17%	78.43%	76.86%	73.91%	69.44%

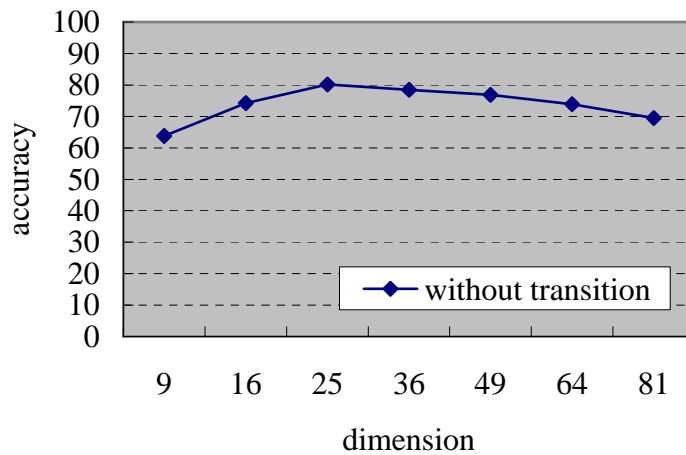


圖 5-3 使用 2DLDA 將高維資料降至各維度之比較

第二部份，採用 K-means 分群法將每個人訓練用之人臉影像，以姿態分群(每個人皆分為 7 群)，再以二維線性鑑別分析法(2DLDA)將資料降至低維的子空間中。章節 4-3.2 中提到，相鄰兩張影像，必須計算出姿勢與姿勢間的遞移機率，我們用 Eq.(22) 估計出屬於此 20 人姿勢遞移機率。表 5-4 可驗證在章節 4-3.2 的論述，「不管是哪一個人，不同姿勢間之遞移機率都會是相似的」，例如：不同人的姿勢遞移機率中，前一個時間點是左臉-90°，下一個時間點是左-45°的機率比下一個時間點是正臉的機率高。

表 5-4 從訓練用之人臉影像所計算出的不同人之姿勢遞移機率表

$\begin{matrix} t \\ \backslash \\ t-1 \end{matrix}$	左-90°	左-45°	正	右-45°	右-90°	上	下	$\begin{matrix} t \\ \backslash \\ t-1 \end{matrix}$	左-90°	左-45°	正	右-45°	右-90°	上	下
左-90°	0.925	0.075	0	0	0	0	0	左-90°	0.9091	0.0909	0	0	0	0	0
左-45°	0.1429	0.5714	0.2857	0	0	0	0	左-45°	0.1111	0.7778	0.1111	0	0	0	0
正	0	0.027	0.8716	0.0473	0	0.0338	0.0203	正	0	0.0042	0.9789	0.0042	0	0.0042	0.0084
右-45°	0	0.0217	0.1304	0.7391	0.087	0.0217	0	右-45°	0	0	0.1429	0.7143	0.1429	0	0
右-90°	0	0	0	0.0758	0.9242	0	0	右-90°	0	0	0	0.0833	0.9167	0	0
上	0	0	0.0833	0.0208	0.0208	0.875	0	上	0	0	0.0357	0	0	0.9643	0
下	0	0.0417	0.0833	0	0	0	0.875	下	0	0	0.1111	0	0	0	0.8889

最後再對 20 人的姿勢遞移機率取平均，即為屬於此 20 人的人臉姿勢變化的遞移機率表，如表 5-5 所示。

表 5-5 人臉姿勢變化的遞移機率表

$\begin{matrix} t \\ \backslash \\ t-1 \end{matrix}$	左-90°	左-45°	正	右-45°	右-90°	上	下
左-90°	0.8144	0.1257	0.0258	0.0078	0.0024	0.0012	0.0227
左-45°	0.0786	0.7158	0.1487	0.0048	0.0035	0.032	0.0166
正	0.0011	0.0235	0.9154	0.0261	0.0018	0.0152	0.0169
右-45°	0.0086	0.0001	0.1588	0.6805	0.1034	0.0392	0.0095
右-90°	0.0001	0.0027	0.0387	0.0894	0.82	0.0443	0.0048
上	0.0024	0.0361	0.0705	0.0322	0.0167	0.821	0.0211
下	0.0044	0.0406	0.1275	0.0218	0.0015	0.0198	0.7843

5-2.2 辨識結果分析

本小節，將我們在第四章提出的論點分成二個部份來驗證。

第一個部份為將章節 4-3.1 所提出之「人與人的遞移機率(person transition)」加入單張人臉影像辨識中。表 5-6 為單張人臉影像辨識(without transition)以及加入人與人的遞移機率(person transition)之比較，當資料採用二維線性鑑別分析法(2DLDA)降至 5×5 維後，若加入「人與人的遞移機率」可將先前時間點的辨識結果保留住並且修正了大部份零星的錯誤，因此提高了 7.16%的辨識率，對一串有

序影像之人臉辨識有著很大的幫助。

表 5-6 單張影像之人臉辨識與加入人和人的遞移機率之比較

transition type (dim=5×5)	without transition	person transition
Accuracy	80.17%	87.33%

第二個部份為將章節 4-3.2 中所提出的「人臉姿勢變化的遞移機率(pose transition)」再加入第一部份的實驗中。研究發現，大部份影像在辨識錯人的情況下，姿勢都不會判錯，也就是說，若出現連姿勢也辨識不正確的重大錯誤情況時，人臉姿勢變化的遞移機率便可將辨識結果修正為正確的結果。

當此串影像在某個時間點產生錯誤的辨識結果時，主要是因為此錯誤的辨識結果與正確的辨識結果很像，我們可以將「很像」分成兩類來說明：(一) 不同人在同個姿勢很像：表示在此時間點，錯誤與正確的辨識結果都可辨識出相同的姿勢，因此需靠此張影像與他們各別的相似度以及人與人的遞移機率來修正結果；(二) 不同人在不同個姿勢很像：表示在此時間點，錯誤與正確的辨識結果會辨識出不同的姿勢，此時即可藉由姿勢轉換的遞移機率來修正結果。由表 5-7 所示，加入人與人的遞移機率(person transition)可以修正大多數單張影像之人臉辨識的錯誤情況，若再加入姿勢轉換的遞移機率(pose transition)，則是在檢查整串影像是否有連姿勢也判斷錯誤的情況發生，而予以修正，因此可再提升 3.34% 的辨識率。實驗結果說明了，加入兩種遞移機率(person and pose transition)可比傳統的單張影像的人臉辨識高出 10.5% 的準確度，也可得知一段影片中，相鄰影像間資料相關性的重要性。

表 5-7 單張影像之人臉辨識與加入兩種遞移機率之比較

transition type (dim=5×5)	without transition	person transition	person and pose transition
Accuracy	80.17%	87.33%	90.67%

我們可藉由 Honda/UCSD 資料庫來驗證出當影片中人臉有姿勢、表情的變化或者是遮蔽的情況產生時，加入本研究方法所提出的兩種遞移機率，一方面可以將先前的辨識結果合理的保留住，另一方面利用兩階段的遞移機率互相牽制，更可以防止有嚴重判錯的情況產生。圖 5-4，在各種維度下，使用本研究所提出的方法皆能比原先沒有加入任何遞移機率更能準確辨識出正確的結果。

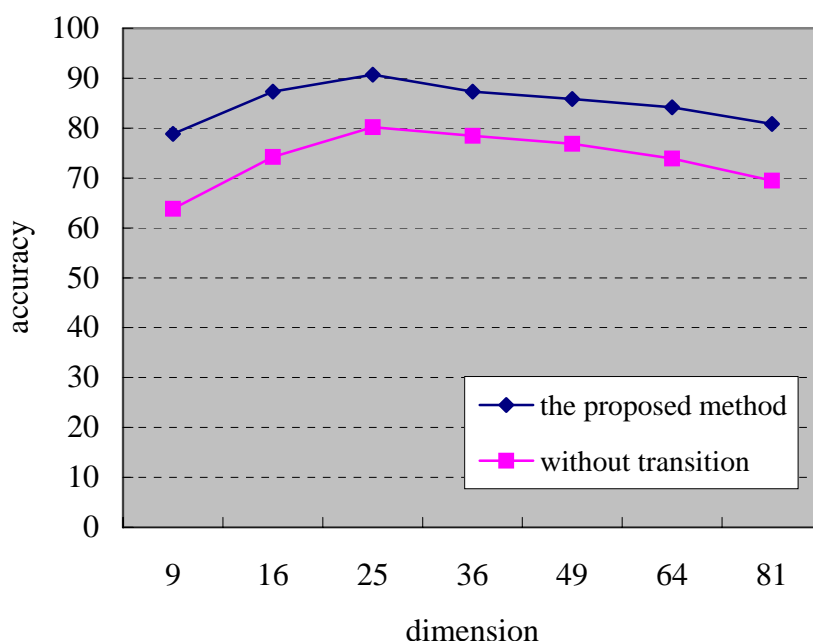


圖 5-4 未加入遞移機率與本研究方法在各維度上之比較

5-2.3 辨識率估算

關於辨識率的計算，由於在我們的方法中是以一段影片為單位，而每段影片

只會有一個辨識結果，因此我們必需對欲測試用之人臉影像以固定的大小裁切成一段段較短影片[27][15]，若裁切的大小(window size)越大，即表示在每小段影片中，保留較多先前影像的辨識結果，因此一般來說會得到較高的辨識率。如表 5-8 可發現隨著裁切大小(window size)變大，辨識率也會逐漸上升，但裁切大小也不能過大，否則會造成少數結果決定整體辨識率的情況，本實驗之辨識率估算皆是採用裁切大小為 110，將 20 段欲測試用之影片分成 58 小段來估算的辨識結果。

表 5-8 不同的 window size 所測出的辨識率

window size	70	90	110
Accuracy	89.58%	90.25%	90.67%

5-2.4 機率分佈圖之探討

本實驗把影片上之人臉影像辨識過程分成三個小部份來研究，在章節 4-2、章節 4-3、章節 4-4 有詳加描述。圖 5-6 為描述此人 M_6 由正臉(Fr)到出現右臉($R90^\circ$)的有序影像，在四個時間點 $t = 52$ ， $t = 55$ ， $t = 57$ ， $t = 59$ 時的機率分部情況。第一行表示當下時間點，此待測之人臉影像和 M_1 ， M_6 ， M_{10} 這三個人的相似度；第二到四行分別表示當下時間點，此待測之人臉影像屬於 M_1 ， M_6 ， M_{10} 的哪一個姿勢；第五行為使用本研究提出之方法，從第一張影像累積到當下時間點之辨識結果。由實驗結果發現，原先，在時間點 $t = 52$ 時，屬於 M_{10} 的機率最高，在時間點 $t = 55$ 時， M_{10} 的姿勢判斷錯誤，由 $R45^\circ$ 判斷成 Fr，而後在時間點 $t = 57$ 時， M_6 和 M_{10} 有相近的相似度之情況下，由於 M_6 有正確的姿勢判斷，因此有較

高的姿勢遞移機率，因此在時間點 $t=57$ 時，屬與 M_6 的機率會最高，即得到正確的辨識結果，而後在時間點 $t=59$ 時，也會藉由人與人之遞移機率保留先前的正確辨識結果，因此屬與 M_6 的機率也會最高。

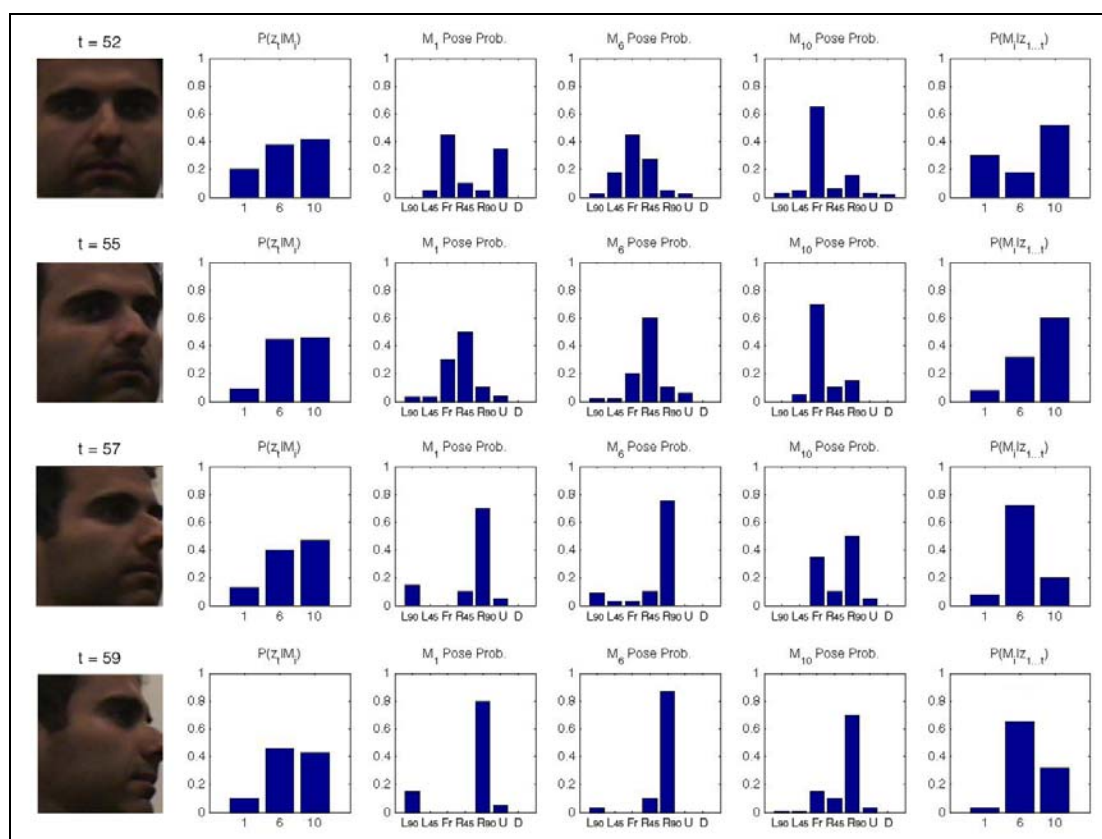


圖 5-5 在四個時間點的機率分部情況

5-3 單張影像上人臉辨識技術之比較

本節的實驗為將傳統的單張影像之人臉辨識技術和本研究方法做比較，以歸納出加入相鄰影像間的時間資訊後，能比單張影像之人臉辨識技術擁有更高的辨識率[22]。而比較的技术其一為廣為大家使用的「特徵臉(EigenFaces)」，使用主要成份分析法(PCA)將所有訓練用之影像從高維(9000 維)降維至 30 維之子空間中；其二為「費雪臉(FisherFaces)」，使用了線性鑑別分析法(LDA)，將所有訓練

用之資料降維至 19 維之子空間中(線性鑑別分析法(LDA)因為不同組間散佈矩陣之限制，最多只能降至類別數減一維)；而最後一種為「最近鄰居法(Nearest Neighbor)」，即是不需要任何降維動作，直接將資料在 9000 維的高維空間中做比對，而本研究所提出之方法，則是採用二維的線性鑑別分析法(2DLDA)將影像降至 5×5 維之子空間中，並且加入遞移機率。表 5-9 為使用 Honda/UCSD 資料庫，在各個技術上與我們提出的方法做比較，驗證出本研究所提出之方法因加入了相鄰影像之重要資訊，因此能比單張影像之人臉辨識技術提升近 10% 的辨識率。

表 5-9 本研究方法和三種單張影像之人臉辨識技術之比較

Method	Accuracy
EigenFaces(30-dim)	69.30%
FisherFaces(19-dim)	74.50%
Nearest Neighbor	81.60%
Only 2DLDA	80.17%
Our propose method(25-dim)	90.67%

5-4 使用 VIPlab 資料庫之辨識結果

VIPlab 資料庫有 20 個人，每個人皆有兩段影片，共 40 段影片。每段影片皆是在室內的環境下由 Panasonic DMC-TZ3 所錄製，每秒有 10 張影像，而影像之解析度為 320×240，每段影片至少 15 秒以上。VIPlab 資料庫之特點在於每段影片除了人的頭部有重大的 2-D 和 3-D 之旋轉外，還在頭部旋轉的過程中加上表情的變化，如圖 5-7(a)、圖 5-7(b)，或者是加上手的遮蔽和物體的遮蔽，如圖

5-7(c)、圖 5-7(d)，以及一些現實中會發生的情況，如圖 5-7(e)。

VIPlab 資料庫和 Honda/UCSD 資料庫有兩個不同的地方：(一) 在 VIPlab 資料庫的測試影片中，某個時間點常會碰到兩個以上問題同時發生的情況，例如：表情的變化與頭部的旋轉、遮蔽與頭部的旋轉等等，而 Honda/UCSD 資料庫當有表情的變化或者是遮蔽的情況時，通常人的頭部都是正臉的；(二) VIPlab 資料庫的所有訓練與測試之影片，都是在同一天，同個環境下完成的，因此降低光線的影響，而 Honda/UCSD 則有一部份人的訓練與測試不是在同天拍攝的。



(a) Asilverfox 之人臉影像



(b) Sibevin 之人臉影像



(c) Jarway 之人臉影像



(d) Yichia 之人臉影像

圖 5-6 在 VIPlab 資料庫中，原始影像的一些例子



(e) Jiya 之人臉影像

圖 5-6 在 VIPlab 資料庫中，原始影像的一些例子

表5-10為本研究方法在VIPlab資料庫與Honda/UCSD資料庫上之辨識率。

由此可看出，在不同條件下，我們依舊可以達到90%以上的辨識率，因此可證明本研究方法的強健性。

表 5-10 使用本研究方法於 VIPlab 資料庫與 Honda/UCSD 資料庫上之辨識率

dataset	VIPlab	Honda/UCSD
Accuracy	92.4%	90.67%

第六章 結論與未來展望

本研究在影片上之人臉辨識上套用本論文所提出的演算法，一方面使用二維影像直接降維來保留原始影像的空間資訊，另一方面，建構 PGFR 模型，並且藉由兩個遞移機率來包含相鄰影像間的時間資訊，使得辨識結果在加入本論文所提出的兩個遞移機率後，皆能穩定提升。本研究碰到像 Honda/UCSD 資料庫[36]與 VIPlab 資料庫中人臉有大角度的姿勢變化、表情變化、與遮蔽的問題時，皆能成功的解決，達到 90% 以上的辨識率。

本研究目前使用直方圖等化的方式來解決光線所造成的問題，然而近幾年有許多論文提出採用 gabor 濾波器解決不同時間點拍攝所造成的光線問題，皆有不錯的成效。本研究目前建立了人臉辨識的模型架構，未來可專注在影像的前處理問題上，當光線的問題也能克服時，本研究之模型架構將能被更廣泛運用。而另一方面，當此人臉模型建立的時間過長，人臉訓練用之資料與測試用之資料彼此可能會有極大的差異，例如：髮型個改變，鬍鬚變長等等，因此如何即時的(on-line)更新人臉模型，並且能夠將自動更新人臉模型的機制加入本架構中，將是未來的一大挑戰。

目前本研究在人臉辨識上還是屬於半自動的狀態，因為一個完整的人臉辨識系統，需與追蹤做結合，才能達到一個全自動之人臉辨識系統的標準，一但追蹤的效果不好，辨識的準確度也會受到影響，期望在未來能夠和追蹤相輔相成，進而製作出一個更人性化的人臉辨識系統。

參考文獻

- [1] A.F. Abate, M. Nappi, D. Riccio, and G. Sabatino, “2D and 3D face recognition: A survey.”, *Pattern Recognition Letters*, vol.28, no.14, pp.1885-1906, Jan. 2007.
- [2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection.”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.19, no.7, pp.711-720, Jul. 1997.
- [3] M.S. Bartlett, H.M. Lades, and T.J. Sejnowski, “Independent component representation for face recognition.”, In *Proceedings of the SPIE Symposium on Electronic Imaging: Science and Technology; Conference on Human Vision and Electronic Imaging II*, pp.528-539, Jan. 1998.
- [4] M.E. Brand, “Incremental singular value decomposition of uncertain data with missing values.”, In *Proceedings of the Seventh European Conference on Computer Vision*, vol.2350, pp.707-720, May 2002.
- [5] S. Chen, H. Zhao, M. Kong, and B. Luo, “2D-LPP:A two-dimensional extension of locality preserving projections.”, *Neurocomputing*, vol.70, pp.912—921, Jan. 2006.
- [6] B.A. Draper, K Baek, M.S. Bartlett, and J.R. Beveridge, “Recognizing Faces with PCA and ICA.”, *Computer Vision and Image Understanding*, vol.91, pp.115-137, Feb. 2003.
- [7] G. Edwards, C. J. Taylor, and T. F. Cootes, “Interpreting face images using active appearance models.”, In *Proceedings of the Three International Conference on Automatic Face and Gesture Recognition*, pp.300-305, Apr. 1998.
- [8] Z. Ghahramani, “An Introduction to Hidden Markov Models and Bayesian Networks.”, *International Journal of Pattern Recognition and Artificial Intelligence*, vol.15, no.1, pp.9-42, 2001.
- [9] A. J. Howell, and H. Buxton, “Towards unconstrained face recognition from image sequences.”, In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, pp.224-229, Oct. 1996.
- [10] B. Heisele, P. Ho, and T. Poggio, “Face recognition with support vector machines: Global versus component-based approach.”, In *Proceedings of the Eighth*

- International Conference on Computer Vision, vol.2, pp.688-694, 2001.*
- [11] X. He, and P. Niyogi, "Locality Preserving Projections.", In *Proceedings of the Conference on Advances in Neural Information Processing System*, 2003.
- [12] Y. Hu, D. Jiang, S. Yan, L. Zhang, and H. Zhang, "Automatic 3D reconstruction for face recognition.", In *Proceedings of the Sixth International Conference on Automatic Face and Gesture Recognition*, pp.843-848, May. 2004.
- [13] P. Hall, D. Marshall, and R. Martin, "Incremental Eigenanalysis for Classification.", In *Proceedings on British Machine Vision Conference, vol.1*, pp.286-295, Sep. 1998.
- [14] P.M Hall, D.R. Marshall, and R.R. Martin, "Merging and splitting eigenspace models.", *IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.22*, no.9, pp.1042-1049, Sep. 2000.
- [15] A. Hadid, and M. Pietikainen, "From Still Image to Video-Based Face Recognition:An Experimental Analysis.", In *Proceedings of the Sixth International Conference on Automatic Face and Gesture Recognition*, pp.813-818, May 2004.
- [16] F.B. ter Haar, and R.C. Veltkamp, "3D Face Model Fitting for Recognition.", In *Proceedings of the Tenth European Conference on Computer Vision, vol.5305*, pp.652-664, Oct. 2008.
- [17] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, "Face Tracking and Recognition with Visual Constraints in Real-World Videos.", *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, Jun. 2008.
- [18] C. Liu, and H. Wechsler, "A shape- and texture-based enhanced fisher classifier for face recognition.", *IEEE Trans. on Image Processing, vol.10*, no.4, pp.598-608, Apr. 2001.
- [19] Y. Li, S. Gong, and H. Liddell, "Constructing facial identity surface in a nonlinear discriminating space.", *IEEE Conference on Computer Vision and Pattern Recognition, vol.2*, pp.258-263, Dec. 2001.
- [20] X. Liu and T. Cheng, "Video-based face recognition using adaptive hidden Markov models.", *IEEE Conference on Computer Vision and*

Pattern Recognition, vol.1, pp.340-345, Jun. 2003.

- [21] K.C. Lee, J. Ho, M.H. Yang, and D. Kriegman, "Video-Based Face Recognition Using Probabilistic Appearance Manifolds.", *IEEE Conference on Computer Vision and Pattern Recognition, vol.1*, pp.313-320, Jun. 2003.
- [22] K.C. Lee, J. Ho, M.H. Yang, and D. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds.", *Computer Vision and Image Understanding, vol.99*, pp.303-331, 2005.
- [23] K.C. Lee, and D. Kriegman, "Online Learning of Probabilistic Appearance Manifolds for Video-based Recognition and Tracking.", *IEEE Conference on Computer Vision and Pattern Recognition, vol.1*, pp.852-859, Jun. 2005.
- [24] A.M. Martínez, and A.C. Kak, "PCA versus LDA.", *IEEE Trans.on Pattern Analysis and Machine Intelligence, vol.23*, no.2, pp.228-233, Feb. 2001.
- [25] K. Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning*, PhD thesis, University of California at Berkeley, Computer Science Division, 2002.
- [26] P.J. Phillips, "Support vector machines applied to face recognition.", In *Proceedings of the Conference on Advances in Neural Information Processing Systems, vol.11*, pp.803-809, 1998.
- [27] G. Shakhnarovich, J.W. Fisher, and T. Darrell, "Face recognition from long-term observations.", In *Proceedings of the Seventh European Conference on Computer Vision, vol.3*, pp.851-865, May 2002.
- [28] 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.2, pp.88-103, 2006.
- [29] M. Turk, and A. Pentland, "Eigenfaces for recognition.", *Journal of Cognitive Neuroscience, vol.3*, no.1, pp.72-86, 1991.
- [30] J. Yang, and D. Zhang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition.", *IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.26*, no.1, pp.131-137, Jan. 2004.
- [31] J. Yang, D. Zhang X. Yong, and J. Yang, "Two-dimensional Discriminant Transform for Face Recognition.", *Pattern Recognition, vol.38*, no.7,

pp.1125-1129, Feb. 2005.

- [32] J. Yang, J. 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.
- [33] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, “Face recognition: a literature survey.”, *ACM Computing Surveys*, vol.35, no.4, pp.399-458, Dec. 2003.
- [34] W.Y. Zhao, and R. Chellappa, “Symmetric shape-from shading using self-ratio image.”, *International Journal of Computer Vision*, vol.45, no.1, pp.55-75, Oct. 2001.
- [35] S. Zhou and R. Chellappa, “Probabilistic human recognition from video.”, In *Proceedings of the Seventh European Conference on Computer Vision*, vol.3, pp.681-697, May 2002.
- [36] The Honda/UCSD Video Database,
<http://vision.ucsd.edu/~leekc/HondaUCSDVideoDatabase/HondaUCSD.html>.