國立交通大學

多媒體工程研究所

碩 士 論 文

以人臉為依據建立視訊影片中
人物出現時間之索引

Video Indexing by Information of Face Images

研 究 生：蘇偉誌

指導教授：王才沛　教授

中 華 民 國　九 十 九 年 七 月

以人臉為依據建立視訊影片中人物出現時間之索引

Video Indexing by Information of Face Images

研 究 生：蘇偉誌　　Student：Wei-chih Su

指導教授：王才沛　　Advisor：Tsai-pei Wang

國 立 交 通 大 學

多 媒 體 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Multimedia Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

July 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年七月

以人臉為依據建立視訊影片中
人物出現時間之索引

學生：蘇偉誌　　指導教授：王才沛

國立交通大學多媒體工程研究所 碩士班

摘要

本篇論文中，我們提出一套視訊中的人物分群流程，從人臉偵測、演員串列的建立、串列間關係的描述、分群方法的設計、到最後的串列擴張，都有清楚地介紹及討論，並針對投影基底、影像前處理、分群方法、描述資訊的選擇等多項議題進行討論與分析，當中又以身體資訊的使用為主要討論對象。

由於演員的臉部姿勢會跟著劇情變化或鏡頭位置而有所不同，造成臉部辨識的困難度，因此在進行演員串列間相似度／相異度之描述時，除了臉部資訊的使用外，我們加入演員的身體（衣服）資訊，希望藉由兩資訊的結合，達到提升描述力之目的。然而在使用身體資訊時，可能會有演員更換服裝之情形發生，因此我們提出“根據串列間的時間差距決定身體資訊權重”進行臉部資訊與身體資訊的結合，藉由可變權重的使用，降低不同服裝所可能產生的錯誤描述。

Video Indexing by Information of Face Images

Student：Wei-chih Su　　Advisor：Tsai-pei Wang

Institute of Multimedia Engineering
College of Computer Science
National Chiao Tung University

Abstract

In this thesis we propose a procedure for human face clustering from video. The procedure consists of human face detection, the construction of actor-sequences, the evaluation of similarity between actor-sequences, face clustering, and actor-sequence extension. The following issues are analyzed and discussed in this paper: the selection of basis vectors for subspace projection, the preprocessing of face images, various clustering algorithms, and the selection of descriptive information used in the clustering processing. Among them we have emphasized on the use of body appearance in the process.

The face pose of an actor/actress changes a lot within a video, and this causes much difficulty for face recognition. To overcome this problem and improve the description of individual actors, while computing the similarity/dissimilarity between actor sequences, we utilize the information of body/clothing appearance in addition to the face images. However, as an actor may change his/her clothes within the video, we propose a weighted summation method that adjusts the relative weighting of face and body appearances according to the time difference between actor-sequences. We find that this can reduce errors caused by changed clothing in the video.

誌謝

這篇論文能夠順利完成，最要特別感謝的是指導教授王才沛老師，對於我的任何問題，老師都會耐心的教導，幫助我瞭解問題所在並解決它，並且適時給予我許多有關研究方的向建議，讓我的論文能夠順利的完成，非常感謝老師。感謝莊仁輝教授及陳祝嵩教授擔任我的口試委員，在口試過程中提供許多寶貴的建議，使得本論文更加完備，特此致謝。此外，感謝在實驗室一起討論與努力的同學們，特別是崇桂與俞邦，有了你們的陪伴,在遇到問題時能夠互相討論及打氣,讓我的兩年研究所生活中不會感覺到孤單。還有學弟良佑、裕傑、俊予和俞丞，感謝你們的幫忙，讓我的研究過程更有效率，以及學長建順、聖文、詩祐、劉強、昇毅、耿維，感謝你們對我的照顧與指導。最後我要感謝我的家人以及朋友，謝謝你們的支持與關心，讓我能在無憂無慮的環境中完成碩士學業。

目錄

圖目錄

表目錄

第一章 簡介

1.1 研究動機

人臉偵測與人臉辨識的研究已有將近三十年的歷史，各式各樣的相關應用也相繼衍生出來，例如：生活中常見的門禁及監視系統，利用人臉辨識機制達到身分識別之目的；在數位相機中加入人臉偵測的功能，輔助鏡頭自動對焦等等，各式各樣以人臉分析為主題的應用，當中又以圖片相關之應用為多數。在此我們提出一以圖片分析為基礎，搭配視訊影片原有特性，將人臉偵測與辨識之應用延伸至影片層面之應用。

在資訊爆炸的今日，視訊影片已成為生活中最為常見，也最為便利的資訊傳輸媒介，不論是新聞報導、人文科技類知識影片、電視影集等等，各式各樣的視訊影片隨時隨地存在於我們的生活當中。由於影片數量相當龐大，以人工方式進行影片內容分析是沒有效率的，因此許多針對影片內容進行分析的演算法相繼被提出，例如視訊瀏覽（video browsing）、影片摘要（video summarization）、影片分類（video classification）等，其目的皆是在影片觀賞前，提供觀賞者想預先得知的影片資訊，作為影片類型篩選或影片中章節選擇之參考。

大多數的影片分析皆以整體畫面為分析依據。不同於此類型的分析，本篇論文將以畫面中人物為分析對象，建立影片中人物出現時間之索引，提供觀賞者影片中演員的相關資訊。所謂的人物出現時間之索引是以影片中個別角色為單位，分別記錄不同角色出現時間，建立出所有角色的出現時間表，如圖 1-1 的例子，其中黑色底線表示影片時間軸，帶狀線條表示個別角色出現的時間。為了索引過程的處理便利性，我們將連續的人臉影像合併為演員串列，接著求取串列之間的關係，最後利用分群演算法將串列分群至我們所指定的群組數量，例如圖 1-1 中我們將四個演員串列分成三群。

關於人物索引的探討，現存的文獻大多只使用人物之臉部特徵來進行演員的分群，除了臉部資訊，我們將加上身體特徵、以及人物出現時間等資訊來幫助分群工作之進行。除了各種資訊的探討外，本篇論文也將對人物分群流程中多項議題進行討論。

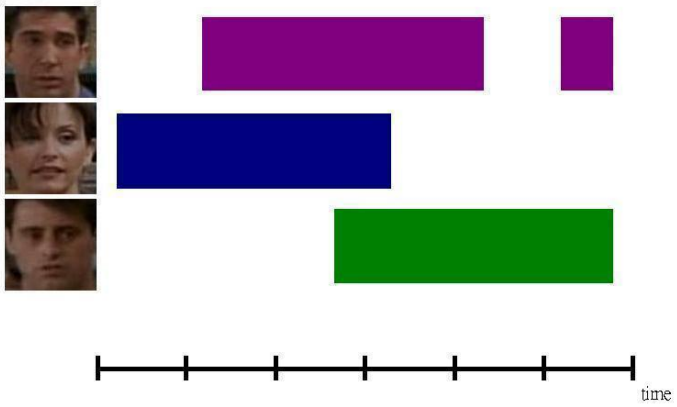


圖 1-1：人物索引範例示意圖

1.2 章節概要

接下來的四章中，第二章介紹影像索引，以及人臉偵測、辨識、分群的相關文獻，在第三章介紹本篇論文所提出的人物分群演算法，第四章中討論各種不同方法與議題之實驗結果，第五章為結論與未來的展望。

第二章 文獻探討

2.1 視訊瀏覽（Video Browsing）

如同 Zhong 等人在[1]所採用，常見的視訊瀏覽包含場景轉換偵測（shot detection）和場景分群（shot clustering）兩大部分。首先利用 RGB 色彩直方圖進行畫面的表示，若相鄰兩畫面之色彩直方圖差異大於給定的數值，則兩畫面屬於不同場景，即發生場景轉換。接著各場景中皆挑選出一色彩直方圖作為場景代表，最後藉由各場景之色彩直方圖搭配分群演算法進行場景分群，將視訊影片當中具有類似特性之畫面或片段分配至相同單元中，提供影片檢視之功能。為了獲得多影片資訊，Zhong 等人在[1]中提出階層性瀏覽（hierarchical video browsing）的概念，針對各場景群組再進行更細部的分群動作，透過多階層的場景分類，提供更多且更精確的視訊資訊。以新聞影片為例，在第一階層中劃分為主播台、人物受訪、運動新聞、氣象報導等不同單元。接下來針對不同單元繼續作分類，例如將主播台單元再分成雙主播螢幕、單一主播等各種播報型態，或是將人物受訪單元進行人物的區分。

不同於上述技術，Tan 等人在[2]中提出利用圖型理論中端點切割（graph partition）之概念進行場景的分群，每一個場景均代表一個端點（node），利用 HSV 色彩空間（hue-saturation-value color space）之色彩直方圖進行相似度（likeness）的評估，若兩場景具有高相似度，則以邊（edge）連結兩場景所對應之端點，最後，對此圖型進行切割達到場景分群之結果。

除了針對畫面內容進行的瀏覽外，許多以演員為瀏覽對象的方法也被廣泛討論，將演員出現的片段標示於時間軸上，提供使用者關於演員的演出資訊。例如 Ma 等人在[3]中提出，針對自製影片內的演員片段進行分類，利用演員的臉部影像作為分群元素，搭配人臉辨識技術進行分群，達到人物索引之目的。Peker 等人在[4]中同樣提出人物索引的分法，將新聞節目以及脫口秀影片進行人物的分群，並將瀏覽單元分為單人物、雙人物、多人物等多種以人物為主要對像的視訊瀏覽功能。

2.2 人臉偵測（Face Detection）

現存的人臉偵測演算法大多都將“膚色像素（skin color pixel）”所組成膚色區塊（skin region）作為人臉偵測的首要處理對象，再搭配不同的機制來進行人臉的偵測。Czirjek 等人在[6]提出以膚色區塊的面積、形狀、旋轉角度等資訊作為第一階段篩選之依據，接著利用 Eigenfaces 對通過第一階段的膚色區塊進行人臉之判定。相似於[6]中所提出的判定方法，Jin 等人在[7]提出先對膚色區塊進行眼睛偵測（eye detection），利用眼睛座標將膚色區塊旋轉至水平的正臉，最後再與資料庫人臉進行比對計算（template matching），決定是否為真實人臉。而 Hsu 等人在[8]中所提出：將膚色區塊分別進行眼睛偵測、嘴巴偵測、以及橢圓偵測，利用這三個偵測結果判定膚色區塊是否為真實人臉，除此之外，這三個偵測結果還能夠提供我們臉部姿勢（face pose）資訊。

除了上述的幾種偵測方法，Hjelmas 等人在[5]中整理了所有常見的臉部偵測方法，主要分成特徵導向（feature-based）和影像導向（image-based）兩種方式，其中特徵導向包含：低階分析（low-level analysis）、特徵分析（feature analysis）和有效形狀分析（active shape models）三種主要形式；而影像導向則分為：線性子空間方法（linear subspace methods）、類神經網路（neural networks）和統計方法（statistical approaches）等。

2.3 人臉辨識（Face Recognition）

在人臉辨識的工作上，由於大部分的人臉影像包含數百甚至數千個像素，以大小70*70的影像為例，就有4900個像素，為了降低高維度所帶來的高度計算量，各種利用較低維度向量表示人臉的方法一一被提出，Turk 等人在[9]中介紹了最為著名且常見的方法—Eigenfaces。

除了Eigenfaces外，Kepenekci在[10]中整理了相當多進行人臉辨識的方法，例如：FLD（Fisher’s Linear Discriminant）、LDA（Linear Discriminant Analysis）、SVD（Singular Value Decomposition）、Hidden Markov Model、Template Based Matching等多種辨識方法，並提出Gabor Wavelet，將人臉以空間頻率（spatial frequency）、空間局部性（spatial locality）、方向選擇性和（orientation selectivity）等特徵來表示。此外LBP（Local Binary

Patterns）也是常被拿來進行臉部影像描述的工具，例如Ahonen等人在[11]即是使用此工具來進行人臉辨識。

Zhao等人在[12]中對臉部辨識的方法進行統整，將辨識方法分類為以下三種：整體匹配方法（holistic matching methods）、基於特徵的匹配方法（feature-based (structural) matching methods）和混合方法（hybrid methods），除了對各方法一一進行說明外，也討論了許多相關議題，例如光影對影像的影響、影片品質的影響、以及圖片或視訊影片來源之人臉影像等進行了探討。

2.4 人臉分群（Face Clustering）

所謂的人臉分群，即是多個臉部影像同時進行人臉辨識之行為，直覺上，我們可以藉由上段所介紹的臉部描述方法，配合分群演算法對量化的臉部資訊進行分群，然而在實際應用時，卻常因為臉部影像的過大變化，造成分群效果的低落，因此各種針對不同特性之人臉影像所產生的分群方法被相繼提出。

如同[12]中所說，來自圖片與來自視訊影片之人臉是不同的，來自視訊影片之臉部影像較來自圖片的影像有更多額外資訊可以使用。例如 Tao 等人在[13]中使用畫面相依性及時間軸兩種資訊，藉由畫面之間的相依特性，將影片中連續出現的人臉結合成為演員串列（Actor Sequence），接著將各串列依照臉部姿勢切割成數個子串列，最後利用“在時間軸上重疊的串列，必屬於不同演員”以及“來自相同串列的子串列，必屬於相同演員”兩項時間軸資訊來幫助分群工作的進行。

上段中，Tao 等人希望透過上述方式[13]，解決臉部姿勢所造成的人臉分群問題，其中最常見的情況為：“不同角色、相同臉部姿勢”較“相同角色、不同臉部姿勢”更為相似，即多視角臉部分群（face clustering with multi-views）問題。針對此問題，Huang 等人在[14]中提出兩階段式分群，首先依據臉部姿勢進行第一階段分群，接著針對不同臉部姿勢的群組，分別進行一般的人臉分群演算法。面對類似問題，Ramanan 等人在[15]中提出以頭髮、衣服等額外資訊進行輔助，避開需要複雜的臉部姿勢矯正。

[15]所提出的方法，除了可以幫助不同臉部姿勢的人臉進行分群，也可藉由資訊權

重的設定，提升人臉間的描述力，例如影集資料中，同一集內的人臉分群和不同集間的人臉分群所使用的資訊權重是不同的，同一集中，可以大量倚重頭髮及衣服的資訊，不同集中，衣服資訊將不被使用。Khoury 等人也在[16]提出臉部資訊搭配衣服資訊的人物分群方法，藉由臉部影像之 SIFT(Scale Invariant Feature Transform)特徵、膚色資訊、身體區塊的色彩直方圖和衣服主要色彩等資訊，搭配其提出的階層式分群方法進行脫口秀節目中的人物分群。除了人物本身的資訊外，Yamamoto 等人在[17]提出場景資訊的使用，藉由人臉所在場景特性的應用，提升人臉分群之效率。

第三章 實驗方法

本章我們將介紹視訊影片內人物分群方法之流程，當中包含各步驟中方法之說明，圖 3-1 為我們提出之視訊影片中人物分群方法之流程圖。

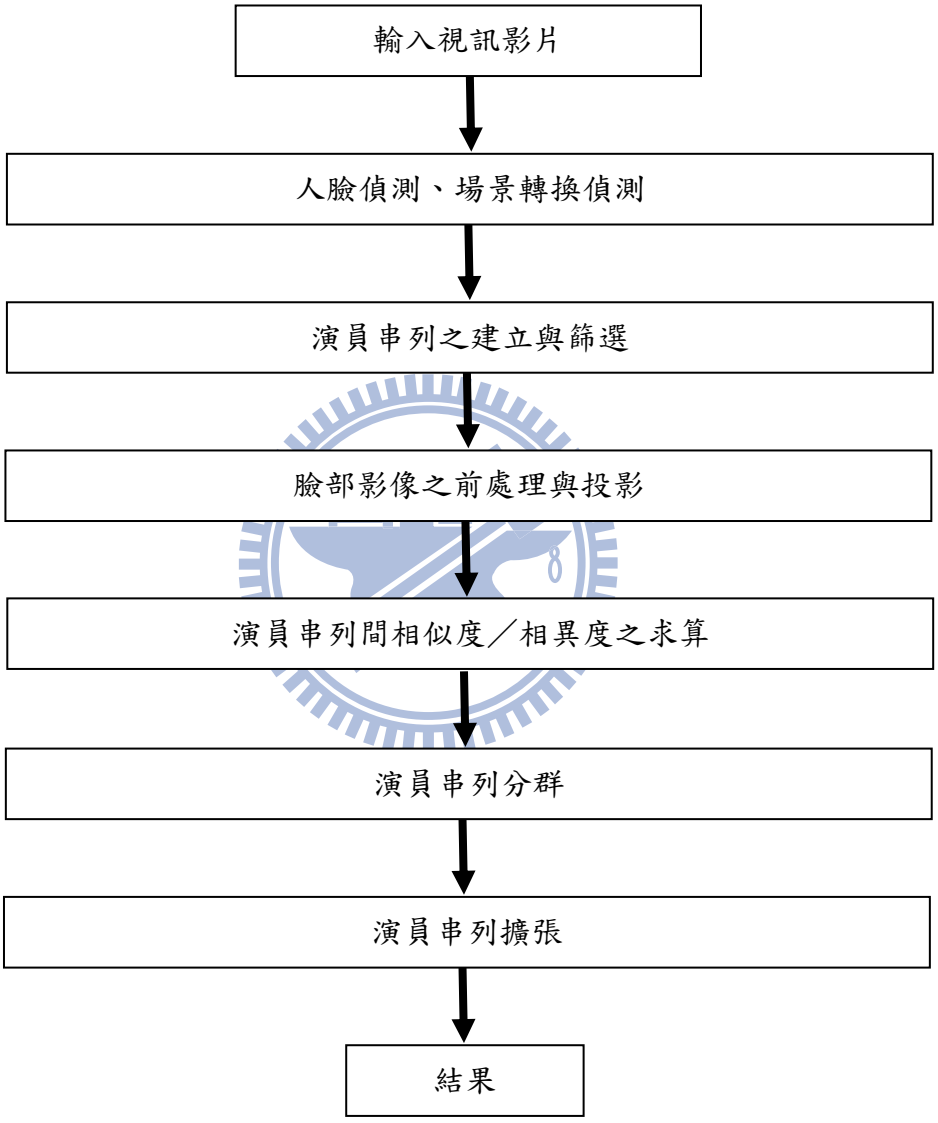


圖 3-1：視訊影片中人物分群演算法流程圖

3.1 前置作業

在正式進行演員分群前，我們必須先進行以下三項前置作業：首先由視訊影片中擷取一系列影像，取得截取影像之後，我們再接著進行人臉偵測、以及場景轉換偵測兩項工作。

3.1.1 影片畫面擷取

一般視訊影片畫面更新率（Frame Rate）大多為 30 FPS（Frames per Second），以長度 20 分鐘的短片為例，就含有三萬六千張畫面，為減少畫面重複率，我們利用 DVDVideoSoft.com 所提供的免費軟體“Free Video To JPG Converter”（http://dvdvideosoft.com/），以畫面更新率 5FPS 進行影片畫面的擷取，達到降低影像處理數量的目的，這些影像也將作為後續方法的分析對象。

3.1.2 人臉偵測

如同第2.2節中所提到，現存的人臉偵測演算法大多以膚色區塊作為人臉偵測的首要處理對象，整體畫面進行膚色偵測將是這類演算法的首要工作，除此之外，許多人臉判定工作也需跟著進行，為節省膚色偵測以及後續判定工作所需的龐大處理時間，我們將直接使用Intel Corporation（http://opencv.willowgarage.com/wiki/Welcome）所開發OpenCV (Open Source Computer Vision) Library中的人臉偵測指令，找出各畫面中近似人臉的物件，作為後續分析之主要依據。

3.1.3 場景轉換偵測

我們利用簡易色彩直方圖（color histogram）差距來進行場景轉換偵測，計算兩連續畫面色彩直方圖之差距，若差距大於給定的臨界值（threshold），則兩畫面屬於不同場景；反之，則為相同場景。

簡易色彩直方圖的求取方法：首先給定區間寬度（interval width），屬於相同區間的灰階值將被視為相同量值，接著分別求取影像 R、G、B 三個部份之區間直方圖，最後

再將三個直方圖合併作為此畫面之色彩直方圖。以區間寬度 64 為例，灰階值將切割成 0~63、64~127、128~191、192~255 四個區間進行像素個數的統計，分別求取 R、G、B 三個部份的區間直方圖後，再將其合併，可得到一個含有 12 個區間之直方圖，如圖 3-2 所示。

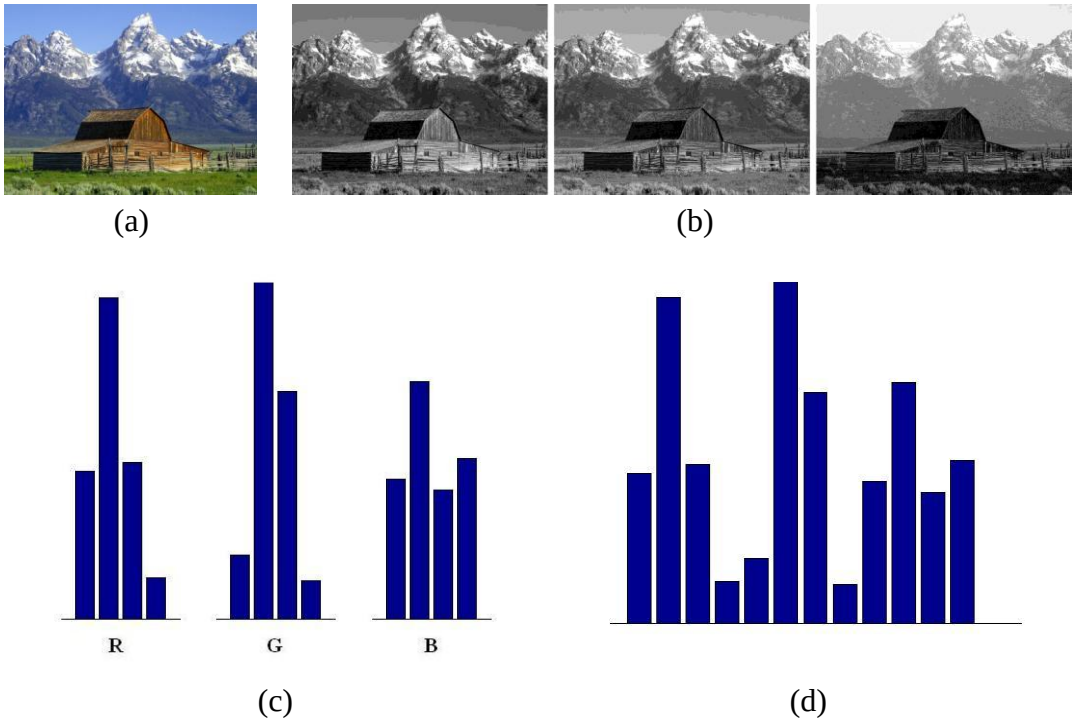


圖 3-2：簡易色彩直方圖示意圖（interval width = 64）
(a) 原圖　(b) R、G、B 三個部份
(c) 三個部份之區間直方圖　(d) 合併後的色彩直方圖

在本論文中，我們使用的區間寬度為 16，所得色彩直方圖為 1*48 之向量，再搭配歐氏距離(Euclidean distance)來計算兩畫面之色彩差距，若差距大於給定之臨界值 50000，則兩畫面為不同場景。

3.2 演員串列之建立與篩選

對於影片中的主要演員，其每次出場皆為主要的拍攝目標，且時間也會在一定的長度之上，因此 OpenCV 將偵測到連續出現的同一演員之臉部位置，為降低眾多人臉影像獨立處理的複雜度，我們將各場景（shot）中屬於同一位演員的連續偵測合併成演員串列，作為接下來分群作業的基本元素。除了人臉影像，我們在建立過程中也一併將人臉下方的身體影像記錄於演員串列當中，其大小將設定為兩倍臉部寬度所形成的方形，如圖 3-3 所示。下面我們介紹演員串列的建立以及篩選方法。

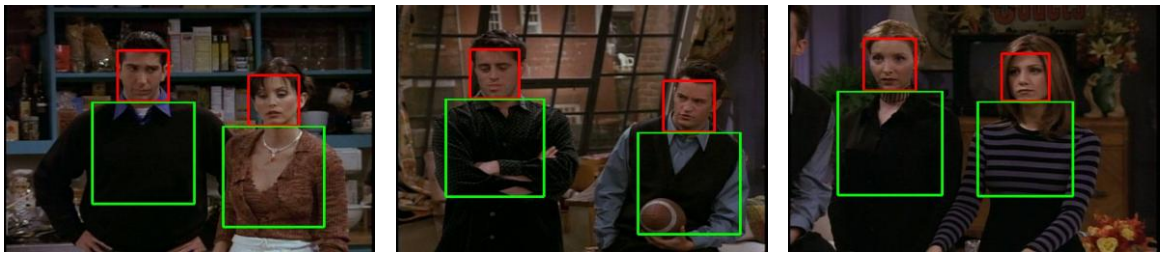


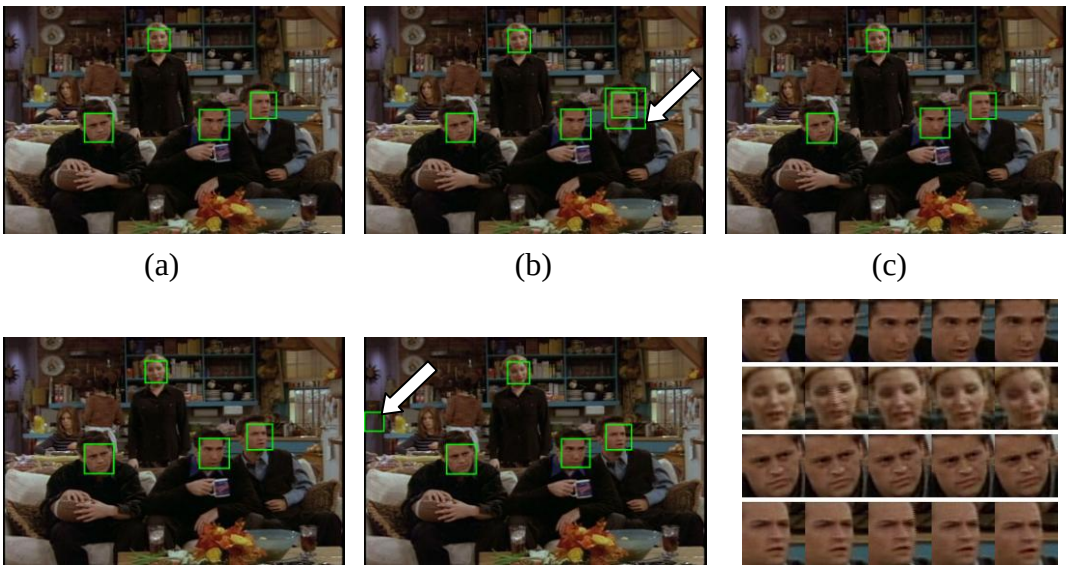
圖 3-3：身體影像與臉部影像之對應關係

3.2.1 演員串列的建立

為了確保演員串列中所有人臉影像同屬相同一位演員，所有演員串列必須滿足下列幾項條件：

- 所有人臉必須在相同場景內、並且為連續出現、
- 人臉的位置差距必須在 35 個 pixel 距離內，即位置相近、
- 人臉影像的大小必須在前一張人臉的 0.5~1.5 倍內，即大小固定、
- 人臉個數必須在 3 個以上。

藉由上述限制，我們可以移除部分 OpenCV 的錯誤偵測，例如出現於背景中的錯誤偵測、人臉鄰近的重覆或錯誤偵測等。如圖 3-4，(a)至(e)為五張連續影像的人臉偵測結果，其中(b)中有鄰近的錯誤偵測、(e)內有背景影像的錯誤偵測，而(f)為四個演員串列的建立結果。



(a) (b) (c)

(d) (e) (f)

圖 3-4：利用演員串列建立條件成功移除錯誤偵測
(a)~(e)為連續畫面的人臉偵測結果，藉由串列建立規則，
我們可將白色箭頭所指示的錯誤偵測移除。
(f)為此 5 個畫面所得的 4 個演員串列。

由於我們採用簡易色彩直方圖進行場景轉換偵測，因此有些顏色較相近的場景轉換可能無法被偵測出來，此時若相同位置上有不同人物之人臉，則這些屬於不同演員的人臉影像將被串聯至同一個串列中。為了確保串列中人物的一致性，我們利用人物之臉部及身體資訊判定串列中是否含有不同演員。針對所有演員串列，依序檢查內部相鄰元素，若兩相鄰元素之臉部相異度大於 2000、且身體相異度大於 2500，則判定兩元素屬於不同演員，並分割此串列。其中，臉部及身體之相異度求算方法將於第 3.4 節中說明。圖 3-5 中為含不同演員之串列成功切割範例。

圖 3-5：含不同演員之串列利用臉部及身體資訊成功切割（黑色方塊為切割點）

3.2.2 演員串列的篩選

當 OpenCV 連續的錯誤偵測滿足串列之建立規則時，則連續的非人臉影像亦會形成演員串列，因此我們將對每一個串列進行檢測，透過影像的膚色資訊，判斷是否為真實人臉所組成的串列。

本論文中，我們使用洪詩祐在[18]所提出的膚色偵測公式與人種參數（如表 3-1），透過色彩空間 YCbCr，將膚色機率大於臨界值的像素視為膚色像素，再將膚色像素個數除以像素總數，得到膚色比例，最後再把膚色比例小於 50%之影像移除。若串列中沒有任何影像被保留，則判斷此串列為非人臉串列。觀察 OpenCV 所偵測得到的臉部影像，僅管是非正臉，其膚色比例仍會超過整體影像的一半，因此我們以 50%作為篩選的臨界值，如圖 3-6，可將非真實人臉的演員串列移除。

[18]中所提出的方法是忽略亮度資訊 Y，僅使用 Cb 與 Cr 資訊進行偵測。公式(1)為膚色機率之計算方法，其中向量 $\vec{x}$ 代表受測像素之 Cb 與 Cr 值、$\vec{\mu}$ 與 $\sum$ 皆為人種參數（Cb、Cr 之 mean 及 covariance matrix），人種膚色參數如表 3-1 所列，在後續實驗中我們僅使用亞洲人以及白種人兩組人種膚色參數，而膚色機率的臨界值為 0.1。

$$P = \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1}(\vec{x}-\vec{\mu})\right) \quad (1)$$

表 3-1：各人種之膚色參數

參數 \ 人種	亞洲人	白種人	黑種人
$\vec{\mu}$	$\begin{bmatrix}111.88\\148.09\end{bmatrix}$	$\begin{bmatrix}113.17\\149.03\end{bmatrix}$	$\begin{bmatrix}111.82\\148.05\end{bmatrix}$
Σ	$\begin{bmatrix}70.20 & -50.91\\-50.91 & 66.47\end{bmatrix}$	$\begin{bmatrix}44.98 & -31.01\\-31.01 & 46.60\end{bmatrix}$	$\begin{bmatrix}75.48 & -64.68\\-64.68 & 76.95\end{bmatrix}$

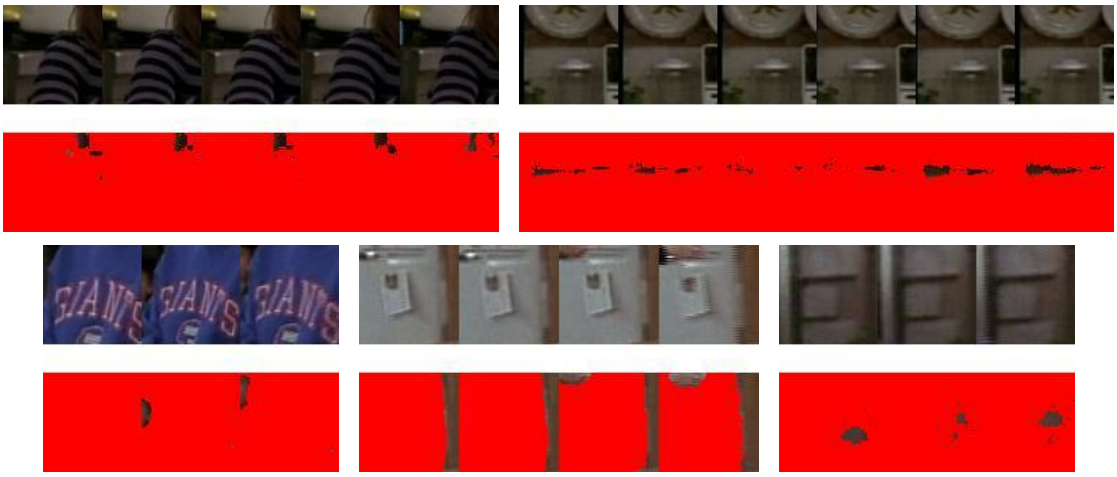
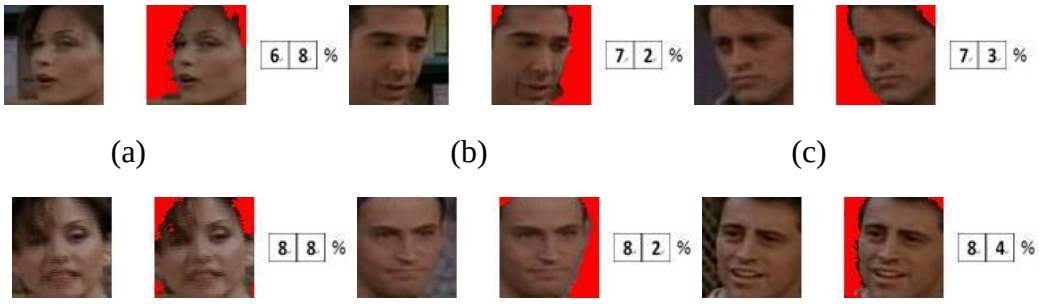


圖 3-6：使用膚色資訊移除的非人臉串列（紅色部分代表非膚色像素）

膚色資訊的應用，除了上述非人臉串列的移除（如圖 3-6），我們也將膚色資訊作為正臉／非正臉（frontal／non-frontal face）的臉部姿勢（face pose）判別依據。由於 OpenCV 偵測所的人臉皆有以鼻子為中心的特性，因此我們利用臨界值 80%來作為臉部姿勢判定依據，若膚色比例超過 80%，則我們將此人臉影像歸類為正臉，相反，歸類為非正臉，如圖 3-7 中包含兩種臉部姿勢的範例圖。



(a) (b) (c)

(d) (e) (f)

圖 3-7：膚色資訊判別正臉／非正臉（紅色部分為非膚色像素）
(a)(b)(c) 膚色比例小於 80%，歸類為非正臉
(d)(e)(f) 膚色比例大於 80%，歸類為正臉

3.3 臉部影像之前處理與投影

影片拍攝時，例如光影的變化、鏡頭焦距調整等，許多環境或攝影器材之因素都可能造成影片品質的差異，為了統一所有人臉影像之特性，所有影像在使用之前皆須進行相同的處理。接著，再使用 Eigenfaces 進行影像之投影，達到降低維度之目的，整體處理流程如圖 3-8 所示。

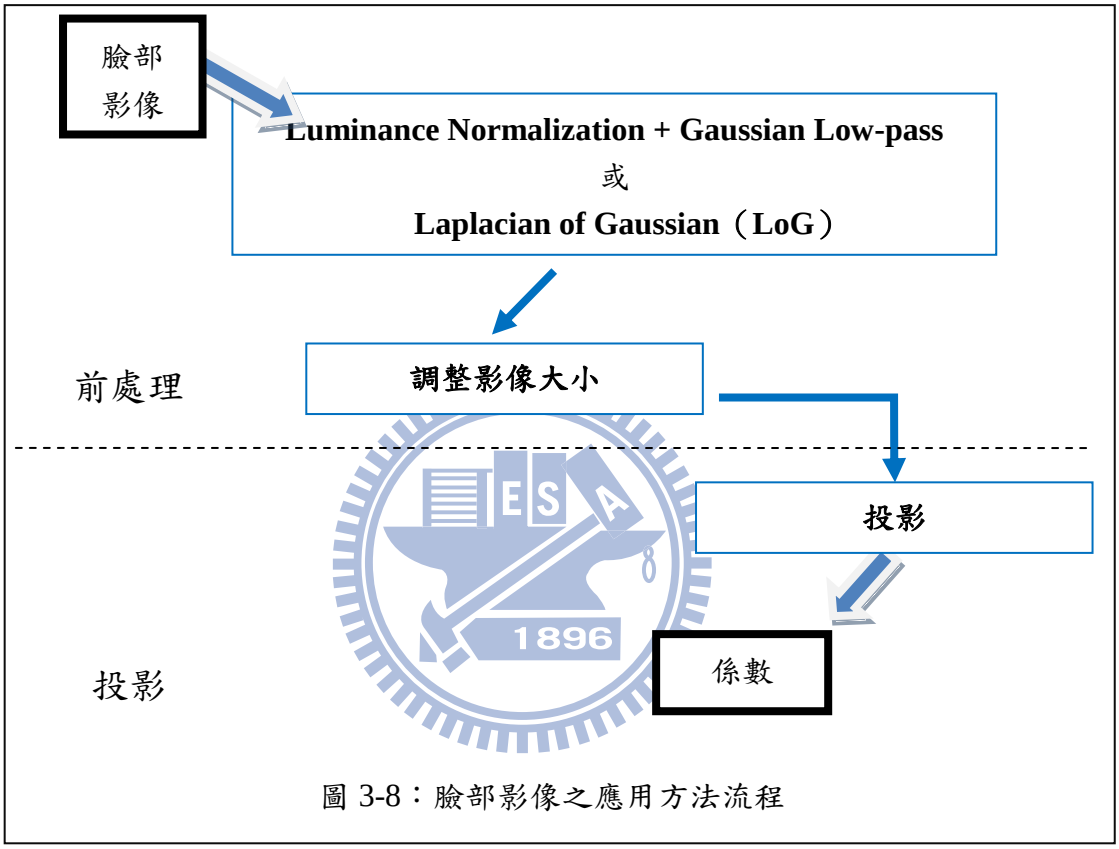


圖 3-8：臉部影像之應用方法流程

3.3.1 臉部影像前處理

除了常見的模糊化（低通運算）處理外，我們也嘗試使用高通運算來統一臉部影像之特性，在第 4.5 節中我們也進行這兩種前處理方法的比較。以下我們先介紹兩種前處理方法。

光影平衡（luminance normalization）顧名思義是為了降低光影變化所產生的影響。我們分別求取 RGB 三個分量的平均值與標準差，並將三組平均值與標準差均調整至特定的 μ_0 及 δ_0，使所有影像達到相同亮度。以單一部份為例，假設其平均值與標準差分

別為 μ 及 δ，我們利用以下公式對量值 x 進行調整：

$$x \to (x-\mu)\cdot\frac{\delta_0}{\delta}+\mu_0 \text{。} \tag{2}$$

接下來將光影平衡後的影像轉換為灰階影像，再利用大小為 5x5、標準差為 1 的高斯低通遮罩（Gaussian low-pass mask，(3)）進行低通運算（low-pass），最後將所有人臉影像調整至相同大小（70*70），以方便後續投影處理之使用。

另一種前處理方式 LoG，則是直接將臉部影像利用 LoG 遮罩(4)進行高通運算（high-pass），同樣的運算後的影像也被調整為相同大小。圖 3-9 為兩種不同前處理方法所得的影像。

$$\frac{1}{273}\begin{bmatrix}1&4&7&4&1\\4&16&26&16&4\\7&26&41&26&7\\4&16&26&16&4\\1&4&7&4&1\end{bmatrix} \tag{3}$$

$$\begin{bmatrix}0&0&-1&0&0\\0&-1&-2&-1&0\\-1&-2&-16&-2&-1\\0&-1&-2&-1&0\\0&0&-1&0&0\end{bmatrix} \tag{4}$$

圖 3-9：兩種前處理所得影像（左：原人臉影像、中：Luminance Normalization + Gaussian Low-pass 結果、右：LoG 結果）

3.3.2 臉部影像投影

經過前處理後，大小 70*70 的影像仍因維度過大不便被直接使用，因此需要其他的描述方法來表示各影像。在本論文中，我們使用 Eigenfaces 來描述臉部影像[9]。我們挑選 OpenCV 用於自行拍攝影片中偵測所得的人臉影像，作為 Eigenfaces 的訓練集（training set）。我們將訓練集內經過前處理的影像轉換成 4900*1 之向量，並計算其平均值 Ψ 與 covariance matrix，接下來運用主成分分析（Principal Component Analysis；PCA）的概念，將 covariance matrix 之最大 70 個特徵值（eigenvalue）所對應之特徵向量（eigenvector）集合成為投影基底（即維度 70 之 Eigenfaces）。其中維度 70 是透過實驗觀察所選定的，詳細介紹請見第 4.4 節。

臉部影像之投影座標求算方法：首先，將影像大小調整至 70*70 並轉換為 4900*1 之向量型式，接著，將此向量減去平均值 Ψ，最後，利用投影基底求算其投影係數。此係數即代表此影像在投影基底下之描述，以下，我們將利用此係數作為臉部運算之元素。

3.4 演員串列間相似度／相異度之求算

本節中我們先介紹演員串列間的相似度／相異度進行求算方法，以作為後續分群法之依據。除了臉部資訊外，我們參考了 Ramanan 等人在[15]中提出的頭髮、衣服等額外資訊之輔助，將衣服資訊也列入相似度／相異度的描述中。以下我們將分別介紹臉部資訊及身體資訊的求算方法、以及如何將兩資訊合併。

3.4.1 臉部相似度／相異度

在此我們使用相異度（dissimilarity）來描述臉部影像間的關係。經過 3.3 中所介紹的處理後，臉部影像將被轉換成一組投影係數，針對兩影像的投影結果，我們使用歐氏距離來計算它們的相異度。兩臉部串列的相異度則定義為兩串列中最接近的兩個臉之相異度，而所有串列間的臉部相異度將被儲存於 D_{Face} 矩陣中。

3.4.2 身體相似度／相異度

在身體資訊的應用上，同樣以相異度來描述身體影像間的關係，在計算歐氏距離之前，必須先取得身體影像的特徵向量。在此我們利用區間寬度 16 之三維色彩直方圖（3-D color histogram）取得身體影像的色彩座標，而身體串列間的相異度求算方法與臉部相異度求算方式相似，類似地 D_{Body} 矩陣中記錄所有串列間的身體相異度。

為了得到比簡易色彩直方圖更精準的影像色彩表示，三維色彩直方圖是將影像之 RGB 值視為單一向量，也就是利用三維空間進行直方圖的求算。以區間寬度（interval width）64 為例，簡易色彩直方圖將色彩空間分割成個 $\frac{256}{64}*3=12$ 子空間（subspace）；而三維色彩直方圖則把色彩空間分割成 $\left(\frac{256}{64}\right)^3=64$ 個子空間。

3.4.3 臉部及身體資訊的合併

在兩資訊合併之前，必須將相異度（dissimilarity；簡寫為 d ）轉換為相似度（similarity；

簡寫為 s），我們利用指數公式分別將臉部以及身體相異度轉換至 $[0,1]$ 區間之相似度：

$$d_{Face} \to e^{-d_{Face}^2 / 2\delta^2} \text{，其中 } \delta = std(D_{Face})\text{；} \tag{5}$$

$$d_{Body} \to e^{-d_{Body}^2 / 2\delta^2} \text{，其中 } \delta = std(D_{Body})\text{。} \tag{6}$$

接下來的相似度合併，我們提出“依時間差距決定權重”的資訊合併方式。隨著劇情變化，演員在一段時間後常會有更換衣服的行為發生，若採取固定比例權重，會對串列相似度的描述產生干擾，因此我們依據串列之間的時間差距來決定身體權重，此機制包含以下兩個參數：決定基本權重之參數 *height*（簡寫為 h）、以及決定權重（隨時間差距）遞減速度之參數 σ 來決定身體資訊權重。其中身體權重最小為 0、最大為 1，因此 h 之範圍在 $[0,1]$；而 $\sigma = \infty$ 時表示權重不隨時間差距遞減，即固定比例的權重方式。兩資訊的合併權重求算方法如下，首先給定 h 與 σ，接著計算兩串列之畫面差距數量 x，再利用下列公式取得身體及臉部資訊權重：

$$w_{Body}(x) = h \cdot e^{-\frac{x^2}{2\sigma^2}} \text{、} \tag{7}$$

$$w_{Face}(x) = 1 - w_{Body}(x)\text{。} \tag{8}$$

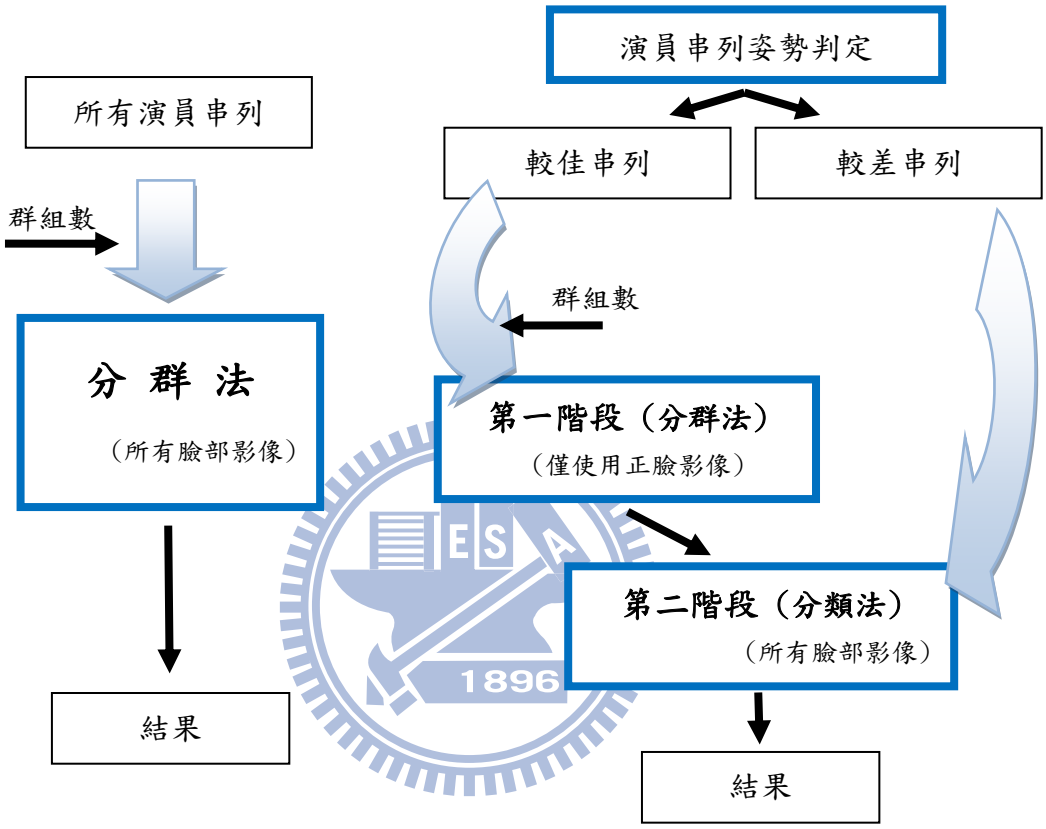
最後利用所得權重進行資訊的合併，並將相似度轉換為相異度：

$$s = w_{Face} \cdot s_{Face} + w_{Body} \cdot s_{Body} \text{、} \tag{9}$$

$$d = 1 - s\text{。} \tag{10}$$

3.5 演員串列分群

本節中將先介紹兩種分群策略。接著介紹論文中所使用之分群法 （clustering method）、和兩階段分群方法之後半段程序—分類法（classification method）。最後說明各種不同方案之選擇。



(a) 所有串列一併分群流程圖　　(b) 兩階段分群方法流程圖

圖 3-10：分群策略流程圖

3.5.1 分群策略

除了將所有串列一併進行分群之外，我們提出兩階段分群方法，其目的在於減低 Huang 等人在[14]中所提到的側臉影像對人臉辨識（Face Recognition）可能造成的影響，即“不同角色、相同臉部姿勢”較“相同角色、不同臉部姿勢”更為相似。首先利用第 3.2.2 節中所得臉部姿勢判定結果，將串列中的非正臉（non-frontal face）暫時移除，若移除後串列中仍含有影像，則將此串列歸類為較佳串列。反之，若串列中不含任何影像，則歸類於較差串列。接下來僅以較佳串列搭配保留的正臉影像進行分群（第一階段），最

後利用串列中的所有臉部影像進行較差串列的群組分類（第二階段）。兩種分群策略之流程如圖 3-10。

3.5.2 分群法（Clustering）

“所有串列一併進行分群”以及“兩階段分群方法的第一階段”所使用的分群法有以下幾種：階層式分群演算法（Hierarchical Clustering Algorithm）[19]、匈牙利演算法（Hungarian Algorithm）[20]、以及 Karypis Lab 所提供的相關性資料（relational data）分群程式[21]。

在分群的過程中我們也將時間資訊作為參考資訊，如同[13]中所提及，兩演員串列於時間軸上發生重疊，則兩串列必定為不同演員，在此我們將此現象稱之為碰撞（collision），若分群過程中採用此資訊，則碰撞的演員串列永遠無法被分類於同一群組中。第四章中我們將針對碰撞資訊的使用進行討論。

◆ **階層式分群演算法：**

在此，我們採用聚合型（agglomerative）階層式分群演算法進行分群，方法如下：

輸入相異度矩陣 M 以及群數 C。
目前群組數 C' = 矩陣之維度。
while($C' \neq C$)
　　尋找 M 中的最小相異度，假設為 m_{ij}。
　　合併群組 C_i 與群組 C_j，$C' = C' - 1$。
　　更新相異度矩陣 M。
End of while

當兩群組 C_i 與 C_j 合併形成群組 C_q 時，其餘群組 C_s 與新形成之群組 C_q 的距離之更新方式有下列幾種，公式中 $d(C_m, C_n)$ 表示群組 C_m 與群組 C_n 之相異度：

Single-link：

$$d(C_q, C_s) = \min\{d(C_i, C_s), d(C_j, C_s)\} \tag{11}$$

Complete-link：

$$d(C_q, C_s) = \max\{d(C_i, C_s), d(C_j, C_s)\} \tag{12}$$

Average-link：

$$d(C_q, C_s) = \frac{n_i}{n_i + n_j} d(C_i, C_s) + \frac{n_j}{n_i + n_j} d(C_j, C_s) \tag{13}$$

其中 n_i 及 n_j 分別為群組 C_i 和群組 C_j 內的元素數量

Ward：

$$d(C_q, C_s) = \frac{n_i + n_s}{n_i + n_j + n_s} d(C_i, C_s) + \frac{n_j + n_s}{n_i + n_j + n_s} d(C_j, C_s) - \frac{n_s}{n_i + n_j + n_s} d_{ij}' \tag{14}$$

$$d_{ij}' = \frac{n_i n_j}{n_i + n_j} d_{ij} \quad \text{其中} \quad d_{ij} = \|m_i - m_j\|^2 \quad \left(m_q = \frac{1}{n_q} \sum_{x \in C_q} x \right) \tag{15}$$

◆ **匈牙利演算法：**

藉由利用匈牙利演算法（Hungarian Algorithm）處理最小成本問題（minimum cost problem）之概念進行分群，此時成本矩陣（cost matrix）為先前所定義之相異度矩陣。分群方式如 Goldberger 等人在[20]中所提，若矩陣之第(i,j)項為構成最小成本之一個元素，則合併群組 i 與群組 j，持續進行合併，直到群組數到達指定個數為止。由於每回合都會有多組群組的合併，為了使演算法收斂至給定的群組數，我們對原程式進行些微修改。若合併後群組數小於我們所給定的群組數時，則重新此回合，並僅進行較相近群組的合併，使群組數達到我們所給定的數值。假設目前群組數為 6 而給定的群組數為 5，若此回合的合併中為兩兩群組進行合併，則新的群組數為 3，由於 3 小於給定的群組數，因此我們重新進行此回合，並僅合併最相近兩群組以達到群組數 5 之目標。

◆ **METIS：**

在此我們將各演員串列視為圖中的端點（node），利用 Karypis Lab 所提供的圖型端點切割（graph partition）程式，輸入端點間相似度（即 similarity matrix）以及群組個數後，即可得到分群結果。此方法之優點在於執行時間非常短；而缺點則是所有端點將被

平均分配至各群組中，即數量較少的演員將會與其他演員分配至同一群組，無法獨自形成一群。

3.5.3 分類法（Classification）

兩階段分群方法的第二階段，我們利用較差串列與各群組內較佳串列之關係，將較差串列分類至最合適的群組中，使用的方法有以下三種：

- **最小距離分類（Minimum Distance Classification）**

  將此較差串列分類至最相似較佳串列所在群組。

- **K-NN（K Nearest Neighbor）**

  相似於一般熟悉之 K-NN，找出與受測串列最相似的 K 個較佳串列，藉由此 K 個串列的所屬群組，決定受測串列應被分類之群組。其中，為了避免有群組元素小於 3 造成檢測過程的缺失（假設群組C_m僅含一個較佳串列，此較佳串列同時為與受測較差串列最相似之串列，此時若第二與第三接近的較佳串列同在群組C_n內，則受測串列將被分類至群組C_n當中），我們用以下公式來決定 K 值：

$$K = \min\left\{3, \min_i |C_i|\right\} \text{。} \tag{16}$$

- **Modified K-NN**

  我們在每個群組內皆找出 K 個最相似的較佳串列並進行相似度之加總。最終此串列將分類至相似度總合最高的群組。K 值求算方法同公式(16)。

在分類法進行時，我們有兩種模式可供選擇，第一種，較差串列僅與上階段中分群完成的較佳串列進行比較；第二種，將完成分類工作的較差串列也列入後續較差串列分類工作的比較對象，稱動態分類（dynamic classification）。假設目前有 15 個較佳串列與 5 個較差串列，若採取第一種分類模式，則 5 個較差串列都只有 15 個比較對象。若採取第二種模式（動態分類），則第 1 個較差串列有 15 個比較對象、第 2 個較差串列有 16

個比較對象（加上第 1 個較差串列）、…、第 5 個較差串列有 19 個比較對象（加上第 1~4 個較差串列）。

3.5.4 Prototype 的使用

當演員串列中的人臉影像大於一給定數量時，串列中可能存在少數幾個相對離群的人臉影像。為防止部分離群影像對相異度求算產生影響，我們利用模糊 C 均值演算法（Fuzzy C-menas Algorithm；FCM）找出特定個數的群聚中心，並以這些群聚中心作為串列的代表影像。此概念僅使用於分群法過程中，因此只有下列兩種狀況發生時須對演員串列元素進行模糊 C 均值演算法，一個是所有串列一併進行分群策略中元素個數超過 10 的演員串列；另一個是兩階段分群方法中元素個數大於 10 的較佳串列（僅含正臉）。

3.6 演員串列擴張

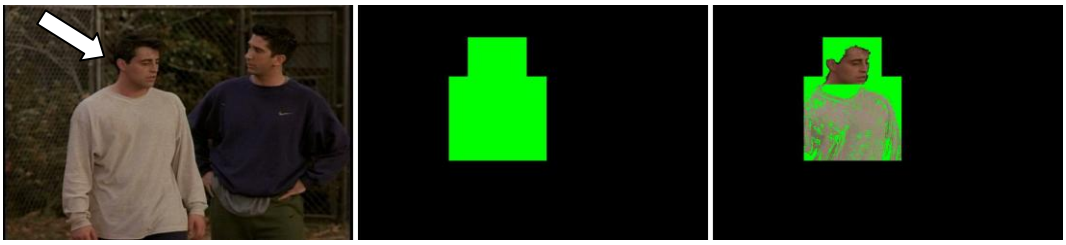
由於 OpenCV 無法偵測姿勢為側面的臉部影像，為了使所有人物的演出片段都被記錄在索引結果中，我們在演員串列分群結束之後，檢查演員串列前後是否有偵測的人臉，並將遺漏的人臉擴充至串列中。然而，為了避免不同演員或非人臉物件被併入串列中，我們對擴張的範圍與對象有嚴格的限制，僅將串列前後兩側屬於相同場景之畫面作為可能擴張之對象，同時，為避免索引標記的重疊，分群至相同群組之演員串列所在的畫面將不在擴張的檢測對象當中。

在進行擴張檢測時，我們針對串列元素出現過的範圍，透過臉部膚色資訊以及身體色彩資訊進行偵測檢查。如上段中提到，為避免錯誤的擴張發生，我們使用更嚴格的膚色臨界值 0.15，利用公式(1)進行膚色像素的判定；身體色彩的部分，我們先對串列中的所有身體影像，進行區間寬度 16 的三維色彩直方圖，統計 $\left(\frac{256}{16}\right)^3 = 4096$ 個顏色子空間的出現次數，並找出統計次數前 5 大的顏色子空間，若受測像素所對應的顏色子空間屬於這 5 個中的任意一個，則視為滿足身體色彩之像素。

利用上段所介紹的顏色檢測方法，我們分別計算臉部範圍內以及身體範圍內滿足顏色要求的像素數量，若兩數量均超過一定比例（實驗中使用 66.7%作為臨界值），則將此畫面擴充至串列中，並繼續往下檢查；相反，則不擴充，並停止此方向的擴張檢測。以下我們用條列方式介紹臉部及身體檢查區塊的設定方式、以及像素比例的計算方法：

- 臉部範圍：串列中所有臉部影像所出現過的區域、
- 身體範圍：串列中所有身體影像所出現過的區域；
- 臉部像素比例：”符合膚色檢測之像素總數”除以”所有臉部影像之平均面積”、
- 身體像素比例：”符合身體色彩之像素總數”除以”所有身體影像之平均面積”。

圖 3-11 為實際檢測範圍及檢測結果之實例，(b)中的兩綠色區塊分別代表臉部範圍以及身體範為，(c)中的非綠色像素即為滿足色彩要求之像素；圖 3-12 為擴張結果的範例圖，包含向前擴張所得的 20 張畫面以及向後擴張所得的 1 張畫面。



(a) (b) (c)

圖 3-11：膚色檢測及身體色彩檢測示意圖

(a) 受測畫面原影像，箭頭所指演員為受測演員

(b) 臉部範圍及身體範圍示意圖

(c) 非綠色像素即為滿足色彩要求之像素

(a)

(b)

(c)

圖 3-12：演員串列擴張範例

(a)(b)(c)分別為：向前擴充、串列元素、向後擴充

第四章 實驗結果與討論

4.1 分群結果評估工具

4.1.1 Adjusted RAND index（ARI）

本論文中我們使用 Hubert 等人在[22]提出的 Adjusted RAND index 公式來進行分群結果之評估。首先介紹分群作業所可能產生的 4 種結果，假設共有 n 個物件，正確切割為集合 U、分群結果為集合 V，則 4 種結果之定義及求算方式如下：

- a = 在 U 中的相同群組內且在 V 中的相同群組內之物件對(pair)數 = $\sum_{i,j}\binom{n_{ij}}{2}$、
- b = 在 U 中的相同群組內而在 V 中的不同群組內之物件對（pair）數 = $\sum_{i}\binom{n_{i.}}{2}-\sum_{i,j}\binom{n_{ij}}{2}$、
- c = 在 V 中的相同群組內而在 U 中的不同群組內之物件對（pair）數 = $\sum_{j}\binom{n_{.j}}{2}-\sum_{i,j}\binom{n_{ij}}{2}$、
- d = 在 U 中的不同群組內且在 V 中的不同群組內之物件對（pair）數，即 a、b、c 以外的情形，因此數值 $d=\binom{n}{2}-a-b-c$，

其中 n_{ij} 為"屬於群組 u_i 中且分群至群組 v_j"之元素個數(如表 4-1 之對角線)、$n_{i.}$ 為"屬於群組 u_i"之元素個數(如表 4-1 之右側欄位)、$n_{.j}$ 為"分群至群組 v_j"之元素個數(如表 4-1 之下方欄位)。則 ARI 值的計算公式可表示為：

$$ARI=\frac{a-(b*c)\Big/\binom{n}{2}}{(b+c)/2-(b*c)\Big/\binom{n}{2}}\quad,\qquad(17)$$

結果值將落於 $[0,1]$ 區間內，數值越接近 1 則表示分群結果越接近正確切割。

以實例來進行說明，假設某一集合含 10 個物件，其正確分割為 $U=\{1,1,2,2,2,3,3,3,3,3\}$、分群結果得到的分割為 $V=\{1,2,1,2,2,2,3,3,3,3\}$，如表 4-1 所示，則

$$a=\sum_{i,j}\binom{n_{ij}}{2}=\binom{2}{2}+\binom{4}{2}=7\ 、$$

$$b=\sum_{i}\binom{n_{i.}}{2}-\sum_{i,j}\binom{n_{ij}}{2}=\binom{2}{2}+\binom{4}{2}+\binom{4}{2}-7=6\ 、$$

$$c=\sum_{j}\binom{n_{.j}}{2}-\sum_{i,j}\binom{n_{ij}}{2}=\binom{2}{2}+\binom{3}{2}+\binom{5}{2}-7=7\ 、$$

又 $a+b+c+d=\binom{n}{2}$，故 $d=\binom{10}{2}-7-6-7=25$，

因此，利用公式(17)，可得 ARI 數值為 $\frac{7-(13*14)/45}{(13+14)/2-(13*14)/45}=0.3126$。

表 4-1：Adjusted RAND index 範例表格

Class \ Cluster	v_1	v_2	v_3	Sums
u_1	1	1	0	2
u_2	1	2	1	4
u_3	0	0	4	4
Sums	2	3	5	$n=10$

4.1.2 Classification via Clustering（CVC）

進行分群結果分析時，除了 Adjusted RAND index 外，各群組內的純淨度（purity）也是我們常用的。其定義為"所有群組中最大種類元素之個數總和"所佔元素總數比例，假設群組內含有 5 個◯、1 個□、及 2 個╳，則此群組的對大種類為◯。與 ARI 值相同，CVC 值同樣落於 $[0,1]$ 區間內，而數值越大代表整體群組純淨度越高。

我們以圖 4-1 之範例進行說明，計算各群組中最大種類元素之個數（cluster 1：5 個╳、cluster 2：4 個◯、cluster 3：3 個□），加總後除以資料總數即可得純淨值 CVC 為 $(5+4+3)/17=0.71$。

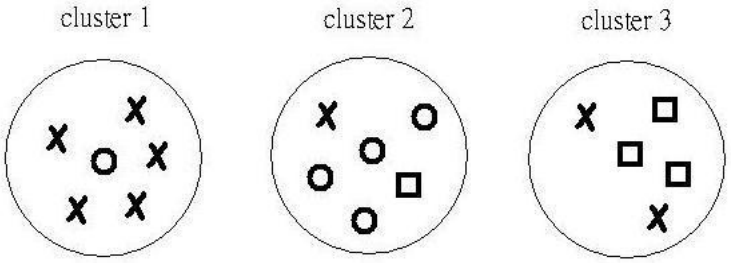


圖 4-1：CVC 範例圖

為了使評估數據更加接近原始資料，在進行 ARI 及 CVC 之計算時，我們是以串列中臉部影像之個數進行元素個數的計算，而非使用人工定義的演員串列。假設某一演員串列含有 15 個人臉影像，則計算 ARI 與 CVC 時，使用的元素個數為 15。

另一方面，除了演員個數外，實驗中也進行了較大群組數的分群，希望透過多個群組數量的分群結果，幫助我們得到更正確的分析。

4.2 測試資料介紹

本篇論文之實驗分為兩大部分，第一部份僅使用演員之臉部影像、第二部份則使用臉部影像以及身體影像，本節中我們將介紹實驗所使用影像資料之來源與特性。

4.2.1 僅含臉部影像之測試資料

在第一部份的實驗中，我們將使用兩組僅含臉部資訊之測試資料進行實驗，測資 1 為本實驗室自行拍攝之影片，其中包含 2765 張影像、7 位演員組成的 129 個演員串列（110 較佳串列 + 19 較差串列）；測資 2 為美國影集“Friends（六人行）”之影片，包含 6520 張影像與 7 位演員組成的 489 個演員串列（299 較佳串列 + 190 較差串列）。

由於測資 1 為自製影片、測資 2 為有劇情變化之影片，因此單一人物的變化度，測資 1 中的人物變化度將較測資 2 中來得小，此外測資 1 的光影條件、演員動作、拍攝角度等也都較測資 2 來得固定和簡單。圖 4-2 為測試資料 1 與測資 2 之部份串列範例。

(a) 測資 1　　(b) 測資 2

圖 4-2：測試資料 1~2 之部分演員串列

4.2.2 含臉部影像與身體影像之測試資料

第二部份之實驗為身體資訊相關應用之實驗，為了避免片頭曲中非本集內容之衣物或場景對實驗產生影響，我們將影片中片頭曲片段移除，移除後得到的三組測試資料分別如下：測資 3（與測資 2 為相同影片來源）、測資 4 均為美國影集“Friends”之影片，分別包含 6294 張影像與 7 位演員組成的 472 個演員串列（288 較佳串列 + 184 較差串列）；、以及 5275 張影像與 6 位演員組成的 318 個演員串列（235 較佳串列 + 83 較差串列）；測資 5 為美國影集“Everybody Loves Raymond”，包含 6270 張影像、8 位演員組成的 440 個演員串列（415 較佳串列 + 25 較差串列）。

我們針對影集“Friends”及“Everybody Loves Raymond”之演員臉部特性進行比較，影集“Friends”為六位年輕男女為主角之影片，而“Everybody Loves Raymond”則有老年人、年輕人、小孩子各種年齡層之演員。

透過演員串列內身體影像的觀察，各測試資料中身體資訊之特性分別如下：測資 3 中，演員僅有進行至多 2 次的服裝更換，即身體資訊較單純；反觀測資 4 以及測資 5，

演員進行了多次的服裝更換，甚至有脫掉外套再穿上的情況，因此兩測資皆擁有較複雜的身體資訊。圖 4-3 為測試資料 3~5 之部份串列範例。

(a) 測資 3　　(b) 測資 4　　(c) 測資 5

圖 4-3：測試資料 3~5 之部分演員串列（臉部+身體）

4.3 不同 Eigenfaces 基底之比較

我們準備兩組投影基底（projection basis），兩者皆將人臉影響投影至維度 70 之座標，其中投影基底 1（basis_1）僅用 15 張人臉影像訓練求得，訓練集如圖 4-4；投影基底 2（basis_2）使用了 538 張影像，如圖 4-5。以下我們藉由真實資料的測試，觀察不同投影基底訓練集對影像投影之影響。

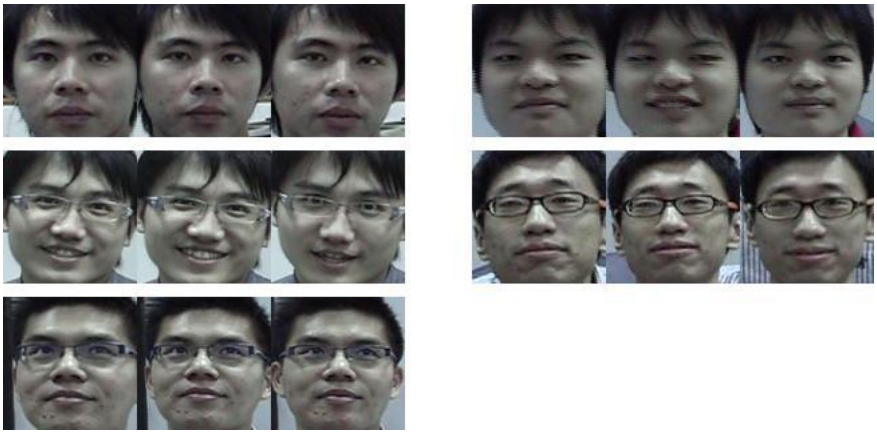


圖 4-4：15 張人臉影像之訓練集

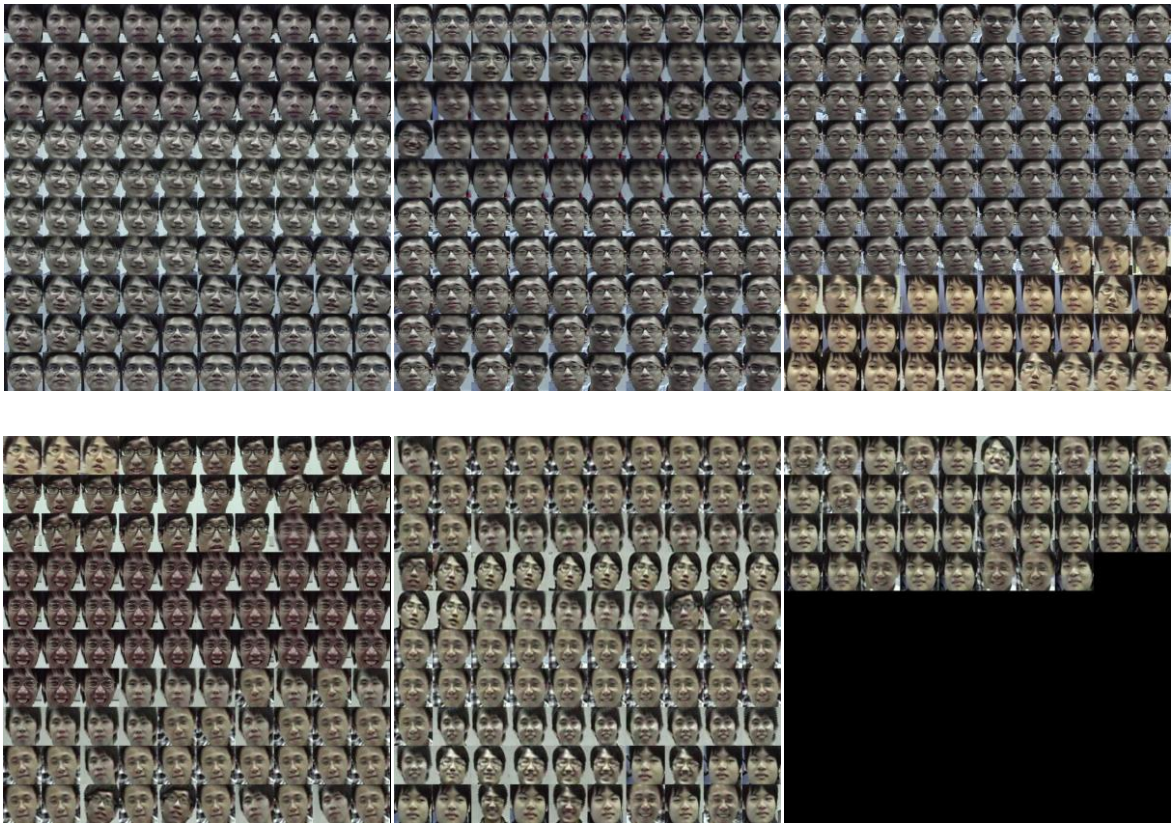


圖 4-5：538 張人臉影像之訓練集

表 4-2：測資 1 在不同 Eigenfaces 基底下的 ARI 值

[實驗採取：兩階段分群方法，搭配最小距離分類]

basis / C = 10	basis_1	basis_2	basis / C = 15	basis_1	basis_2
Average-link	**0.315**	**0.647**	Average-link	**0.363**	**0.502**
Complete-link	**0.296**	**0.391**	Complete-link	**0.317**	**0.393**
Ward	**0.316**	**0.525**	Ward	**0.328**	**0.402**

表 4-2 中為測資 1 在不同 Eigenfaces 基底下分群結果之 ARI 值，我們可明顯觀察出使用基底 2 所得的分群結果比使用基底 1 來得好。透過 PCA 過程中 covariance matrix 所求得特徵值之觀察，基底 1 僅有 14 個整數部分大於零的特徵值，而基底 2 則有 536 個，故基底 1 所得的投影係數僅有前面十幾項具有描述功用，其餘將只是細部微調；相較於基底 1 所得係數，利用基底 2 所得的 70 個投影係都是具有描述力的，因此串列間相異度的求算上基底 2 較基底 1 更具分辨力，故可得到較佳的分群結果。在往後的實驗中，將全部使用投影基底 2。

4.4 不同投影維度之比較

利用訓練集 2 內的 538 張人臉影像作為投影基底訓練集，我們求算出四組分別將人臉影像投影至維度 70、100、130 與 160 的投影基底，利用測資 1 搭配階層式分群演算法（使用 Average-link 合併公式）及匈牙利演算法，並記錄不同群組個數下的 ARI 與 CVC 於表 4-3、表 4-4 中。

表 4-3：測資 1 使用階層式分群演算法（Average-link 合併公式）
[實驗採取：兩階段分群方法，搭配最小距離分類]

Hierarchical ~ Average-link	C = 7		C = 10		C = 15	
	ARI	CVC	**ARI**	CVC	**ARI**	CVC
Dimension = 70	**0.607**	0.806	**0.633**	0.837	**0.487**	0.891
Dimension = 100	**0.607**	0.806	**0.633**	0.837	**0.528**	0.899
Dimension = 130	**0.624**	0.814	**0.653**	0.844	**0.507**	0.907
Dimension = 160	**0.624**	0.814	**0.594**	0.845	**0.509**	0.915

表 4-4：測資 1 使用匈牙利演算法
[實驗採取：兩階段分群方法，搭配最小距離分類]

Hungarian	C = 7		C = 10		C = 15	
	ARI	CVC	**ARI**	CVC	**ARI**	CVC
Dimension = 70	**0.513**	0.760	**0.454**	0.814	**0.439**	0.930
Dimension = 100	**0.508**	0.729	**0.552**	0.860	**0.478**	0.915
Dimension = 130	**0.466**	0.721	**0.386**	0.775	**0.437**	0.907
Dimension = 160	**0.524**	0.736	**0.567**	0.868	**0.490**	0.922

觀察兩表格，我們說四組基底並無明顯的優劣之分。同樣透過 PCA 過程中之特徵值，計算各投影基底中所用特徵值之總和，再除上所有特徵值總和，此數值表示各基底內之向量所佔整體變異量比例。維度 70 至 160 之基底所占比例分別為 0.955、0.974、0.984、0.990，由於此四組基底之特徵值比例差距有限，使得此四組基底所得之分群結果並無明顯的優劣之分。在後續的實驗中，我們固定使用維度 70 之投影基底。

4.5 人臉影像前處理之比較

在第 3.3 節中我們介紹了兩種臉部影像的前處理方法，“Luminance Normalization + Gaussian Low-pass”以及“Laplacian of Gaussian（LoG）”，本節我們利用測資 1 以及階層式分群演算法進行兩方法之比較。觀察表 4-5 中分群結果之 ARI 值，我們發現使用“Luminance Normalization + Gaussian Low-pass”的分群結果明顯較“LoG”好，兩方法的特性說明如下，Low-pass 可忽略影像間的細微差異，使用明顯的差異來描述影像之相似度，因此可得到較佳的分群結果。反觀 LoG 是將影像進行特徵之強調，影像間的些微差異會變得明顯，使得相同演員之串列間的相似度因此降低，造成分群不佳之結果。故 Luminance Normalization + Gaussian Low-pass 為較佳的臉部影像前處理方式，我們在後續實驗將全部採用此前處理方法。

表 4-5：測資 1 進行前處理比較之 ARI 值
[實驗採取：兩階段分群方法，搭配最小距離分類]

Preprocessing / C = 10	Luminance Normalization + Gaussian Low-pass	Laplacian of Gaussian (LoG)
Average-link	**0.647**	**0.35**
Complete-link	**0.403**	**0.3**
Ward	**0.477**	**0.325**
C = 15		
Average-link	**0.502**	**0.445**
Complete-link	**0.341**	**0.361**
Ward	**0.402**	**0.385**

4.6 分群法、分類法、碰撞資訊、動態分類、Prototype 等討論

本節中我們將依序進行“分群法比較”、“碰撞資訊的使用”、“分類法比較”、“動態分類的使用”、以及“使用 Prototype 與否”五個主題之討論，希望透過各種方法的比較與討論，找出最佳的分群方法和分類方法。

4.6.1 分群法比較

我們利用測資 1 進行各種分群方法之比較，如表 4-6 結果所示，“階層式分群演算法配合 Average-link 合併公式”所得的分群結果為最佳。

表 4-6：測資 1 進行分群法之討論

[實驗採取：兩階段分群方法，搭配最小距離分類]

群組個數 / 分群法		C = 7		C = 10		C = 15		C = 20	
		ARI	CVC	**ARI**	CVC	**ARI**	CVC	**ARI**	CVC
階層式分群演算法	Single-link	**0.130**	0.442	**0.108**	0.496	**0.261**	0.690	**0.304**	0.744
	Complete-link	**0.401**	0.651	**0.403**	0.729	**0.334**	0.783	**0.306**	0.853
	Average-link	**0.607**	0.806	**0.633**	0.837	**0.487**	0.891	**0.504**	0.985
	Ward	**0.346**	0.636	**0.342**	0.729	**0.350**	0.837	**0.323**	0.868
Hungarian method		**0.513**	0.760	**0.454**	0.814	**0.439**	0.930	**0.339**	0.930
METIS		**0.293**	0.589	**0.236**	0.643	**0.195**	0.682	**0.112**	0.659

4.6.2 碰撞資訊的使用

由 4.6.1 節中的實驗得知，階層式分群演算法配合 Average-link 合併公式為最佳的分群方法，我們再加上串列間的時間資訊（碰撞資訊），使互相碰撞之串列無法合併，利用測資 1 以及測資 2 來進行測試，所得結果分別為表 4-7、表 4-8。

表 4-7：測資 1 進行碰撞資訊使用之討論
[實驗採取：兩階段分群方法，搭配最小距離分類]

測資 1	Average-link		Average-link with Collision	
	ARI	CVC	**ARI**	CVC
C = 7	**0.607**	0.806	**0.648**	0.829
C = 10	**0.633**	0.837	**0.633**	0.837
C = 15	**0.487**	0.891	**0.487**	0.891
C = 20	**0.504**	0.985	**0.504**	0.985

表 4-8：測資 2 進行碰撞資訊使用之討論
[實驗採取：兩階段分群方法，搭配最小距離分類]

測資 2	Average-link		Average-link with Collision	
	ARI	CVC	**ARI**	CVC
C = 7	**0.046**	0.302	**0.134**	0.421
C = 10	**0.068**	0.334	**0.124**	0.434
C = 15	**0.114**	0.409	**0.123**	0.447
C = 20	**0.114**	0.449	**0.171**	0.551

透過以上兩組數據觀察，加上碰撞資訊後的分群結果將有明顯的提升，原因在於碰撞資訊被採用前，階層式分群法過程中可能會有時間重疊之演員串列遭合併，若採用碰撞資訊，則可避免這兩個不同演員之串列被併入同一群組中，因此，碰撞資訊的應用可提升分群準確度。綜合以上兩個實驗，我們說“Average-link with Collision”為最佳的分群方法。

4.6.3 分類法比較

此段落中，我們利用測資 1 與測資 2 探討最小距離分類、K-NN、以及 Modified K-NN 何者為兩階段分群方法中最佳的分類方法。由於 K=1 時三者皆為最小距離分類，因此表 4-9 中僅記錄 K 值大於 1 之比較數據。由結果數據的觀察我們發現 Modified K-NN 將帶來比其他兩者較好的分類結果。分析三種分類方法，由於 Modified K-NN 具有整體性比較之特性，因此不易受到 local minimum 影響，故優於其它兩方法。

表 4-9：測資 1 與測資 2 進行分類方法討論

Stage 1		**Stage 2**					
（測資 1）Average-link with C=7 → K = 3							
ARI	CVC	最小距離分類		K-NN		Modified K-NN	
0.627	0.8	**ARI**	CVC	**ARI**	CVC	**ARI**	CVC
		0.607	0.806	**0.613**	0.809	**0.624**	0.814
（測資 1）Hungarian method with C=15 → K =3							
0.456	0.945						
		0.439	0.93	**0.441**	0.92	**0.463**	0.946
（測資 1）“Average-link with Collision” with C=7 → K = 3							
0.667	0.827						
		0.648	0.829	**0.652**	0.833	**0.666**	0.837
（測資 1）“Hungarian method with Collision” with C=7 → K = 3							
0.516	0.727						
		0.528	0.736	**0.535**	0.739	**0.543**	0.744
（測資 2）“Hungarian method with Collision” with C=7 → K = 3							
0.133	0.469						
		0.096	0.408	**0.113**	0.429	**0.113**	0.431

4.6.4 動態分類的使用

接下來，我們探討動態分類對於分類結果之影響，表 4-10 記錄測資在未採取動態分類、以及採用動態分類之結果，由實驗結果得知，動態分類將有助於分類正確率之提升。因此綜合以上兩個實驗，我們說"動態 Modified K-NN"為最佳的分類方法。

表 4-10：測資 2 進行動態分類應用之比較

Average-link with Collision	**Stage 1**		**Stage 2**（Modified K-NN）	
			Modified K-NN	
	ARI	CVC	**ARI**	CVC
C = 7	**0.108**	0.39	**0.087**	0.355
			↓↓↓ Dynamic ↓↓↓	
			0.088	0.361
C = 10	**0.131**	0.452	**0.106**	0.405
			↓↓↓ Dynamic ↓↓↓	
			0.107	0.416
C = 20	**0.136**	0.528	**0.114**	0.482
			↓↓↓ Dynamic ↓↓↓	
			0.116	0.49
C = 30	**0.151**	0.587	**0.126**	0.541
			↓↓↓ Dynamic ↓↓↓	
			0.133	0.556

4.6.5 Prototype 的使用

最後，探討 Prototype 之使用對於整體分群之影響，除了兩階段分群方法搭配動態 Modified K-NN 分類，我們也進行另一種分群策略─所有串列一併進行分群，希望透過更多策略的進行幫助我們觀察 Prototype 之特性。表 4-11 中記錄測資 2 進行 Prototype 探討之結果。由實驗結果得知，Prototype 之使用並無法達到第 3.4.5 節中所說提升分群正確率的假設，因此在之後的實驗中，我們將不進行 Prototype 之測試。

表 4-11：測資 2 進行 Prototype 應用之比較

Average-link with Collision	兩階段分群方法		所有串列一併分群	
	ARI	CVC	**ARI**	CVC
C = 7	**0.088**	0.361	**0.1**	0.397
	↓↓↓ Prototype ↓↓↓			
	0.097	0.38	**0.073**	0.361
C = 10	**0.107**	0.416	**0.105**	0.448
	↓↓↓ Prototype ↓↓↓			
	0.103	0.423	**0.12**	0.452
C = 20	**0.116**	0.49	**0.132**	0.533
	↓↓↓ Prototype ↓↓↓			
	0.118	0.501	**0.109**	0.463
C = 30	**0.133**	0.556	**0.138**	0.56
	↓↓↓ Prototype ↓↓↓			
	0.12	0.558	**0.118**	0.507

截至目前為止，我們已經探討了投影基底及維度；人臉的前處理方式；同時也得知“Average-link with Collision”為最佳的分群方法，另一方面，在兩階段分群方法中，“動態 Modified K-NN”則是最佳的串列分類方法。而針對兩分群策略的優劣比較上，兩策略間沒有明顯的優劣差距，因此在下一個討論議題實驗中，將同時採用兩分群策略之結果。

4.7 合併臉部資訊及身體資訊

我們在第 3.4.3 節提到臉部資訊與身體資訊的合併有依時間差距決定權重與固定權重比例($\sigma=\infty$)兩種方式。為比較各種策略在不同測資中的特性，我們利用“Average-link with Collision”分群法和“動態 Modified K-NN”分類法，分別對 h=0、0.2、0.4、0.6、0.8、1.0 與 σ=1500、3000、6500、∞共 24 種組合，皆進行“所有串列一併進行分群”及“兩階段分群方法”。其中 h=0 表示僅使用臉部資訊、$\sigma=\infty$ 代表權重為固定比例、h=1.0 搭配 $\sigma=\infty$ 為完全使用身體資訊之測試。以下我們將分別對 h 與 σ 兩參數之變化來帶來之影響進行討論。

4.7.1 身體資訊比重之探討 －h 參數

首先，探討不同的身體資訊比重對整體分群結果之影響，為了方便觀察，針對各 h，我們將四個不同 σ 所得數據進行平均值計算，作為各 h 之結果。以下依序為測試資料 3~5 的實驗結果，希望藉由多部影片的觀察，觀察不同的 h 對分群工作之影響。

表 4-12：測資 3 進行 *h* 變化分析

策略	群組數 \ *h*		0.0	0.2	0.4	0.6	0.8	1.0
兩段式分群方法	C = 7	ARI	**0.237**	**0.311**	**0.338**	**0.37**	**0.295**	**0.276**
		CVC	0.619	0.599	0.616	0.648	0.585	0.574
	C = 10	ARI	**0.221**	**0.315**	**0.364**	**0.389**	**0.311**	**0.305**
		CVC	0.493	0.637	0.693	0.713	0.658	0.655
	C = 20	ARI	**0.219**	**0.335**	**0.376**	**0.363**	**0.34**	**0.34**
		CVC	0.524	0.765	0.813	0.8	0.77	0.775
	C = 30	ARI	**0.24**	**0.305**	**0.326**	**0.33**	**0.312**	**0.306**
		CVC	0.6	0.824	0.852	0.845	0.839	0.825
所有串列一併分群	C = 7	ARI	**0.197**	**0.44**	**0.378**	**0.313**	**0.32**	**0.299**
		CVC	0.507	0.663	0.644	0.603	0.612	0.573
	C = 10	ARI	**0.204**	**0.434**	**0.385**	**0.363**	**0.36**	**0.352**
		CVC	0.527	0.696	0.665	0.677	0.684	0.671
	C = 20	ARI	**0.214**	**0.416**	**0.41**	**0.396**	**0.396**	**0.381**
		CVC	0.542	0.797	0.818	0.794	0.796	0.774
	C = 30	ARI	**0.252**	**0.399**	**0.368**	**0.354**	**0.359**	**0.352**
		CVC	0.633	0.853	0.839	0.84	0.842	0.826

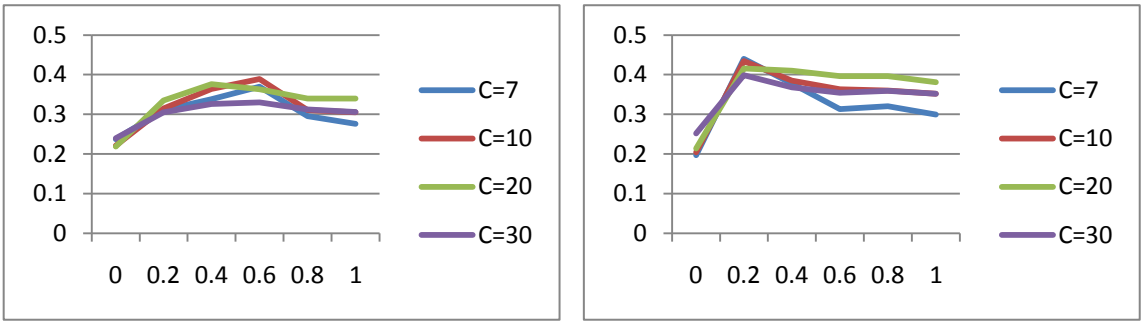


圖 4-6：表 4-12 之 ARI 變化曲線圖（左為兩階段分群方法、右為所有串列一併分群）

藉由圖 4-6 的輔助觀察，我們說加上身體資訊對於分群準確性有顯著的提升，其中在 h=0.2~0.6 時分群效果為最佳；另一方面，當 h 持續加大至 0.8、1.0 時，並無法產生更準確的分群結果，甚至使得結果更差。藉由測資 3 的資料特性進行分析，由於所有演員都只更換過一次服裝，這使得串列之間的身體資訊較不複雜，因此加上身體資訊的使用可提供更好的辨別作用，進而提升分群的準確率。

表 4-13：測資 4 進行 h 變化分析

策略	群組數 \ h		0.0	0.2	0.4	0.6	0.8	1.0
兩段式分群方法	C = 6	**ARI**	**0.116**	**0.148**	**0.136**	**0.159**	**0.153**	**0.173**
		CVC	0.415	0.465	0.449	0.458	0.457	0.471
	C = 10	**ARI**	**0.134**	**0.158**	**0.15**	**0.177**	**0.18**	**0.191**
		CVC	0.445	0.505	0.485	0.513	0.527	0.531
	C = 20	**ARI**	**0.15**	**0.274**	**0.247**	**0.243**	**0.251**	**0.251**
		CVC	0.483	0.733	0.7	0.701	0.703	0.708
	C = 30	**ARI**	**0.162**	**0.258**	**0.246**	**0.257**	**0.254**	**0.24**
		CVC	0.531	0.803	0.777	0.781	0.777	0.768
所有串列一併分群	C = 6	**ARI**	**0.133**	**0.176**	**0.15**	**0.155**	**0.156**	**0.133**
		CVC	0.432	0.49	0.462	0.468	0.493	0.449
	C = 10	**ARI**	**0.13**	**0.209**	**0.183**	**0.173**	**0.17**	**0.145**
		CVC	0.448	0.549	0.535	0.526	0.55	0.482
	C = 20	**ARI**	**0.135**	**0.274**	**0.238**	**0.241**	**0.237**	**0.235**
		CVC	0.503	0.738	0.702	0.706	0.691	0.674
	C = 30	**ARI**	**0.115**	**0.269**	**0.234**	**0.228**	**0.229**	**0.222**
		CVC	0.546	0.832	0.803	0.791	0.762	0.752

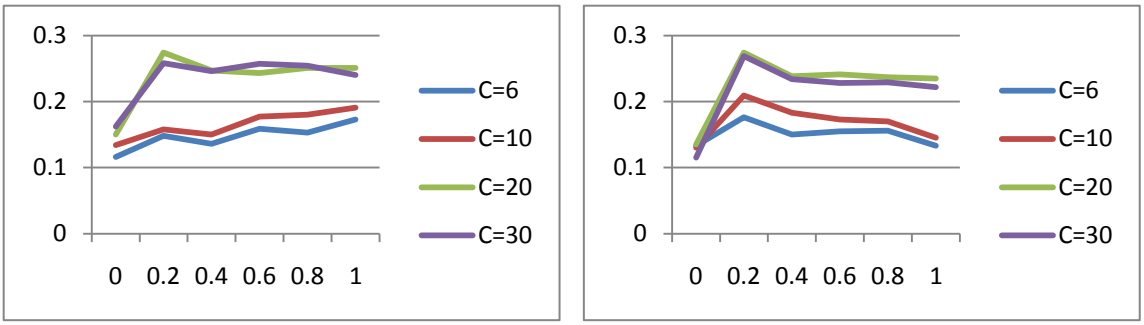


圖 4-7：表 4-13 之 ARI 變化曲線圖（左為兩階段分群方法、右為所有串列一併分群）

觀察圖 4-7 的 ARI 變化曲線，我們發現其變化情形與圖 4-6 相似，都在加入身體資訊後對分群結果有明顯提升，其中 h=0.2 時有最佳的分群結果；另一方面，持續加重 h 同樣無法產生更好的結果。此外我們也發現在測資 3 中使用身體資訊可提升大約 0.2 的準確率，而測資 4 中僅提升約 0.1 的準確率，這是由於測資 4 中演員更換衣服的次數較測資 3 來得頻繁（僅一次），使得串列之間的身體資訊較為複雜，因此加上身體資訊後分群準確度雖有提升，但幅度卻不如測資 3 明顯。

表 4-14：測資 5 進行 *h* 變化分析

策略	群組數 \ *h*		0.0	0.2	0.4	0.6	0.8	1.0
兩段式分群方法	C = 8	ARI	**0.55**	**0.458**	**0.246**	**0.193**	**0.21**	**0.241**
		CVC	0.748	0.652	0.561	0.508	0.513	0.547
	C = 15	ARI	**0.582**	**0.38**	**0.237**	**0.179**	**0.206**	**0.262**
		CVC	0.809	0.707	0.624	0.564	0.59	0.645
	C = 20	ARI	**0.571**	**0.384**	**0.242**	**0.193**	**0.254**	**0.289**
		CVC	0.809	0.73	0.65	0.606	0.667	0.696
	C = 30	ARI	**0.541**	**0.391**	**0.249**	**0.197**	**0.225**	**0.265**
		CVC	0.825	0.782	0.69	0.645	0.69	0.721
所有串列一併分群	C = 8	ARI	**0.592**	**0.428**	**0.256**	**0.294**	**0.286**	**0.312**
		CVC	0.76	0.664	0.57	0.56	0.581	0.602
	C = 15	ARI	**0.596**	**0.37**	**0.251**	**0.277**	**0.284**	**0.253**
		CVC	0.793	0.738	0.645	0.627	0.626	0.63
	C = 20	ARI	**0.546**	**0.377**	**0.259**	**0.284**	**0.299**	**0.274**
		CVC	0.823	0.761	0.668	0.658	0.673	0.668
	C = 30	ARI	**0.532**	**0.393**	**0.263**	**0.292**	**0.265**	**0.257**
		CVC	0.838	0.8	0.706	0.707	0.699	0.701

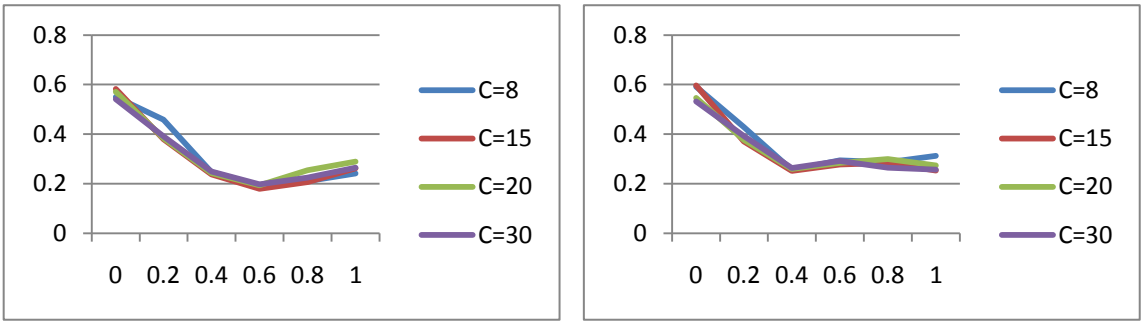


圖 4-8：表 4-14 之 ARI 變化曲線圖（左為兩階段分群方法、右為所有串列一併分群）

藉由表 4-14 與圖 4-8 的觀察，我們發現加上身體資訊反而使得分群準確性降低。我們已知測資 5 為多個不同年齡層演員共同演出的，這使得演員串列間的臉部差異性較為明顯，即較容易進行辨識，因此僅使用臉部資訊進行分群即可獲得很棒的結果。

根據$h=0$之結果（即僅使用臉部資訊進行分群），我們觀察到影集“Everybody Loves Raymond”之臉部分辨度明顯較“Friends”佳，造成此項差異的原因在於影集內的演員差異性，由於“Everybody Loves Raymond”之主要演員有老年人、年輕人、小孩子各種年齡層，臉部影像的差異性較大，而“Friends”主要演員皆為年輕男女，臉部影像的差異性較小，因此在人臉辨識的進行上，具有較大差異性的臉部影像獲得較佳的辨識結果是必然的。

藉由身體資訊的使用，測資 3 與測資 4 皆獲得比僅使用臉部資訊更佳的分群結果；另一方面不論是身體資訊簡單的測資 3 或者是身體資訊複雜的測資 4，若將身體資訊權重持續增加，並不會產生更好的分群結果，甚至有不升反降的情形，因此使用少許的身體資訊（Ex：$h=0.2\sim0.4$）將可獲得最佳的分群結果。

4.7.2 身體資訊比重之探討 －σ參數

如同第 3.4.3 節所述，演員服裝會隨著時間的進行而有更換，因此我們提出公式(7)來決定身體資訊權重，其中 σ 之目的在於決定權重的遞減幅度，σ 值越大則表示遞減速度越慢，反之則越快，以抽樣比例 5 FPS 計算，$\sigma=1500$為 5 分鐘、$\sigma=3000$為 15 分鐘、$\sigma=6500$約為 21.5 分鐘。與第 4.7.1 節中相同，我們同時觀察兩種分群策略在不同 σ 情形下的結果，以下我們僅針對測資 3 及測資 4 在$h=0.2$和$h=0.4$時之分群結果，希望藉此觀察不同 σ 所帶來之影響。

表 4-15：測資 3 進行 σ 變化分析

$h=0.2$			$\sigma=1500$	$\sigma=3000$	$\sigma=6500$	$\sigma=\infty$
兩段式分群方法	C = 7	ARI	**0.378**	**0.237**	**0.324**	**0.303**
		CVC	0.671	0.515	0.602	0.609
	C = 10	ARI	**0.344**	**0.279**	**0.319**	**0.316**
		CVC	0.666	0.594	0.642	0.646
	C = 20	ARI	**0.38**	**0.327**	**0.306**	**0.328**
		CVC	0.763	0.75	0.758	0.79
	C = 30	ARI	**0.294**	**0.327**	**0.295**	**0.305**
		CVC	0.811	0.852	0.802	0.83
所有串列一併分群	C = 7	ARI	**0.478**	**0.467**	**0.411**	**0.406**
		CVC	0.677	0.655	0.675	0.645
	C = 10	ARI	**0.481**	**0.39**	**0.439**	**0.427**
		CVC	0.698	0.669	0.724	0.692
	C = 20	ARI	**0.4**	**0.406**	**0.442**	**0.417**
		CVC	0.788	0.762	0.838	0.798
	C = 30	ARI	**0.342**	**0.394**	**0.444**	**0.418**
		CVC	0.815	0.834	0.9	0.861
$h=0.4$			$\sigma=1500$	$\sigma=3000$	$\sigma=6500$	$\sigma=\infty$
兩段式分群方法	C = 7	ARI	**0.387**	**0.279**	**0.363**	**0.324**
		CVC	0.653	0.584	0.637	0.589
	C = 10	ARI	**0.4**	**0.315**	**0.368**	**0.371**
		CVC	0.703	0.686	0.681	0.702
	C = 20	ARI	**0.37**	**0.374**	**0.373**	**0.388**
		CVC	0.824	0.784	0.809	0.834
	C = 30	ARI	**0.307**	**0.317**	**0.328**	**0.353**
		CVC	0.836	0.839	0.875	0.858
所有串列一併分群	C = 7	ARI	**0.36**	**0.423**	**0.359**	**0.371**
		CVC	0.615	0.678	0.641	0.641
	C = 10	ARI	**0.359**	**0.418**	**0.377**	**0.385**
		CVC	0.615	0.692	0.676	0.677
	C = 20	ARI	**0.41**	**0.43**	**0.396**	**0.404**
		CVC	0.79	0.808	0.834	0.839
	C = 30	ARI	**0.331**	**0.388**	**0.373**	**0.381**
		CVC	0.822	0.85	0.844	0.839

表 4-16：測資 4 進行 σ 變化分析

$h=0.2$			$\sigma=1500$	$\sigma=3000$	$\sigma=6500$	$\sigma=\infty$
兩段式分群方法	C = 7	ARI	**0.118**	**0.183**	**0.137**	**0.155**
		CVC	0.458	0.502	0.437	0.462
	C = 10	ARI	**0.126**	**0.204**	**0.168**	**0.133**
		CVC	0.501	0.552	0.505	0.462
	C = 20	ARI	**0.262**	**0.304**	**0.28**	**0.249**
		CVC	0.726	0.758	0.728	0.719
	C = 30	ARI	**0.276**	**0.266**	**0.254**	**0.234**
		CVC	0.821	0.823	0.808	0.758
所有串列一併分群	C = 7	ARI	**0.117**	**0.237**	**0.201**	**0.149**
		CVC	0.447	0.52	0.533	0.459
	C = 10	ARI	**0.155**	**0.266**	**0.236**	**0.18**
		CVC	0.545	0.584	0.574	0.493
	C = 20	ARI	**0.286**	**0.308**	**0.275**	**0.225**
		CVC	0.752	0.764	0.741	0.693
	C = 30	ARI	**0.248**	**0.301**	**0.263**	**0.264**
		CVC	0.817	0.874	0.822	0.816
$h=0.4$			$\sigma=1500$	$\sigma=3000$	$\sigma=6500$	$\sigma=\infty$
兩段式分群方法	C = 7	ARI	**0.104**	**0.141**	**0.174**	**0.124**
		CVC	0.41	0.457	0.468	0.46
	C = 10	ARI	**0.137**	**0.16**	**0.176**	**0.125**
		CVC	0.507	0.496	0.473	0.465
	C = 20	ARI	**0.252**	**0.252**	**0.241**	**0.244**
		CVC	0.722	0.707	0.683	0.688
	C = 30	ARI	**0.286**	**0.274**	**0.222**	**0.2**
		CVC	0.819	0.807	0.773	0.708
所有串列一併分群	C = 7	ARI	**0.078**	**0.217**	**0.156**	**0.148**
		CVC	0.419	0.505	0.463	0.462
	C = 10	ARI	**0.103**	**0.236**	**0.199**	**0.193**
		CVC	0.515	0.543	0.553	0.529
	C = 20	ARI	**0.22**	**0.273**	**0.265**	**0.193**
		CVC	0.703	0.728	0.725	0.652
	C = 30	ARI	**0.242**	**0.251**	**0.223**	**0.219**
		CVC	0.831	0.821	0.786	0.774

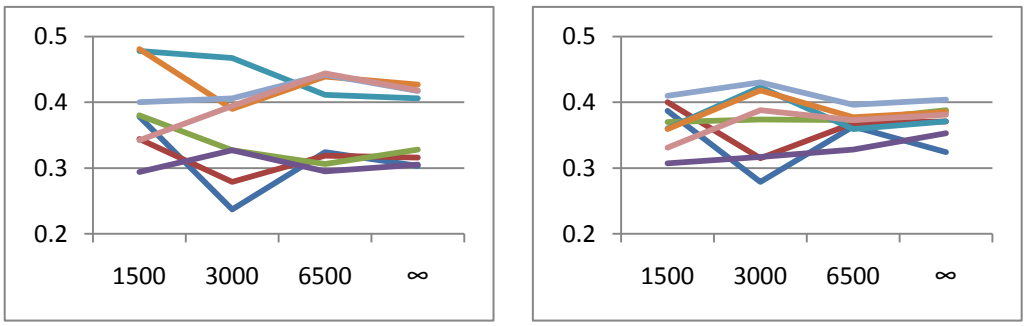


圖 4-9：表 4-15 之 ARI 變化曲線圖（左為 h=0.2、右 h=0.4）

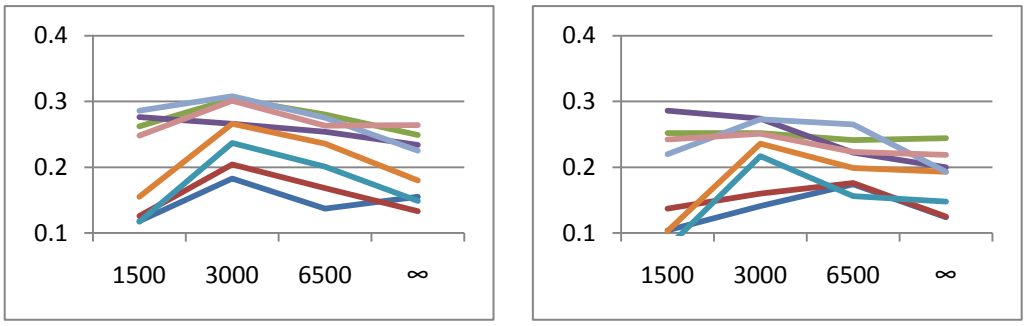


圖 4-10：表 4-16 之 ARI 變化曲線圖（左為 h=0.2、右 h=0.4）

藉由圖 4-9 與圖 4-10 的輔助觀察，“可變權重的資訊合併方式”比“固定權重的資訊合併方式（$\sigma=\infty$）”在身體資訊的使用上有更好的表現，其原因在於演員服裝會隨時間改變所造成，因此透過可變權重的方式，時間差距大的串列之間將得到較低的身體資訊權重，避免不同服裝所可能產生之影響。另一方面，針對參數 σ 的設定方式，不論是僅更換一次服裝的測資 3、或是更換多次服裝的測資 4，採取折衷的權重遞減速率（$\sigma=3000$）都可產生很好的分群結果。

透過以上的觀察，臉部資訊與身體資訊的運用方法有以下幾個要點：對於演員臉部差異明顯的影片，僅透過臉部資訊進行分群即可，由於臉部資訊即有足夠的辨識功能，因此為了避免可能的干擾，不需要加入身體資訊；反之，當影片演員之臉部差異較不顯著時，加入身體資訊可帶給分群工作極大的幫助，其中可變權重的資訊合併方式比固定權重的資訊合併方式來得好。

第五章 結論與未來展望

本篇論文中，我們提出了完整的視訊內人物分群流程，並針對人臉影像的前處理、Eigenfaces 的使用、人物間關係的求算、人物的分群、以及人物擴張等各階段之方法進行討論。另一方面，我們探討各種資訊之使用所帶來的影響，由實驗發現，針對臉部辨識度較差的影片，碰撞資訊以及可變權重式身體資訊的使用可大幅提升人物分群的準確度，以測資 3 為例，完全採用臉部資訊進行人物分群，其準確度僅有 0.2（ARI 值）左右，若加上身體資訊的輔助後，準確率可提升至 0.45 左右。兩對照圖如下頁圖 5-1 與圖 5-2。

在第 4.3 至 4.5 節的實驗中，我們探討了投影基底訓練集、投影維度等 Eigenfaces 相關議題，以及人臉影像的前處理方法。當基底訓練集內的臉部影像含有足夠變異度時，僅須足夠數量的投影維度即可達到準確描述的目的，過大的維度並無法帶來更準確的結果；相反的，若訓練集合內資訊之變異度過小時，則再大的投影維度都是不具描述力的。此外，在人臉影像進行投影前，使用光影平衡及高斯低通將所有影像調整至相同狀態，為較佳的臉部影像前處理方式。

關於分群策略的選擇上，由數據結果我們說所有串列一併進行分群與兩階段分群方法並沒有明顯的優劣之分，在不同影片中，兩策略的結果不盡相同。在分群法與分類法的討論中，階層式分群演算法配合 Average-link 合併方法為最佳的分群法，加上碰撞資訊的運用，可獲得到更好的分群結果；另一方面，動態的 Modified K-NN 為最佳分類方式。最後我們也進行 Prototype 的實驗，由結果得知，Prototype 概念的運用並無法提升演員串列間的描述力。

在資訊的使用選擇上，當處理臉部差異較不明顯的串列分群時（如測資 3、測資 4），透過身體資訊的輔助，將可大大提升分群的準確度，其中也證明“依時間差距決定身體資訊權重”較“固定權重比例”更能有效利用身體資訊，提升整體分群準確度。反之，在處理臉部差異很明顯之串列時（如測資 5），僅使用臉部資訊即可獲得很棒的分群結果，

身體資訊的使用將對分群產生反效果。

論文中使用的階層式分群演算法在進行串列合併時，當錯誤合併發生即無法再被分開，造成分群準確度的低落，未來我們希望尋找更好的分群方法，提升分群作業的準確性。除了分群方法的改進外，希望透過更好的臉部辨識技術來進行臉部影像的描述，例如2-D PCA、FLD、LDA、SVD等；以及，餘弦距離（cosine distance）等不同的距離求算方式，都是可進行的嘗試。

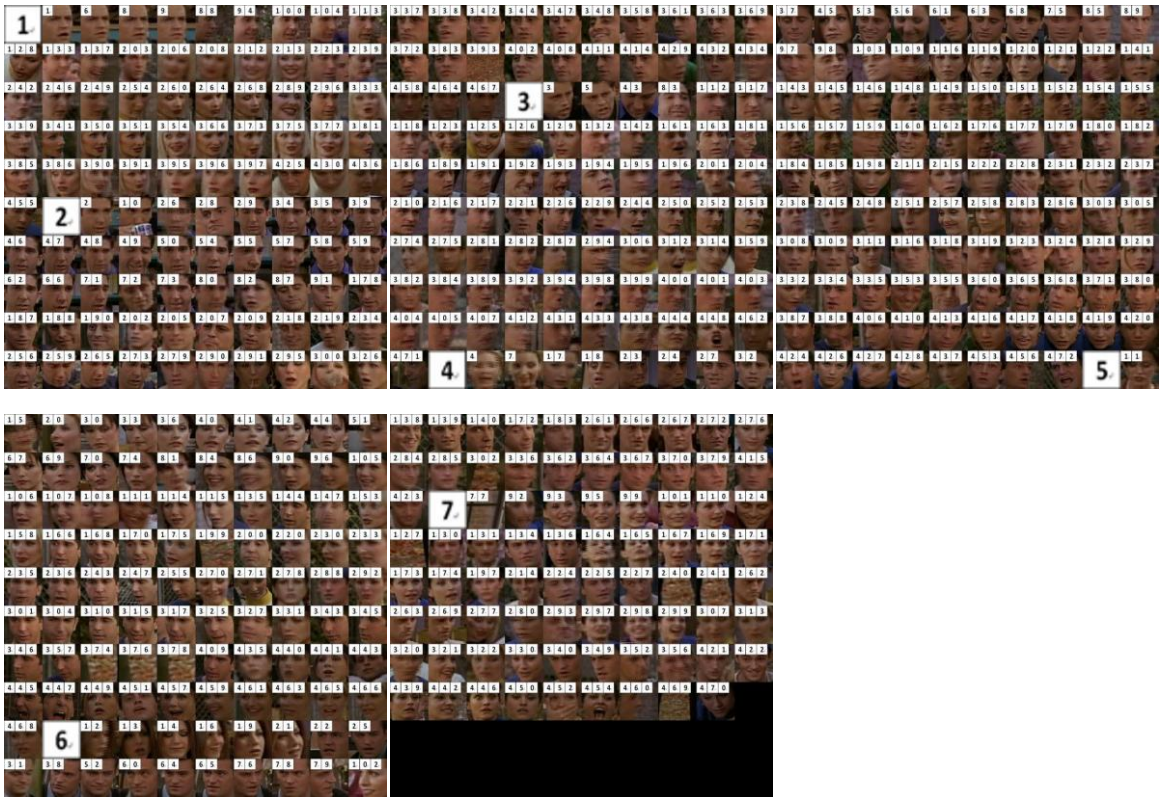


圖 5-1：測資 3 僅使用臉部資訊進行分群之結果圖
[實驗採取：所有串列一併進行分群、Average-link with Collision]
（ARI=0.197、CVC=0.507）

圖 5-2：測資 3 使用臉部資訊及身體資訊進行分群之結果圖（$h=0.2$、$\sigma=3000$）
[實驗採取：所有串列一併進行分群、Average-link with Collision]
（ARI=0.467、CVC=0.655）

參考文獻

[1] D. Zhong, H. J. Zhang, and S. F. Chang, "Clustering Methods for Video Browsing and Annotation," Proc. Storage and Retrieval for Image and Video Databases IV, vol. 2670, pp. 239-246, 1996

[2] Y. P. Tan and H. Lu, "Video Scene Clustering by Graph Partitioning," Electronics Letters, vol. 39, no. 11, pp. 841- 842, 2003

[3] W. Y. Ma and H. J. Zhang, "An Indexing and Browsing System for Home Video," Proc. European Conference on Signal Processing, pp. 131–134, 2000

[4] K. A. Peker, A. Divakaran, and T. Lanning, "Browsing News and Talk Video on a Consumer Electronics Platform Using Face Detection," Proc. SPIE Multimedia Systems and Applications VIII, vol. 6015, 2005

[5] E. Hjelmas and B. K. Low, "Face Detection: A Survey," Computer Vision and Image Understanding, vol. 83, pp. 236–274, 2001

[6] C. Czirjek, N. O'Connor, S. Marlow, and N. Murphy, "Face Detection and Clustering for Video Indexing Applications," Proc. Advanced Concepts for Intelligent Vision Systems, 2003

[7] Z. Jin, Z. Lou, J. Yang, and Q. Sun, "Face Detection Using Template Matching and Skin-color Information," Neurocomputing, vol. 70, pp. 794-800, 2007

[8] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face Detection in Color Images," IEEE Trans on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 696-707, 2002

[9] M. Turk and A. Pentland, “Eigenfaces for Recognition,” Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991

[10] B. Kepenekci, "Face Recognition Using Gabor Wavelet Transform," PhD thesis, The Middle East Technical University, 2001

[11] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Recognition with Local Binary Patterns," Proc. European Conference on Computer Vision, vol. 3021, pp. 469–481, 2004

[12] W. Y. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," ACM Computing Surveys (CSUR), vol. 35, no. 4, pp. 399-458, 2003

[13] J. Tao and Y. P. Tan, "Efficient Clustering of Face Sequences with Application to Character-based Movie Browsing," Proc. IEEE International Conference on Image

Processing, pp. 1708-1711, 2008

[14] P. Huang, Y. Wang, and M. Shao, "A New Method for Multi-view Face Clustering in Video Sequence," Proc. IEEE International Conference on Data Mining Workshops, pp. 869-873, 2008

[15] D. Ramanan, S. Baker, and S. Kakade, "Leveraging Archival Video For Building Face Datasets," Proc. IEEE International Conference on Computer Vision, pp. 1-8, 2007

[16] E. El-Khoury, C. Senac, and P. Joly, "Face-and-Clothing Based People Clustering in Video Content," Proc. International Multimedia Conference on Multimedia Information Retrieval, pp. 295-304, 2010

[17] K. Yamamoto, O. Yamaguchi, and H. Aoki, "Fast Face Clustering Based on Shot Similarity for Browsing Video," Progress in Informatics, pp. 53-62, 2010

[18] 洪詩祐,“使用臉部訊息輔助自動化膚色偵測,”交通大學多媒體工程研究所碩士論文, 2009

[19] S. Theodoridis and K. Koutroumbas, *Pattern Recognition (4th Edition)*, Academic Press, 2008

[20] J. Goldberger and T. Tassa, "A Hierarchical Clustering Algorithm Based on the Hungarian Method," Pattern Recognition Letters, vol. 29, no. 11, pp. 1632-1638, 2008

[21] G. Karypis and V. Kumar, "METIS: A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices Version 4.0," University of Minn. Dept. of Comp. Sci., Minneapolis, 1997

[22] L. Hubert and P. Arabic, "Comparing Partitions," Journal of Classification, vol. 2, no. 1, pp.193-218, 1985