

# 國立交通大學

## 多媒體工程研究所

### 碩士論文

視訊人物分群方法之分析與改進

The Analysis and Improvements of People Clustering in Video

研究生：許經國

指導教授：王才沛 教授

中華民國 一 百 零 三 年 九 月

視訊人物分群方法之分析與改進

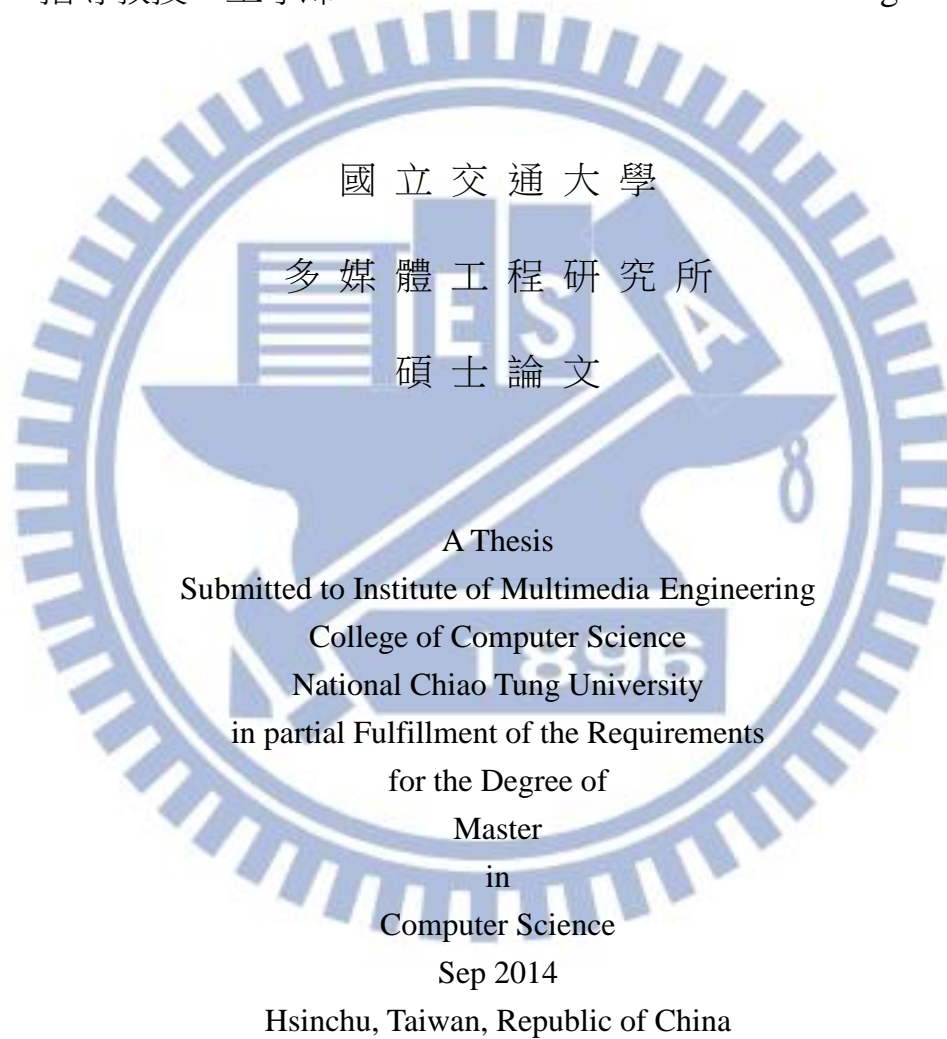
The Analysis and Improvements of People Clustering in Video

研究生：許經國

Student：Ching-Kuo Hsu

指導教授：王才沛

Advisor：Tsai-Pei Wang



中華民國一百零三年九月

# 視訊人物分群方法之分析與改進

學生：許經國 指導教授：王才沛

國立交通大學多媒體工程研究所

## 摘要

人臉分群是近年研究廣泛的課題，近三十幾年出現了許多此領域的論文與相關研究。用於保安方面的人臉辨識系統，以及監視器偵測人臉，也是這方面的主要應用。目前還沒有一個完美的方法被提出來，所以未來此領域仍是多媒體工程的發展主流。

近年許多的研究，偏重於一段影片而非一張張影像上，把影片中連續出現的人臉影像作成串列，有助於人臉分群。本文以人臉影像串列之間的相似度做為主要研究方向，對多種人臉分群方法進行實驗與分析，並提出各種情況適用哪一種分群方法與參數，以及改進這些分群方法的策略。

# The Analysis and Improvements of People Clustering in Video

Student : Ching-Kuo Hsu Advisor : Tsai-Pei Wang

Institute of Multimedia Engineering

College of Computer Science

National Chiao Tung University

## Abstract

Face clustering has been widely studied in recent years. In the past thirty years, there have been developed many papers and applications about this subject. Face recognition for security system and face detection for monitors are primary applications of this subject. For now, there does not have a perfect method, so this topic is a hot development direction for multimedia engineering in the future.

Recent research focuses more on a video instead of images. We can extract facial images which appear sequentially in a video, and convert these images into image sequences. By this way, we can improve face clustering techniques. We focus on face image sequence similarity as the main research topic. We experiment on many face clustering algorithms and analyze experimental results. We illustrate how each method works in any cases, and some ways to improve these algorithms.

## 致謝

這篇論文能夠順利完成，首先要感謝我的指導教授王才沛老師。剛進研究室時很多事都不會，老師先訓練我的研究能力，之後一步步地引導我進行論文研究，才有今天的成果。可以說沒有老師的幫助，我不可能完成這篇論文。

剛進入交大時，耿德、向德、育任、瀚賢、堡評等學長幫助我順利適應交大生活，並給了我一些修課與研究方面的建議與經驗，讓我能夠快速適應碩士班生活。兆祥、台盛、柏淵等同學和我一起修課、擔任助教、進行研究計畫、論文研究，能和同學一起成長茁壯，是這兩年最大的喜悅。子韻、俊宇、嘉暉、博裕等學弟在助教工作與研究計畫幫助我，讓我體認到資工人必須終生學習與重視分工合作的精神。此外，好友証揚在學業方面給我許多建議與幫助，讓我這兩年持續進步。

同時也要感謝我的祖父母、父母和妹妹，在生活及經濟上提供很多協助，使我可以無後顧之憂地順利完成碩士學業，同時也支持我繼續攻讀博士學位。

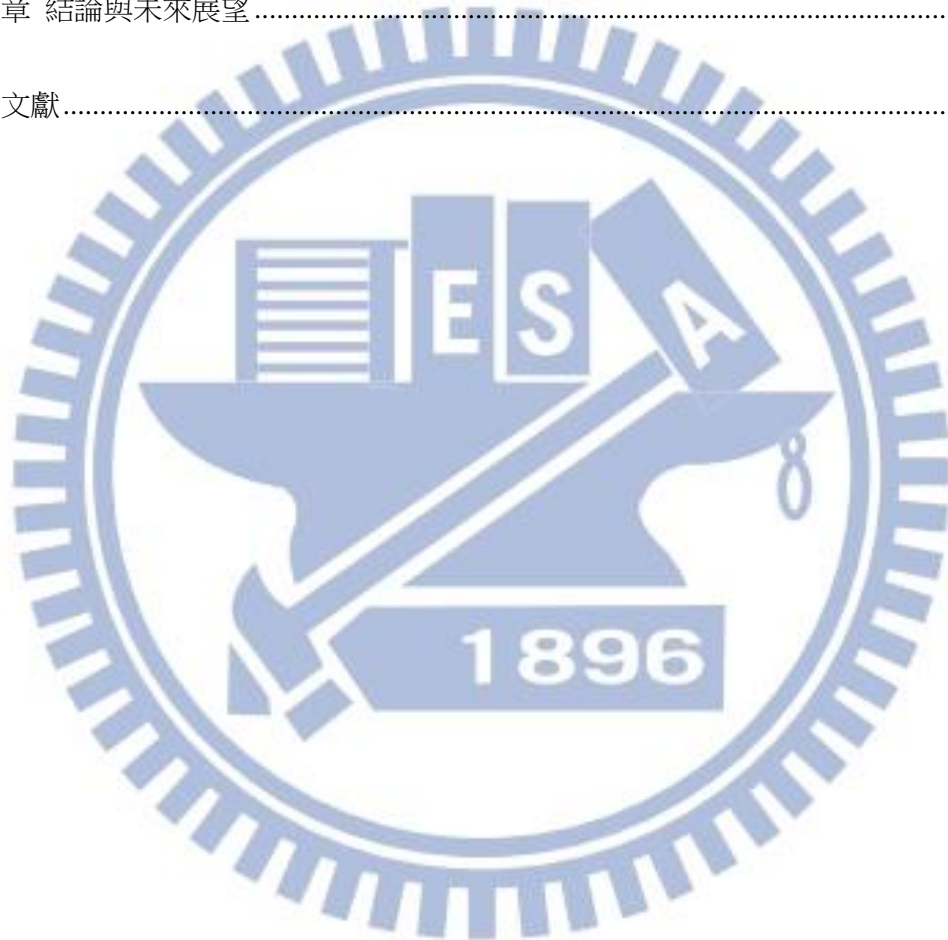


# 目錄

摘要.....	I
Abstract .....	II
致謝.....	III
目錄.....	IV
圖例.....	VII
表格.....	VIII
第一章 簡介.....	1
1.1 研究動機 .....	1
1.2 論文架構 .....	2
第二章 文獻探討.....	4
2.1 人臉辨識(Face recognition) .....	4
2.2 人臉分群(Face clustering) .....	5
2.3 近鄰傳播(Affinity Propagation).....	6
2.4 叢集整合(Clustering Ensemble) .....	7
第三章 實驗方法.....	8
3.1 前置作業.....	8
3.1.1 產生臉部影像.....	8
3.1.2 建立人臉串列.....	8

3.1.3 臉部影像前處理.....	9
3.1.4 距離矩陣之計算.....	9
3.2 分群方法.....	9
3.2.1 Agglomerative Hierarchical Clustering (階層式分群法) .....	9
3.2.2 Affinity Propagation (AP, 近鄰傳播) .....	10
3.2.3 Clustering Ensemble (叢集整合) .....	11
3.3 分群效能評估方法.....	11
3.3.1 ARI(Adjusted Rand Index) .....	11
3.3.2 NMI(Normalized Mutual Information).....	12
3.4 LBP(Local Binary Pattern)與 HOG(Histogram of Gradient)介紹.....	12
3.4.1 LBP 介紹 .....	12
3.4.2 HOG 介紹.....	12
第四章 實驗結果.....	13
4.1 三國影片串列分群概述.....	13
4.1.1 實驗對象.....	13
4.1.2 Ground Truth(基準).....	13
4.2 階層式分群法(Agglomerative Hierarchical Clustering)分析 .....	14
4.3 Single Run 分析.....	16
4.3.1 近鄰傳播 Single Run 分析.....	16

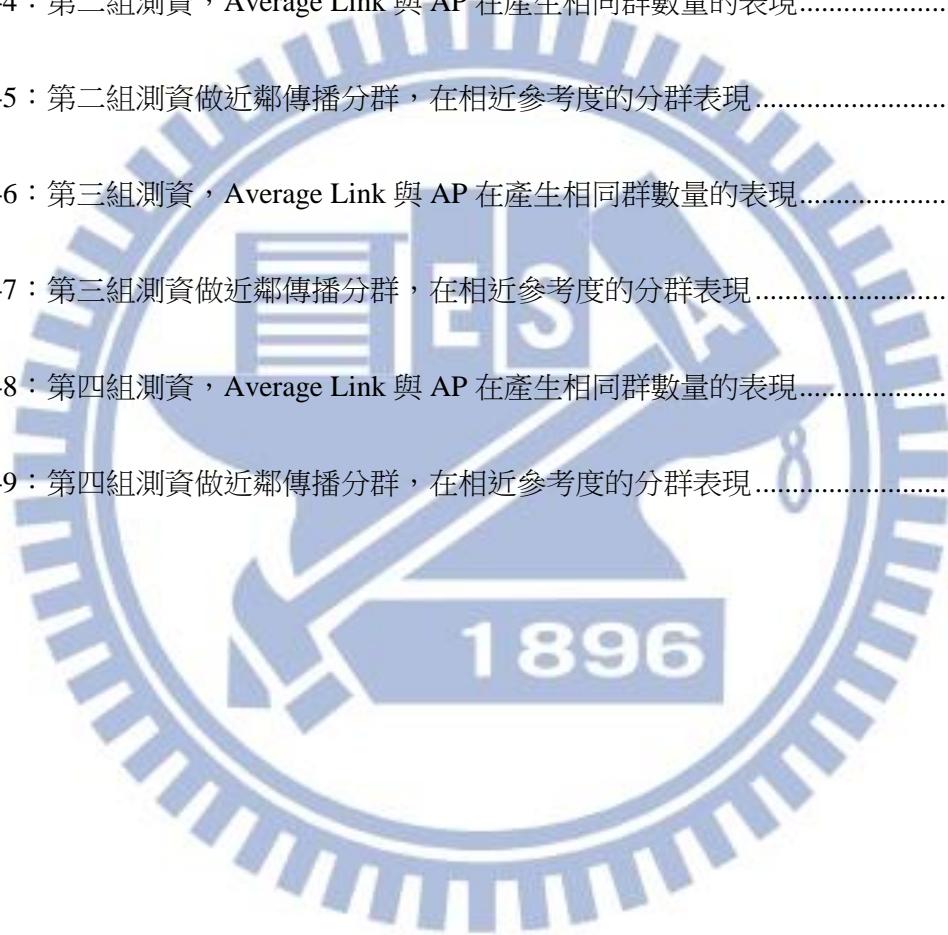
4.3.2 Average Link 與近鄰傳播的 Single Run 分析 .....	18
4.4 叢集整合(Clustering Ensemble)的 Average Link 分析.....	23
4.5 叢集整合(Clustering Ensemble)的近鄰傳播分析 .....	25
4.6 近鄰傳播參考度與分群結果之關係.....	28
第五章 結論與未來展望 .....	29
參考文獻.....	30





## 圖例

圖 4-1：經前處理 1 的灰階串列做三種階層式分群的實驗結果曲線.....	15
圖 4-2：經前處理 1 的灰階串列，Average Link 與 AP 在產生相同群數量的表現.....	19
圖 4-3：經前處理 1 的灰階串列做近鄰傳播分群，在相近參考度的分群表現.....	19
圖 4-4：第二組測資，Average Link 與 AP 在產生相同群數量的表現.....	20
圖 4-5：第二組測資做近鄰傳播分群，在相近參考度的分群表現.....	21
圖 4-6：第三組測資，Average Link 與 AP 在產生相同群數量的表現.....	21
圖 4-7：第三組測資做近鄰傳播分群，在相近參考度的分群表現.....	22
圖 4-8：第四組測資，Average Link 與 AP 在產生相同群數量的表現.....	22
圖 4-9：第四組測資做近鄰傳播分群，在相近參考度的分群表現.....	23



## 表格

表 4-1：經前處理 1 的灰階串列、LBP、HOG 做三種階層式分群的實驗結果 .....	15
表 4-2：經前處理 2 的灰階串列、LBP、HOG 做三種階層式分群的實驗結果 .....	16
表 4-3：經前處理 1 的 HOG 串列做近鄰傳播分群，參考度為相似度矩陣中位數及其倍數 .....	17
表 4-4：經前處理 1 的 HOG 串列做近鄰傳播分群，參考度為均一初值及其倍數 .....	18
表 4-5：經前處理 1 的灰階串列做 Ensemble AL 的實驗結果以及與 Single Run AL 之比較 .....	25
表 4-6：經前處理 1 的灰階串列做 Ensemble AP(均一值做為參考度)的實驗結果以及與 Single Run AP(均一初值做為參考度)之比較.....	26
表 4-7：經前處理 1 的灰階串列做 Ensemble AP(相似度矩陣中位數之倍數做為參考度)的實驗結果以及與 Single Run AP(相似度矩陣中位數做為參考度)之比較.....	26
表 4-8：經前處理 1 的灰階串列以不同參考度做 Ensemble AP 的實驗結果 .....	27
表 4-9：經前處理 1 的灰階串列以不同參考度做 Ensemble AP 的實驗結果以及與 Ensemble AL 之比較.....	27

# 第一章 簡介

## 1.1 研究動機

人臉辨識是近年研究相當熱門的領域，近三十幾年出現了許多此領域的論文與相關研究。近代在保安方面的人臉辨識系統，以及監視器偵測人臉，也都是被廣為應用的方向。

近年許多的研究，偏重在一段影片而非一張張影像上，當人們欣賞一段影片時，想要搜尋特定腳色出現的場景，必須透過遙控器來快轉或倒帶，如果對於影片內容不甚熟悉，勢必得花費許多時間造成事倍功半。如果先整理出所有腳色出現的片段來做分群，會有助於觀眾回顧與搜尋，而這些片段必須是連續的，所以一個腳色可能會有多個片段。因此，我們想要藉由將每一腳色出現的片段，計算彼此之間的相似度再做分群，以助於人們在影片中搜尋自己喜愛的腳色。

如果只是做一次分群，就會缺乏隨機性與知識重用(knowledge reuse)，使得分群過程完全取決於演算法的設計，得到的結果不一定會是最好的。因此我們想利用近年常被探討的叢集整合(Clustering Ensemble)，來輔助分群演算法。叢集整合的概念，是先隨機取樣部分物體作分群，再合併剩下的物體至各群，採用多回合(turn)的方式，所以原先為同一群的物體，在此種方式下，被分到同一群的次數較高。因此我們會計算這些物體被分到同一群的次數，根據次數多寡重作分群。如此分群的正確率相信會提升不少。

近鄰傳播(Affinity Propagation)是近幾年才被提出的演算法，根據資料點(data

point)之間的相似度(similarity)分群，同時考慮各點為潛在的群中心(exemplar)，其優點是降低錯誤率並節省大量的時間。近鄰傳播的參考度(preference)通常設為相似度矩陣的中位數(median)，我們猜想有些均一(uniform)的參考度值可以讓近鄰傳播有更好的表現，此外調整參考度的倍數可以讓群數量產生變化，也許可得到更好的分群結果。

同時為了直接比較分群方法的優缺點，我們對於“產生相同群數量的 Average Link 與近鄰傳播”及“參考度值相近的近鄰傳播(相似度矩陣中位數的某個倍數對比與其平均值相近的均一參考度值)”的 Single Run，也做了探討與分析，讓讀者了解在任何情況下，採用何種分群方法，再搭配何組實驗參數，可以達到優化分群的功效。

我們的分群評估標準是 ARI(Adjusted Rand Index)與 NMI(Normalized Mutual Information)，因為二者評估的方向不同，所以我們想要探討在產生不同群數量、極小的群數量多寡等多種情況下，二者數值會如何變化，並指出哪些參數調整可以讓他們進步。

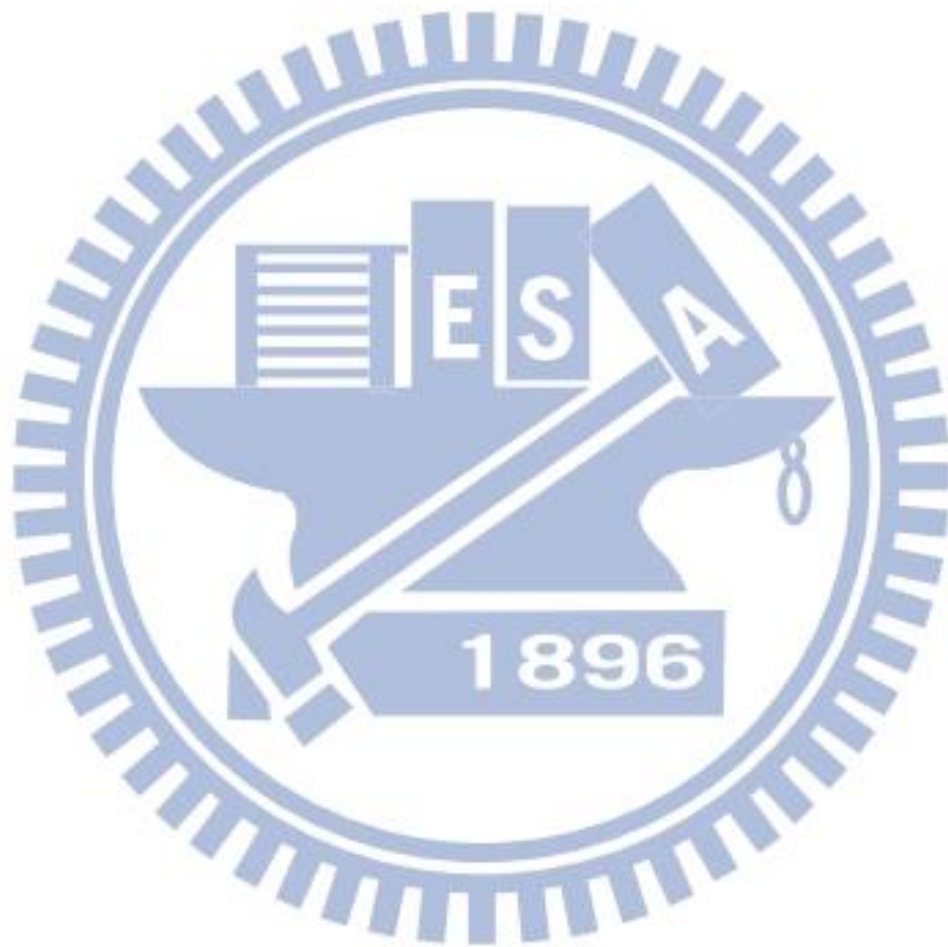
本論文的主要貢獻，在於分析各種分群方法在哪些情況表現較好，哪些情況較不適用，並提出改進分群的措施，例如近鄰傳播採用不同參考度可以得到比原作者更好的分群表現，也對多組資料進行分析，以強化我們的論述。

## 1.2 論文架構

本論文的架構如下：第二章是文獻探討，第三章是實驗方法介紹，第四章是



實驗結果與比較，第五章是結論與未來展望。





## 第二章 文獻探討

### 2.1 人臉辨識(Face recognition)

Everingham 和 Zisserman [1]提出自動化視覺辨認情境喜劇中腳色的方法，給定一個特定人物與一段未被標記的(unlabeled)影片，可以偵測出這個人物是否出現在這段影片內。Everingham 和 Zisserman [2]提出在影片中自動辨識人物的方法，結合了電腦視覺(computer vision)與機器學習(machine learning)的方法來達到目的。

Hu 等[3][4]提出用 SANP(Sparse Approximated Nearest Points)來做影像集辨識(image set classification)。Turk 和 A. P. Pentland [5]提出使用 Eigenface 來做人臉辨識的方法，將人臉辨識視為二維辨識的問題，再利用臉部多為正向(upright)，可以用二維特性曲線圖(characteristic view)描述的既定事實來分群。Liu 和 Chen [6]提出用適應性的(adaptive)的隱馬可夫模型(HMM, Hidden Markov Models)來做視訊人臉辨識(video-based face recognition)，可以達到優於多數決(majority voting)的表現。Chu 等[7]提出核心鑑別式轉換(KDT, kernel discriminant transformation)來做影像集人臉辨識(image set-based face recognition)，實驗顯示其辨識表現優於現行的靜態影像人臉辨識方法(still-image-based face recognition method)與影像集人臉辨識方法。缺點是當訓練影像(training image)增加時，KDT 演算法的時間複雜度將會大幅增加。Harandi 等[8]提出將直推式學習(transductive learning)運用在影像集人臉辨識的方法，可以將影像集與單一影像的配對問題(image-set to single

image matching problem)轉換為典型相關的集合配對(set matching using canonical correlation)。

Ahonen 等[9][10]提出用局部二元特徵(LBP, Local Binary Pattern)來代表一張臉部影像(face image), 加強人臉辨識的正確率。LBP 的概念是將灰階像素用該值與周圍鄰居的相對關係來取代, 例如取鄰近 8 個鄰居, 則該 LBP 值為 0 至 255。

Cevikalp 和 Triggs [11]提出了使用 affine hull 或 convex hull 代表影像集(image set), 以他們之間的幾何距離當作影像集之間的歧異度(dissimilarity), 來對影像集分群。

我們的實驗直接使用[11]的作者提供的程式。Yang 等[12]提出了使用 2DPCA(two-dimensional principal component analysis)來代表影像的方法, 並比較 2DPCA 與 PCA(eigenface), 並分析前者優於後者的原因。

## 2.2 人臉分群(Face clustering)

Yamamoto 等[13]提出兩種以鏡頭相似度為依據的分群方法(SSC, similar shot-based clustering), 第一種方法只使用依據鏡頭相似度的分群, 第二種方法增加使用縮小版臉部的分群(FTC, face thumbnail clustering)。Le 等[14]用

SIFT(scale-invariant feature transform)產生的影像特徵(image feature)進行階層人臉分群, El Khoury 等提出可用臉部與服裝綜合資訊分群的方法, 先找出影片中的關鍵臉(keyface), 使用 SIFT 進行臉部匹配(face matching), 再用三維直方圖(three dimensional histogram)與主色(dominant color)分析做服裝匹配(clothing matching), 最後結合 SIFT 特徵(feature)、膚色、三維直方圖與服裝質地(feature)

等做三階段階層分群，Zhang 等提出結合背景(context)輔助人臉分群的方法，Chu 等[17]提出結合背景資訊與局部特徵點(local feature point)輔助人臉分群的方法。Huang 等[18]提出“相同姿勢，不同人物”的相似度比“相同人物，不同姿勢”更好，所以先依據姿勢分群，再做各姿勢的人物分群。Tao 和 Tan [19]提出了當人們搜尋影片中特定腳色時，可以極有效率的臉部串列分群方法。先將臉部串列分割成相同姿勢為一組的子串列，再導入實驗所需的限制使用近鄰傳播分群，最後再用作者提出的查詢瀏覽系統達到目的，Kayal [20]將各種人臉分群方法對新聞影片的影像集進行實驗，並比較這些方法之間的優劣關係。Kayal [21]和 Zhang 等[22]提出用時間資訊輔助人臉分群的方法，Foucher 和 Gagnon [23]以及 Orfanidis 等[24]提出可將光譜分群(spectral clustering)應用於人臉分群，Czirjek 等 [25]提出人臉偵測與分群在影片索引(video indexing)方面的應用，Cui 等[26]根據人臉分群技術開發了互動式相片註記系統(interactive photo annotation system)，Li 和 Tang [27]提出可用支持向量機(SVM, support vector machine)與人臉分群的技術改進人臉辨識。

### 2.3 近鄰傳播(Affinity Propagation)

Frey 和 Dueck [28]率先提出近鄰傳播分群方法，指出此方法相較於其他分群方法，可以減少錯誤率，且節省 99%以上的時間。Wang 等[29]指出了近鄰傳播的兩個限制：我們很難知道怎樣的參考度(preference)可以找出最好的分群結果，而且震動(oscillation)一旦發生就不能消除。因此提出了適應性的(adaptive)的近鄰



傳播，適度調整參考度找出最好的分群結果，與阻尼係數(damping factor)以避免震動。Lu 和 Carreira-Perpinan [30]指出了成對的限制(pairwise constraints)可區分兩個不同物體是否屬於同一群，所以透過成對的限制改善近鄰傳播分群。Givoni 和 Frey [31]提出了二維變數模型(binary variable model)簡化近鄰傳播分群，可以簡化訊息更新的推導。

## 2.4 叢集整合(Clustering Ensemble)

Strehl 和 Ghosh [32]提出了叢集整合的概念，藉由整合多次的分群結果以找出一個共識解(consensus solution)。叢集整合屬於強韌叢集演算法(robust clustering)，可以提升分群結果的品質(quality of clustering)，並達到知識重用(knowledge reuse)的好處。Topchy 等[33]指出叢集整合是結合多個分割(partition)以找出一個最好品質的解，提出了適應性的取樣(sampling)與分群(clustering)的概念。Minaei-Bidgoli 等[34]提出藉由資料重複取樣(data resampling)來做分割整合(ensembles of partitions)的方法，即為叢集整合的精髓。實驗發現每一組件(component)分割的群數量與整合的分割數量之間有著權衡(tradeoff)的關係。

## 第三章 實驗方法

### 3.1 前置作業

#### 3.1.1 產生臉部影像

首先須挑選實驗用的影片，我選擇陸劇“新三國”第 83 集，選擇的原因是本集影片人臉在影像中比例較大，且沒有激烈動作如戰爭、打鬥等。先轉成 360p 的.wmv 檔，以 3fps 的速度，利用“Free Video To JPG Converter”軟體，從影片擷取了 7427 張影像（實驗時排除片頭的 15 張影像），再用 OpenCV 偵測出每一張影像中可能的人臉位置。

在 Single Run 的分析實驗，我們也加入了另外三組測試資料：

- 第二組資料有 529 個串列，基準分為 8 群。
- 第三組資料有 463 個串列，基準分為 10 群。
- 第四組資料有 1000 個串列，基準分為 16 群。

#### 3.1.2 建立人臉串列

藉由人工標記鏡頭轉換的時刻，加上 OpenCV 得到的人臉資訊，我們可以建立人臉串列，需要符合以下條件：

- 人臉串列的影像須為同一人(或物體)
- 人臉串列的每一影像必須是同一場景且連續出現
- 每個人臉大小必須至少  $40 \times 40$  pixels

符合上述條件，人臉串列的長度必須介於 4 至 40，每張人臉影像的大小調



整到 $40 \times 40$ 。

之後再將彩色影像轉換為灰階影像。

### 3.1.3 臉部影像前處理

我們使用的影像前處理方式有二種：

#### 1. Histogram Equalization

#### 2. Histogram Equalization + Gaussian Bandpass Filtering

先校正影像的亮度，再用兩組反向的二維 Gaussian 函數(如式(1)、(2))組成帶通濾波器(bandpass filter)，過濾影像中低頻及高頻的部分。

$$f_{low} = \exp\left(-\frac{(x-c_x)^2+(y-c_y)^2}{2\sigma_{low}^2}\right), \text{size} = 5 \times 5, \sigma_{low} = 1 \quad (1)$$

$$f_{high} = 1 - \exp\left(-\frac{(x-c_x)^2+(y-c_y)^2}{2\sigma_{high}^2}\right), \text{size} = 9 \times 9, \sigma_{high} = 2 \quad (2)$$

### 3.1.4 距離矩陣之計算

#### LAHISD (Linear Affine Hull based Image Set Distance)

用一個仿射包(affine hull)代表一個影像串列(image sequence)，串列中的影像(image)是仿射包的特徵向量(feature vector)，串列之間的歧異度(dissimilarity)由仿射包之間的幾何距離(geometric distance)來決定。我們使用文獻[11]的作者提供的程式碼。

## 3.2 分群方法

### 3.2.1 Agglomerative Hierarchical Clustering (階層式分群法)

一開始將每一物體視為一群，每一回合找出距離最小的兩群 $C_i$ 和 $C_j$ 進行合併，

持續進行合併直到群的數量滿足我們所要求的為止。每一回合合併產生的群 $C_q$ 和其他各群 $C_s$ 的更新公式有以下三種：

- Single Link :

$$d(C_q, C_s) = \min\{d(C_i, C_s), d(C_j, C_s)\} \quad (3)$$

- Complete Link :

$$d(C_q, C_s) = \max\{d(C_i, C_s), d(C_j, C_s)\} \quad (4)$$

- Average Link :

$$d(C_q, C_s) = \frac{n_i}{n_i+n_j} d(C_i, C_s) + \frac{n_j}{n_i+n_j} d(C_j, C_s) \quad (5)$$

其中 $n_i$ 和 $n_j$ 分別代表 $C_i$ 和 $C_j$ 包含的物體個數。

### 3.2.2 Affinity Propagation (AP, 近鄰傳播)

這是一種用於生物醫學的演算法，根據所有物體之間的相似度進行分群，最初將所有的物體都視為潛在的群中心。

用到的一些名詞如下：

- Exemplar : 群中心
- Similarity :  $s(i, j)$ 表示物體  $i$  和  $j$  的相似度
- Preference :  $p(i)$ 表示物體  $i$  作為群中心的參考度
- Responsibility :  $r(i, j)$ 表示物體  $j$  適合作為物體  $i$  的群中心的程度
- Availability :  $a(i, j)$ 表示物體  $i$  選擇物體  $j$  作為其群中心的適合程度
- $r$  和  $a$  的更新公式：

$$r(i, j) = s(i, j) - \max_{k \neq j} (a(k, i) + s(i, k)) \quad (6)$$

$$a(j, j) = \sum_{k \neq j} \max\{0, r(k, j)\} \quad (7)$$

$$a(j, i) = \min(0, r(j, j) + \sum_{k \neq j, i} \max\{0, r(k, j)\}) \quad (8)$$

- 各物體  $i$  所屬的群中心  $c_i$  更新公式：

$$c_i^* = \operatorname{argmax}_j r(i, j) + a(j, i) \quad (9)$$

程式中止條件：重複做一千回合或收斂條件達成

收斂條件：連續一百回合群中心沒有更新

### 3.2.3 Clustering Ensemble (叢集整合)

因為 Average Link 與近鄰傳播沒有隨機性，所以我們想要藉由叢集整合來改善分群。叢集整合在多次不同的隨機取樣中，兩個物體  $i$  和  $j$  被分在同一群的次數越高，代表將所有物體分群，二者被分在同一群的機會也越高，這是叢集整合的核心概念：善用先前實驗的結果達到更優化的分群。實驗中每個回合(turn)採用 Average Link 或近鄰傳播分群。我們建立一個大小為  $n \times n$  的矩陣，稱為 co-association matrix，其中編號  $(i, j)$  的元素表示串列  $i$  與  $j$  在實驗中有多少個回合被分在同一群。在叢集整合的實驗中，co-association matrix 作為第二階段分群的歧異度矩陣(dissimilarity matrix)。

## 3.3 分群效能評估方法

### 3.3.1 ARI(Adjusted Rand Index)

給定一個含有  $n$  物的集合， $X = \{X_1, X_2, \dots, X_r\}$  為 ground truth， $Y =$

$\{Y_1, Y_2, \dots, Y_s\}$  為被評估的分群結果，令  $n_{ij}$  為同時屬於  $X_i$  和  $Y_j$  的物體數量， $a_i$  和  $b_j$  分別代表在  $X_i$  和  $Y_j$  內的物體數量，則

$$\text{ARI} = \frac{\sum_{i,j} c_2^{n_{ij}} - [\sum_i c_2^{a_i} \sum_j c_2^{b_j}] / c_2^n}{\frac{1}{2}[\sum_i c_2^{a_i} + \sum_j c_2^{b_j}] - [\sum_i c_2^{a_i} \sum_j c_2^{b_j}] / c_2^n} \quad (10)$$

### 3.3.2 NMI(Normalized Mutual Information)

給定一個 ground truth X 和一個被評估的分群結果 Y，先計算他們的分布(即各群有多少比例的物體)，再計算他們的熵(entropy,  $H(X) = -\sum_i P(x_i) \ln P(x_i)$ )和聯合熵(joint entropy,  $H(X, Y) = -\sum_x \sum_y P(x, y) \ln P(x, y)$ )，最後得到

$$\text{MI(Mutual Information)} = H(X) + H(Y) - H(X, Y) \quad (11)$$

$$\text{NMI} = \frac{\text{MI}}{\sqrt{H(X)H(Y)}} \quad (12)$$

## 3.4 LBP(Local Binary Pattern)與 HOG(Histogram of Gradient)介紹

### 3.4.1 LBP 介紹

對於每一像素，比較它的 8 個鄰居。如果此像素值大於鄰居的值，給 1，否則給 0，形成一個 8 bits 的二進位數。之後做 0~255 的直方圖，所得即為該影像的特徵向量。

### 3.4.2 HOG 介紹

對於每一  $6 \times 6$  的方格單元 (排除最外層的兩圈)，做每一像素的梯度方向邊緣方向直方圖，之後整合所有直方圖，所得即為該影像的特徵向量。



## 第四章 實驗結果

### 4.1 三國影片串列分群概述

#### 4.1.1 實驗對象

1. 灰階影像的人臉串列，經過 3.1.3 提到的兩種影像前處理。
2. LBP (Local Binary Pattern) 特徵串列，經過 3.1.3 提到的兩種影像前處理。
3. HOG (Histogram of Gradient) 特徵串列，經過 3.1.3 提到的兩種影像前處理。

#### 4.1.2 Ground Truth(基準)

我們實驗有 2 組 Ground Truth，本實驗的串列包含 18 位角色，以及 15 個非人臉的物品（如演員的衣服或柱子等），因此 2 組 Ground Truth 分別為 19 群（非人臉的串列分至同一群）與 33 群（非人臉的不同物品各自為一群）。取兩種基準的動機是非人臉有非常多種物體，這些物體差別很大，而後面的實驗結果，除了以比較這兩者為目的之部分以外，我們的表格只列第二種基準(33 群)的實驗數據。

我們用下列敘述簡化論文分析：

- 前處理 1：使用 Histogram Equalization 為前處理方式
- 前處理 2：使用 Histogram Equalization + Gaussian Band Pass Filtering 為前處理方式
- ARI<sub>1</sub> 與 NMI<sub>1</sub>：使用第一組基準(19 群)，在 4.3 表示 Single Run 的 ARI 與 NMI，在 4.2 與 4.4 表示分 10~72 群的平均 ARI 與 NMI



- ARI\_2 與 NMI\_2：使用第二組基準(33 群)，在 4.3 表示 Single Run 的 ARI 與 NMI，在 4.2 與 4.4 表示分 10~72 群的平均 ARI 與 NMI

## 4.2 階層式分群法(Agglomerative Hierarchical Clustering)分析

我們對經兩種前處理的灰階影像串列，以及他們的 LBP 與 HOG 特徵串列，做三種階層式分群法，分 10~358 群，並進行比較與分析。

使用前處理 1 的影像串列的實驗結果如圖 4.1 與表 4.1：實驗發現三種階層分群都是 HOG 串列表現最好，灰階串列其次，LBP 串列表現最差，因此我們認為在這樣的前處理方式下，HOG 特徵更能代表一張影像。LBP 串列用 Complete Link 的表現最好，灰階串列與 HOG 串列用 Average Link 的 ARI 表現最好，Complete Link 的 NMI 表現最好，三者都是用 Single Link 的表現最差。

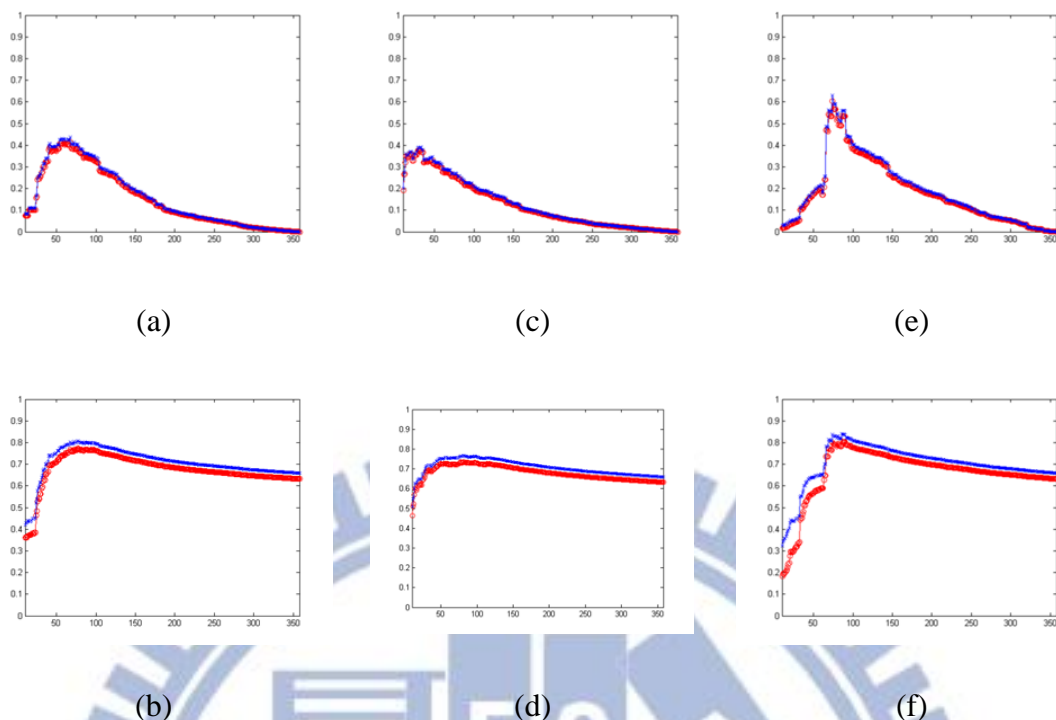


圖 4-1：經前處理 1 的灰階串列做三種階層式分群的實驗結果曲線，(a)經前處理 1 的灰階串列做 Average Link 的 ARI 曲線(兩組基準)，(b)經前處理 1 的灰階串列做 Average Link 的 NMI 曲線(兩組基準)，(c)經前處理 1 的灰階串列做 Complete Link 的 ARI 曲線(兩組基準)，(d)經前處理 1 的灰階串列做 Complete Link 的 NMI 曲線(兩組基準)，(e)經前處理 1 的灰階串列做 Single Link 的 ARI 曲線(兩組基準)，(f)經前處理 1 的灰階串列做 Single Link 的 NMI 曲線(兩組基準)，紅線表示對照第一組基準的數值，藍線表示對照第二組基準的數值

表 4-1：經前處理 1 的灰階串列、LBP、HOG 做三種階層式分群的實驗結果

	ARI_2(AL)	ARI_2(CL)	ARI_2(SL)	NMI_2(AL)	NMI_2(CL)	NMI_2(SL)
灰階	0.309	0.295	0.176	0.663	0.669	0.566
LBP	0.231	0.262	0.086	0.604	0.643	0.484
HOG	0.333	0.331	0.233	0.676	0.701	0.618

使用前處理 2 的影像串列的實驗結果如表 4.2：實驗發現不論採用哪一組基準，灰階串列在 Average Link 表現最好，HOG 串列在 Single Link 表現最好，Complete Link 則是 LBP 串列的 ARI 表現最好，灰階串列的 NMI 表現最好。我們認為是使用了 Gaussian Band Pass Filtering，濾掉了高頻與低頻的部分，影響了 LBP 與 HOG 的元素成分，使得 HOG 串列不是三種方法都表現最好。同前處理

1 的結果，LBP 串列用 Complete Link 的表現最好，灰階串列與 HOG 串列用 Average Link 的 ARI 表現最好，Complete Link 的 NMI 表現最好，三者都是用 Single Link 的表現最差。

整體而言，前處理 1 在 Single Link 表現較佳，前處理 2 在 Average Link 與 Complete Link 表現較佳。

表 4-2：經前處理 2 的灰階串列、LBP、HOG 做三種階層式分群的實驗結果

	ARI_2(AL)	ARI_2(CL)	ARI_2(SL)	NMI_2(AL)	NMI_2(CL)	NMI_2(SL)
灰階	0.390	0.330	0.168	0.701	0.709	0.569
LBP	0.209	0.337	0.061	0.578	0.674	0.438
HOG	0.366	0.318	0.205	0.686	0.691	0.609

### 4.3 Single Run 分析

#### 4.3.1 近鄰傳播 Single Run 分析

近鄰傳播分群，選定較大的參考度，代表各串列被選為為群中心的機會就越高，會產生較多的群，反之則會產生較少的群。因此我們可以試著調整參考度的倍數，讓分群結果更好。實驗結果如表 4-3：我們發現將參考度調整為 2 倍或 4 倍時，分群結果非常糟糕，群數目是串列數的相近至一半。而將參考度調整為二分之一或四分之一時，兩種基準的 ARI 都進步，但是 NMI 都退步。因為參考度不做倍數調整時，產生的群較多，且極小的群(包含串列數不超過 5)也較多，所以 NMI 較大，ARI 較小。

表 4-3：經前處理 1 的 HOG 串列做近鄰傳播分群，參考度為相似度矩陣中位數及其倍數

倍數調整	ARI_1	ARI_2	NMI_1	NMI_2	群數量
1/4	0.364	0.381	0.712	0.747	32
1/2	0.362	0.379	0.719	0.753	35
1	0.333	0.349	0.745	0.774	45
2	0.085	0.090	0.685	0.714	189
4	0.001	0.001	0.633	0.660	354

傳統上近鄰傳播的參考度是一個向量，其實也可以取純量，即所有的串列參考度均相同，但是此純量的取法對於分群結果影響極大。如果均一參考度高於相似度矩陣最大值，會使得每一串列自成一群，沒有分群的效果。若是均一參考度接近 0，則只會產生一群。因此我們想要經由實驗來找出何種均一參考度值可以讓分群結果更好？

我們想要讓均一參考度的初值，比相似度矩陣絕大多數的正數值要小，因此我們取二十分之一做為 HOG 串列首次實驗的均一參考度值，此值產生的群數量會略少於真實分群。之後我們再做倍數調整以達到更好的分群表現。實驗結果如表 4-4：我們發現這樣設定初值時，ARI 的表現比使用相似度矩陣中位數做為參考度要好，而 NMI 的表現則相反，因為這樣設值會造成各群大小相近，極小的群較少而且群數量接近實際的分群。此外經倍數調整後，ARI 呈現進步，NMI 呈現退步，因為適中的均一參考度會使得產生的群數量與實際的分群數量相近，所以 ARI 表現較好，而相似度矩陣中位數會產生很多極小的群，所以 NMI 表現較好。



表 4-4：經前處理 1 的 HOG 串列做近鄰傳播分群，參考度為均一初值及其倍數

初值倍數調整	ARI_1	ARI_2	NMI_1	NMI_2	群數量
1/4	0.351	0.366	0.690	0.722	29
1/2	0.351	0.366	0.690	0.722	29
1	0.346	0.362	0.691	0.726	31
2	0.358	0.375	0.704	0.740	32
4	0.355	0.371	0.711	0.746	34

#### 4.3.2 Average Link 與近鄰傳播的 Single Run 分析

我們想要探討 Average Link 和近鄰傳播在 Single Run 產生相同群數量時，何者的分群表現較好，所以我們拿 Single Run AL 與分群結果具相同群數量的 Single Run AP 進行比較。實驗結果如圖 4-2：實驗發現在群數量小於 40 的時候，近鄰傳播的表現較好，反之則是 Average Link 稍微好一點，推測是因為近鄰傳播所取用的參考度較小時，能夠產生較少的群數量，而在產生較少的群數量時，Average Link 是直接取兩群最小的距離作為分群依據，近鄰傳播則是用整體的相似度作為分群依據，所以前者在產生較少的群數量時比較容易偏離真實情況。而近鄰傳播取用參考度較大時，可能被選為群中心的資料點會增加，使得產生的群會變多，產生偏差的機會增加，與 Average Link 的表現就會互有勝敗了。

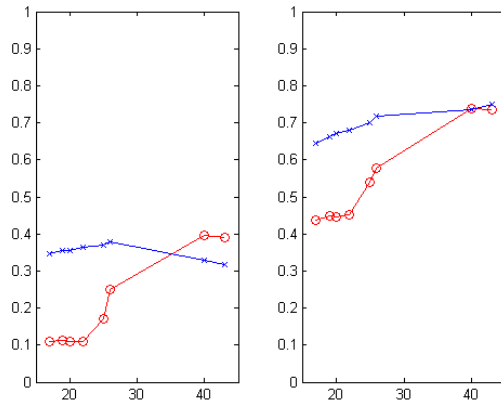


圖 4-2：經前處理 1 的灰階串列，Average Link 與 AP 在產生相同群數量的表現，左圖是 ARI 的表現，右圖是 NMI 的表現，紅線表示 AL 的結果，藍線表示 AP 的結果，橫軸表示群數量，縱軸表示數值

我們對於一個問題感到興趣：如果取用的均一參考度，與相似度矩陣中位數某個倍數的平均值相似，何者的分群表現較好？因此我們將相似度矩陣中位數做倍數調整，取各倍數的平均值做為與其比較的均一參考度值。實驗結果如圖 4-3：實驗發現相似度矩陣中位數系列的表現略優於均一參考度系列，因此我們認為這組串列若有適當且相異的參考度可以找出較好的群中心，讓分群結果更好。

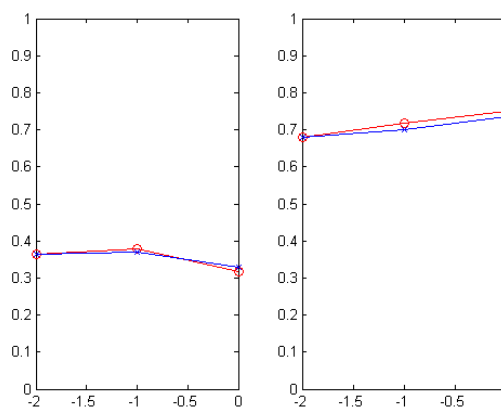


圖 4-3：經前處理 1 的灰階串列做近鄰傳播分群，在相近參考度的分群表現，左圖是 ARI 的表現，右圖是 NMI 的表現，紅線表示相似度矩陣中位數系列的結果，藍線表示均一參考度系列的結果，橫軸表示相似度矩陣中位數的倍數調整為 2 的多少次方，縱軸表示數值

接下來我們對其他三組資料進行相同的分群實驗：

我們對第二組資料做 Average Link 與近鄰傳播的 Single Run 實驗，實驗結果

如圖 4-4：實驗發現產生相同群數量時，近鄰傳播的表現遠優於 Average Link，

我們認為原因是近鄰傳播產生的群數量普遍較少，如同三國影片串列，這組資料

在產生群數量較少時，近鄰傳播可以表現較好。

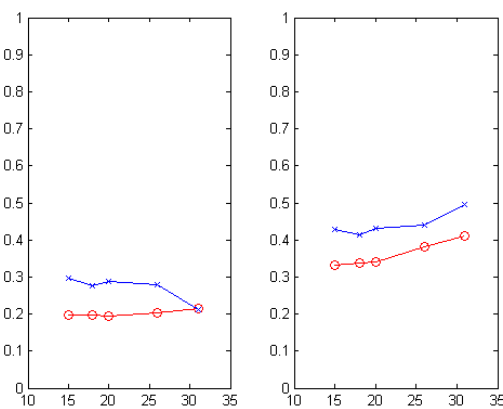


圖 4-4：第二組測資，Average Link 與 AP 在產生相同群數量的表現，左圖是 ARI 的表現，右圖是 NMI 的表現，紅線表示 AL 的結果，藍線表示 AP 的結果，橫軸表示群數量，縱軸表示數值

我們同樣將相似度矩陣中位數做倍數調整，取各倍數的平均值做為與其比較的均一參考度值。實驗結果如圖 4-5：實驗發現分群結果互有勝敗，因此這組資料可能兩系列都適用。

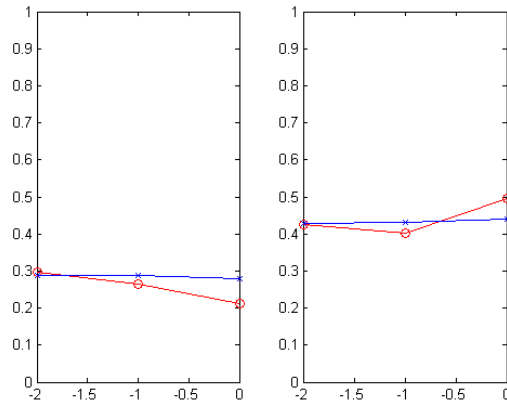


圖 4-5：第二組測資做近鄰傳播分群，在相近參考度的分群表現，左圖是 ARI 的表現，右圖是 NMI 的表現，紅線表示相似度矩陣中位數系列的結果，藍線表示均一參考度系列的結果，橫軸表示相似度矩陣中位數的倍數調整為 2 的多少次方，縱軸表示數值

我們對第三組資料做 Average Link 與近鄰傳播的 Single Run 實驗，實驗結果如圖 4-6：實驗發現產生相同群數量時，Average Link 的表現遠優於近鄰傳播，我們認為原因是這組資料用歧異度分群的效果，比用相似度分群更佳，所以 Average Link 表現較好。

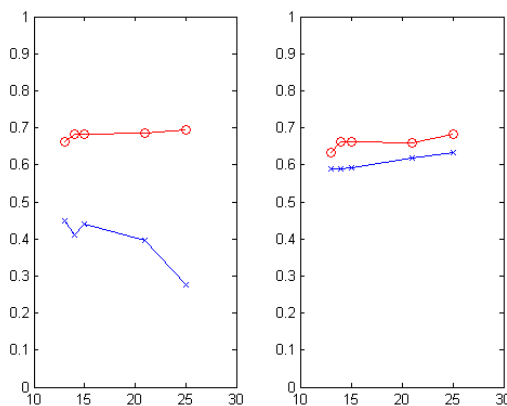


圖 4-6：第三組測資，Average Link 與 AP 在產生相同群數量的表現，左圖是 ARI 的表現，右圖是 NMI 的表現，紅線表示 AL 的結果，藍線表示 AP 的結果，橫軸表示群數量，縱軸表示數值

我們同樣將相似度矩陣中位數做倍數調整，取各倍數的平均值做為與其比較的均一參考度值。實驗結果如圖 4-7：實驗發現分群結果互有勝敗，因此這組資



料可能兩系列都適用。

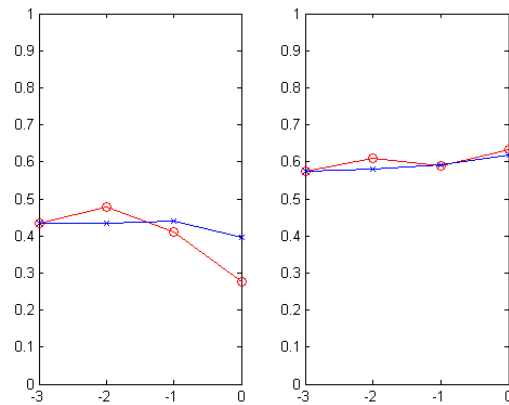


圖 4-7：第三組測資做近鄰傳播分群，在相近參考度的分群表現，左圖是 ARI 的表現，右圖是 NMI 的表現，紅線表示相似度矩陣中位數系列的結果，藍線表示均一參考度系列的結果，橫軸表示相似度矩陣中位數的倍數調整為 2 的多少次方，縱軸表示數值

我們對第四組資料做 Average Link 與近鄰傳播的 Single Run 實驗，實驗結果如圖 4-8：實驗發現產生相同群數量時，Average Link 的表現遠優於近鄰傳播，我們認為原因同第三組測資，這組資料用歧異度分群的效果，比用相似度分群更佳，所以 Average Link 表現較好。

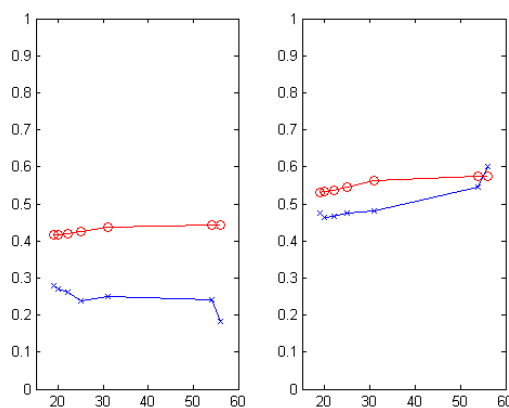


圖 4-8：第四組測資，Average Link 與 AP 在產生相同群數量的表現，左圖是 ARI 的表現，右圖是 NMI 的表現，紅線表示 AL 的結果，藍線表示 AP 的結果，橫軸表示群數量，縱軸表示數值

我們同樣將相似度矩陣中位數做倍數調整，取各倍數的平均值做為與其比較的均一參考度值。實驗結果如圖 4-9：實驗發現倍數較高時，均一參考度表現較好，倍數較低時，均一參考度表現較差，因此我們認為這組資料在調低參考度倍數讓分群改進時，讓各串列參考度相異較好。

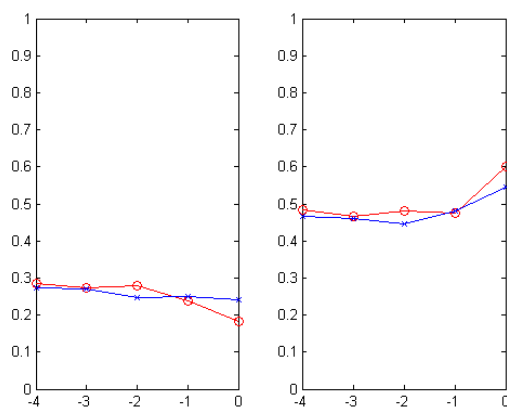


圖 4-9：第四組測資做近鄰傳播分群，在相近參考度的分群表現，左圖是 ARI 的表現，右圖是 NMI 的表現，紅線表示相似度矩陣中位數系列的結果，藍線表示均一參考度系列的結果，橫軸表示相似度矩陣中位數的倍數調整為 2 的多少次方，縱軸表示數值

#### 4.4 叢集整合(Clustering Ensemble)的 Average Link 分析

我們實驗的架構如下：

設定一個循環(cycle)的叢集整合做 20 個回合，取 10 個循環的結果平均。隨機取樣的比例分別為  $1/2$  (179/358)、 $3/5$  (215/358)、 $2/3$  (239/358)、 $3/4$  (269/358) 和  $4/5$  (286/358)。

1. 隨機挑選出來的串列，做 Average Link (4.5 用 Affinity Propagation) 分群，

其中 Average Link 分 30 群。

2. 剩下未被取樣的串列，根據其與步驟 1 所分之群之平均距離，取最小者加入

該群中。

步驟 1~2 為一回合，重複 20 次。

3. 根據這 20 個回合的分群結果，計算任二串列  $i$  和  $j$ ，分至同一群的次數，建立 co-association matrix。

4. 用這個 co-association matrix，將這些串列用 Average Link 分 10~358 群。算出 ARI 和 NMI 之曲線。

步驟 1~4 為一循環，重複 10 次。

5. 取 10 個循環的 ARI 與 NMI 之曲線平均，即為實驗結果。

我們比較 Ensemble AL 和 Single Run AL 之間的表現。實驗結果如表 4-5：實驗發現只有取樣比例不超過 2/3 的 ARI 以及取樣比例 3/5 的 NMI，和 Single Run AL 相比有進步，所以我們認為這樣的方法，在取樣比例適中時可以得到較大的進步。ARI 和 NMI 都是取樣 3/5 表現最好，1/2 與 2/3 居次，3/4 與 4/5 最差。我們認為過高的取樣比例，在叢集整合實驗架構步驟 2 的分群效果較差，所以適中的取樣比例可以達到較好的分群效果。因為 LBP 串列在 Single Run AL 表現遠不如灰階串列，因此這裡我們只與 HOG 串列進行比較，發現 HOG 串列做 Ensemble AL 的表現不如表 4-5 所列的值(同一取樣比例之間做比較)。

表 4-5：經前處理 1 的灰階串列做 Ensemble AL 的實驗結果以及與 Single Run AL 之比較

取樣比例	ARI_2	NMI_2
1/2	0.318	0.656
3/5	0.327	0.666
2/3	0.315	0.662
3/4	0.284	0.645
4/5	0.270	0.629
Single Run AL	0.309	0.663

#### 4.5 叢集整合(Clustering Ensemble)的近鄰傳播分析

一般傳統的叢集整合演算法中，是採用 Average Link 作為每一回合抽樣分群的方法，而 Single Link 和 Complete Link 被驗證應用在叢集整合的表現不如 Average Link。那麼近鄰傳播是否也能應用在人臉串列叢集整合分群？這是本文想要探討的一個重要方向。

首先比較 Ensemble AP 與 Single Run AP 之間的表現，我們將均一參考度初值與相似度矩陣中位數做倍數調整，在這裡 Ensemble AP 僅採用與 Single Run AP 相同群數量的數值，實驗結果如表 4-6 與 4-7：表 4-6 可以看出，只有取樣比例 4/5 的分群表現有比 Single Run AP 好，因為 Single Run AP 採用均一參考度，產生的群數量較少，而 Average Link 在產生群數量較少時表現較差，所以我們認為在均一參考度系列，選用較高的取樣比例會有較好的表現。表 4-7 可以看出，只有採用相似度矩陣中位數做為參考度的三個取樣比例 ARI 比 Single Run AP 好，推測是因為該 Single Run AP 產生的群數量偏多，ARI 表現較差，若 Average Link 產生群數量稍微偏多，ARI 表現較好，所以 Ensemble AP 在這個條件下表現較好。



表 4-6: 經前處理 1 的灰階串列做 Ensemble AP(均一值做為參考度)的實驗結果以及與 Single Run AP(均一初值做為參考度)之比較

取樣比例	ARI_2(初值)	ARI_2(初值 4 倍)	NMI_2(初值)	NMI_2(初值 4 倍)
1/2	0.295	0.323	0.602	0.648
3/5	0.320	0.333	0.622	0.649
2/3	0.340	0.333	0.648	0.658
3/4	0.354	0.367	0.655	0.686
4/5	0.365	0.380	0.667	0.703
Single Run AP	0.356	0.370	0.670	0.700

表 4-7: 經前處理 1 的灰階串列做 Ensemble AP(相似度矩陣中位數之倍數做為參考度)的實驗結果以及與 Single Run AP(相似度矩陣中位數做為參考度)之比較

取樣比例	ARI_2(原值)	ARI_2(1/2)	NMI_2(原值)	NMI_2(1/2)
1/2	0.352	0.320	0.726	0.647
3/5	0.305	0.327	0.699	0.647
2/3	0.314	0.349	0.722	0.674
3/4	0.330	0.365	0.745	0.695
4/5	0.331	0.374	0.747	0.702
Single Run AP	0.318	0.380	0.750	0.719

接下來比較不同取樣比例與參考度做 Ensemble AP 的表現，實驗結果如表

4-8: 實驗發現用相似度矩陣中位數做為參考度，各項目的表現都是最差的，所以我們可以藉由調整 AP 的參考度來改進分群結果。ARI 的部分，選用均一參考度表現較佳，NMI 的部分，選用相似度矩陣中位數的二分之一做為參考度表現較佳。

表 4-8: 經前處理 1 的灰階串列以不同參考度做 Ensemble AP 的實驗結果, median(s) 表示相似度矩陣中位數, “初值” 表示均一參考度初值

取樣比例	3/5	3/4	4/5
ARI_2(median(s))	0.271	0.242	0.236
ARI_2(median(s)/2)	0.343	0.337	0.336
ARI_2(初值)	0.350	0.339	0.339
ARI_2(初值 4 倍)	0.335	0.340	0.338
NMI_2(median(s))	0.655	0.660	0.658
NMI_2(median(s)/2)	0.683	0.696	0.692
NMI_2(初值)	0.680	0.678	0.682
NMI_2(初值 4 倍)	0.676	0.691	0.692

最後來比較選用各種參考度的 Ensemble AP 與 Ensemble AL 的表現, 實驗結果如表 4-9: 實驗發現選用均一值或相似度矩陣中位數的二分之一做為參考度的 Ensemble AP, 分群結果都優於 Ensemble AL, 因此我們認為我們取用參考度的方法, 可以讓 Ensemble AP 的表現勝過 Ensemble AL。ARI 的部分, 選用均一參考度在取樣比例較高時進步較多, NMI 的部分, 選用相似度矩陣中位數的二分之一做為參考度在取樣比例較高時進步較多。

表 4-9: 經前處理 1 的灰階串列以不同參考度做 Ensemble AP 的實驗結果以及與 Ensemble AL 之比較, median(s)表相似度矩陣中位數, “初值” 表均一參考度初值, AL 列表示 Ensemble AL 之實驗結果

取樣比例	1/2	3/5	2/3	3/4	4/5
ARI_2(median(s))	0.315	0.271	0.246	0.242	0.236
ARI_2(median(s)/2)	0.323	0.343	0.338	0.337	0.336
ARI_2(初值)	0.348	0.350	0.349	0.339	0.339
ARI_2(初值 4 倍)	0.336	0.335	0.337	0.340	0.338
ARI_2(AL)	0.318	0.327	0.315	0.284	0.270
NMI_2(median(s))	0.682	0.655	0.651	0.660	0.658
NMI_2(median(s)/2)	0.675	0.683	0.691	0.696	0.692
NMI_2(初值)	0.678	0.680	0.687	0.678	0.682
NMI_2(初值 4 倍)	0.680	0.676	0.684	0.691	0.692
NMI_2(AL)	0.656	0.666	0.662	0.645	0.629

#### 4.6 近鄰傳播參考度與分群結果之關係

最後我們來總結：近鄰傳播實驗中參考度值、產生的群數量以及 ARI 和 NMI 之間的相應關係。

整體的實驗結果顯示：當取用的參考度較小時，會使得可能作為群中心的資料點較少，所以產生的群較少。當產生的群數量越接近基準的群數量時，ARI 大致呈現遞增，當產生的群數量越多，NMI 大致呈現遞增，這些現象與大多數的近鄰傳播實驗相同。



## 第五章 結論與未來展望

我們的實驗驗證了影像保留高頻及低頻部分時，HOG 特徵更能代表一張影像，並且對叢集整合近鄰傳播提出了改進方法，包括挑選均一的參考度與做參考度的倍數調整，使得分群結果進步。

同時我們也對 Single Run 做了深入的分析，指出 Average Link 和近鄰傳播分別適用於哪些條件的分群，近鄰傳播實驗中不同的參考度取法會對分群結果有什麼樣的影響，並對幾組不同的資料進行實驗，以驗證我們的論述。

本論文的主要貢獻，在於分析各種分群方法在哪些情況表現較好，哪些情況較不適用，並提出改進分群的措施。

人臉分群是現今研究廣泛的領域，未來也會是多媒體工程的主要發展方向。我們探討了一些增進人臉分群的方法，其中近鄰傳播採用均一參考度算是比較少見的觀點。最後我們想拋出一個問題：如果叢集整合近鄰傳播實驗選用“相似度矩陣最小的正數值”矩陣作為參考度，能否讓分群結果更好呢？因為近鄰傳播應用在叢集整合的研究，算是比較新穎的方向，所以我們的這個猜測，或許不久之後會有相關的新發現問世。



## 參考文獻

- [1] M. Everingham and A. Zisserman, "Automated visual identification of characters in situation comedies," *Proceedings of the 17th IEEE International Conference on Pattern Recognition*, 2004, vol.4., pp. 983-986
- [2] M. Everingham and A. Zisserman, "Automated person identification in video," in *Image and Video Retrieval*, ed:Springer Berlin Heidelberg, 2004, pp. 289-298.
- [3] Y. Hu, A. S. Mian and R. Owens, "Face recognition using sparse approximated nearest points between image sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, vol. 34.10, pp. 1992-2004.
- [4] Y. Hu, A. S. Mian and R. Owens, "Sparse approximated nearest points for image set classification," *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 121-128.
- [5] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586-591.
- [6] X. Liu and T. Chen, "Video-based face recognition using adaptive hidden markov models," *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003, vol.1, pp. I-340-I-345
- [7] W. S. Chu, J. C. Chen and J. J. J. Lien, "Kernel discriminant transformation for image set-based face recognition," *Pattern Recognition*, 2011, vol. 44.8, pp. 1567-1580,.
- [8] M. T. Harandi, A. Bigdeli and B. C. Lovell, "Image-set face recognition based on transductive learning". *IEEE International Conference on Image Processing (ICIP)*, 2010, pp. 2425-2428.
- [9] T. Ahonen, A. Hadid and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, vol. 28.12, pp. 2037-2041.
- [10] T. Ahonen, A. Hadid and M. Pietikainen, "Face recognition with local binary patterns", *Computer Vision*, ed:Springer Berlin Heidelberg, 2004, pp. 469-481.
- [11] H. Cevikalp and B. Triggs, "Face recognition based on image sets," *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2567-2573.
- [12] J. Yang, D. Zhang, A. F. Frangi and J. Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, vol. 26.1, pp. 131-137.
- [13] K. Yamamoto, O. Yamaguchi and H. Aoki, "Fast face clustering based on shot similarity for browsing video," *Proceedings of the Progress in Informatics*,

*Special issue: 3D image and video technology*, 2010, vol. 7, pp. 53-62.

- [14] Q. V. Le, W. Y. Zou, S. Y. Yeung and A. Y. Ng, "Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis," *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 3361-3368.
- [15] E. El Khoury, C. Senac and P. Joly, "Face-and-clothing based people clustering in video content," *Proceedings of the ACM international conference on Multimedia information retrieval*, 2010, pp. 295-304.
- [16] L. Zhang, D. V. Kalashnikov and S. Mehrotra, "A unified framework for context assisted face clustering," *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, 2013, pp. 9-16.
- [17] W. T. Chu, Y. L. Lee and J. Y. Yu, "Using context information and local feature points in face clustering for consumer photos," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 1141-1144.
- [18] P. Huang, Y. Wang and M. Shao, "A new method for multi-view face clustering in video sequence," *IEEE International Conference on Data Mining Workshops*, 2008, pp. 869-873.
- [19] J. Tao and Y. P. Tan, "Efficient clustering of face sequences with application to character-based movie browsing," *IEEE International Conference on Image Processing*, 2008, pp. 1708-1711.
- [20] S. Kayal, "Face Clustering Experiments on News Video Images," *Proceedings of the International Conference on Intelligent and Automation Systems, Vietnam*, 2013, pp. 213-216.
- [21] S. Kayal, "Face clustering in videos: GMM-based hierarchical clustering using spatio-temporal data," *IEEE UK Workshop on Computational Intelligence*, 2013, pp. 272-278.
- [22] T. Zhang, D. Wen and X. Ding, "Person-based video summarization and retrieval by tracking and clustering temporal face sequences," *International Society for Optics and Photonics on IS&T/SPIE Electronic Imaging*. 2013, pp. 866400-866400-6.
- [23] S. Foucher and L. Gagnon, "Automatic detection and clustering of actor faces based on spectral clustering techniques," *IEEE Canadian Conference on Computer and Robot Vision*, 2007, pp. 113-122.
- [24] G. Orfanidis, N. Nikolaidis and I. Pitas, "Facial image clustering in single channel and stereo video content," *IEEE International Workshop on Biometrics and Forensics*, 2013, pp. 1-4.
- [25] C. Czirjek, N. E. O'Connor, S. Marlow and N. Murphy, "Face detection and clustering for video indexing applications," 2003.

- [26] J. Cui, F. Wen, R. Xiao, Y. Tian and X. Tang, "EasyAlbum: an interactive photo annotation system based on face clustering and re-ranking," *Proceedings of the ACM SIGCHI conference on Human factors in computing systems*, 2007, pp. 367-376.
- [27] Z. Li and X. Tang, "Bayesian face recognition using support vector machine and face clustering," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II-374.
- [28] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, 2007, vol. 315.5814, pp. 972-976.
- [29] K. Wang, J. Zhang, D. Li, X. Zhang and T. Guo, "Adaptive affinity propagation clustering," *arXiv preprint arXiv:0805.1096*, 2008.
- [30] Z. Lu and M. A. Carreira-Perpinan, "Constrained spectral clustering through affinity propagation", *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.
- [31] I. E. Givoni and B. J. Frey, "A binary variable model for affinity propagation," *Neural computation*, 2009, vol. 21.6, pp. 1589-1600.
- [32] A. Strehl and J. Ghosh, "Cluster ensembles-a knowledge reuse framework for combining partitionings," *AAAI/IAAI*, pp. 93-99, 2002.
- [33] A. Topchy, B. Minaei-Bidgoli, A. K. Jain and W. F. Punch, "Adaptive clustering ensembles," *Proceedings of the 17th IEEE International Conference on Pattern Recognition*, vol. 1, pp. 272-275.
- [34] B. Minaei-Bidgoli, A. Topchy and W. F. Punch, "Ensembles of partitions via data resampling," *Proceedings of the IEEE International Conference on Information Technology: Coding and Computing*, 2004, vol. 2, pp. 188-192.