

國立臺北大學資訊工程學系110學年度專題

相似系列圖片自動整理系統

簡介

現在大部分使用者對於圖片的使用習慣，常常是將許多圖片儲存到同一資料夾中卻無法將其好好分類，例如會將偶像的重複照片存到裝置內的追星族，或是在日常對話中會加入梗圖的梗圖使用者，都會有圖片過於雜亂且未經過整理而找不到圖的窘境。因此我們以這類使用者客群為出發點，做出一個在電腦端執行的相似系列圖片整理系統。

問題與目的

使用者的裝置裡不乏有以下幾種類型的多張圖片，而我們系統的主要功能就是將這些性質相同的圖片分類到一起。

一、部分相同或經裁切



二、人物背景、姿勢幾乎一樣



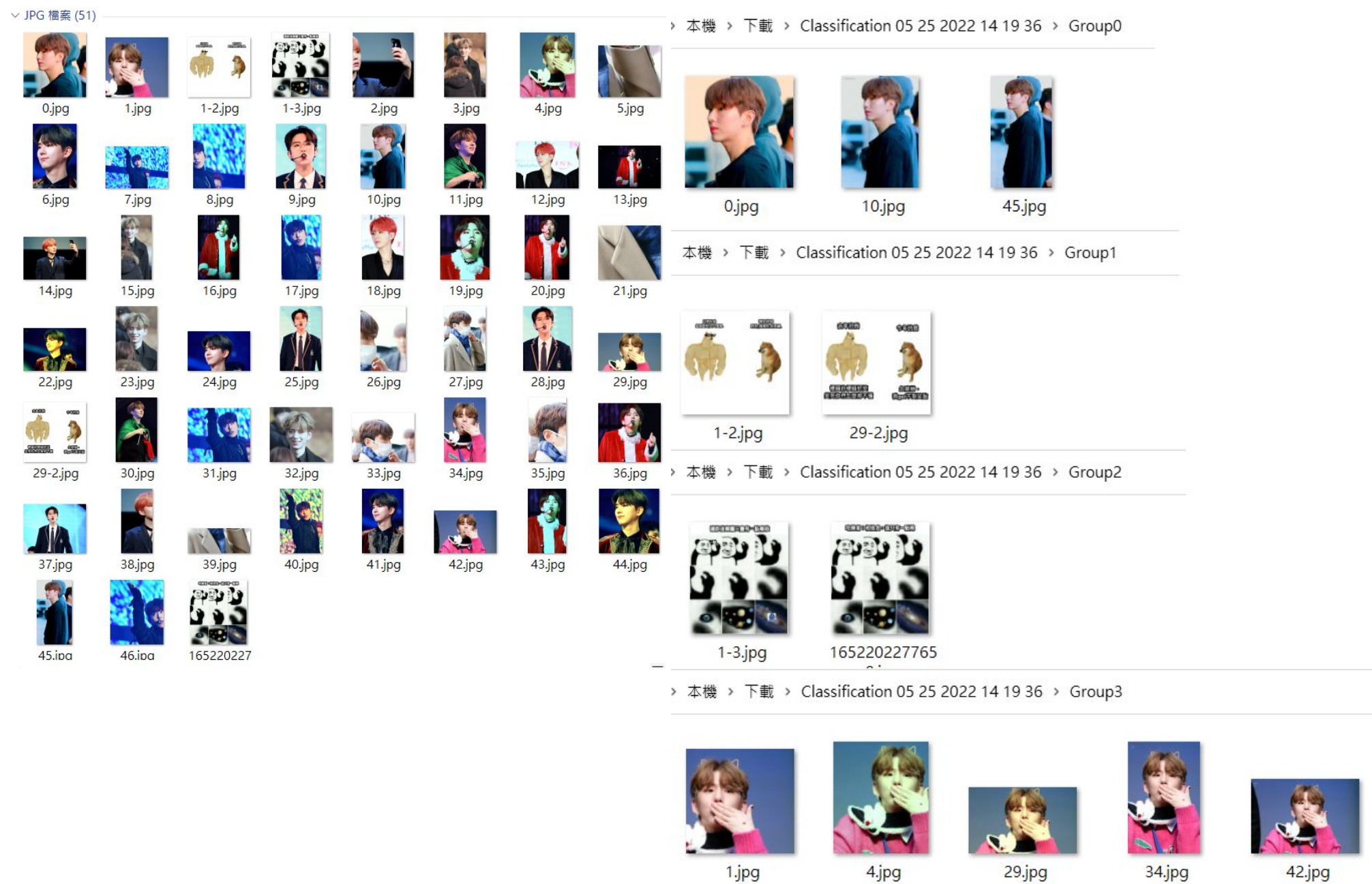
三、調整過色調、濾鏡



四、有模板的、文字代換過的梗圖



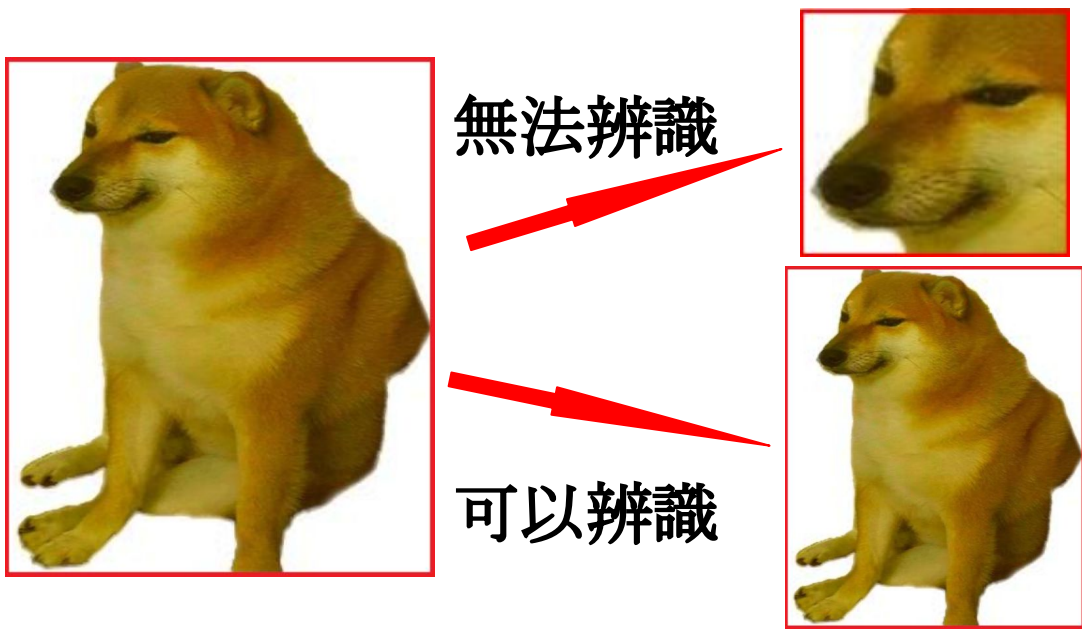
成果:能將指定資料夾內混亂的圖片自動分於各系列資料夾



一般方法

▶兩圖檢測

AHash 算法：
只能辨識相同區塊、
比例的圖片。



SIFT演算法:通常用在找出圖片中關鍵點，做拼接圖片。



▶圖庫檢測

因相似系列圖片間的多樣關係，通常使用暴力法將圖片以兩兩對比，共需進行 C_2^N 次，相當耗時，且會有圖片間相似度的問題。

方法概述



▶兩圖檢測

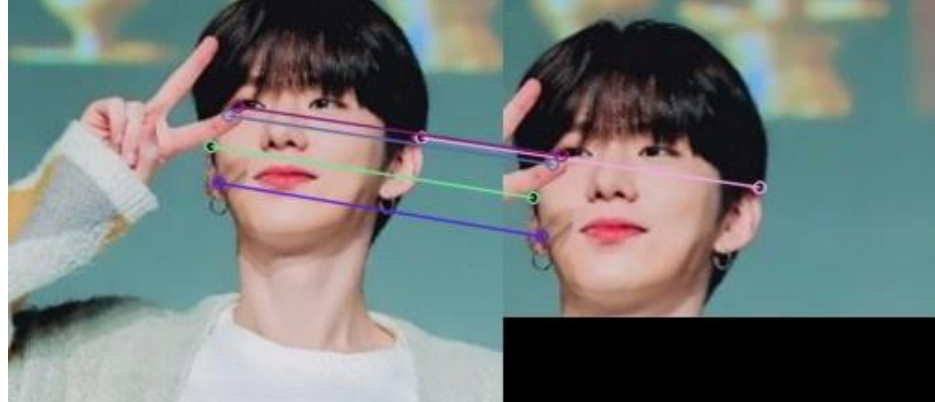
以兩張部分相同的圖片為例。

0. 做Size-Reduce，將長邊超過1K/2K/4K的圖片縮小2/4/8倍

1. 用SIFT找出所有關鍵點



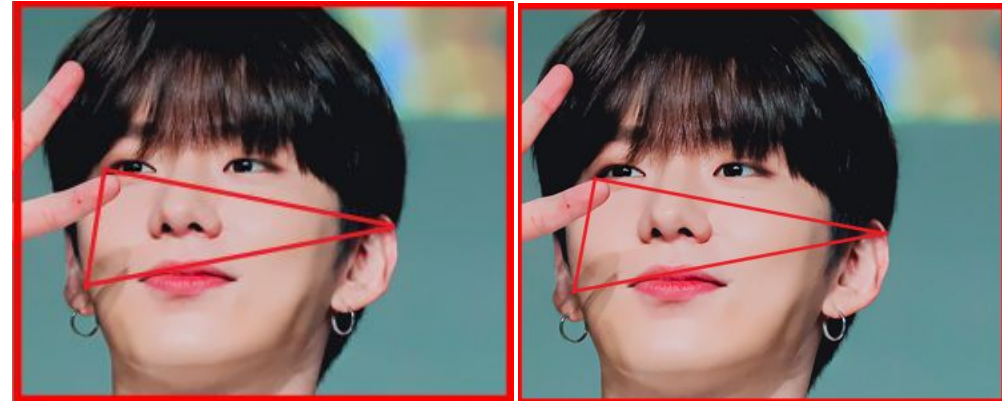
2. 找出最相近的五個關鍵點



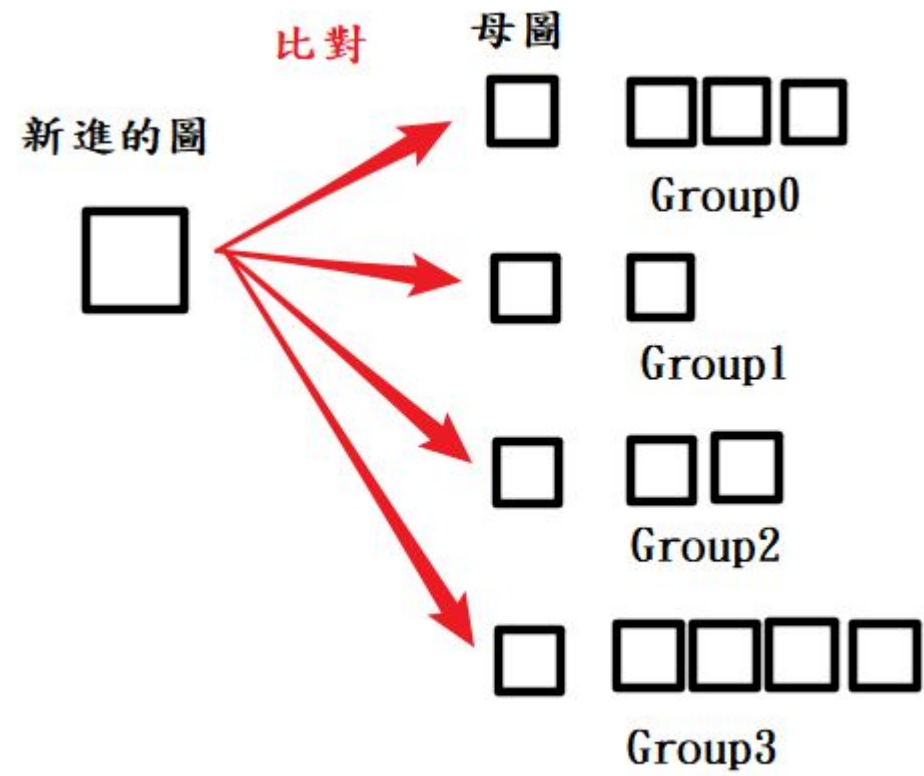
3. 找相似三角形，按比例縮放



4. 截取三角形延伸出去之方形區塊



5. 將截取之方形區塊以aHash比對。若其漢明距離小於12，兩原始圖片視為『相似系列圖片』。將大圖視為『母圖』，以改善圖庫檢測效率。



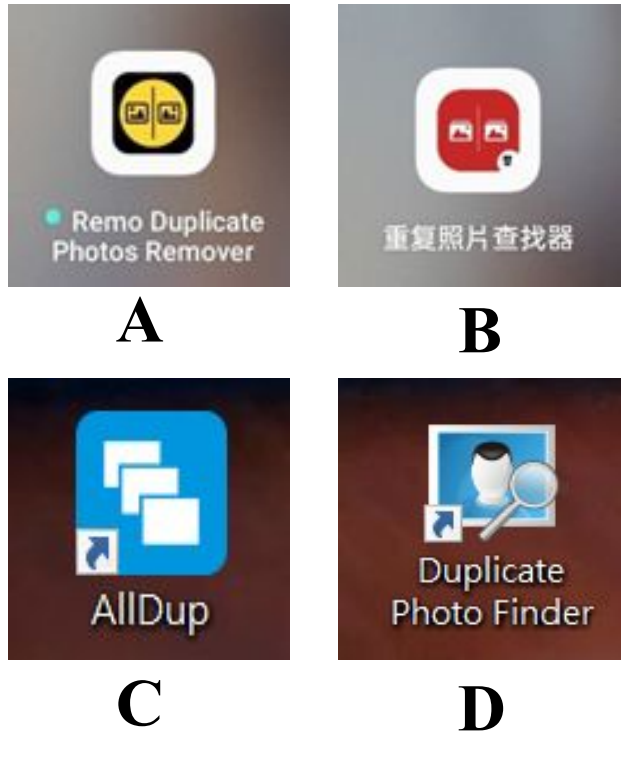
▶圖庫檢測

1. 新進圖依序與『母圖』比對。
2. 比對成功，加入該群，若新進圖較原母圖大，更新新進圖為新母圖；皆比對失敗，新進圖自成一群，且為該群母圖。
3. 若一群母圖發生更新，比對新母圖與後續群母圖之關係，若檢測為相似系列圖片，兩群合併。

實驗結果對比

以100組成對照片測試：

使用軟體	原圖比對	經裁剪圖比對	經調色圖比對	調色&裁剪圖比對
A	63	0	21	0
B	77	0	24	0
C	90	0	88	0
D	0	0	0	0
ours	100	100	99	98



$$MR = \frac{\#G1 \cap T0}{\#G1}$$
$$IMR = \frac{\#G0 \cap T1}{\#G0}$$

若兩張圖片應為一組/群，其關係以布林值1表示，反之，以布林值0表示。以G表示實際關係，T表示檢測關係。

以167張41群照片測試：

	Off-Time	On-Time	MR	IMR
A	3.475s	0.075s	99%	0.4%
B	54s	3402s	52%	10.1%
C	31s	1066s	62%	7.3%
D	52s	3126s	20%	0.42%
E	61s	3720s	62%	0.5%
F	30s	1087s	66%	0.49%
ours	31s	118s	0.35%	0.022%

* A ~ F 皆採暴力法做圖庫檢測。

兩圖檢測：

- A - aHash
- B - Ours w/o 0, r/match-template 4&5
- C - Ours r/match-template 4&5
- D - Ours w/o 0
- E - Ours w/o 0, r/match-template 4
- F - Ours r/match-template 4