

# 國立交通大學

## 機械工程學系

### 碩士論文



應用等階集合法手勢操控系統的研究與實作  
The study and application of level set method on  
gesture control system

研究生：余涵宇

指導教授：鄭璧瑩 博士

中華民國一百零二年七月

應用等階集合法手勢操控系統的研究與實作  
The study and application of level set method on  
gesture control system

研究生：余涵宇

Student : Han-Yu Yu

指導教授：鄭璧瑩 博士

Advisor : Dr. Pi-Ying Cheng

國立交通大學

機械工程學系

碩士論文

A Thesis

Submitted to Department of Mechanical Engineering  
College of Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements  
for the Degree of  
Master  
In

Mechanical Engineering

July 2013

Hsinchu, Taiwan, Republic of China

中華民國一百零二年七月

# 應用等階集合法手勢操控系統的研究與實作

學生：余涵宇

指導教授：鄭璧瑩 博士

## 摘要

本論文建立一個可以辨識靜態手勢，並進一步控制機械手臂的系統，全部過程可分為取出手勢、學習及手勢辨識。

在取出手勢的步驟中，先利用連續影像相減找出變動的物體部分，再以此為影像分析的主要區域(ROI, Region of interest)，進行應用等階集合法的手勢分割，以提高效率。為了手勢分割較完整，本論文提出新型的影像切割法則，其所定義的曲面隱函數為能量模型採用的參數，依循零階集合(Level set zero)移動後取得內外圈的資訊作為邊緣資訊，再利用尤拉格朗日方程式(Euler-Lagrange equation)與梯度下降法(Gradient descent method)建立隱函數演化式，使初始輪廓收斂到手掌邊緣，完成手勢影像切割的目的。為了摒除手掌角度的影響，手勢分割後再利用手掌的慣性積矩陣，利用特徵值分解找出慣性積矩陣的特徵向量，找到主軸並完成手掌轉正的程序。最後為了消除較不重要的特徵，再利用傅立葉描述子，將輪廓座的點標作為參數轉至頻率空間，並取前 40 項係數再反轉換回座標空間，使手勢輪廓較平滑，消除高頻雜訊部分。

在學習步驟中，本論文針對每種手勢事先蒐集 300 張訓練手勢影像，並以每張訓練手勢的形心為起點，向手掌輪廓劃分 360 度，每隔一度取距離並標準化作為特徵。將每張的手掌影像特徵蒐集成資料庫並做奇異值分解，取出前 50 項最重要的特徵，並保留投射矩陣 (projection matrix)，待欲辨識特徵與之相乘。

在線上(on-line)辨識時，在手勢影像分割出後，取出距離特徵並利用主成份分析降低特徵維度，從資料庫中尋找與之最相近的影像，即達到辨識的目的。

最後本論文將本系統應用在滑鼠游標控制上，並與本實驗室研究計畫的伺服馬達所發展之機械手臂做連結，使用者將可以非接觸式控制伺服馬達動作，進而方便遙控遠端設備。此系統辨識之速度與正確率足可滿足非接觸控制機械手臂的目的。

關鍵字：手勢辨識、等階集合法、傅立葉描述子、奇異值分解、主成份分析

# The study and application of level set method on gesture control system

Student: Han-Yu Yu

Advisor: Dr. Pi-Ying Cheng

## ABSTRACT

This study establishes a system that can recognize ten different gestures. The total process can be separated into three steps: getting gesture, learning and recognizing.

In getting gesture step, we first use frame differencing method to find the region with pixel data changed, the area is called ROI(region of interest) for applying level set method to segment gesture in image. In order to segment more complete shape of gesture, this study establishes a new model of level set algorithm for image segmentation. This method use an implicit function as a variable of a defined energy model, and according to the movement of zero level set to get interior and exterior information and determine the edge position. This study use Euler-Lagrange equation and gradient descent method to get evolution of implicit function. We accomplish this step as soon as the zero level set converges on the edge of hand shape with specified gesture.

In order to avoid the influence from the hand angle, this study get inertia matrix of hand and use eigenvalue/eigenvector decomposition to get direction of principle axis, then rotating hand on the vertical plane. Finally, for the purpose of eliminating unimportant features, we use Fourier descriptor to afford a more smooth hand shape.

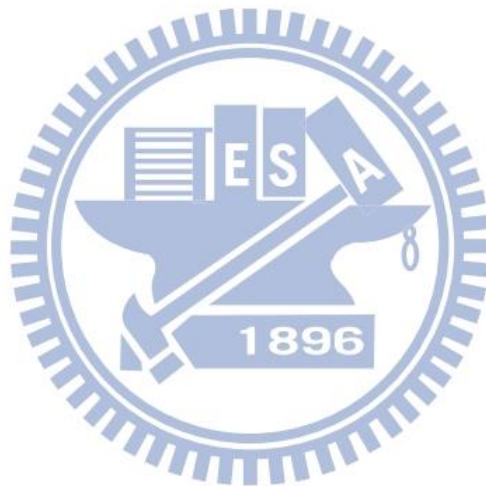
In learning step, this study collect about sample 300 frames for each gesture, and define the distance between centroid and each edge point as features, then use singular value decomposition and retain projective matrix to increase the recognition efficiency.

When the system works on-line, it uses principle value analyze to reduce dimension of feature matrix, the most similar image is defined as the least average distance one, to reach for the result of pattern recognition.

Finally, this study applies this system to control cursor, and combines this system with CIDM robot arm. Users can only use hand gestures to control the action of the arm.

The result of some typical experimental cases show that the proposed method and hand gesture recognition based

control system is successful in the application of remote control system.



**Keywords:** gesture recognition ,level set, Fourier descriptor, singular value decomposition(SVD) , principle value analyze(PCA)

## 誌 謝

本論文經不斷地討論與編修後，終於得以付梓，此時我首先必須感謝我的指導教授—鄭璧瑩博士。我在兩年的碩班研究期間中，每次與老師的討論中都能獲益不少，無論是做研究的態度或思考的方法，老師都會孜孜不倦的教導並引領我們；而在撰寫論文時，老師繁忙之中還願意多次撥冗與我討論與更改內容，這也是我特為感激的地方。

另外我在此特別感謝 CIDM 實驗室中的各成員：學長姐：嗣展、承憲、仲翔、惠君，同學新喆、晏樟、國霖，學弟傳之、皓儀、鍾宇、竑廷。謝謝大家在這兩年間的陪伴，無論是課業與論文上的幫助，或是研究之餘放鬆心情的時候，都使我感到並不孤單。另外我特別感激我的同學兼月下老人—呈祥，沒有您的搭訕舉動與牽線，我就不會認識我現在的女朋友。

再來是我的女友佩芳，雖然妳對論文內容是零貢獻，但是！很感謝妳在我碩班期間的陪伴，尤其在我低潮的時候，幫我加油打氣，給了我不少鼓勵與正面的能量，謝謝妳。

最後必須感謝我的家人，從我小時候到現在即將畢業，你們一直是我背後最強而有力的推手：經濟上總是讓我無後顧之憂，生活上總是一直關心我，而當我遇到學業或生活上的困境時，你們總會一起提

供經驗，幫助我解決當下的難題。我想對你們說：謝謝你們二十幾年來的辛勞，沒有你們就沒有這篇論文，更沒有現在的我。

我要感謝的人太多無法一一列舉，畢竟得之於人者太多；但你們的幫助我一定會銘記在心，謝謝！

余涵宇 謹誌

2013.7 交通大學





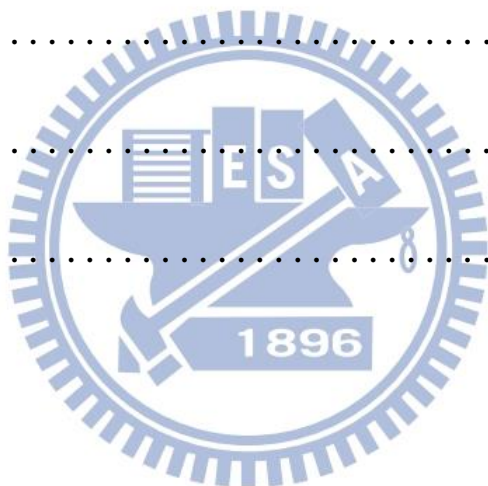
## 目錄

摘要.....	I
ABSTRACT.....	III
誌謝.....	V
目錄.....	VII
第一章 緒論 .....	1
1.1 研究背景與動機 .....	1
1.2 文獻回顧 .....	2
1.2.1 基於分類的切割方法 .....	2
1.2.2 Chan-Vese 切割模型 .....	3
1.2.3 粒子濾波器(Particle filter) .....	5
1.2.4 支持向量機(Support Vector Machine ,SVM) .....	6
1.2.5 倒傳遞類神經網路(Back Propagation Neural Network ) .....	7
1.2.6 均值移動(Mean shift) .....	9
1.2.7 近年手勢辨識方法 .....	11

第二章 研究理論與方法 .....	17
2.1 <b>YCbCr</b> 色彩空間.....	17
2.2 連續影像相減法(Frame differencing) .....	18
2.3 主動式輪廓(Active contour) .....	20
2.4 等階集合法(Level set method) .....	22
2.4.1 邊界(Interface) .....	23
2.4.2 距離函數(Distance function).....	25
2.4.3 移動水平集 .....	27
2.5 最佳設計 .....	28
2.5.1 梯度下降法(Gradient descent method) .....	29
2.5.2 歐拉-拉格朗日方程式(Euler Lagrange equation) ...	30
2.6 手掌轉正.....	31
2.6.1 影像矩(Moment) .....	31
2.6.2 質量慣性矩(Moment of inertia).....	32
2.6.3 質量慣性積(Product of inertia).....	32
2.6.4 轉主軸(Principal axes).....	33
2.6.5 旋轉矩陣(Rotation matrix) .....	35

2.7 傅立葉描述子(Fourier descriptors).....	36
2.8 奇異值分解(Singular value decomposition).....	38
2.9 主成份分析(Principal component analysis).....	39
第三章 手勢辨識系統設計 .....	44
3.1 手掌分割 .....	44
3.1.1 連續背景影像相減 .....	44
3.1.2 主動式輪廓模型推導 .....	45
3.1.3 主動式輪廓模型切割手勢 .....	50
3.2 手掌轉正.....	52
3.3 傅立葉描述子 .....	54
3.4 蒐集訓練影像 .....	55
3.5 以 SVD 轉換訓練蒐集影像特徵 .....	56
3.6 利用主成份分析辨識手勢 .....	58
第四章 實驗分析與應用 .....	62
4.1 輸入手勢.....	63
4.2 辨識速度 .....	64
4.3 手勢分割實驗.....	65

4.4 特徵維度對辨識率的影響 .....	67
4.5 其他受試者的辨識率 .....	72
4.6 手勢控制滑鼠游標 .....	74
4.7 手勢控制機械手臂 .....	75
4.8 實驗結論.....	85
第五章 結論與未來目標 .....	87
5.1 結論.....	87
5.2 未來展望 .....	88
參考文獻.....	90



# 第一章 緒論

## 1.1 研究背景與動機

近年來由於電腦硬體以及影像處理技術的長足進步，影像辨識技術已廣泛出現在人們的日常生活中，且確實已能應用在各個領域，例如車牌辨識、人臉辨識和指紋辨識，目前我們已利用電腦視覺代替人們解決這些過程較單一且不斷重複的事情。而在辨識技術中，手勢辨識是較熱門的研究主題之一。一般手勢辨識可分為靜態手勢辨識與動態手勢辨識，前者目的為找出手勢的形狀，並判別出手勢代表意義；後者為判斷手掌移動的軌跡方向再進行動作的辨識。透過電腦精準的手勢辨識，我們可以不用任何接觸式的控制器，或配戴任何物件，而以直覺的方式，達到人機互動的目的。由於手勢辨識不但省時省力，亦可廣泛應用於各種平台上，例如遊樂器、機器人、電視等，所以勢必會成為未來主流的人機互動方式之一。

目前手勢辨識技術遇到較困難的問題為，在較複雜的場景較難正確找出手勢位置並正確分割出來，因此造成辨識上的困難。

本論文目標為靜態且基於形狀的手勢辨識，而為了解決上述難題，試圖應用較強健的影像切割技術，因為一旦切割後的手勢形狀較完整，正確的辨識便將變得相對容易許多。

## 1.2 文獻回顧

靜態手勢辨識的步驟可大略分為手勢分割或追蹤、辨識，以下介紹目前常用的幾種方法：

### 1.2.1 基於分類的切割方法

以影像中各點的灰階值做分類的依據，在此介紹常用的 K means 方法，可把影像分成 K 類區域，即是切割出 K 塊區域。其演算法如下

1. 設定分成 K 類，並隨機或主動選 K 個點當各類的初始中心 ( $X_1 \sim X_k$ )。
2. 其它各點算 1~K 類中心的距離，若該點到  $X_m$  的距離最短則把該點分到 m 類。
3. 各點分類完畢，重新計算各類的平均值當作中心，重複步驟 2。
4. 算到各點中心不再變動為止(即收斂)，分類結束。

K means 可保證收斂，但是其分類結果不一定是最佳解，初始中心位置及 K 值都會影響分類效果及迴圈次數。若使用 K means 執行影像分

割，我們可用各點的灰階值做分類的依據，分出  $K$  塊有區別的區域。

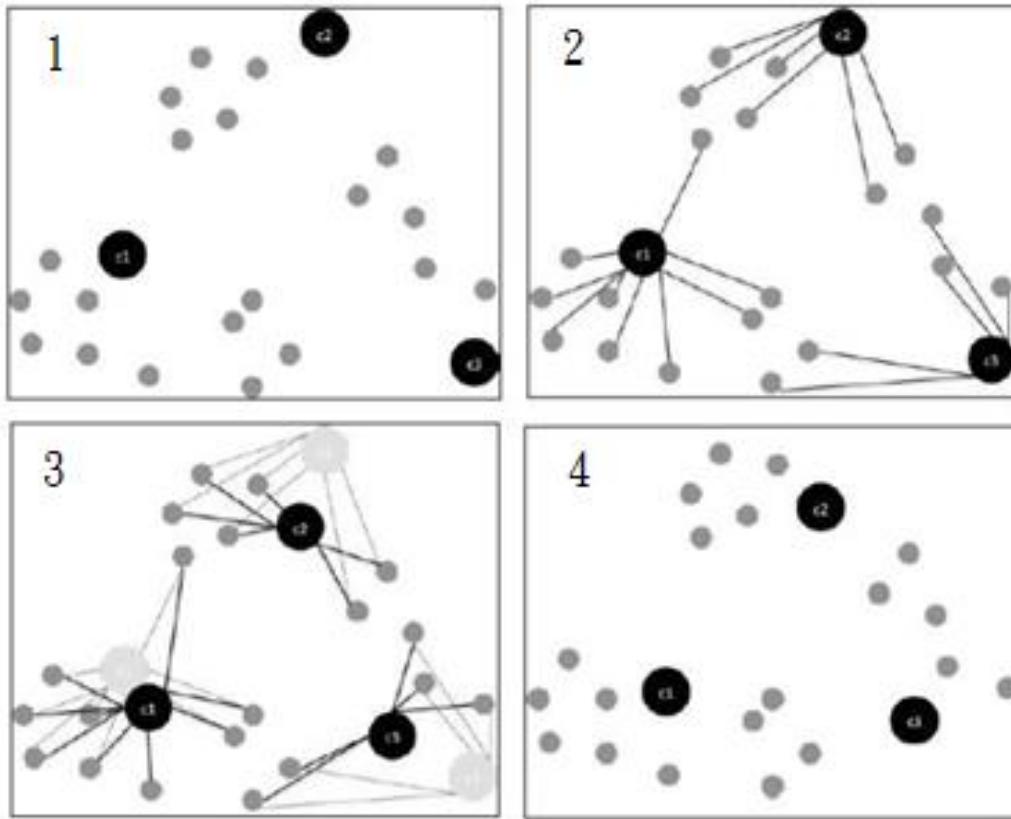


圖 1.1 k means 的收斂過程(1→2→3→4)

### 1.2.2 Chan-Vese 切割模型

Tony Chan 與 Luminita Vese[6]提出的主動式輪廓模型：

Chan-Vese Model。這個模型以水平集(Level set zero)作為參數，並參考影像中區域的資訊定義能量，有別於在此之前其他需要影像梯度作為參數的主動式輪廓的能量模型。

定義影像空間為  $\Omega$ ，曲線  $C$  為  $\Omega$  的子空間  $\omega$  的邊界( $\omega \subset \Omega$ ， $C = \partial\omega$ )，接下來我們以  $\text{inside}(C)$ 表示區域  $\omega$ ，以  $\text{outside}(C)$ 表示

區域  $\Omega \setminus \bar{\omega}$ 。Chan-Vese 模型定義一能量函數並使之最小化。假設影像  $u_0$  由兩個區域構成，灰階值為片段連續，能以  $u_0^i$  和  $u_0^o$  表示這兩個區域的灰階值。為了說明這個模型中 Fitting term 能量的含意，接著考慮

$$F_1(C) + F_2(C) = \tag{1.1}$$

$$\int_{\text{inside}(C)} |u_0(x, y) - C_1|^2 dx dy + \int_{\text{outside}(C)} |u_0(x, y) - C_2|^2 dx dy$$

其中  $C_1$  與  $C_2$  為曲線  $C$  內、外區域的平均灰階值。考慮四種情況(圖 1.2)

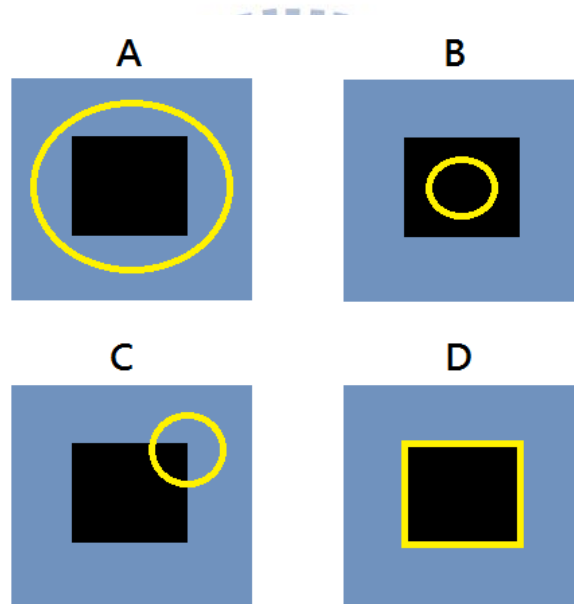


圖 1.2 黃色封閉曲線為  $C$ ，只有當右下情形時，Fitting  $\approx 0$

顯而易見地，在圖 1.2 中的 A、B、C 三種情況中， $F_1(C)$  或  $F_2(C)$  會大於零，只有在 D 時  $F_1(C) + F_2(C) = \text{Fitting} \approx 0$ 。因此我們若最小化 Fitting term 的能量，可以得到在前景邊緣的  $C$

$$\text{Inf}\{F_1(C) + F_2(C)\} \approx 0 \approx F_1(C_0) + F_2(C_0) \tag{1.2}$$



除了 Fitting term 外，再考慮曲線的長度、梯度及其所包圍的面積，則可定義出 Chan-Vese 的能量函數  $F(C_1, C_2, C)$

$$F(C_1, C_2, C) = \mu \cdot \text{Length}(C) + \nu \cdot \text{Area}(\text{inside}(C)) \quad (1.3)$$
$$+ \lambda_1 \int_{\text{inside}(C)} |u_0(x, y) - C_1|^2 dx dy$$
$$+ \lambda_2 \int_{\text{outside}(C)} |u_0(x, y) - C_2|^2 dx dy$$

其中  $\mu \geq 0$ ， $\nu \geq 0$ ， $\lambda_1, \lambda_2 > 0$ ，都為權重。

利用等階集合法表示(1.3)後，便能利用 Euler-Lagrange equation 找出  $F$  的梯度方向  $\nabla F$ ，再以此反方向  $-\nabla F$  和定義時間  $t$ ，帶入梯度下降法，得到輪廓的演化式，而當輪廓演化到物體邊緣上時，能量值就會最小，即停止收斂。



### 1.2.3 粒子濾波器(Particle filter)

粒子濾波器是基於貝氏機率的預測模型，使用影像中的顏色及影像紋理為特徵，學習動態模型後找出事前機率(prior probability)，再利用測量模型，產生一樣本集合，估測事後機率，以此相似性決定下一個時間的狀態，並更新目標物體的模型。

應用在追蹤物件時，藉由視訊中目標物件的附近，隨機設置  $N$  個粒子(particle)，且每一個粒子皆為一個樣本(sample)集合。每一個

粒子代表著一個樣本集中的狀態向量，利用不同的測量方法，給予每個粒子不同的權重。

1. 取樣:測量步驟後，每一個粒子會有不同的權重及不同機率分布，取樣目的為將機率密度較小的粒子去除，並將每一個例子機率密度值一致。

2. 預測:追蹤物件的過程中，物件可能會隨著時間變化而改變移動方向或移動速度，因此預測的主要目的是為了估測出每一個粒子移動後可能的位置。

3. 測量:以物件特徵(ex:顏色)給予每一個粒子相對應的權重，初始化時，產生粒子並建立樣板。



#### 1.2.4 支持向量機(Support Vector Machine ,SVM)

SVM 目的在找出兩群資料間的分界，以區分兩類。對於一群在空間中的樣本，我們希望能夠在該空間之中找出一 Hyper plan(或 decision boundary)，並且，希望此 Hyper plan 可以將這群資料切成兩群，即兩類分別在 Hyper-plan 的兩端，沒有同時在兩端的情形，如圖 1.3

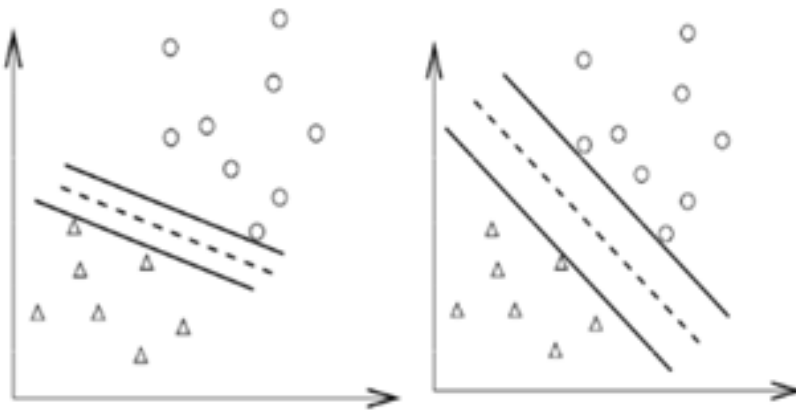


圖 1.3 虛線為  $w^T x + b = [+1 \ 0 \ -1]^T$  的決策邊界

由圖 1.3 可看出，左圖兩點與 decision boundary 平行的切線間距比右圖兩點與 decision boundary 平行的切線間距小，雖然都能分成兩類，但由於右圖分得較開，故為較理想的 decision boundary。

### 1.2.5 倒傳遞類神經網路 (Back Propagation Neural Network)

倒傳遞類神經網路為監督式的學習，讓電腦了解輸入與輸出的關係，進而達到辨識的作用。其可分為兩個階段，學習：找到合適的權重讓輸入特徵可以計算出目標結果；分類：利用學習完的特徵計算結果。

類神經網路為了模仿人腦神經元的工作，其模式設計如圖 1.4，由圖 1.4 可知神經元先將輸入各自乘上一個連接權值，再將所有的乘積加總，最後由活化函數決定臨界值是否被輸出。

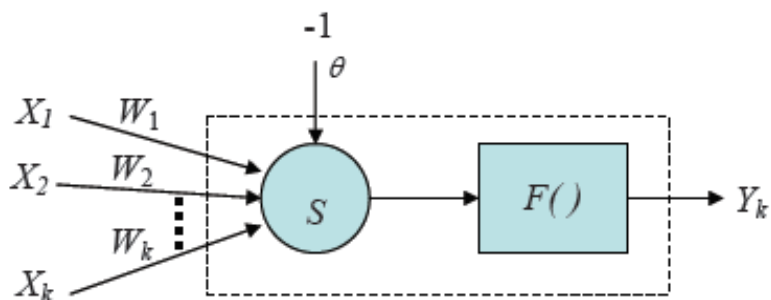


圖 1.4 單輸入單輸出類神經網路

倒傳遞類神經網路的基本架構包括輸入層、隱藏層和輸出層，其中隱藏層與外界資訊隔離，只透過輸入層與輸出層作為對外的介面，用來表現輸入單元間的交互影響，網路可以不只一層隱藏層，也可以沒有隱藏層，視問題的複雜程度而決定其網路架構，如圖 1.5 所示。

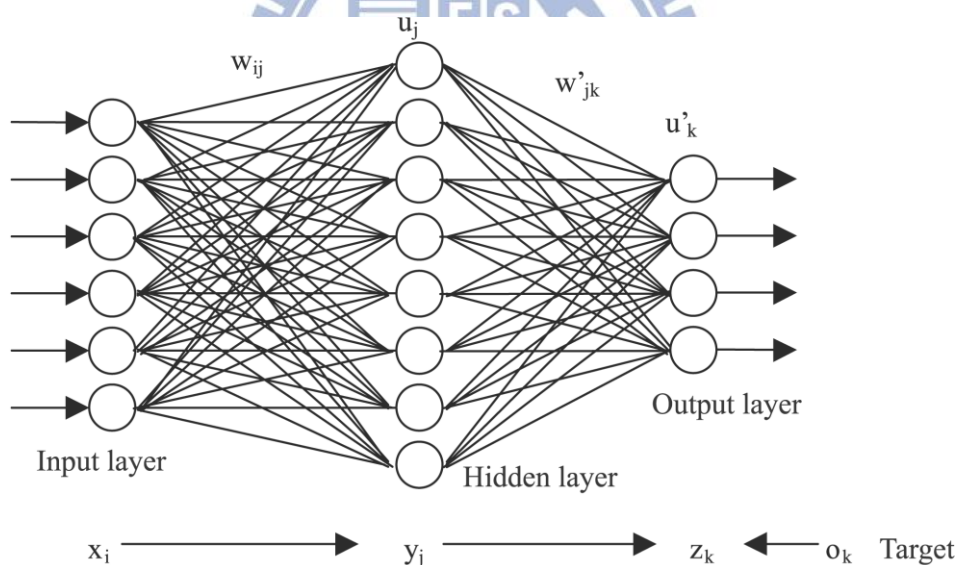


圖 1.5 類神經網路結構

在學習階段，我們可輸入資料且設定對應的輸出，再計算誤差再回傳，反覆修正權重  $w_{ij}$ 。辨識時再將輸入資料乘上利用學習時所得到的權重，得到輸出。

### 1.2.6 均值移動(Mean shift)

Mean shift 為一種追蹤物體的方法，以迭代的方式，先算出當前的偏移值，以該偏移值移動該點，再重新計算偏移值。大略的步驟如下

- (1) 選擇任意區域，窗口大小任意設定
- (2) 計算窗口內的質心位置
- (3) 調整窗口中心到質心位置
- (4) 重複 2 到 3，直到收斂為止

過程如下圖所示。



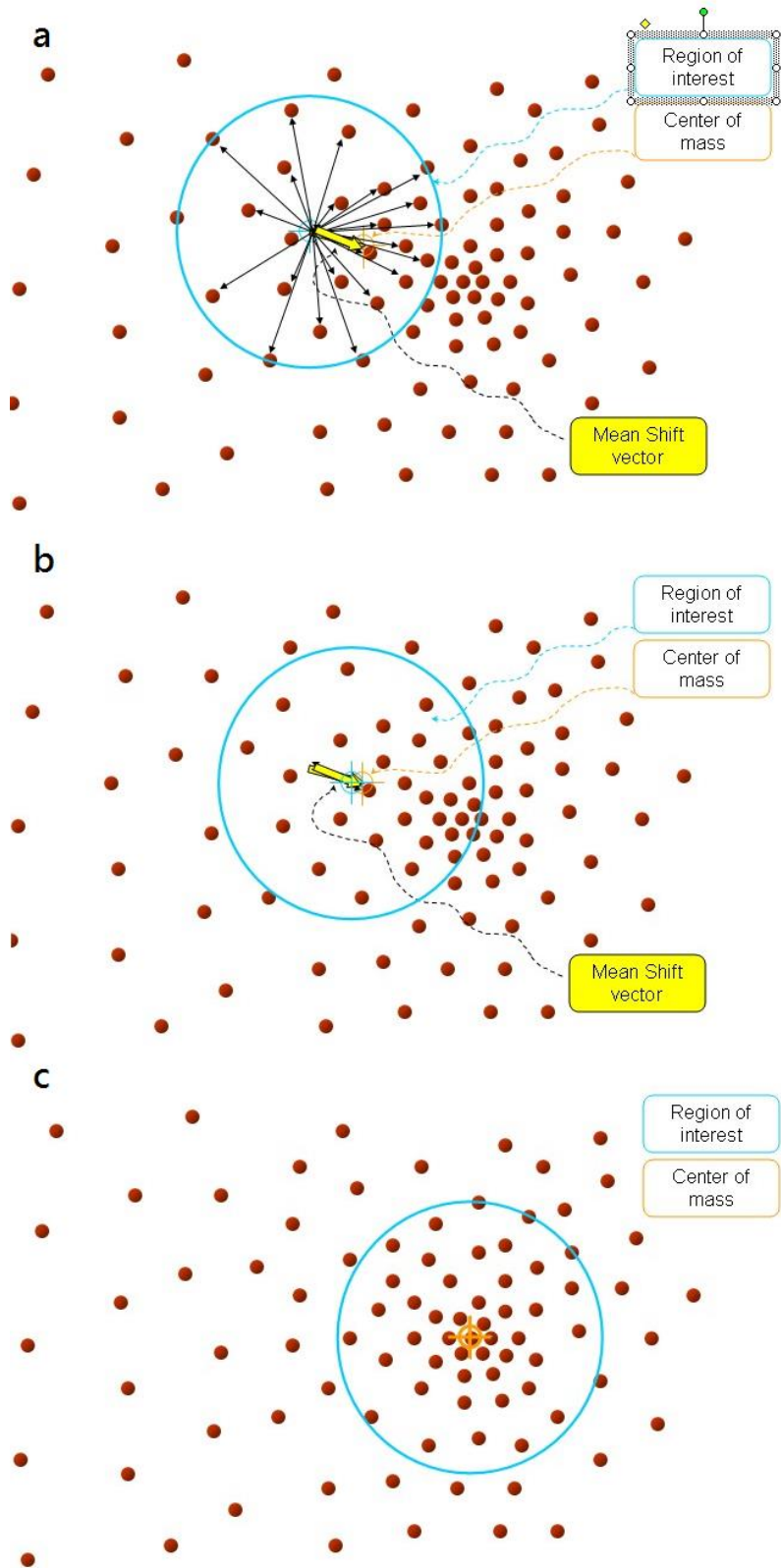


圖 1.6 Mean shift 迭代過程，a→b→c

### 1.2.7 近年手勢辨識方法

在[3]中，利用背景相減找出影像中變動的區域，再將 RGB 影像轉換為 YCbCr 影像，設 Cb 及 Cr 門檻值將膚色部份去掉。再以手掌面心向手部輪廓取角度距離特徵(圖 1.7)。

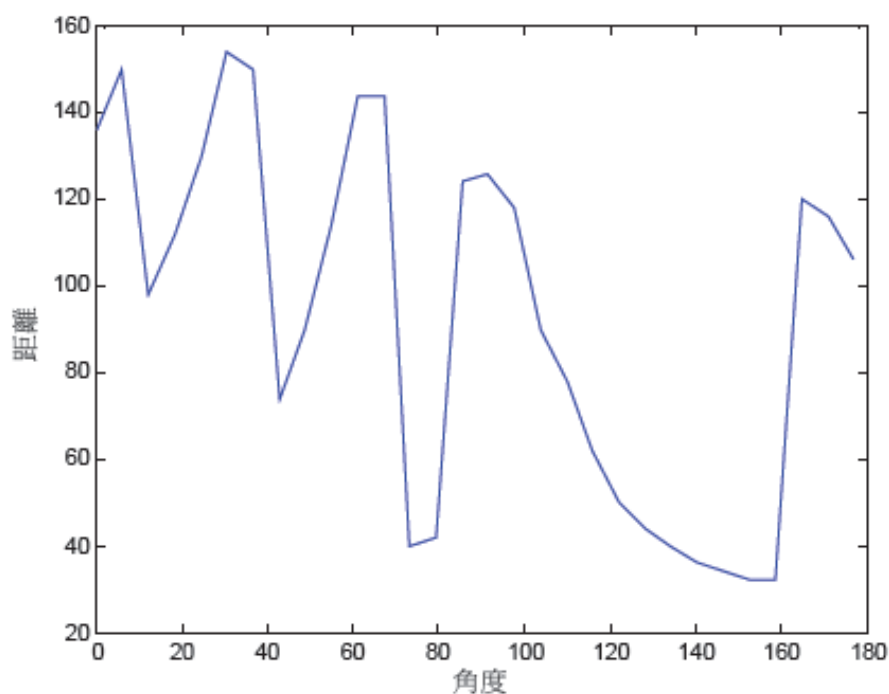


圖 1.7 角度距離圖

學習及辨識則利用倒傳遞類神經網路，先利用訓練影像訓練出權重，辨識時則輸入特徵進類神經網路，利用該權重得到對應手勢。此方法的辨識結果可能因手掌角度不同而有不同結果，且背景不能有太大變化。

在[4]中，先利用 kmeans 把影像分為兩類：手及背景，

再利用最小的矩形框住手掌計算矩形的長寬並計算比例，得知手掌角度。辨識方法為：先用最左及最右的矩形裡包含的手掌比例，判斷是否有大拇指(圖 1.8):

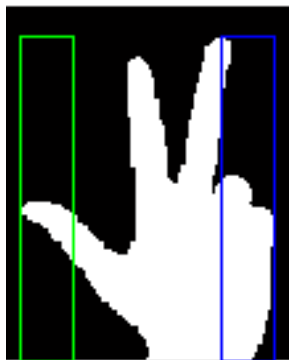


圖 1.8 計算左右兩個矩形內的白色面積比例

再尋找手指峰點，並把較不突出的峰點去掉，完成辨識(圖 1.9)。此辨識方法雖較快速，但缺點為只能在單色背景且手掌角度必須水平或垂直。



圖 1.9 尋找手指峰點

在[1]中，作者先把 RGB 影像轉至 HSV 空間，因為在 HSV 中較好描述膚色。在手勢分割時，作者利用反向投影的概念，即先統計出膚色的出現在影像中的概率，畫出直方圖，在其



他影像中在逐像素查找此色調出現在直方圖的位置，分割出手勢(圖 1.10)。

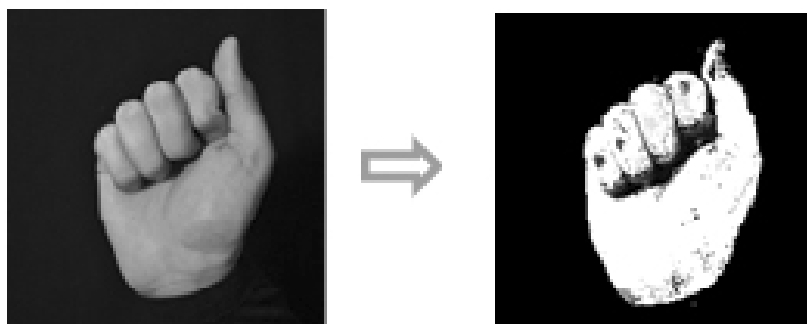


圖 1.10 右圖為膚色的機率圖

再利用多邊形圈住手勢，以多邊形到凹處的距離標準化並算平均深度為特徵一，在算出多邊形面積對手部面積的比值作為特徵二。

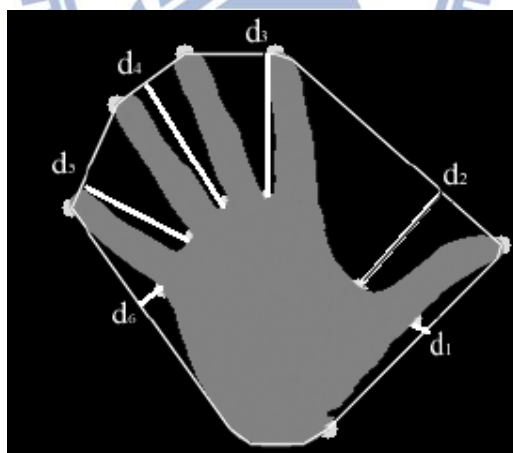


圖 1.11 多邊形取特徵

辨識時，利用影像不變矩找出與之最匹配的輪廓。此方法優點為較快速，但因沒有學習過程，所以無法適應每個人的手型。

在[2]中，作者一開始先定義膚色在RGB空間中的分布，再利用類神經網路辨識是否為膚色。訓練方法為輸入訓練影像 RGB 的值，使之輸出對應為 1 或 0。(圖 1.12)



圖 1.12 訓練膚色

辨識膚色方法時即輸入每個像素 RGB 值，經過類神經網路判斷為 1 或 0。(圖 1.13)

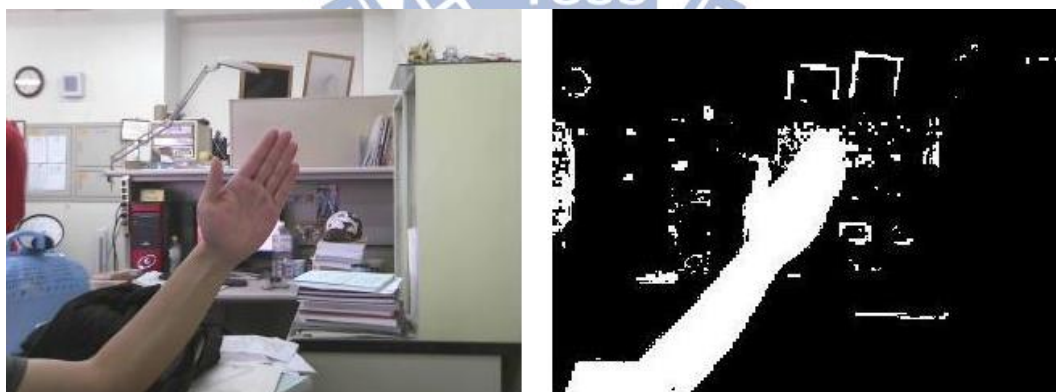


圖 1.13 類神經網路判斷膚色

得到一開始的膚色位置後，再去掉手臂部分，便利用 Mean Shift 追蹤演算法追蹤手掌位置。

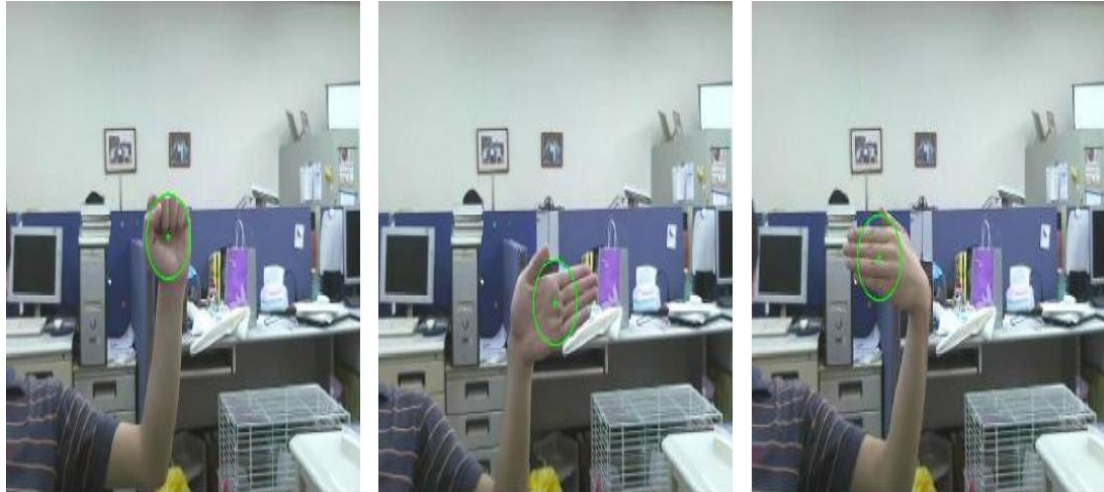


圖 1.14 跟蹤手勢

特徵抽取則計算手掌影像的區塊平均值，和大小平移不變量的影像矩作為特徵，再利用 Kmeans 做不同狀態的分類。最後再將特徵向量送入類神經網路做辨識。此篇方法在辨識上有不錯的表現，缺點為對光線較為敏感。

### 1.3 論文架構

本論文為了發展一套手勢控制的介面，以及改善目前因手勢分割的不完整性引起辨識的誤差，建立一套新的手勢辨識系統，並以此系統建立一個控制介面。

本論文架構如下：

第一章 「緒論」：說明本論文研究之背景、動機、文獻的方法及回顧、研究方向。

第二章 「研究理論及方法」：介紹本論文手勢辨識系統所應

用的各種理論。

第三章 「手勢辨識系統設計」:利用第二章的理論，設計手勢辨識系統的架構:包括手勢分割、學習及辨識。

第四章 「實驗分析與應用」:本章先測試手勢分割的完整性，再針對主成份分析中，不同的降維度對辨識率的影像，並實驗不同受試者(不在訓練影像內)的辨識率。最後再將本系統應用在電腦鼠標的控制上，將不用滑鼠或任何實體接觸控制鼠標。

第五章 「結論與未來目標」:本章歸納本系統適用的條件，並指出本系統未來繼續改進的方向。



## 第二章 研究理論與方法

這個章節將介紹本論文為完成手勢辨識，其所用到的各種技巧與理論，而由於本論文較著重在影像切割上，所以會較詳細介紹此理論。

### 2.1 $YCbCr$ 色彩空間

本論文為了偵測膚色，大部分的影像處理手續都在  $YCbCr$  色彩空間上。

$YCbCr$  為色彩空間的一種，其中  $Y$  為顏色的亮度， $C_b$  指藍色色度， $C_r$  指紅色色度。 $YCbCr$  不是絕對的色彩空間，而是一種對 RGB 空間的編碼方法，所以其值為由 RGB 的值計算而來，如式(2.1)、式(2.2)、式(2.3)所示：

$$Y = (0.257 \times R) + (0.504 \times G) + (0.098 \times B) + 16 \quad (2.1)$$

$$C_b = -(0.148 \times R) - (0.291 \times G) + (0.439 \times B) + 128 \quad (2.2)$$

$$C_r = (0.439 \times R) - (0.368 \times G) - (0.071 \times B) + 128 \quad (2.3)$$

以上三式為在數位影像(八位元深度)中， $YCbCr$  與 RGB 空間中的轉換公式。有別於廣泛使用的 RGB 空間， $YCbCr$  特別將亮度獨立出來，

且從 Cb 和 Cr 空間即可判斷顏色性質。

## 2.2 連續影像相減法(Frame differencing)

連續影像相減法為目前影像與特定影像相減，以找出畫面中變動的區域。一般而言，會以第一張讀入的影像作為背景影像，接下來讀入的影像與之相減，便可知道變動區域。然而實際應用上，會由於雜訊或曝光的影響或環境的變動，使這種方法效果不佳。為了避免這個問題，本論文會不斷地更新背景影像，方法如下：

1. 找背景影像，預設  $t = 0$ ，需累積 3 張變動較小的影像，才認定找到背景影像

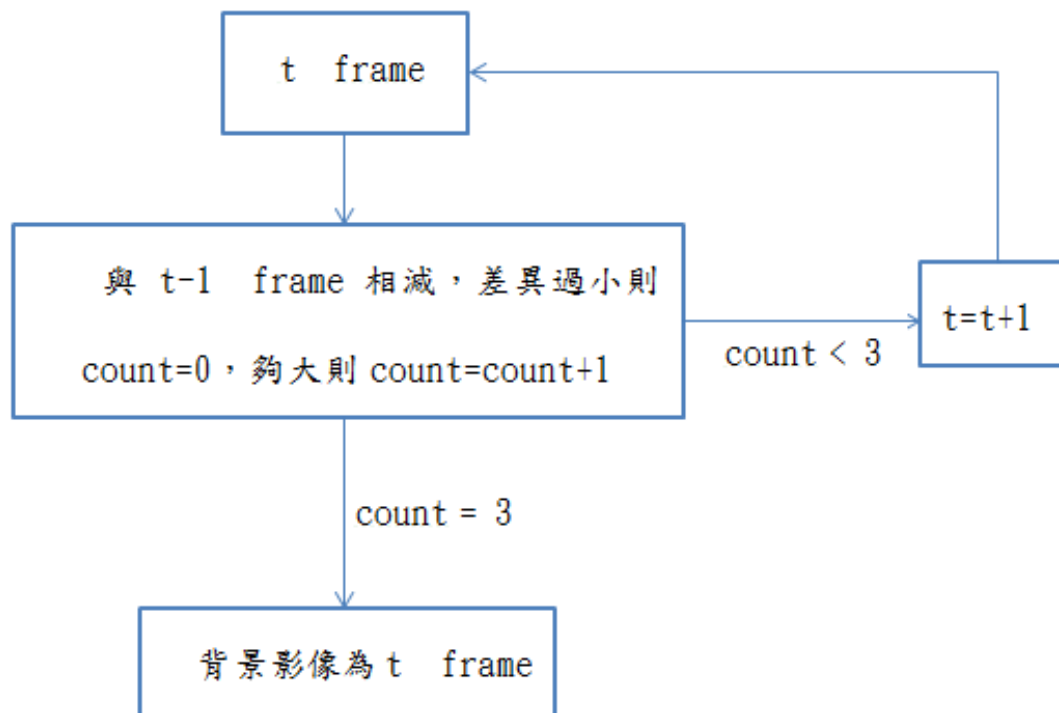


圖 2.1 找背景影像的流程

2. 找到背景影像後，便進行背景相減

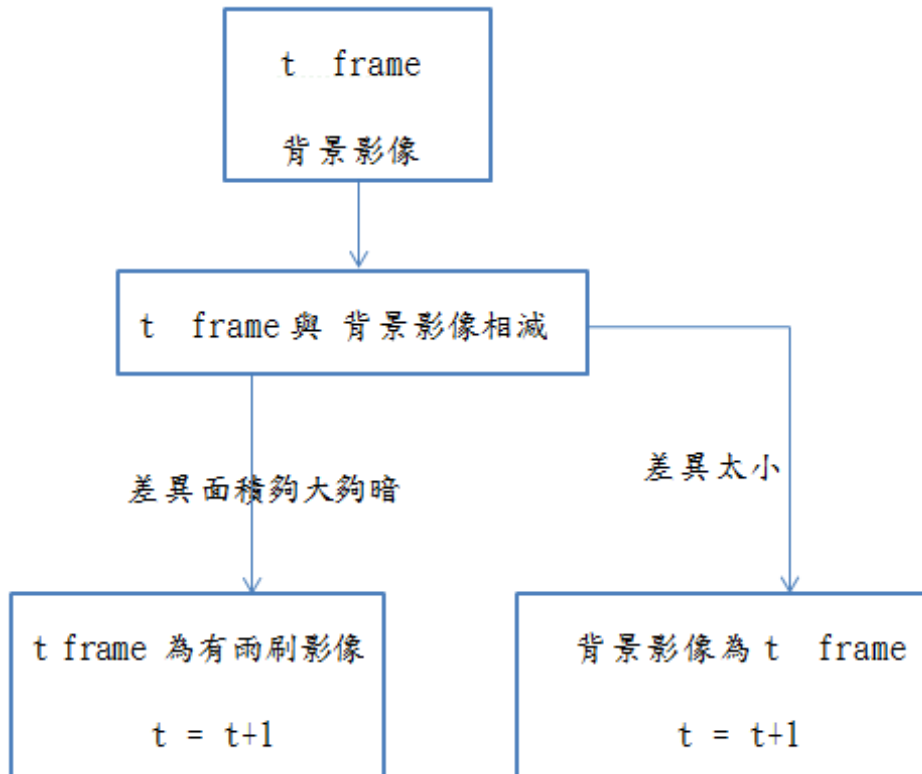


圖 2.2 連續影像相減流程

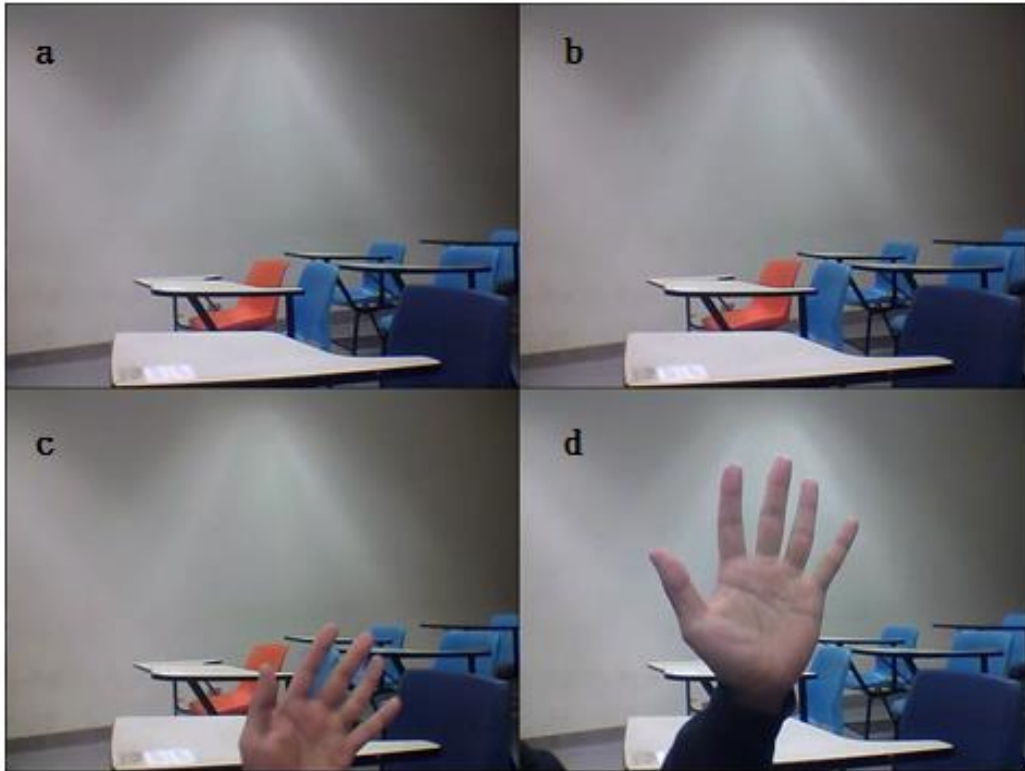


圖 2.3，a-b-c-d 為四張連續影像

以圖 2.3 為例，a 原為背景影像，背景影像再更新至 b，c 和 d 與 b 的差異較大，所以背景影像維持在 b，待手勢離開後再更新背景影像。

## 2.3 主動式輪廓(Active contour)

主動式輪廓發展至今已有眾多模型被提出，其共同特徵為不斷變形一曲線(或封閉輪廓)，使其停留在物體輪廓上：例如定義一初始輪廓環繞一物體，這條曲線會變形至包覆物體為止。主動式輪廓有許多優點：

(1). 其精確度高於以像素為單位的傳統影像切割方法。



(2). 主動式輪廓的能量模型可以參考影像中許多資訊來設計，例如輪廓形狀或影像灰階等。

(3). 此方法可以提供封閉且平緩的輪廓。

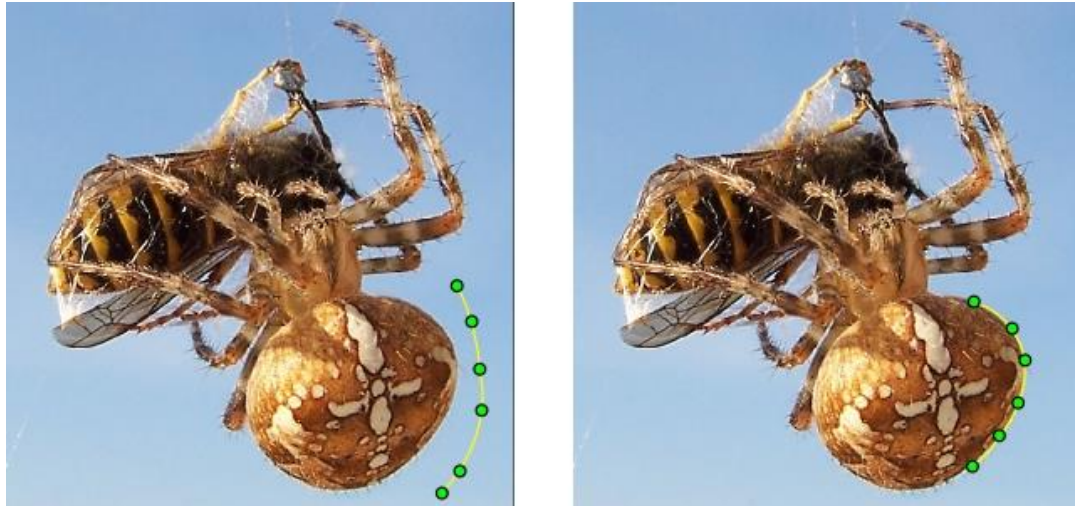


圖 2.4 主動式輪廓收斂過程

Snakes 為主動式輪廓的常見範例，最初是以基於邊緣的模型被提出的 [1]: 定義  $\Omega$  在  $\mathbb{R}^2$  為物體， $\partial\Omega$  為物體邊界， $u_0: \mathbb{R}^2 \rightarrow \mathbb{R}$  為影像，

$C(s): [0, 1] \rightarrow \mathbb{R}^2$  為參數式曲線。

$$J(C) = \alpha \int_0^1 |C'(s)|^2 ds + \beta \int_0^1 |C''(s)| ds - \lambda \int_0^1 |\nabla u_0(C(s))|^2 ds \quad (2.4)$$

其中  $\alpha, \beta, \lambda$  都是正數。第一及第二項控制曲線的平緩程度，又稱作內部能量 (Internal energy)，第三項吸引輪廓往物體靠近，又稱作外部能量 (External energy)。由 (2.4) 可發現經由最小化

$J(C)$ ，可使曲線  $C$  停留在  $\nabla u_0$  最大的地方，使曲線  $C$  像邊緣偵測子

，且同時保持曲線平緩。

## 2.4 等階集合法(Level set method)

本論文利用主動式輪廓切割手勢，且以等階集合法的水平集 0 (Level Set zero) 為參數描述輪廓，所以這節將先介紹等階集合法 (Level set method)。等階集合法這裡用作描述輪廓變化及移動的方法，其原理為把  $n$  維的描述視為  $n+1$  維空間中的一個水平，所以它的運算過程是在  $n+1$  維中進行，我們再取關注的水平集即可(圖 2.5)。藉由等階集合法我們很容易描述輪廓的變化，且在低維時輪廓變化需要不斷的參數化，但在高維就不用。而本章後面部分將介紹如何利用水平集作為主動式輪廓中能量的參數，和最小化能量的方法。

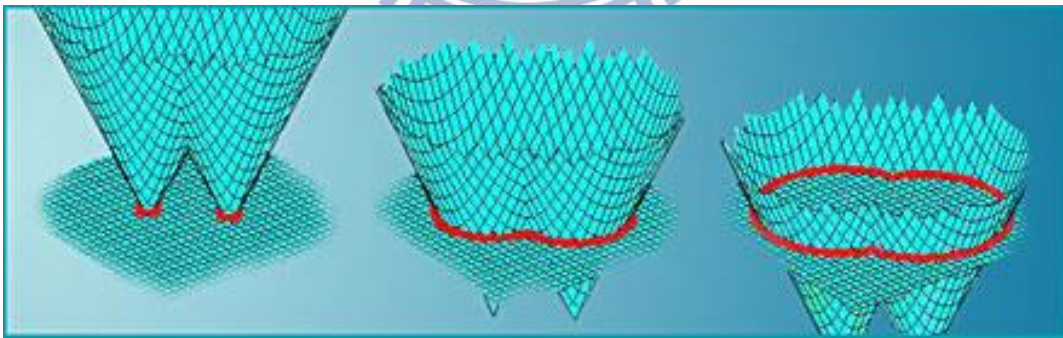


圖 2.5 等階集合法

首先介紹等階集合法如何以隱式函數定義邊界(Interface)，以及為何要把該隱函數設為正負距離函數，以及描述水平集移動的方

法。

### 2.4.1 邊界(Interface)

為了方便說明邊界概念，我們先以點邊界 (Points) 為例子。假設在一維空間中，我們用兩點  $x = 1, x = -1$  將一線段切成三個子空間，即  $(-\infty, -1), (1, -1), (1, \infty)$ 。我們視  $\Omega^- = (1, -1)$  為裡面， $\Omega^+ = (-\infty, -1) \cup (1, \infty)$  為外面，以兩點  $\partial\Omega = \{-1, 1\}$  為邊界。在一維空間中，可發現子空間都為一維，邊界為零維度。事實上在  $\mathbb{R}^n$  空間中，子空間為  $n$  維度，邊界為  $n-1$  維度。

上述描述邊界的方法為顯式(Explicit)法，即我們明確地寫下邊界上的點  $\partial\Omega = \{-1, 1\}$ ；而隱式(Implicit)法則是利用定義函數的某水平集視為邊界：例如我們定義一函數  $\varphi(x) = x^2 - 1$  再取  $\varphi(x) = 0$  的集合，恰可得到  $x = 1$  及  $x = -1$ ，同樣得到邊界  $\partial\Omega = \{-1, 1\}$  (圖 2.6)

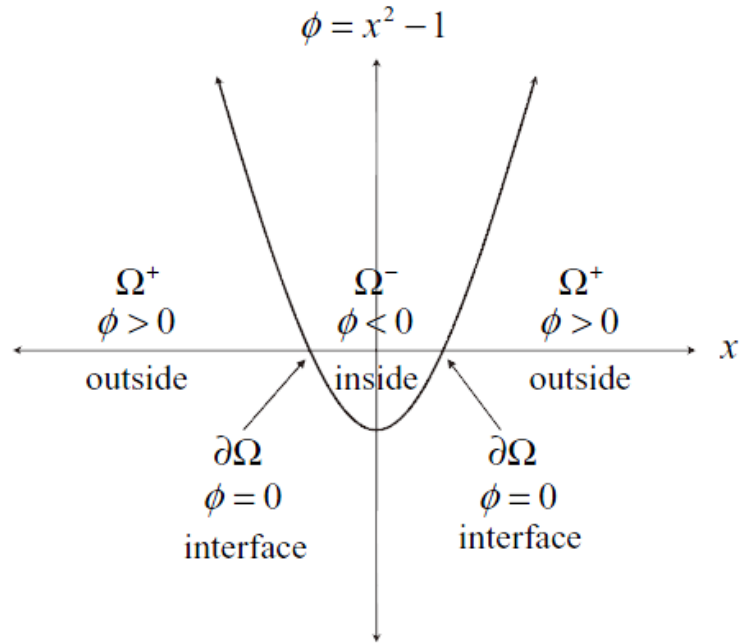


圖 2.6 利用隱函數  $\phi(x) = x^2 - 1$  的  $\phi(x) = 0$  的集合定義邊界

值得注意的是在得到同樣的邊界時，隱式表示方法比顯示法多了一個維度，事實上若用隱式法在  $\mathbb{R}^n$  空間定義邊界，需要定義  $n$  維的隱函數再取  $n-1$  維的水平集作為邊界。

利用上述的定義，可知在二維空間中，利用一維曲線邊界可區分出二維的子空間。這裡我們只考慮封閉曲線，這樣才能容易判別裡面與外面。這裡以一封閉單位圓  $x^2 + y^2 = 1$  作為一平面的邊界作為例子：若用顯示邊界表示法，我們須清楚描述這條封閉曲線上的所有點，即  $\partial\Omega = \{\vec{x} \mid |\vec{x}| = 1\}$ ；而若用隱式法，我們定義函數  $\phi(\vec{x}) = x^2 + y^2 - 1$ ，再取  $\phi(\vec{x}) = 0$  的水平集，得到一樣的邊界(圖 2.7)

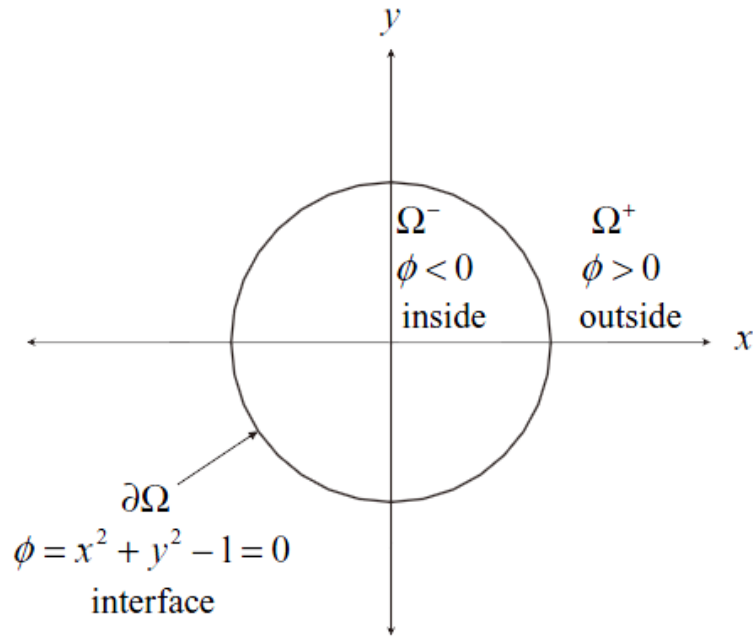


圖 2.7 隱式邊界表示法描述  $x^2 + y^2 = 1$

在二維空間中以顯示法表示邊界需定義曲線上所有的點，這在大部分的情況下很難做到，只能參數化曲線，若曲線移動或變形時則須不斷重新參數化曲線；而隱式法以高維函數定義邊界的方法雖看似浪費，但在這裡就能看處它的好處，所以本論文即應用二維的隱函數表示曲線變化。

#### 2.4.2 距離函數(Distance function)

在定義隱函數  $\phi$  的水平集為邊界的同時，也希望  $\phi$  也為正負距離函數(Signed distance function)，因為這種函數有些良好性質可簡化等階集合法，使運算過程更順利。首先距離函數定義  $d(\vec{x})$  如下

$$d(\vec{x}) = \min(|\vec{x} - \vec{x}_I|) \quad \text{for all } \vec{x}_I \in \partial\Omega \quad (2.5)$$

其意謂  $\vec{x} \in \partial\Omega$  時， $d(\vec{x})=0$ 。而對於其它  $\vec{x}$ ， $d(\vec{x})$  的值為  $\partial\Omega$  上與  $\vec{x}$  最相近的點  $\vec{x}_I$  與  $\vec{x}$  的距離  $|\vec{x} - \vec{x}_I|$ 。

另外距離函數可寫成以下形式

$$d(\vec{x}) = |\vec{x} - \vec{x}_I| = ((\Delta x)^2 + (\Delta y)^2)^{1/2} \quad (2.6)$$

其中  $\Delta x$  與  $\Delta y$  為兩點間水平及垂直分量的距離差，而  $d(\vec{x})$  的梯度為

$$\begin{aligned} \nabla d &= \left( \frac{\partial d}{\partial x}, \frac{\partial d}{\partial y} \right) \\ &= \left( \frac{\Delta x}{((\Delta x)^2 + (\Delta y)^2)^{1/2}}, \frac{\Delta y}{((\Delta x)^2 + (\Delta y)^2)^{1/2}} \right) \end{aligned} \quad (2.7)$$

所以合併(2.6)及(2.7)可得到

$$|\nabla d| = 1 \quad (2.8)$$

此為距離函數的特殊性質之一。

而正負距離函數(Signed distance function)  $\varphi(\vec{x})$  為一種隱函數，對所有的  $\vec{x}$  而言， $|\varphi(\vec{x})| = d(\vec{x})$ 。因此  $\vec{x} \in \partial\Omega$  時， $\varphi(\vec{x}) = d(\vec{x}) = 0$ ； $\vec{x} \in \Omega^-$  時， $\varphi(\vec{x}) = -d(\vec{x})$ ；而  $\vec{x} \in \Omega^+$  時， $\varphi(\vec{x}) = d(\vec{x})$ 。

正負距離函數也具有同樣性質

$$|\nabla \varphi(\vec{x})| = 1 \quad (2.9)$$

上節曾以隱函數  $\varphi(x) = x^2 - 1$  定義出邊界  $\partial\Omega = \{-1, 1\}$ ，而現在也可以此為例子，重新定義正負距離函數  $\varphi(x) = |x| - 1$ ，定義出同樣的邊界(圖 2.8 所示)。

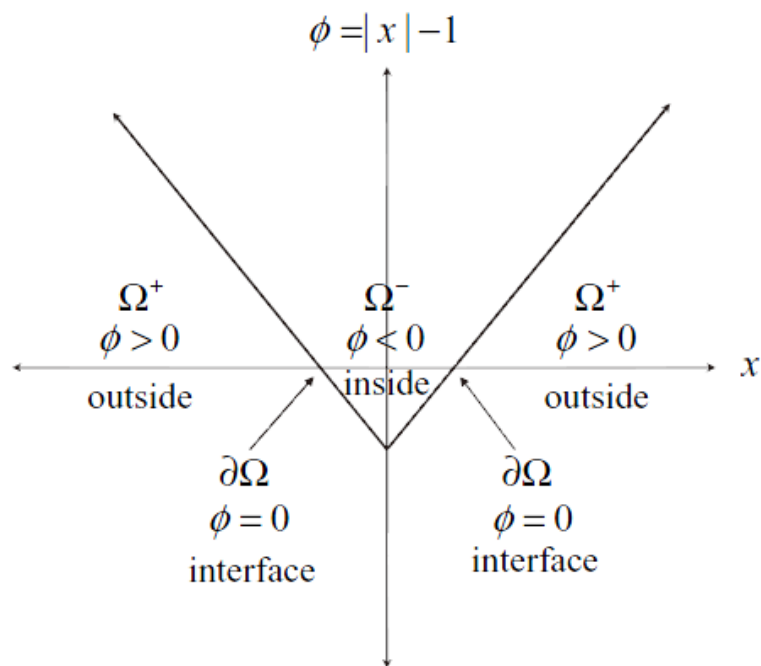


圖 2.8 利用正負距離函數區分出  $\Omega^+$  和  $\Omega^-$  及  $\partial\Omega$

這裡  $X = 0$  時  $\varphi(x)$  的導數並無定義，但因我們將  $\varphi(x)$  應用在卡式座標中，即應用在有限的取樣點，所以在這種情形下  $\varphi(0)$  的值會是有限的。

### 2.4.3 移動水平集

我們將討論如何描述水平集(level set zero)的移動，即  $\varphi(\vec{x}) = 0$  上各點的變化。其變化其實為高維曲面移動後取水平集的結果，這裡我們先以外在產生的速度場作為水平集移動的例子，不過速度場當然也有可能由水平集自身定義，所以我們再用曲率作為第二個例子，而在本論文中，我們將會以影像中的區域資訊定義速度場。

我們以  $\vec{V}(\vec{x})$  表示在  $\varphi(\vec{x})$  上每個點  $\vec{x}$  的速度。若設速度場  $\vec{V} = (u, v, w)$ ，我們希望  $\varphi(\vec{x})$  每點以此速度場移動，最簡單描述每點移動的方法為解決以下的 ODE

$$\frac{d\vec{x}}{dt} = \vec{V}(\vec{x}) \quad (2.10)$$

而接下來我們希望描述  $\varphi(\vec{x})$  的水平集的移動，表示如下

$$\varphi(\vec{x}(t)) = 0 \quad (2.11)$$

其中  $t$  參數表示點  $\vec{x}$  隨時間變化，利用連鎖率法則(chain rule)可推導出(2.11)隨  $t$  的變化率

$$\frac{d\varphi(\vec{x}(t))}{dt} = \frac{\partial\varphi}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial\varphi}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial\varphi}{\partial z} \frac{\partial z}{\partial t} + \frac{\partial\varphi}{\partial t} \quad (2.12)$$

其中

$$\frac{\partial\varphi}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial\varphi}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial\varphi}{\partial z} \frac{\partial z}{\partial t} = u\varphi_x + v\varphi_y + w\varphi_z = \vec{V} \cdot \nabla\varphi \quad (2.13)$$

由(2.13)套入(2.12)得到

$$\varphi_t = \vec{V} \cdot \nabla\varphi \quad (2.14)$$

(2.14)描述  $\varphi(\vec{x})$  的水平集的移動，又稱為等階集合方程式(Level set equation)。

## 2.5 最佳設計

為了最小化各種不同形式的能量模型，最常使用的方法為歐拉-



拉格朗日方程式與梯度下降法。

### 2.5.1 梯度下降法(Gradient descent method)

梯度下降法為最佳化技巧之一，可用數值化方法尋找區域性的最小值。若函數  $f(x)$  在  $x = a$  可微分，利用梯度  $\nabla f(a)$  為函數  $f$  在  $x = a$  變化率最大的方向的性質，可得知沿  $-\nabla f(a)$  下降最快，因此如果

$$b = a - s\nabla f(a) \quad (2.15)$$

其中  $s$  為步伐大小，則  $f(b) \geq f(a)$ 。

因此令  $f$  從  $x_0$  出發，則

$$x_{n+1} = x_n - s_n \nabla f(x_n) \quad (2.16)$$

可得到

$$f(x_{n+1}) \leq f(x_n) \quad (2.17)$$

即隨著步伐進行，函數  $f(x)$  值會越來越小，因此可設定某一收斂值為停止迭代的標準，得到函數的  $f$  的最小值。

梯度下降法的優點為保證收斂，其表現較為強健；缺點為其收斂速度會隨迭代次數增加而越來越慢。若函數為凸函數(Convex function)時則可保證找到的區域最小值(Local minimum)同時也為廣域最小值(Global minimum)。

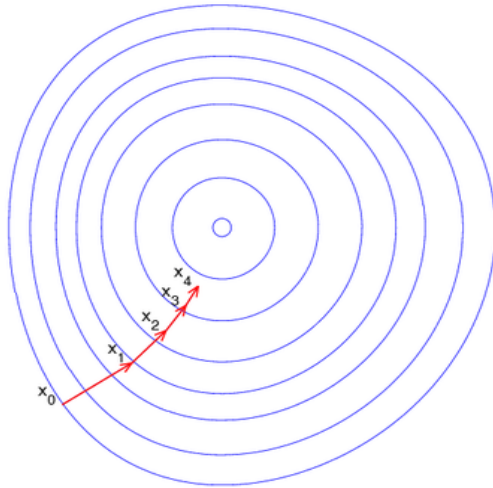


圖 2.9 梯度下降法的迭代過程

### 2.5.2 歐拉-拉格朗日方程式(Euler Lagrange equation)

考慮一能量泛函數(functional) $E$

$$E = \int_a^b L(x, f(x), f'(x)) dx \quad (2.18)$$

利用 first variation，泛函數  $E$  的變分可寫成以下形式，其中  $\varepsilon$  為微小量， $\eta$  為可微分函數

$$\delta E = \int_a^b \left( \frac{\partial L}{\partial f} - \frac{d}{dx} \frac{\partial L}{\partial f'} \right) \eta(x) dx \quad (2.19)$$

而  $\delta E = 0$  為  $E$  有最小值的必要條件，所以

$$\frac{\partial L}{\partial f} - \frac{d}{dx} \frac{\partial L}{\partial f'} = 0 \quad (2.20)$$

為  $E$  有極值的必要條件，又稱為 Euler - Lagrange equation。另外

值得注意的是  $\delta E$  可寫成內積形式

$$\delta E = \int_a^b g \cdot \eta \, dx = \langle g, \eta \rangle \quad (2.21)$$

根據 Cauchy Schwarz 不等式

$$|\langle g, \eta \rangle|^2 \leq \langle g, g \rangle \cdot \langle \eta, \eta \rangle \quad (2.22)$$

因為  $g$  為已知，所以  $\eta=g$  時會使  $\langle g, \eta \rangle$  有最大值，因此  $g$  可視為  $E$  之梯度，即

$$\nabla E = \frac{\partial L}{\partial f} - \frac{d}{dx} \frac{\partial L}{\partial f'} \quad (2.23)$$

## 2.6 手掌轉正

為了避免手掌角度對辨識的影響，本論文將手掌視為對稱的剛體並利用質量慣性矩陣(inertia matrix)，計算手掌主軸的角度，再將手掌每點乘上旋轉矩陣，將手掌轉至垂直，俾利尋找手指峰點。

### 2.6.1 影像矩(Moment)

影像矩由特定的權重與影像位置組合而成，各種矩對影像各有不同的描述，依需求可利用不同的矩。本論文為了取得質量慣性矩陣，需用列矩(Raw moments)。

$p, q = 0, 1, 2, \dots$ ，對灰階影像  $I(x, y)$  而言，列矩  $M_{ij}$  為

$$M_{ij} = \sum_x \sum_y x^i y^j I(x, y) \quad (2.24)$$

利用列矩可以初步計算簡單的手掌幾何性質，例如手面積為  $M_{00}$ ，手掌面心  $(\bar{x}, \bar{y})$  則為  $(M_{10}/M_{00}, M_{01}/M_{00})$ 。

### 2.6.2 質量慣性矩(Moment of inertia)

對於一剛體  $B$ ， $L$  為空間中一軸線，則可定義剛體  $B$  對  $L$  軸的質量慣性矩為

$$I_L = \int r^2 dm \quad (2.25)$$

其中  $r$  為每個質量元素  $dm$  到  $L$  軸線的垂直距離，圖 2.10 以原點為軸。

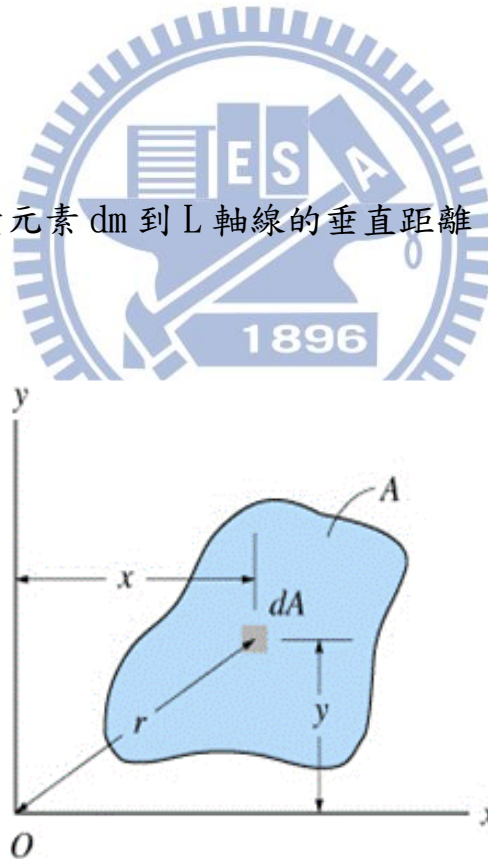


圖 2.10 計算質量慣性矩

### 2.6.3 質量慣性積(Product of inertia)

對於兩正交平面 X 與 Y，每個質量元素 dA 至各平面的垂直距離的乘積，為其元素的質量慣性積  $dI_{xy}$ ，再對整個質量積分後得質量慣性積  $I_{xy}$

$$I_{xy} = I_{yx} = \int_A xy dA \quad (2.26)$$

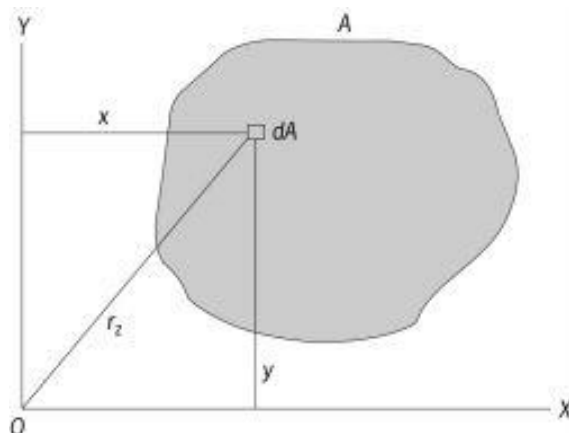


圖 2.11 計算質量慣性積

令用上述的慣性矩及慣性積，在卡式座標可組成矩陣形式，即為質量慣性矩陣  $I$  (inertia matrix)，本論文對質心取質量慣性矩陣  $I_O$

$$I_O = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{yx} & I_{yy} & -I_{yz} \\ -I_{zx} & -I_{zy} & I_{zz} \end{bmatrix} \quad (2.27)$$

#### 2.6.4 轉主軸(Principal axes)

在(2.27)中， $I_O$ 矩陣為九個元素的矩陣，但若適當的選取坐標系，則可將 $I_O$ 化簡為只有三個元素的對角矩陣，即是轉主軸的目的之

一。

主軸為剛體三面均對稱的軸，而由質量慣性積的定義，會導致剛體的 $I_O$ 的質量慣性積部分均為零。以下圖為例，此剛體對 $x'$ 平面對稱，所以 $I_{xy}$ 、 $I_{xz}$ 、 $I_{yx}$ 、 $I_{zx} = 0$ ，故 $x'$ 為主軸之一。

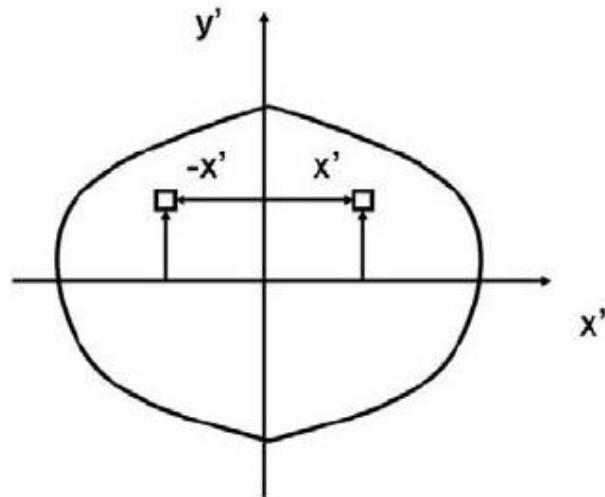


圖 2.12 主軸

在一般情況下，除非物體完美對稱，否則想直接發現主軸位置並不是件容易的事。以圖 2.13 為例，

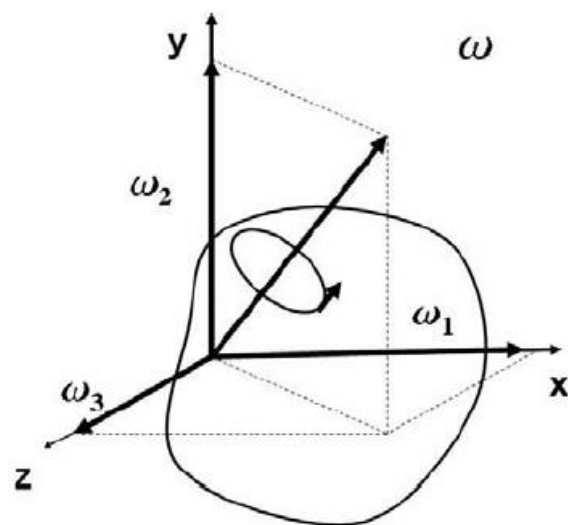


圖 2.13 三維空間上的主軸

我們想尋求一座標系  $x$ 、 $y$ 、 $z$ ，且三軸上的角速度  $w$  分別為

$w_x$ 、 $w_y$ 、 $w_z$ ，分別與三方向角動量平行，即

$$\begin{bmatrix} H_{Gx} \\ H_{Gy} \\ H_{Gz} \end{bmatrix} = \begin{bmatrix} I_{wx} \\ I_{wy} \\ I_{wz} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix} \quad (2.28)$$

若以此座標系，用原始的方法表示角動量，則為

$$\begin{bmatrix} H_{Gx} \\ H_{Gy} \\ H_{Gz} \end{bmatrix} = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{yx} & I_{yy} & -I_{yz} \\ -I_{zx} & -I_{zy} & I_{zz} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix} \quad (2.29)$$

利用上述兩式，可寫成下式以尋求  $x$ 、 $y$ 、 $z$  的方向

$$\begin{bmatrix} I_{xx} - I & -I_{xy} & -I_{xz} \\ -I_{yx} & I_{yy} - I & -I_{yz} \\ -I_{zx} & -I_{zy} & I_{zz} - I \end{bmatrix} \begin{bmatrix} w_x \\ w_y \\ w_z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (2.30)$$

則找尋主軸的問題即為解特徵值的問題，解出的特徵值即為三個主軸的質量慣性矩，特徵向量即為主軸三個正交的方向。

### 2.6.5 旋轉矩陣 (Rotation matrix)

在二維平面上，任何旋轉矩陣  $R$  可寫成以下形式

$$R(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (2.31)$$

在平面上的點向量乘上  $R$  矩陣後，會對原點產生角度  $\theta$  的旋轉(圖

2.14):

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.32)$$

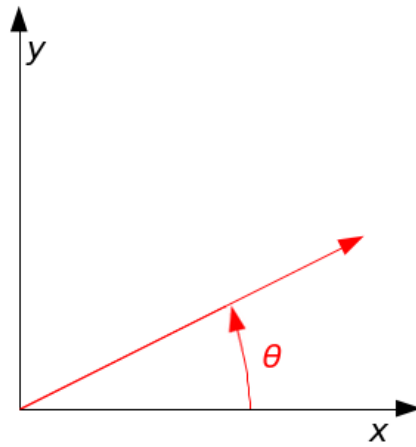


圖 2.14 向量乘上 R 矩陣後，會對原點產生角度 $\theta$ 的旋轉

若欲使點向量繞點 $(x_0, y_0)$ 旋轉，則可寫成下列關係式

$$\begin{bmatrix} x' - x_0 \\ y' - y_0 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \quad (2.33)$$

## 2.7 傅立葉描述子(Fourier descriptors)

在完成影像切割的步驟後，所得到的切割影像往往仍不夠平滑，所以本論文對影像切割後的輪廓做傅立葉描述，並只取前幾項係數再反轉換，這項作法能夠得到較平滑的輪廓，並保留較重要的特徵(手指峰點)。

在二維平面上，若任意輪廓起點為 $(x_0, y_0)$ ，並依序為



$(x_1, y_1)$  ,  $(x_2, y_2)$  ,  $(x_3, y_3)$ ... , 則為了方便做一維的傅立葉轉換，我們可以用複數座標表示之，即

$$s(k) = x(k) + jy(k) \quad (2.34)$$

其中  $k = 0, 1, 2 \dots K-1$  表示輪廓點數， $x$  及  $y$  為原本輪廓的  $x, y$  座標。

將輪廓寫成一維複數形式後， $s(k)$  的離散傅立葉轉換為

$$a(u) = \sum_{k=0}^{K-1} s(k)e^{-j2\pi uk/K} \quad (2.35)$$

其中  $u = 0, 1, 2 \dots K-1$ 。 $a(u)$  稱為此輪廓線的傅立葉描述子，

$a(u)$  的傅立葉反轉換能夠重建輪廓線  $s(k)$ ，即

$$s(k) = \frac{1}{K} \sum_{u=0}^{K-1} a(u)e^{j2\pi uk/K} \quad (2.36)$$

其中  $k = 0, 1, 2 \dots K-1$ 。

在一般情況下，我們並不會取全部的傅立葉係數，反之只取前  $P$  項係數再做反轉換，得到近似的  $s(k)$

$$\tilde{s}(k) = \frac{1}{P} \sum_{u=0}^{P-1} a(u)e^{j2\pi uk/P} \quad (2.37)$$

其中  $k = 0, 1, 2 \dots K-1$ ，表示雖只取前  $P$  項係數，反轉後的輪廓點數仍是  $K$  點。

由於在影像傅立葉轉換中，高頻為表示細節的部分，低頻為則表示大略形狀的部分，所以  $P$  越小則表示會遺失掉較細節的部分，如圖 2.15 所示為長度  $K = 1174$  的輪廓做傅立葉描

述子

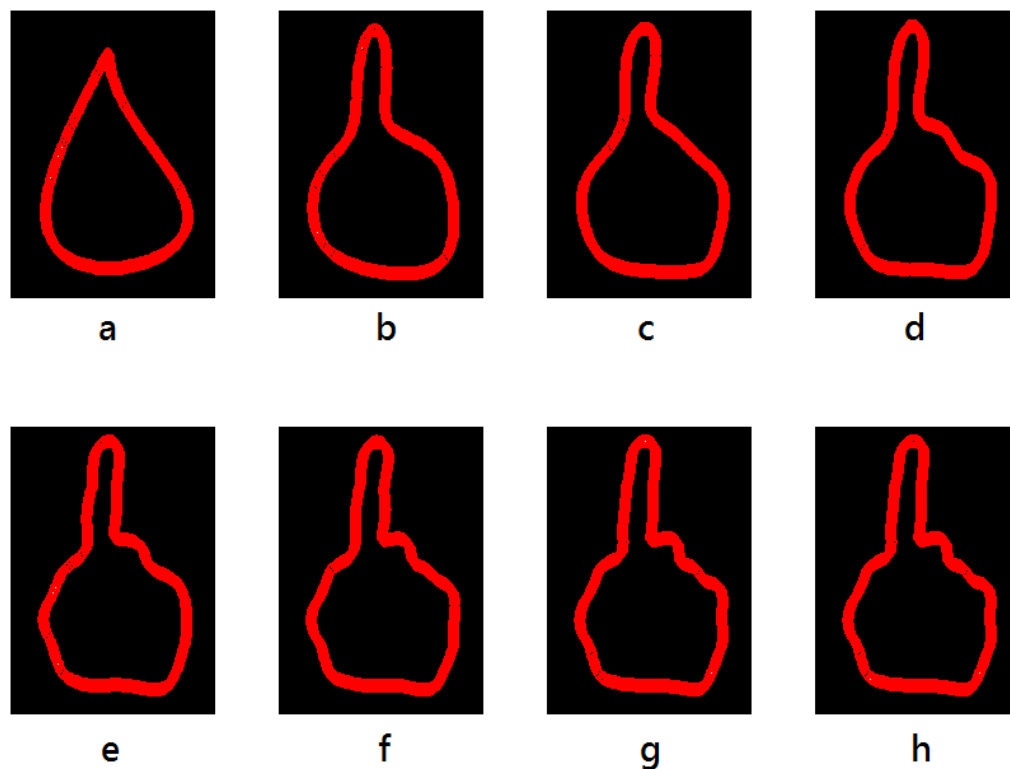


圖 2.15 (a)到(g)為取前 5~35 項係數，(h)為原本的輪廓，可以看出大約取前 10 項係數就可以得到想要的特徵形狀，30 項就幾乎足以表示原本輪廓線。

## 2.8 奇異值分解(Singular value decomposition)

設  $A$  為  $m \times n$  階的矩陣， $\text{rank}(A)=r$ ，則  $A$  可分解為

$$A = U \Sigma V^* \quad (2.38)$$

其中  $U$  為  $m \times m$  的正交矩陣， $V$  為  $n \times n$  的正交矩陣， $\Sigma$  為主對角矩陣

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots & \cdots & 0 \\ 0 & 0 & \sigma_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (2.39)$$

主對角元 $\sigma_i > 0$ ,  $i = 1, 2, \dots, r$  稱為奇異值(singular values)，通常我們會由大至小排序，即 $\sigma_1 \geq \sigma_2 \geq \sigma_3 > 0$ ，而且大部分情況下，少部分的奇異值大於其他奇異值的總和。SVD的特色是 $\Sigma$ 矩陣的多數元為零，下圖白色區塊為 0:

$$m \begin{bmatrix} A \\ n \end{bmatrix} = \begin{bmatrix} U \\ m \end{bmatrix} \begin{bmatrix} \Sigma \\ n \end{bmatrix} \begin{bmatrix} V^* \\ n \end{bmatrix}$$

圖 2.16

這表示 A 僅由 U 的前個行向量， $V^*$  的前個列向量以及  $\Sigma$  的左上  $r \times r$  的子矩陣決定，當  $r$  遠小於  $m$  及  $n$  時，利用 SVD 可大幅減少儲存量。

$$m \begin{bmatrix} A \\ n \end{bmatrix} = m \begin{bmatrix} U_r \\ r \end{bmatrix} [\Sigma_r] \begin{bmatrix} V_r^* \\ n \end{bmatrix}$$

圖 2.17

## 2.9 主成份分析(Principal component analysis)

在本論文中，由於每個手勢樣本採集的特徵數較多，但並不是每個特徵都同等重要，所以有必要從原本特徵中挑出較重要性的特徵再比較，即降低維度。利用主成份分析，我們能將原本特徵矩陣投射至維度較小且變異數較大的空間，達到降維的目的。

主成份分析(或稱PCA)是一種分析、簡化數據集的方法，利用基底變換，使變換後的數據有最大的變異數，可以用來剔除沒有代表性的數據，保留有代表性的樣本。以圖 2.18為例子。

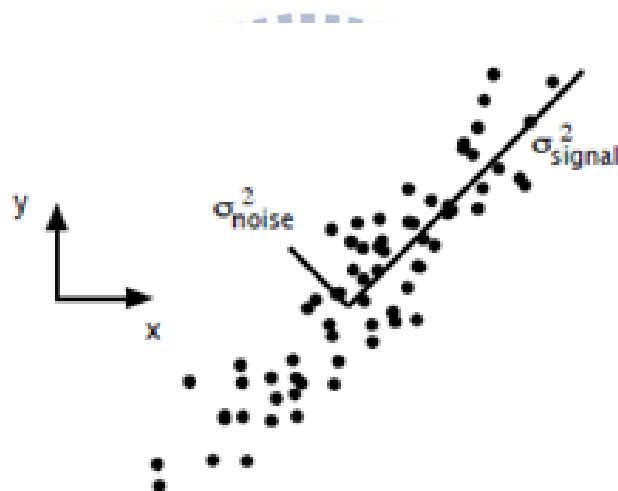


圖 2.18 以一維軸觀測點

若我們想用一軸去近似這些點(即以二維降到一維)，則將每點投射到y或x座標上會得到相當的變異數(因為數據大約呈45度分佈)，但若投射到圖中signal軸，則可得到最大的變異數；投射到noise軸會得到較小的變異數。一般而言，變異數大的方向為訊號的方向，較小則是雜訊的方向，我們可將原數據以此訊號軸近似，即可以用PCA在N維空

間取 $r$ 個軸描述( $r \leq N$ )。

設蒐集到 $n$ 個 $p$ 維的樣本數據 $\{o_1, o_2, o_3 \cdots o_n\}$ ，並扣除每個數據點的平均向量 $\mu$ ，得到離差矩陣 $O$

$$O = \begin{bmatrix} (o_1 - m)^T \\ \vdots \\ (o_n - m)^T \end{bmatrix} \quad (2.40)$$

$O$ 矩陣中，每一列為一個特徵，每一行為一個變數。將此矩陣標準化後可得

$$\tilde{O} = \begin{bmatrix} (o_1 - m)^T \\ \vdots \\ (o_n - m)^T \end{bmatrix} \begin{bmatrix} \frac{1}{s_1} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \frac{1}{s_p} \end{bmatrix} \quad (2.41)$$

其中 $s_j^2 = \frac{1}{n-1} \sum_{k=1}^n (o_{kj} - m_j)^2$ 是第 $j$ 特徵的變異數， $m_j$ 是第 $j$ 特徵的樣本平均數。設 $D$ 為對角矩陣，對角線上依序為 $s_1, s_1 \cdots s_p$ ，則 $\tilde{O}$ 可表示為

$$\tilde{O} = OD^{-1} \quad (2.42)$$

定義共變異數矩陣(covariance matrix)為

$$S = \frac{1}{n-1} O^T O \quad (2.43)$$

$S$ 可以正交對角化為

$$S = WHW^T \quad (2.44)$$

其中 $W = \{w_1, w_2, \cdots w_p\}$ 為標準正交(orthonormal)主成份(為特徵向量)構成的 $p \times p$ 的正交矩陣，即 $W^T W = W W^T = I$ ， $H$ 為特徵值矩

陣。令  $z_{kj}$  是數據點  $o_k - \mu$  所含主成份  $w_j$  的主成份係數，由正交投影得

$$z_{kj} = (o_k - \mu)^T w_j \quad (2.45)$$

其中  $k = 1 \cdots n$ ,  $j = 1 \cdots p$

將這些係數合成  $n \times p$  階主成份係數矩陣

$$Z = \begin{bmatrix} (o_1 - m)^T \\ \vdots \\ (o_n - m)^T \end{bmatrix} [w_1 \quad \cdots \quad w_p] = O W \quad (2.46)$$

再考慮矩陣  $O$  的奇異值分解

$$O = U \Sigma V^T \quad (2.47)$$

其中  $U$  為正交的左奇異矩陣， $\Sigma$  為對角奇異值矩陣， $V$  為正交右奇異矩陣，將奇異值的結果帶入(2.43)式可得到

$$S = \frac{1}{n-1} O^T O = \frac{1}{n-1} V \Sigma^2 V^T = V \left( \frac{1}{n-1} \right) \Sigma^2 V^T \quad (2.48)$$

與(2.44)對應，可得特徵值矩陣  $H$

$$H = \left( \frac{1}{n-1} \right) \Sigma^2 \quad (2.49)$$

主成份矩陣

$$W = V \quad (2.50)$$

主成份係數矩陣

$$Z = O V = U \Sigma V^T V = U \Sigma \quad (2.51)$$

由此可看出奇異值分解出的正交矩陣  $V$  即為主成份矩陣  $W$ ，奇異值  $\sigma_j$

可算出特徵值  $\lambda_j = \left( \frac{1}{n-1} \right) \sigma_j^2$ ，正交矩陣  $U$  和奇異值矩陣  $\Sigma$  可決定主成

份矩陣 $Z$ 的係數。



## 第三章 手勢辨識系統設計

### 3.1 手掌分割

在進行手勢辨識前，如何將手掌從複雜背景正確地擷取出來為相當重要的課題。本論文將此步驟分為：1. 連續背景影像相減 2. 膚色濾除 3. 主動式輪廓切割手勢。

#### 3.1.1 連續背景影像相減

為了找出可能的手勢位置，本模組先利用連續背景影像相減找出是否有物體進入。因為僅偵測是否有膚色物體進入，且欲屏除亮度影響所造成的誤判，所以將 RGB 影像轉至 YCbCr 空間，不考慮 Y 通道的變化(即亮度的變化)，並取 Cr 通道，作為相減的參數。

當前影像與背景影像相減出變動的區域後，若變動太小則更新當前影像為背景影像。若區域夠大，則再進一步設定較寬鬆的門檻值濾除非膚色部分，本論文設定  $77 < Cb < 150$ ， $135 < Cr < 177$ 。

在上述步驟完成後若仍有區域夠大，則視為膚色物體進



入畫面，將此膚色區域設置 ROI(region of interest)，待更進一步的手勢分割，如圖 3.1 與圖 3.2 所示。

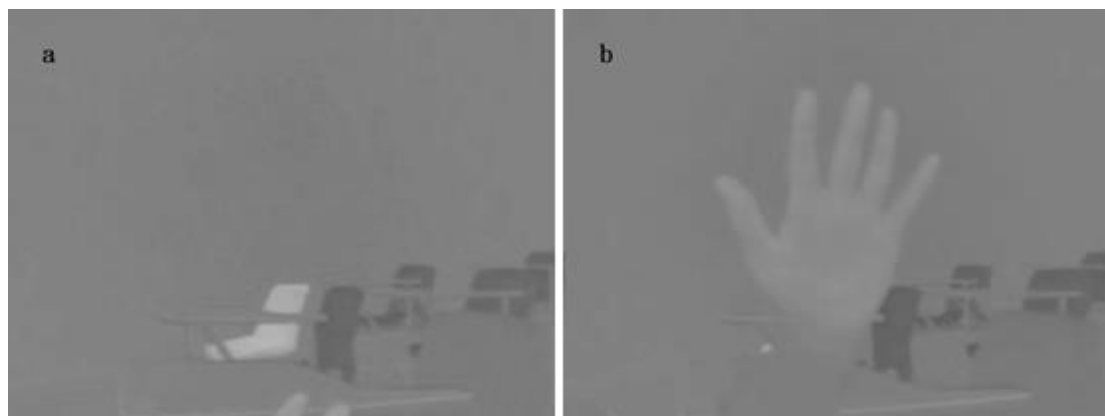


圖 3.1，保留的背景影像與當前影像的 Cr 通道

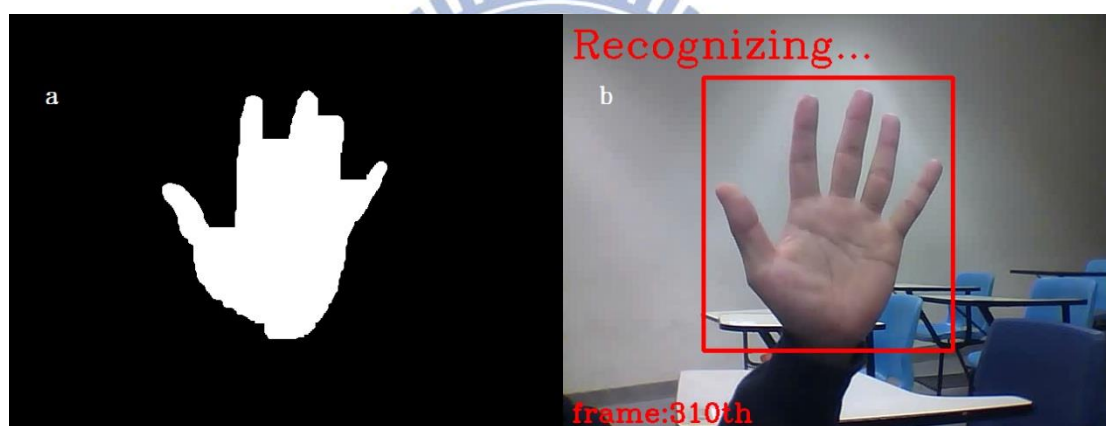


圖 3.2，兩張相減後並二值化，濾波後取出 ROI

### 3.1.2 主動式輪廓模型推導

為了切割較完整的手勢，本論文發展利用等階集合法的主動式輪廓模型。因為欲以形狀做特徵以進行辨識，所以希望此模型的輪廓僅會停在物體的外輪廓上，且較不受雜訊干擾，如圖 3.3 所示。

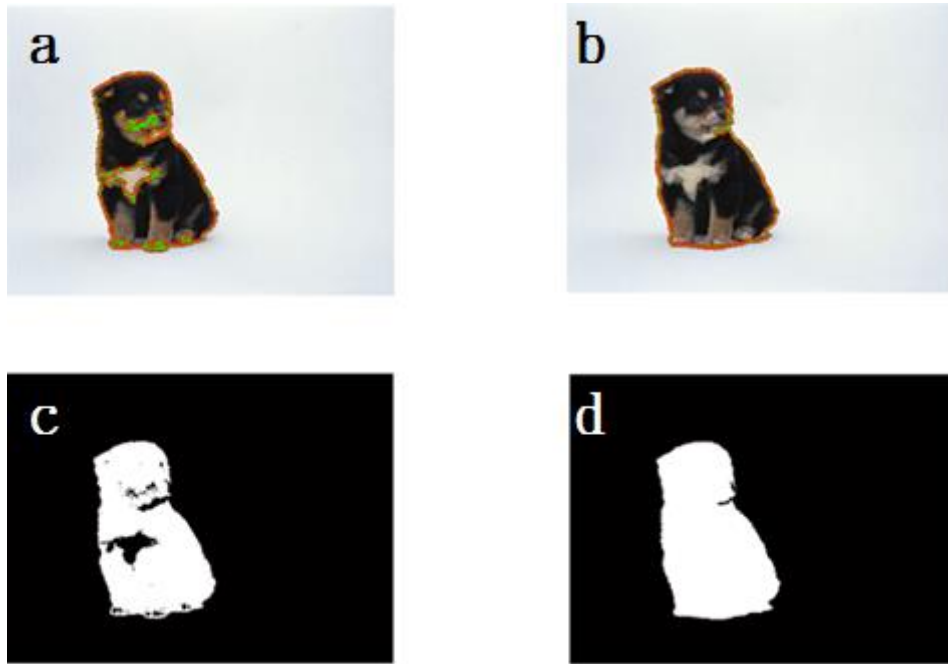


圖 3.3 a 為分割，c 為其二值化後的結果；b 為分割，d 為其二值化後的結果，可發現 b、d 的分割較能描述物體的形狀。

由於影像邊緣部分通常是灰階差異較大的地方，所以以輪廓內外的灰階差當作邊緣能量：

$$E_{edge} = -\frac{1}{2}(u - v)^2 \quad (3.1)$$

其中  $u$  為輪廓上的平均灰階值， $v$  為輪廓外一圈的平均灰階值。通常  $E$  在輪廓位在物體邊界的地方會有最小值，以二值影像(圖 3.4)為例，其  $E$  的最小值

$$\inf(E_{edge}) = -\frac{1}{2}(0 - 255)^2 \quad (3.2)$$

只有輪廓停留在每個連通域的邊緣時，能量才會達到最小值，

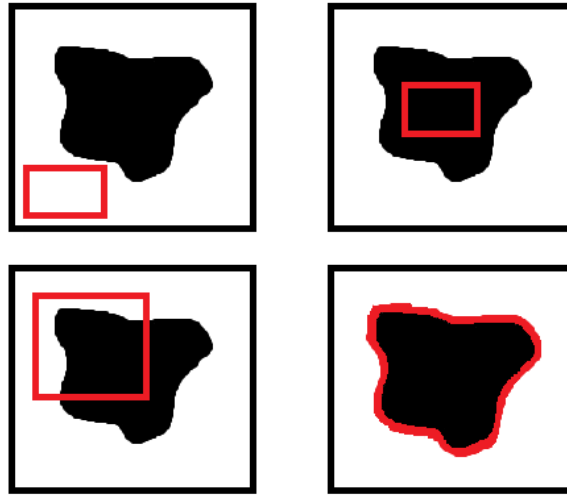


圖 3.4 (左上) $E = 0$  (右上) $E = 0$  (左下) $E = [0, \inf(E)]$  (下) $E = \inf(E)$

為了方便描述輪廓變化(尤其是分離或融合)，本論文利用等階集合

法作為輪廓參數。以  $C \subset \Omega$  表示隱函數  $\varphi = 0$  的集合

$$\begin{cases} C = \partial w = \{(x, y) \in \Omega, \varphi(x, y) = 0\} \\ \text{inside}(C) = w = \{(x, y) \in \Omega, \varphi(x, y) > 0\} \\ \text{outside}(C) = \Omega \setminus \bar{w} = \{(x, y) \in \Omega, \varphi(x, y) < 0\} \end{cases}$$

為了取得輪廓外一圈的資訊，我們以  $C$  演化，向其法向方向演化一格

(圖 3.5)，其中  $\varphi^n = 0$  為目前輪廓， $\varphi^{n+1} = 0$  為其外一圈的輪廓

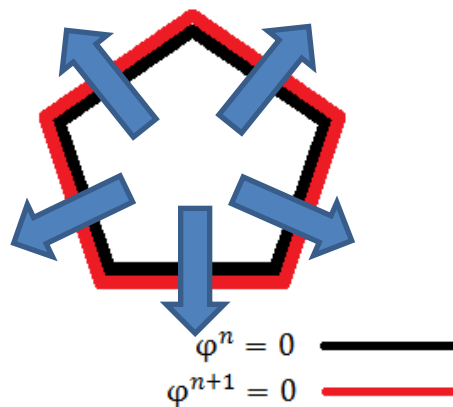


圖 3.5 以  $\varphi^n$  向其法向演化

為了取得輪廓外一圈的數學式，利用描述水平集移動的 Level set

equation

$$\varphi_t = \vec{V} \cdot \nabla \varphi \quad (3.3)$$

上式的數值形式：

$$\frac{\varphi^{n+1} - \varphi^n}{\Delta t} = \vec{V} \cdot \nabla \varphi \quad (3.4)$$

設 $\Delta t=1$  及 $\vec{V}=\vec{N}$ ，得到

$$\varphi^{n+1} = \varphi^n + |\nabla \varphi| \quad (3.5)$$

由於在本論文實驗中會不斷的重新初始化 $\varphi$ ，所以 $\varphi^{n+1}$ 可以寫成更精簡的形式

$$\varphi^{n+1} = \varphi^n + 1 \quad (3.6)$$

由於需在離散空間中表示輪廓長度，我們用正規化(regularized)的脈衝函數 $\delta_r$ ，其由正規化的一階單位步階函數微分求得

$$H_r(z) = \frac{1}{2} \left( 1 + \frac{2}{\pi} \arctan\left(\frac{z}{\varepsilon}\right) \right) \quad (3.7)$$

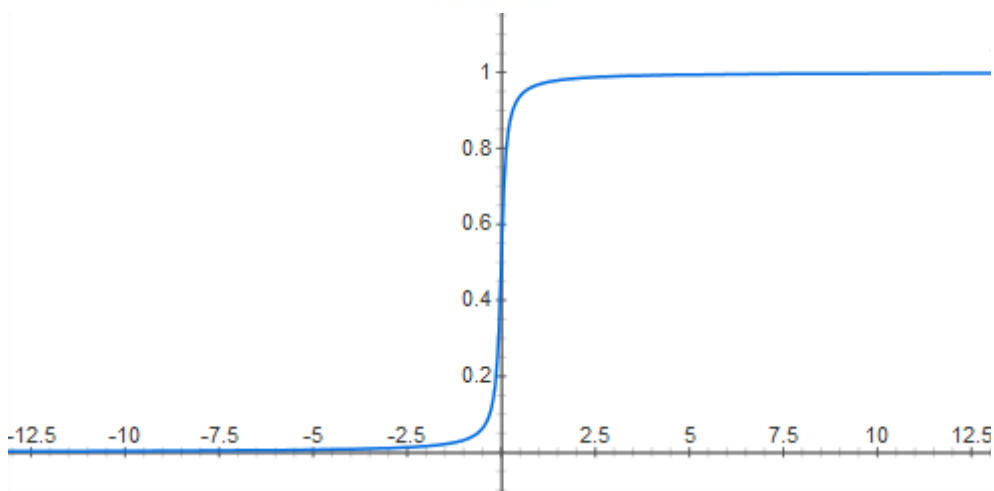


圖 3.6 正規化單位步階函數， $\varepsilon=0.5$

因此可以將 $E_{edge}$  以 Level set method 表示

$$E_{edge} (u, v, \varphi) = -\frac{1}{2}(u - v)^2 \quad (3.8)$$

其中  $u$  及  $v$  為

$$u(\varphi) = \frac{\int_{\Omega} u_0(x,y) \delta_r(\varphi(x,y)) dx dy}{\int_{\Omega} \delta_r(\varphi(x,y)) dx dy} \quad (3.9)$$

$$v(\varphi) = \frac{\int_{\Omega} u_0(x,y) \delta_r(\varphi(x,y) + \nabla\varphi) dx dy}{\int_{\Omega} \delta_r(\varphi(x,y) + \nabla\varphi) dx dy} \quad (3.10)$$

若針對(3.8)做最小化，則會發生  $\varphi$  不會變動的情形，因為(3.8)只參考兩條線的資訊，所以該式並不知道下一步該如何走，故本論文決定改寫[7]提出的分割能量以實現邊緣的想法。利用分割能量

$$E_{separation} (u, v, \varphi) = -\frac{1}{2}(u - v)^2 \quad (3.11)$$

其中  $u$  為  $\varphi \geq 0$  區域的平均灰階值， $v$  為  $\varphi < 0$  區域的平均灰階值，透過最小化 $E_{separation}$  可以以  $\varphi = 0$  分割出灰階差異最大的兩塊區域。

為了求得 $E_{separation}$  的最小值，我們先求得 $\nabla E_{separation}$  再帶入梯度下降法。

令 $S_u = \int_{\Omega} u_0(x,y) H_r(\varphi(x,y)) dx dy$  及  $A_u = \int_{\Omega} H_r(\varphi(x,y)) dx dy$ ，其

中 $u_0(x,y)$  為影像；則根據 Euler-lagrange equation 得到 $\nabla S_u =$

$u_0(x,y)$  及  $\nabla A_u = 1$ ， $\nabla u = \nabla(S_u/A_u)$  為

$$\nabla u = \frac{L_u \nabla S_u - S_u \nabla L_u}{A_u^2} \vec{N} = \frac{L_u I - S_u}{A_u^2} \vec{N} = \frac{I - u}{A_u} \vec{N} \quad (3.12)$$

利用類似的計算 $\nabla v$  為

$$\nabla V = -\frac{I-u}{A_v} \vec{N} \quad (3.13)$$

其中 $L_v$ 為輪廓 $c$ 外一圈的長度。利用以上式子可以寫出 $\nabla E_{separation}$ 為

$$\nabla E_{separation} = -(u-v) \left( \frac{I-u}{A_u} + \frac{I-u}{A_v} \right) \vec{N} \quad (3.14)$$

再利用 $\nabla E_{separation}$ 的負方向帶入梯度下降法，可以得到 $\varphi$ 的演化式

$$\frac{\partial \varphi}{\partial t} = (u-v) \left( \frac{I-u}{A_u} + \frac{I-u}{A_v} \right) \quad (3.15)$$

，而為了只參考邊緣資訊，利用(3.9)及(3.10)算出的 $u$ 及 $v$ 帶入

(3.15)，則可利用(3.15)實現最小化 $E_{edge}$ 。根據此式子， $\varphi=0$ 可演

化到區域性灰階變化劇烈的地方。

### 3.1.3 主動式輪廓模型切割手勢

利用連續影像相減後，可得到大略的變動區域，以此為主動

式輪廓的初始輪廓，開始計算能量的最小值，如圖 3.7 所示。



圖 3.7 以此紅框為初始輪廓

由於此模型僅能切割出內、外兩塊區域，所以以膚色為主動式輪廓的停留基準。在影像相減出的區域中，我們先計算此區塊內的平均 Cr 值，再將 ROI 內的 Cr 通道與之相減，化為

灰階圖並將對比度拉高。如圖 3.8 所示

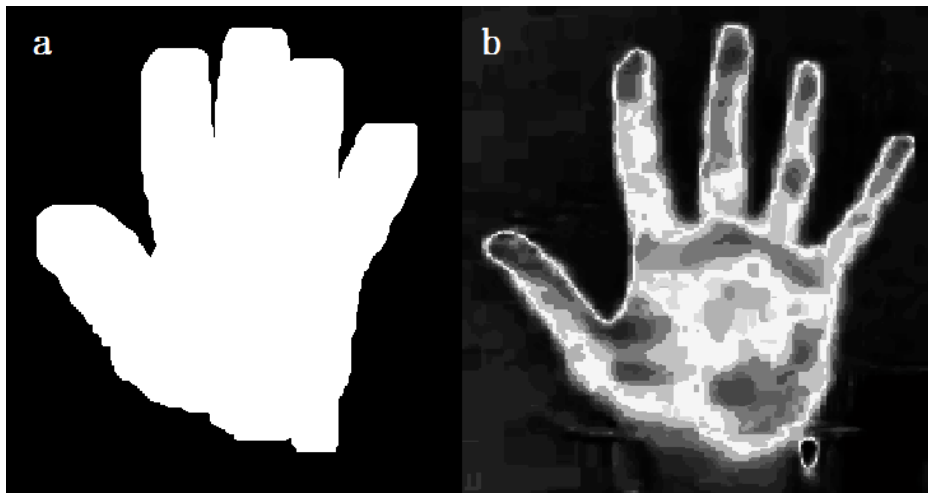


圖 3.8 a 為相減後的區域，以此區域為遮罩，計算此區域的平均 Cr 值，照上述方法處理後即為 b。

在得到 Cr 的差異灰階圖後，便將此圖與初始輪廓帶入主動式輪廓計算(圖 3.9)。在輪廓收斂後，再利用型態學及連通域處理，去掉小塊的面積，得到最後的分割結果。

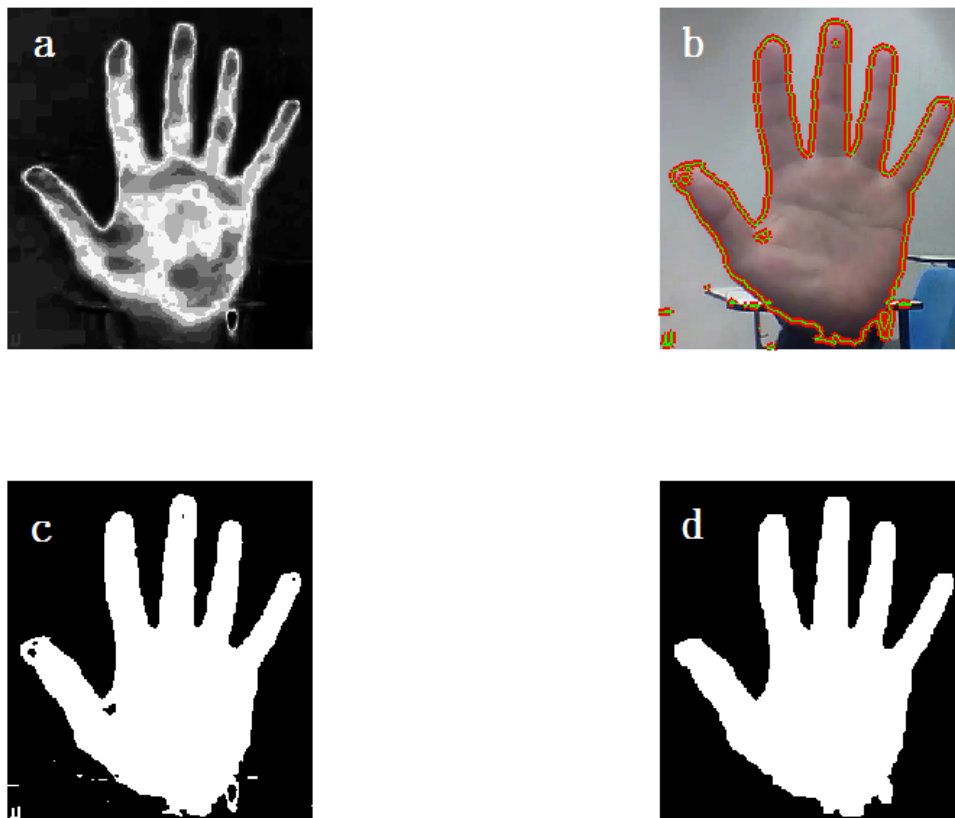


圖 3.9 a:輸入灰階圖 b:等階集合法做切割的結果 c:二值化 d:挑掉較小面積的部分

### 3.2 手掌轉正

為了避免手掌角度影響辨識結果，分割後的手勢均做轉正處理。先利用影像矩，得到手掌重心 $C_0$ 位置，再計算手掌對重心的質量慣性矩陣：

$$I_O = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{yx} & I_{yy} & -I_{yz} \\ -I_{zx} & -I_{zy} & I_{zz} \end{bmatrix} \quad (3.16)$$

由於本論文均處理二維影像，所以上式可簡化為



$$I_O = \begin{bmatrix} I_{xx} & -I_{xy} & 0 \\ -I_{yx} & I_{yy} & 0 \\ 0 & 0 & I_{xx} + I_{yy} \end{bmatrix} \quad (3.17)$$

對 $I_O$ 做特徵值分解可得特徵向量 $V$ ， $V$ 的前兩項 $V_1$ 與 $V_2$ 即為兩個正交軸，也是主軸的方向：

$$\theta = \tan^{-1}(V_2/V_1) \quad (3.18)$$

其中 $\theta$ 為主軸的方向，即手掌在這兩軸上兩側對稱。將垂直角度 $\pi/2$ 與 $\theta$ 相減後便得到待旋轉角度 $\theta_{rotated}$

$$\theta_{rotated} = \pi/2 - \theta \quad (3.19)$$

得到 $\theta_{rotated}$ 後，便可得出旋轉矩陣 $R$

$$R = \begin{bmatrix} \cos \theta_{rotated} & -\sin \theta_{rotated} \\ \sin \theta_{rotated} & \cos \theta_{rotated} \end{bmatrix} \quad (3.20)$$

將手掌原本的每點位置矩陣 $x_o$ 、 $y_o$ ，與 $R$ 相乘後便可得到轉正後的位置矩陣 $x_{rotated}$ 、 $y_{rotated}$ (圖 3.10):

$$\begin{bmatrix} x_{rotated} \\ y_{rotated} \end{bmatrix} = \begin{bmatrix} \cos \theta_{rotated} & -\sin \theta_{rotated} \\ \sin \theta_{rotated} & \cos \theta_{rotated} \end{bmatrix} \begin{bmatrix} x_o \\ y_o \end{bmatrix} \quad (3.21)$$

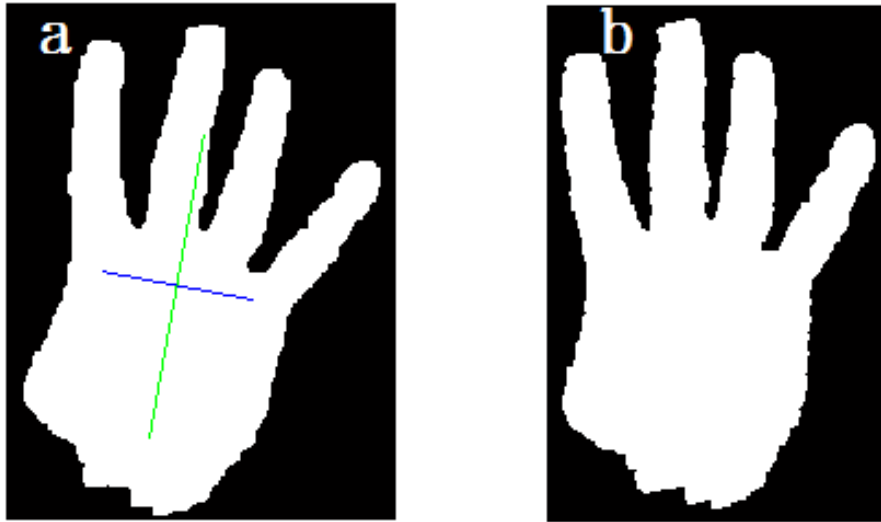


圖 3.10，a 為切割後的手勢，在偵測主軸方向後，便可轉至

垂直如 b

### 3.3 傅立葉描述子

切割後的影像往往仍不夠平滑，為了消除同手勢間因切割的不完整造成的差異，所以用傅立葉描述子平滑化。本論文將原始輪廓做傅立葉轉換後，取其前 40 項係數再轉換回來。圖 3.11 為三個分割的結果作傅立葉描述，a 對應 d、b 對應 e、c 對應 f，利用傅立葉描述使三張的形狀特徵更為相近，去除掉較不重要的形狀凹陷。

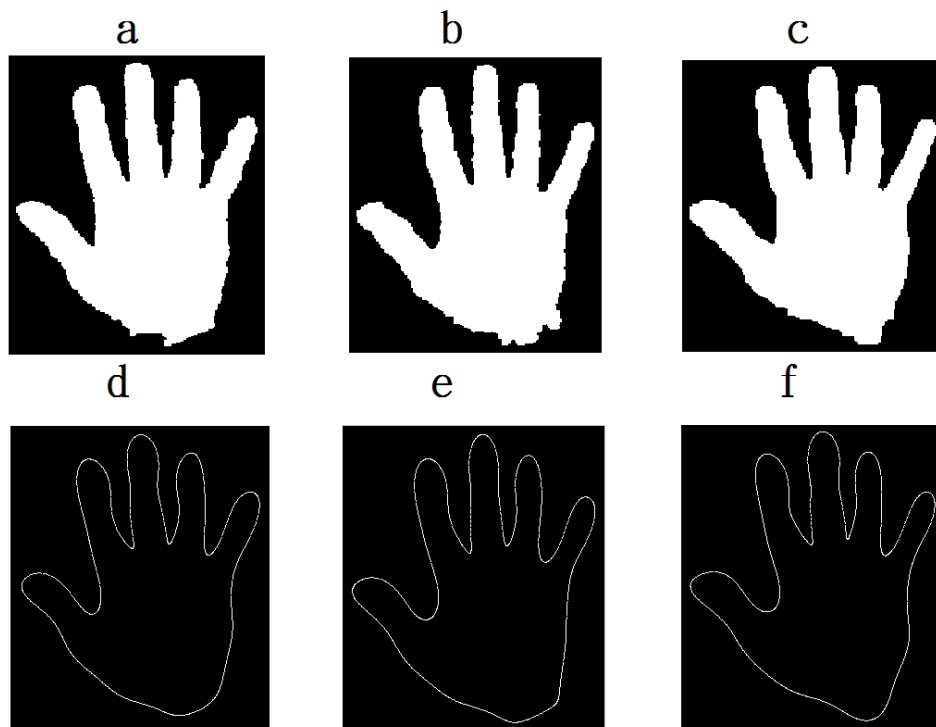


圖 3.11 上排為三個輸入手勢，下排為對應的傅立葉描述子處理

過後的結果

### 3.4 蒐集訓練影像

本論文目標為判斷十種手勢，分別蒐集以下手勢各 70 張，作為訓練的資料庫，蒐集方法為在簡單的場景中，做二值化的結果。由圖 3.12 所示，左上到右下分別為 Gesture0~Gesture9 共十種手勢。

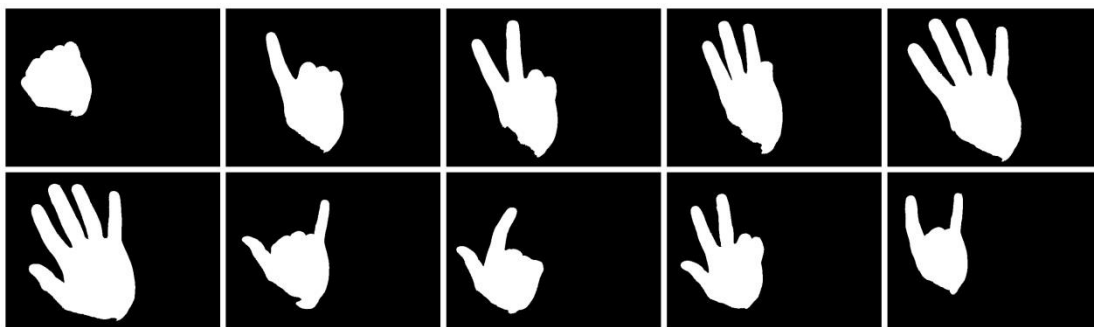


圖 3.12 Gesture0~ Gesture9

為了避免同種手勢間因角度影響的差異，所以拍攝各種角度一併放入資料庫中，如圖 3.13

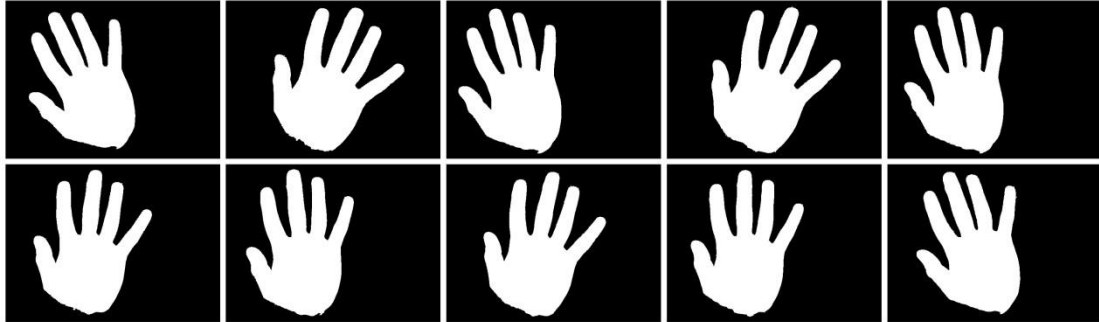


圖 3.13 同個手勢，各種角度

### 3.5 以 SVD 轉換訓練蒐集影像特徵

在蒐集完所有訓練影像後，便將每張所有的影像特徵取出來，以區別每張影像的不同。在取特徵之前，每張訓練影像先做轉正及傅立葉描述，減少角度以及雜訊的影響。

本論文取每張訓練影像的形心，以其為中心，取 1 到 360 度、中心到輪廓的距離，如圖 3.14 所示，取紅點(輪廓上)到藍點(形心)的距離，且除以 360 個距離中的最大值，以標準化摒除手掌大小問題。

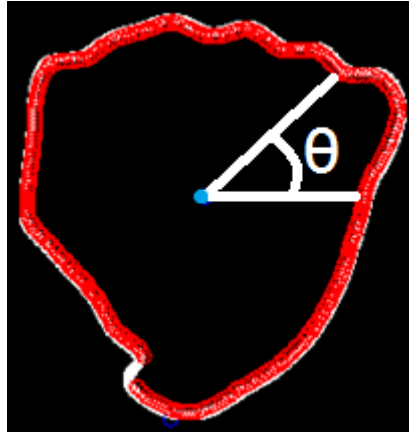


圖 3.14 取角度距離特徵圖

以此方法得到每張影像  $1 \times 360$  的特徵距離向量  $Distance_i$  後

( $i = 1 \sim 3000$ )，蒐集出訓練矩陣  $O$

$$O = \begin{bmatrix} Distance_1 \\ Distance_2 \\ Distance_3 \\ Distance_4 \\ \vdots \\ Distance_{3000} \end{bmatrix} \quad (3.22)$$

針對  $O$  的每一行計算平均值  $\mu$ ，再將  $O$  的每一行向量扣掉  $\mu$

並轉置後，便可得到待分解的矩陣  $O^T$

$$O^T = \begin{bmatrix} Distance_1 - \mu \\ Distance_2 - \mu \\ Distance_3 - \mu \\ Distance_4 - \mu \\ \vdots \\ Distance_{3000} - \mu \end{bmatrix}^T \quad (3.23)$$

對  $O^T$  做奇異值分解可得：

$$O^T = U\Sigma V^* \quad (3.24)$$

其中  $U$  為  $360 \times 360$  的正交矩陣， $V$  為  $3000 \times 3000$  的正交矩陣， $\Sigma$  為  $360 \times 3000$  的主對角矩陣。利用前幾項代表較大變異數空間的性質，本論文取  $\Sigma$  的前  $r$  個奇異值，並去掉  $U$  及  $V$  對應的向量

$$O^T \approx U_{360 \times r} \Sigma_{r \times r} V_{r \times 3000}^* \quad (3.25)$$

其中  $r$  有待實驗分析辨識率再決定其值，表示將 360 個特徵減少至  $r$  個。 $O^T$  奇異值分解後保留訓練影像特徵轉換空間後的矩陣  $Coef$  和投射矩陣  $U_{360 \times r}$ ，待辨識時讀入影像特徵並投射至此空間與之比較：

$$Coef = \Sigma V^* \quad (3.26)$$

上式  $Coef$  即為  $O^T$  降維度後的結果。

### 3.6 利用主成份分析辨識手勢

將每個讀入並切割出手勢的影像，轉正並利用傅立葉描述子平滑化，取出距離特徵的矩陣  $Distance_{1 \times 360}$ ，並且利用訓練階段得到的投射矩陣  $U_{360 \times r}$ ，將此特徵矩陣降低維度

$$Coef\_read = Distance_{1 \times 360} U_{360 \times r} \quad (3.27)$$

其中  $Coef\_read$  為特徵矩陣降維後的結果。將此矩陣再與之

前蒐集並降維後的矩陣 Coef 逐行計算歐幾里得距離 (Euclidean distance)，分別對三千張影像(每個手勢有訓練影像 300 張)計算距離，距離最近的行即代表最像的訓練影像所在的位置，即可知道特徵最相像的訓練影像，如下

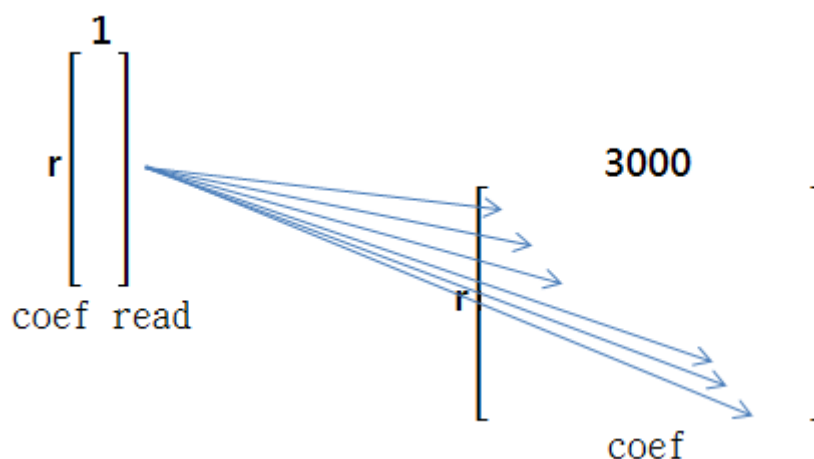
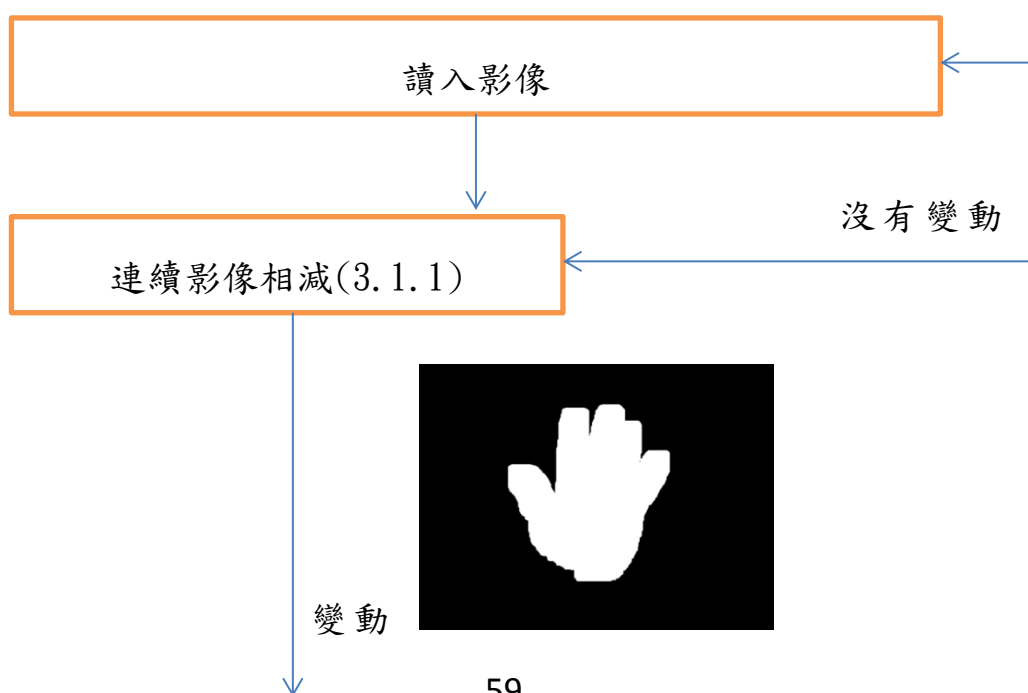


圖 3.15 尋找距離最近的訓練影像

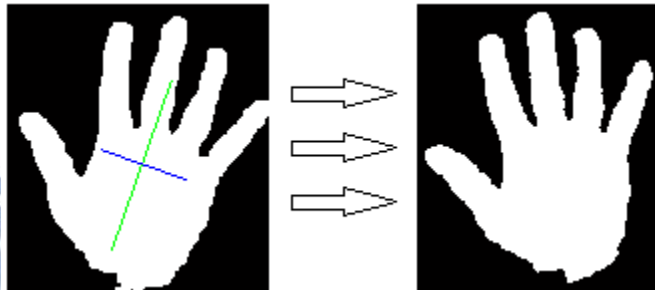
整個手勢辨識系統流程如下：



主動式輪廓切割手勢(3.1.3)



手掌轉正(3.2)



傅立葉描述子(3.3)

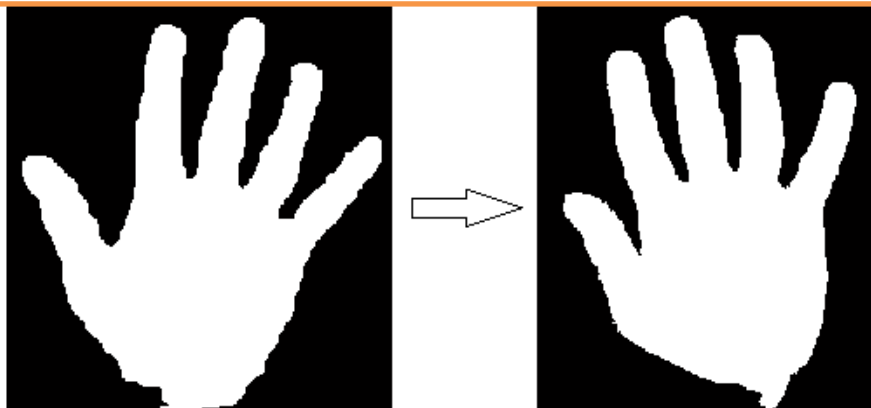


從輪廓取出距離特徵





主成份分析，從資料庫中尋找最像的訓練影像(3.6)



## 第四章 實驗分析與應用

本論文利用 MatLab 及其影像處理工具箱 (Image processing toolbox)，以及利用 C++ 的 Opencv Library 寫成的 mex function 實作本系統，並在 MatLab 下用 visual studio 2010 編譯 mex function。本論文應用此系統的環境為 Intel Core i5-2410M CPU，2.30GHz，讀入影像大小為 640x480x24 bits，八百萬畫素。機械手臂示意模型及實體由本實驗室同仁提供，示意模型及動作利用 Solidworks 的 Motion 模式表現。

本章先分析本論文的影像切割方法之正確性。方法為每個手勢建立十張絕對正確的影像，如圖 4.1 所示：



圖 4.1 左為測試影像，右為本論文建立的絕對正確影像(手勢紅色部分)

再將本論文切割出的手勢影像，與之比較，計算出重疊比率。  
再針對再主成份分析中，取不同的特徵數分別測試，找到足夠的特徵數以進行辨識，並對不同環境對辨識率造成的影響作分析，再測試本系統辨識資料庫以外的受試者的辨識率，最後利用本套系統控制機械手臂

#### 4.1 輸入手勢

本論文將以辨識以下十種手勢為目標：



圖 4.2 從左上到右下分別為 Gesture0~ Gesture9，共十種手勢

辨識環境：交通大學機械系電腦整合設計與製造實驗室。



圖 4.3 交通大學機械系電腦整合設計與製造實驗室一隅

## 4.2 辨識速度

本節量測本論文為了完成手勢辨識而使用的各個步驟所執行的時間，以執行 100 張後的時間取平均並做記錄。由於本系統為 matlab 與 c 的混合編譯，以下將標明每個步驟所耗的時間及開發的平台以供參考。

表格 4.1 執行時間表

	平均執行時間(ms)	開發平台
中值濾波平滑	1	C
連續影像相減	0.7	C
主動式輪廓	80	Matlab
手掌轉正	60	Matlab

傅立葉描述子	20	Matlab
特徵抽取	50	Matlab
主成份分析作辨識	2	C
整體流程(一張)	220	

### 4.3 手勢分割實驗

本節實驗本論文提出的影像分割方法的正確率，方法為計算 Region1，即人為標記正確的區域，和 Region2，即手勢切割出的區域，並計算 R1 和 R2，選數值較小者。當值越接近 1，代表兩區域越吻合；越接近 0，代表兩區域距離越分開。

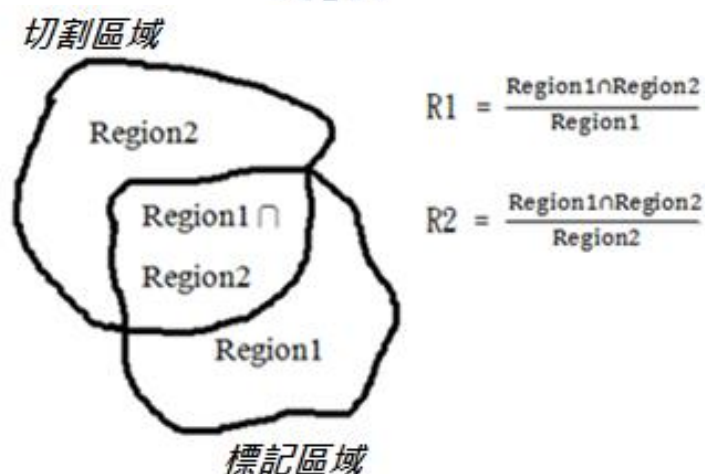


圖 4.4 計算重疊區域

每個手勢測試 10 張，如下圖：

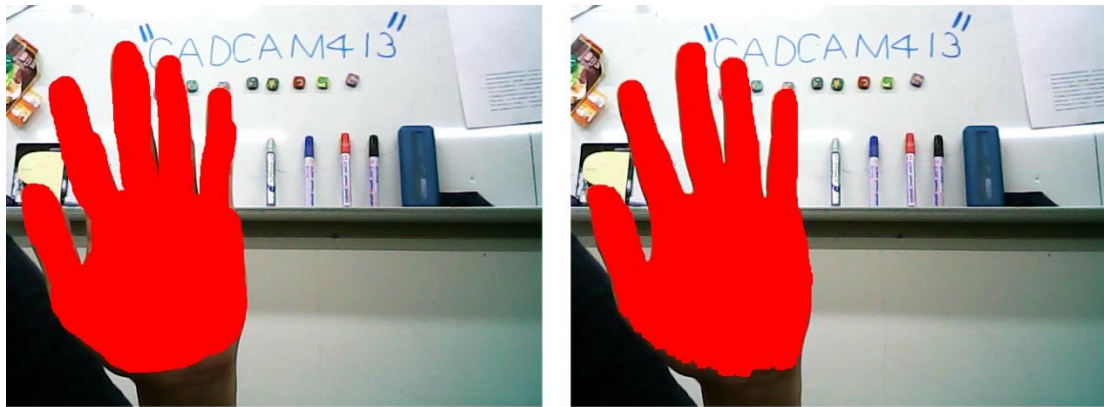


圖 4.5 左邊為標記的正確手勢區域，右為手勢分割出的區域

實驗結果如下：

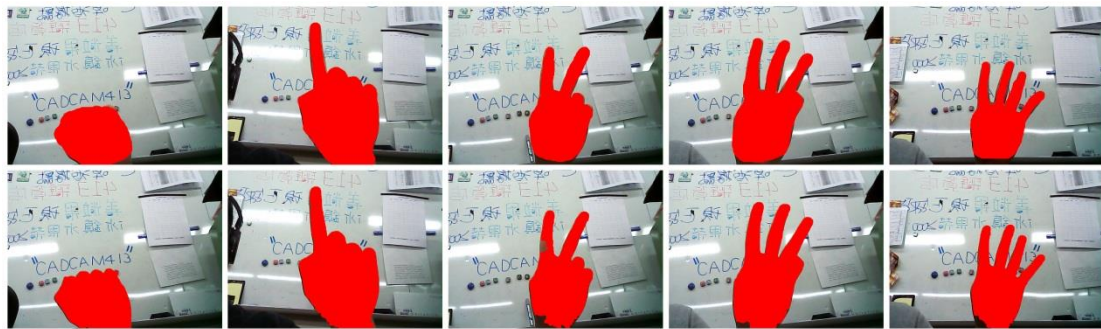


圖 4.6 Gesture0~ Gesture4 的處理結果，上排為切割出的手勢，下排為人工標記的基準

表格 4.2 Gesture0~ Gesture4 手勢切割實驗結果

輸入手勢 張	Gesture0	Gesture1	Gesture2	Gesture3	Gesture4
1	0.956977	0.960756	0.933509	0.883736	0.938445
2	0.915992	0.932259	0.936525	0.95107	0.930126
3	0.921623	0.964721	0.886009	0.927498	0.947728
4	0.910617	0.94508	0.918229	0.952986	0.90137
5	0.887476	0.958874	0.91427	0.913908	0.912591
6	0.926035	0.958273	0.90879	0.928908	0.943179
7	0.916859	0.971879	0.934648	0.960511	0.942588
8	0.78022	0.955317	0.92318	0.939999	0.941194
9	0.82999	0.97207	0.942977	0.944068	0.940063
10	0.803103	0.954972	0.929696	0.939103	0.911221

平均重疊率(%)	88.5	95.7	92.2	93.4	93.1
----------	------	------	------	------	------

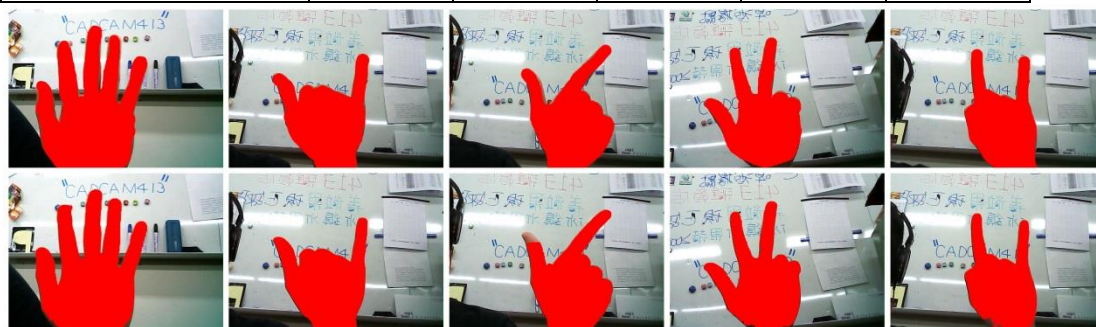


圖 4.7 Gesture5~ Gesture9 的處理結果，上排為切割出的手勢，下排為人工標記的基準

表格 4.3 Gesture5~ Gesture9 手勢切割實驗結果

輸入手勢 張	Gesture5	Gesture6	Gesture7	Gesture8	Gesture9
1	0.873782	0.95425	0.937382	0.952996	0.947762
2	0.865159	0.940463	0.935318	0.955841	0.952597
3	0.900553	0.936061	0.9415	0.932651	0.966456
4	0.925366	0.946238	0.949762	0.944771	0.957252
5	0.888229	0.937159	0.925465	0.939605	0.959938
6	0.902285	0.959843	0.9552	0.956347	0.943578
7	0.928902	0.918256	0.936715	0.959783	0.931788
8	0.953487	0.933581	0.946787	0.948852	0.933056
9	0.941815	0.890242	0.929873	0.957226	0.950251
10	0.936973	0.946598	0.949488	0.955092	0.94119
平均重疊率(%)	91.2	93.6	94.1	95.0	94.9

#### 4.4 特徵維度對辨識率的影響

本論文在辨識過程中利用主成份分析(PCA)，降低特徵矩陣的維度。在這節我們針對  $r$ ，即特徵矩陣的維度，對

辨識率的影響作實驗。我們將  $r$  設為

$$\begin{cases} r = 360 \\ r = 100 \\ r = 50 \\ r = 25 \end{cases}$$

其中  $r = 360$  意謂不做主成份分析，即不做降維，直接比較

距離最相近的特徵矩陣。實驗方法為：Gesture0~ Gesture9

每個手勢各辨識一百張，並記錄輸入的手勢與辨識出的手勢。

以下為分別的測試結果：

表格 4.4 辨識率， $r = 360$

輸入手勢 辨識結果	Gesture0	Gesture1	Gesture2	Gesture3	Gesture4
Gesture0	99	0	3	0	0
Gesture1	0	93	24	1	1
Gesture2	0	1	72	0	5
Gesture3	0	0	0	83	0
Gesture4	0	0	0	3	88
Gesture5	0	0	0	7	3
Gesture6	1	0	0	0	1
Gesture7	0	6	0	2	1
Gesture8	0	0	1	0	1
Gesture9	0	0	0	4	0
辨識率(%)	99	93	72	83	88
輸入手勢 辨識結果	Gesture5	Gesture6	Gesture7	Gesture8	Gesture9
Gesture0	0	6	0	0	0
Gesture1	0	0	14	0	5
Gesture2	0	0	0	5	0
Gesture3	0	0	0	0	0
Gesture4	0	0	0	0	0



Gesture5	100	0	0	0	0
Gesture6	0	89	2	0	1
Gesture7	0	5	83	0	0
Gesture8	0	0	0	95	0
Gesture9	0	0	1	0	94
辨識率(%)	100	89	83	95	94

表格 4.5 辨識率， $r = 100$

輸入手勢 \ 辨識結果	Gesture0	Gesture1	Gesture2	Gesture3	Gesture4
Gesture0	100	1	1	0	0
Gesture1	0	99	12	0	1
Gesture2	0	0	86	0	4
Gesture3	0	0	0	93	0
Gesture4	0	0	0	6	92
Gesture5	0	0	0	1	3
Gesture6	0	0	1	0	0
Gesture7	0	0	0	0	0
Gesture8	0	0	0	0	0
Gesture9	0	0	0	0	0
辨識率(%)	100	99	86	93	92
輸入手勢 \ 辨識結果	Gesture5	Gesture6	Gesture7	Gesture8	Gesture9
Gesture0	0	0	0	0	0
Gesture1	1	0	4	0	0
Gesture2	0	0	0	1	0
Gesture3	0	0	0	0	0
Gesture4	0	0	0	0	0
Gesture5	99	0	0	0	0
Gesture6	0	84	8	0	0
Gesture7	0	16	88	0	0
Gesture8	0	0	0	99	0

Gesture9	0	0	0	0	100
辨識率(%)	99	84	88	99	100

表格 4.6 辨識率， $r = 50$

輸入手勢 \ 辨識結果	Gesture0	Gesture1	Gesture2	Gesture3	Gesture4
Gesture0	100	0	0	0	0
Gesture1	0	93	4	0	0
Gesture2	0	0	96	0	5
Gesture3	0	0	0	96	0
Gesture4	0	0	0	4	91
Gesture5	0	0	0	0	4
Gesture6	0	5	0	0	0
Gesture7	0	2	0	0	0
Gesture8	0	0	0	0	0
Gesture9	0	0	0	0	0
辨識率(%)	100	93	96	96	91
輸入手勢 \ 辨識結果	Gesture5	Gesture6	Gesture7	Gesture8	Gesture9
Gesture0	0	0	0	0	0
Gesture1	0	0	4	0	0
Gesture2	0	0	0	8	0
Gesture3	0	0	0	0	0
Gesture4	0	0	0	0	0
Gesture5	100	0	0	0	0
Gesture6	0	81	0	0	0
Gesture7	0	19	96	0	0
Gesture8	0	0	0	92	0
Gesture9	0	0	0	0	100
辨識率(%)	100	81	96	92	100

表格 4.7 辨識率， $r = 25$

輸入手勢 \ 辨識結果	Gesture0	Gesture1	Gesture2	Gesture3	Gesture4
Gesture0	100	6	0	0	1
Gesture1	0	78	0	1	1
Gesture2	0	0	97	1	3
Gesture3	0	0	0	92	1
Gesture4	0	0	0	0	87
Gesture5	0	0	0	4	6
Gesture6	0	6	0	1	1
Gesture7	0	10	0	0	0
Gesture8	0	0	3	1	0
Gesture9	0	0	0	0	0
辨識率(%)	100	78	97	92	87
輸入手勢 \ 辨識結果	Gesture5	Gesture6	Gesture7	Gesture8	Gesture9
Gesture0	1	5	0	0	1
Gesture1	4	0	14	0	0
Gesture2	0	0	0	15	0
Gesture3	0	0	0	0	0
Gesture4	0	0	0	0	0
Gesture5	93	0	0	0	0
Gesture6	1	94	8	0	0
Gesture7	0	1	76	0	0
Gesture8	0	0	0	85	0
Gesture9	1	0	2	0	99
辨識率(%)	93	94	76	85	99

由實驗數據中可以發現，若辨識不做 PCA，則整體辨識率比做 PCA 的辨識率來的低。而進一步觀察 PCA 中效果較好的為  $r = 100$  及  $r = 50$ ，因此本論文將以  $r = 50$  作為預設的維度，

以下實驗也以此維度做為固定參數。

#### 4.5 其他受試者的辨識率

本節對未經訓練的他人手勢進行辨識，一樣實驗 Gesture0~ Gesture9 十種手勢，觀察本人與他人的辨識率差異。實驗方法為：受試者 Gesture0~ Gesture9 每個手勢各辨識一百張，並記錄輸入的手勢與辨識出的手勢。

辨識結果

表格 4.8 他人辨識率

輸入手勢 \ 辨識結果	Gesture0	Gesture1	Gesture2	Gesture3	Gesture4
Gesture0	100	0	0	0	0
Gesture1	0	86	0	41	13
Gesture2	0	0	83	4	2
Gesture3	0	0	0	53	0
Gesture4	0	0	0	0	70
Gesture5	0	0	0	1	15
Gesture6	0	14	0	0	0
Gesture7	0	0	0	1	0
Gesture8	0	0	17	0	0
Gesture9	0	0	0	0	0
辨識率(%)	100	86	83	53	70
輸入手勢 \ 辨識結果	Gesture5	Gesture6	Gesture7	Gesture8	Gesture9
Gesture0	0	2	0	0	0
Gesture1	1	0	0	0	8

Gesture2	0	0	0	0	0
Gesture3	0	0	0	0	0
Gesture4	0	0	0	1	0
Gesture5	94	0	0	0	0
Gesture6	3	98	21	0	2
Gesture7	1	0	78	0	0
Gesture8	0	0	0	99	0
Gesture9	1	0	1	0	90
辨識率(%)	94	98	78	99	90

從結果可歸納出，雖然大部分辨識率有 80%以上，但因為 Gesture3 與 Gesture4 受試者的手型輪廓不符訓練影像，所以造成辨識率過低的情形。這表示 Gesture3 與 Gesture4 僅以一個人蒐集樣本是不夠的，不符實際上的應用要求，需再蒐集其他人的手勢。圖 4.8 與圖 4.9 所示：

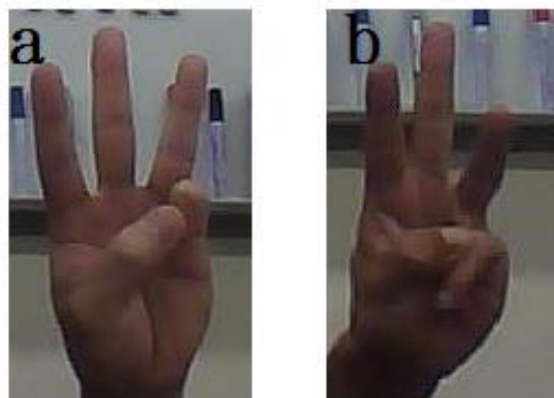


圖 4.8 左:訓練者的 Gesture3，右:他人的 Gesture3

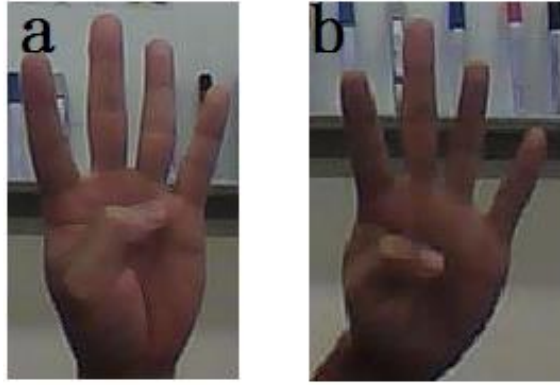


圖 4.9 左:訓練者的 Gesture4，右:他人的 Gesture4

#### 4.6 手勢控制滑鼠游標

為求廣泛性，本論文將本系統應用在控制滑鼠游標的介面上，分別為

Gesture1:鼠標向上移動

Gesture3:鼠標向右移動

Gesture6:鼠標向下移動

Gesture9:鼠標向左移動

Gesture0:選取

Gesture5:取消選取

Gesture2:鼠標移動速度變緩

Gesture4:鼠標移動速度加快

Gesture8:啟動手勢辨識功能

## Gesture7:關閉手勢辨識功能

### 4.7 手勢控制機械手臂

目前機械手臂在工業界早已成為不可或缺的幫手，自從機器手臂技術開始發展，機器手臂已成功的應用於汽車製造業等產業，而許多工業危險之繁重工作如組裝、噴漆、焊接、高溫鑄鍛等，皆能以機器手臂取代人工作業。目前機械手臂在機器人技術領域中得到最廣泛實際應用的自動化機械裝置，除了主要用於工業製造上，商業農業、醫療救援、娛樂服務、軍事保全甚至在太空探索等領域都可以發現其應用裝置。

本實驗室完成之機械手臂模型(圖 4.10)，可模擬一般機械手臂的應用，例如舉重、搬運貨物、甚或遞取物品等，能夠替代操控者，尤其是行動不便的使用者，達到功能需求。

本論文完成一控制機械手臂之介面，將此介面與本實驗室提供之機械手臂結合，可利用手勢控制可選取伺服馬達角速度，及動作的馬達(一號或二號或一號和二號同時)和往上或往下的模式，三個參數設定好後，選取 Start 鍵即開始動

作。

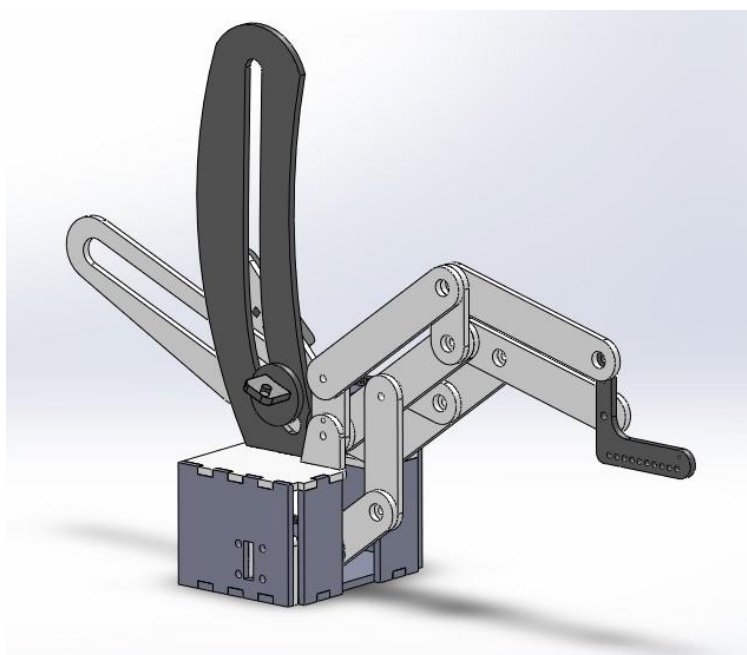


圖 4.10 本實驗室提供機械手臂模型圖

本機械手臂為兩個自由度的系統，利用不同的馬達動作可以達到不同的目的：

動作 1： 2 號伺服馬達順時針旋轉

此動作可以控制該機械手臂末端平緩的往上圓弧運動，並因平行四連桿機構，末端元件只會有平移，不會有角速度。



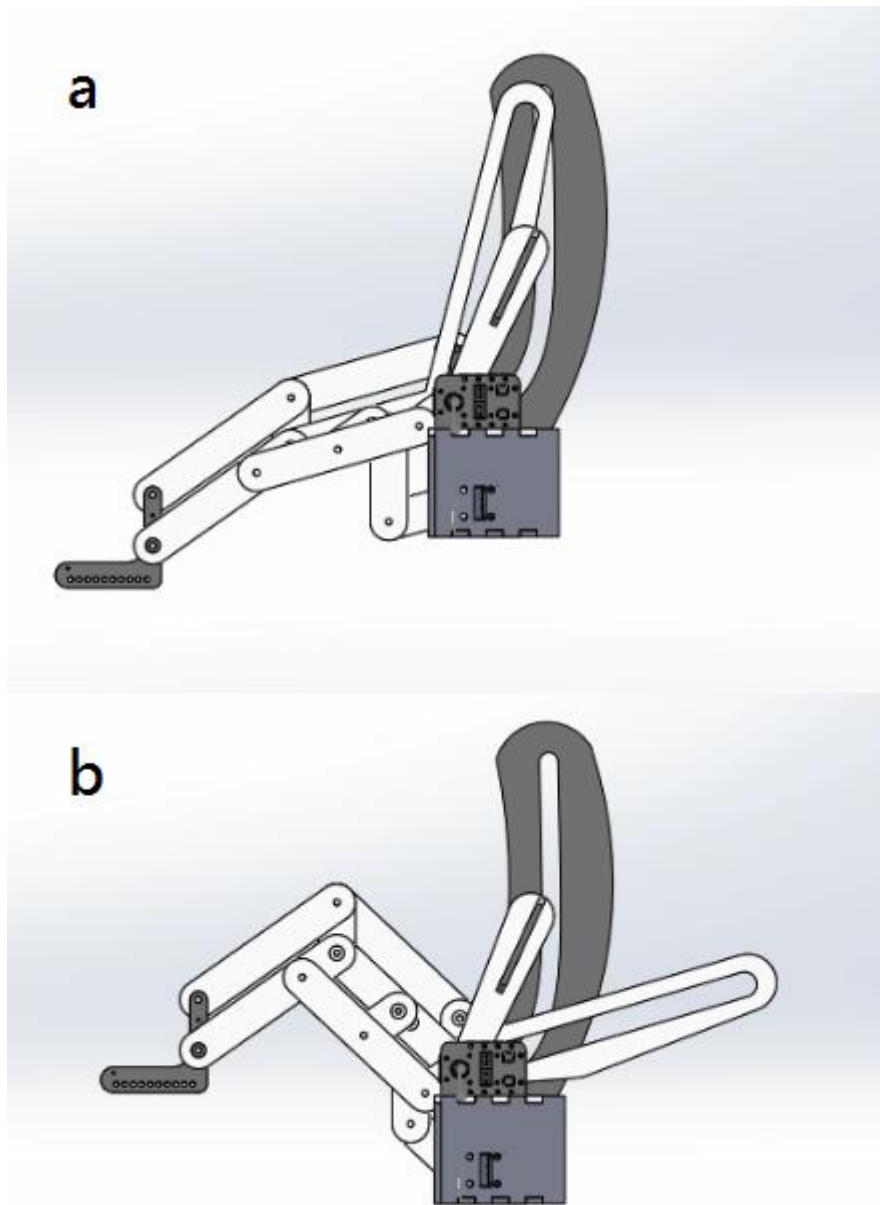


圖 4.11 由 a 圖至 b 圖為 2 號伺服馬達順時針旋轉

動作 2: 1 號伺服馬達順時針旋轉

此動作可以控制該機械手臂末端較快速的往上圓弧運動，並且往上幅度較動作 1 大。因這裡也包含平行四連桿機構，末端元件只會有平移，不會有角速度。

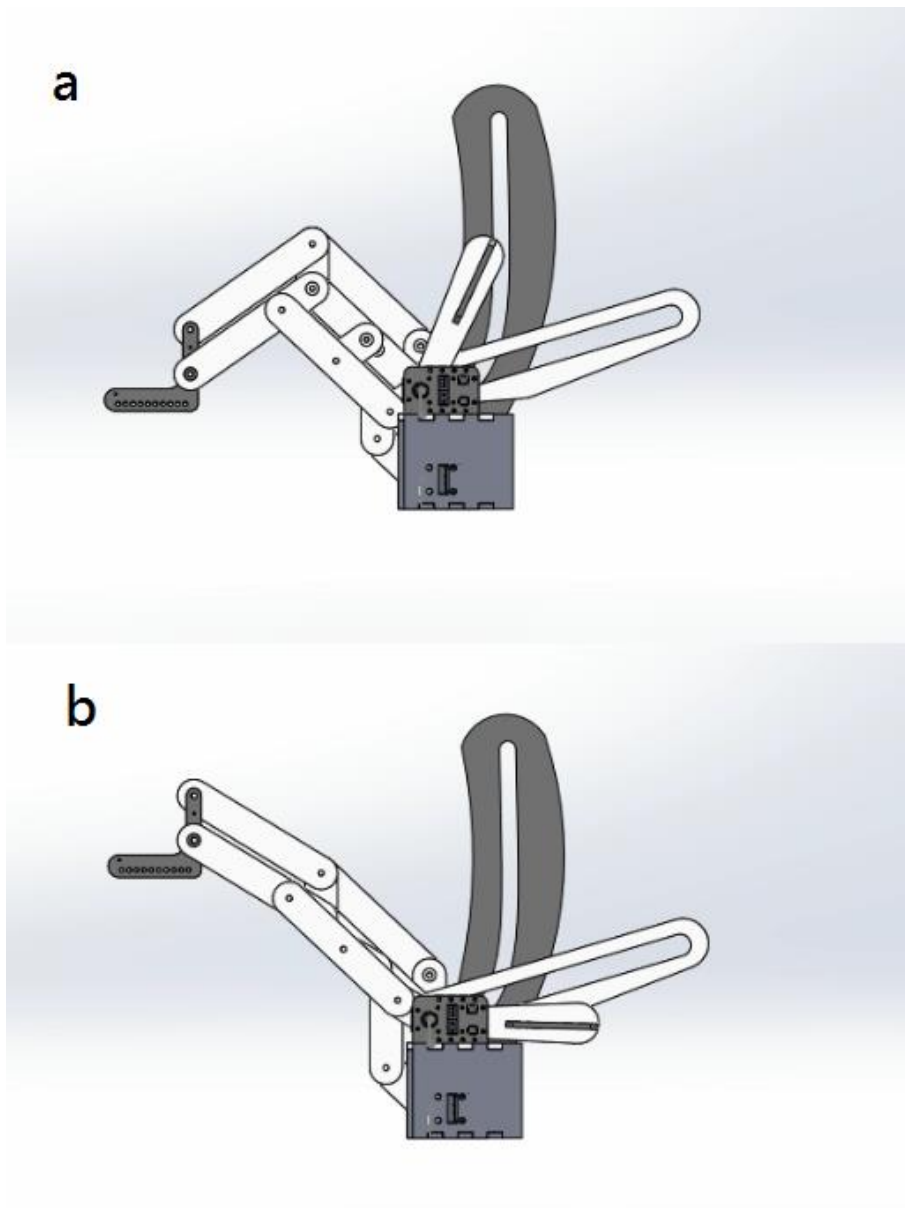


圖 4.12 由 a 圖至 b 圖為 1 號伺服馬達順時針旋轉

動作 3: 動作 2 的逆向

動作 4: 動作 1 的逆向

設備與軟硬體：

(1). 個人電腦

(2). 控制板，Innobasic education board(圖 4.13)

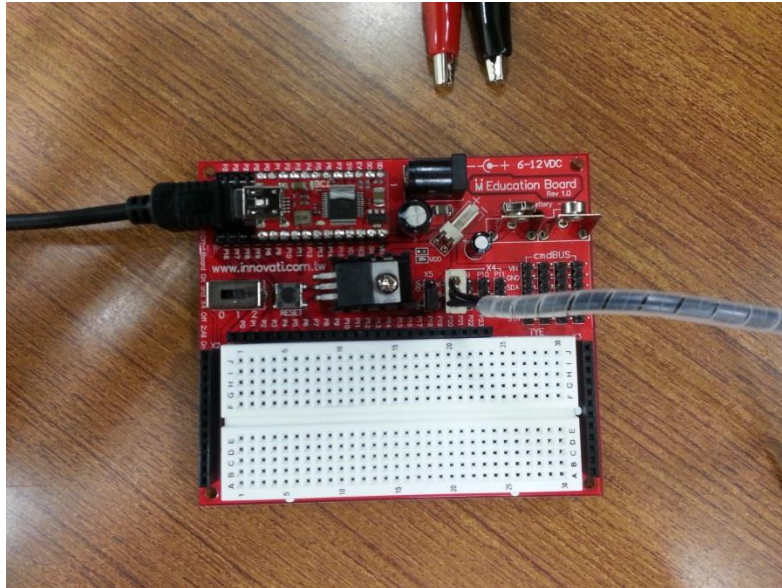


圖 4.13 伺服馬達控制板

(3). 伺服馬達外接電源供應器

(4). 機械手臂主體(圖 4.14)

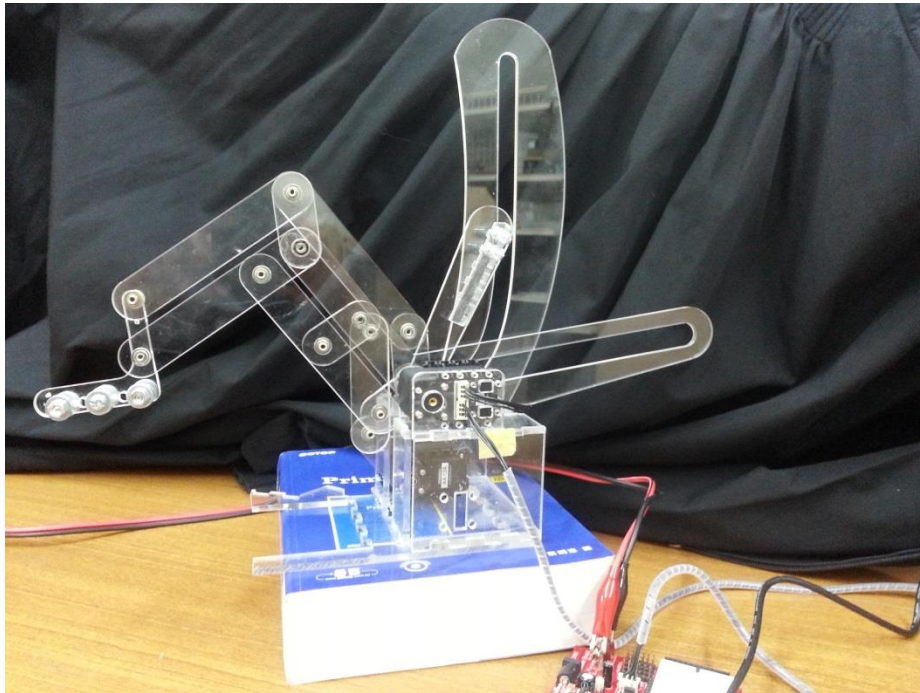


圖 4.14 機械手臂實體

(5). 控制程式: innoBASIC Workshop2

控制介面如下:



圖 4.15 專題系統識別視窗

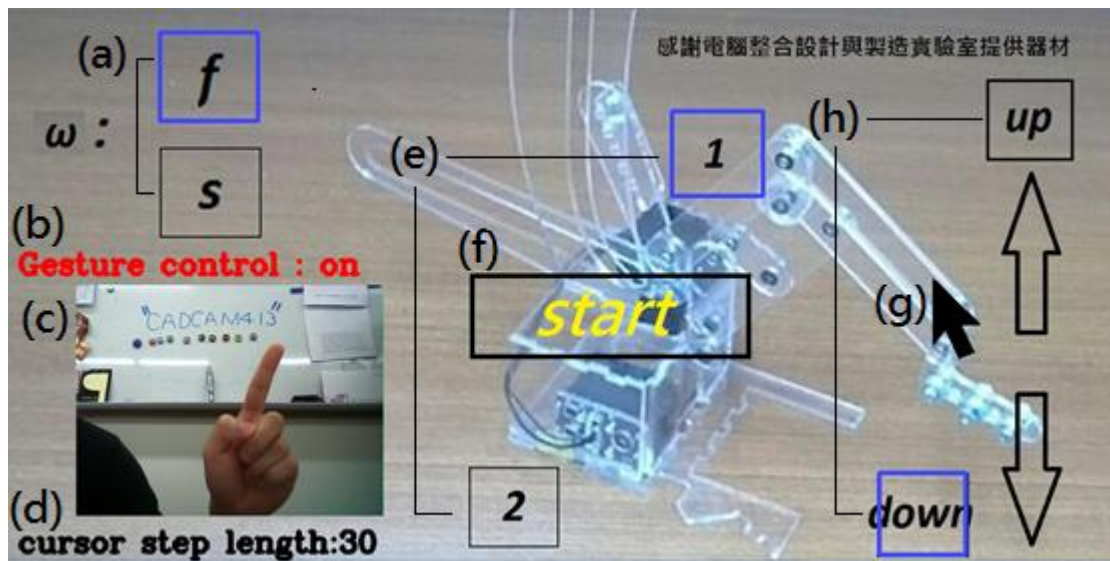


圖 4.16 伺服馬達控制圖形介面

在圖 4.16 中，各標記意義如下

(a): 選取伺服馬達角速度，可選擇快[f]、慢[s]兩種模式。

(b): 是否有啟動手勢控制，若有，顯示[on]；沒有則顯示 [off]，可用 Gesture7 和 Gesture8 開關。

(c): 顯示輸入畫面

(d): 鼠標每次移動的距離，可用 Gesture2 和 Gesture4 調校。

(e): 選取動作的馬達，可選取一號伺服機動作或二號伺服機動作或兩者同動，

(f): 將所有的按鈕確定後，按下此鍵即開始動作。

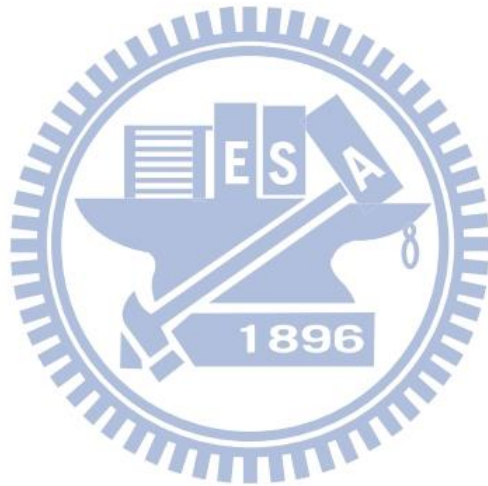
(g): 選取馬達的動作為上[up]或下[down]。

控制方法：

利用 C++ 的 API (Application programming interface)，將選取完的參數確認並按下 start 後，將指令送到 innoBASIC Workshop2 的命令視窗上，對照預先寫的腳本動作。

實驗結果：

執行動作 1：



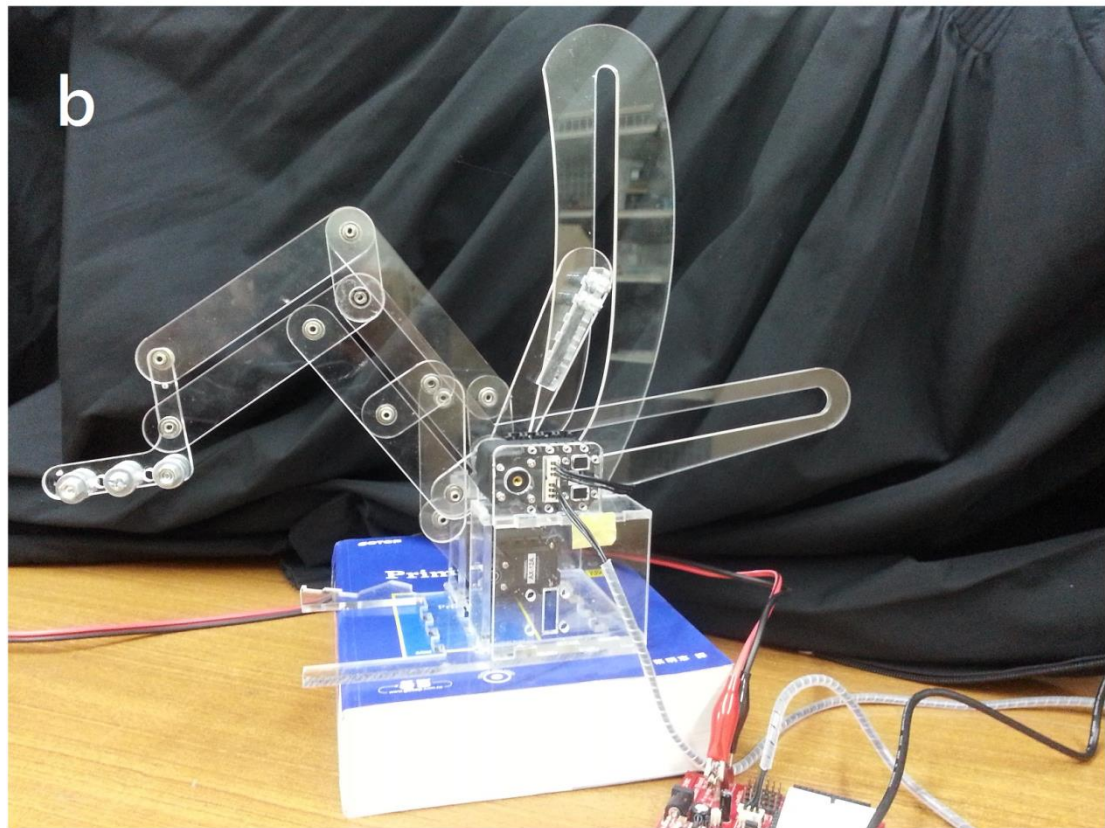
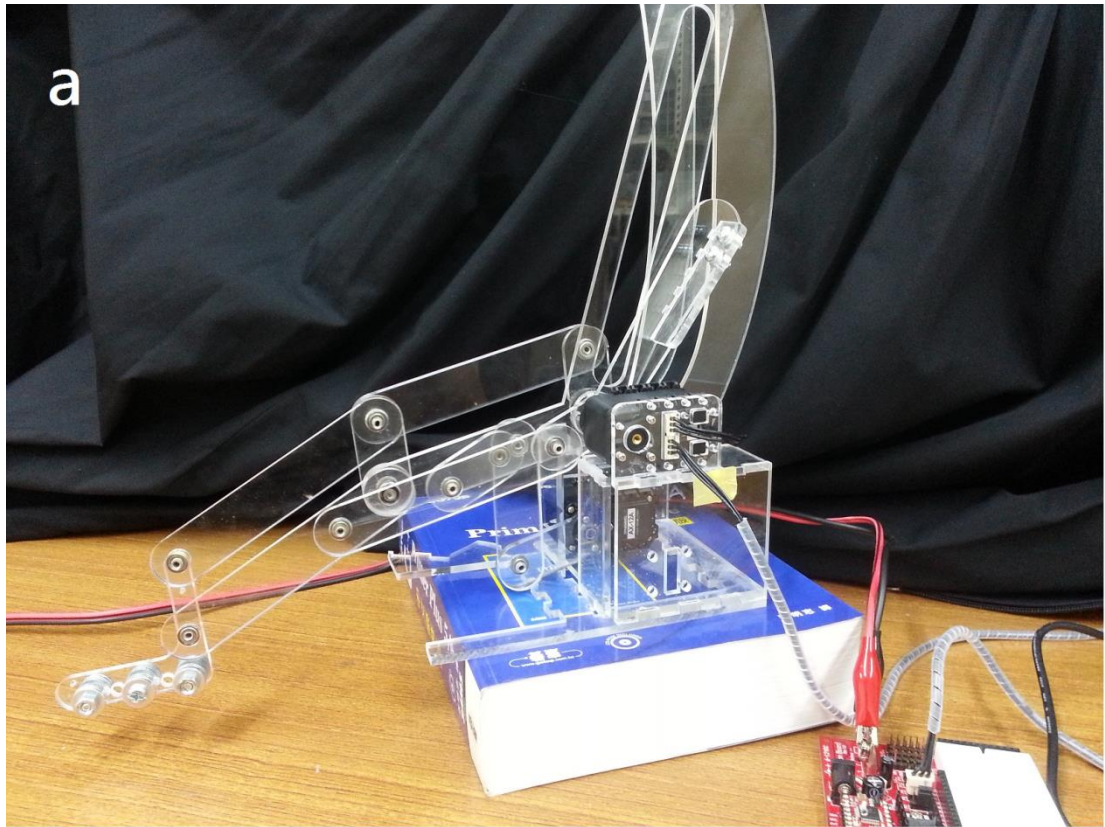


圖 4.17 動作 1

執行動作 2:

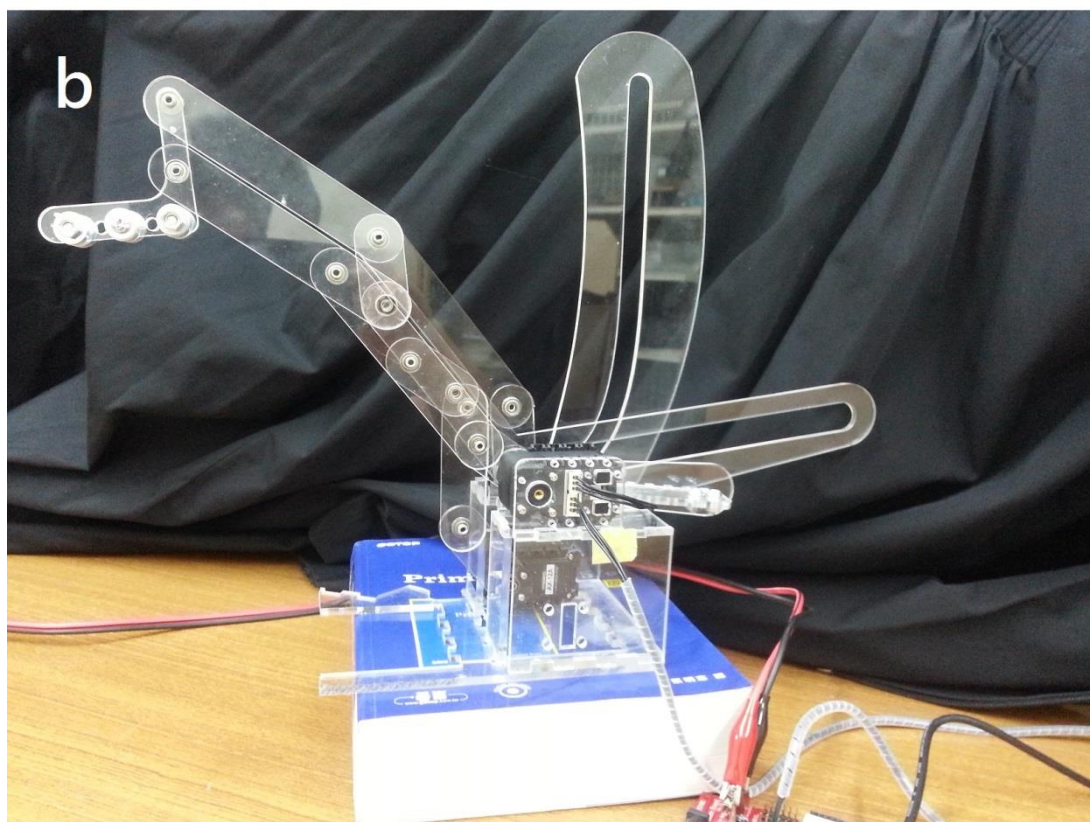
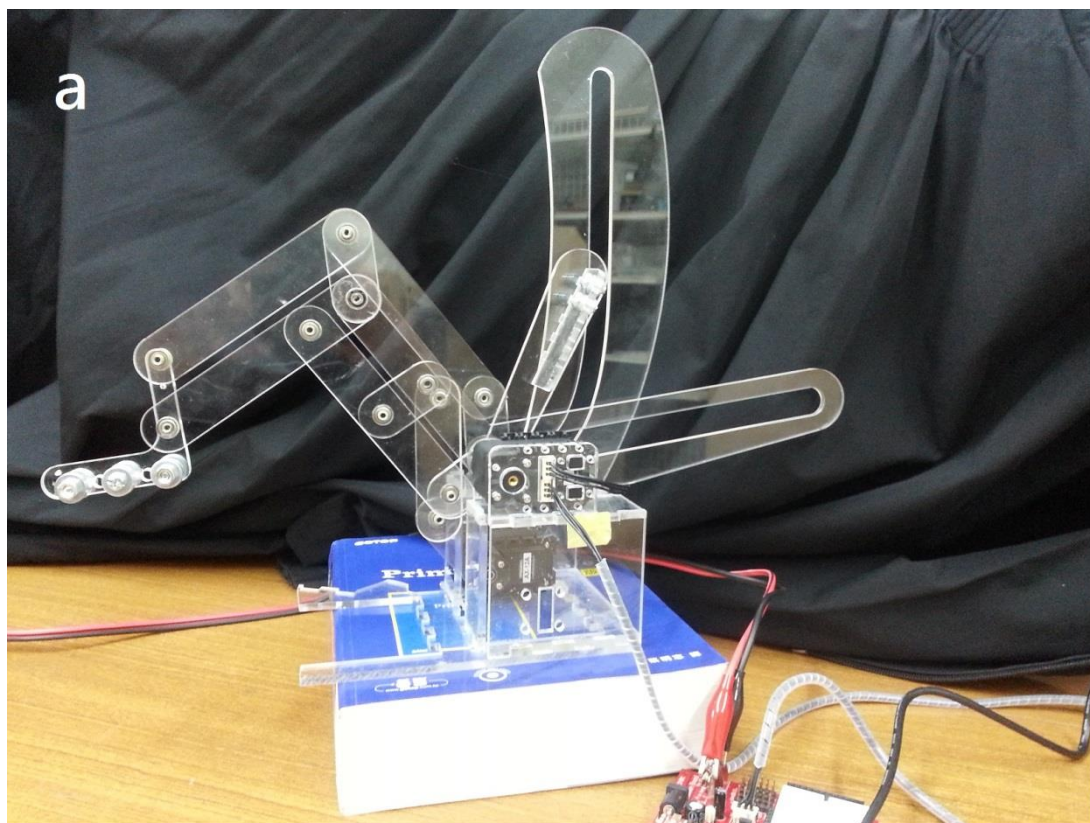




圖 4.18 動作 2

## 4.8 實驗結論

在手勢分割的步驟中，本論文先對每個手勢建立十張完全正確的比較基準影像，再利用等階集合法對這些手勢影像做手勢分割，將分割出的手勢與基準影像比較並計算重疊率。實驗結果為每個手勢的重疊率至少都有 88% 以上，足以提供辨識階段做正確的判斷。

在手勢辨識的步驟中，本研究就是否加入影像特徵的主成份分析程序進行實驗分析與比較。而由主成份分析實驗數據可看出，經由僅挑出變異數較大的特徵(即降維度法則)再與訓練用之影像特徵比較，其正確率會比不採用主成份分析程序來的高。而利用主成份分析進行辨識時，由於本研究希望以最少的特徵比對次數達成可接受的辨識率，所以另又規劃維度的大小對辨識率影響的實驗分析。結果顯示維度從 360 降到 50 時，其辨識率大約是 85% 以上，為可接受的範圍；然而降到 25 便開始出現辨識率降低的情形，故將維度大小  $r$  確定為 50。在學習步驟中，便把前 50 大的奇異值保留，並保留對應的投射矩陣  $U$ 。辨識手勢時，取出的特徵矩陣與  $U$  矩陣相乘，藉以降低特徵維度為 50，將前述特徵矩陣分別對所有訓練影像特徵矩陣的各對應單元做距

離長度的計算；尋找歐幾里得距離最小，即影像特徵最相近的影像，完成辨識。

而對未經訓練的手勢進行辨識時，會發現有少部分手勢有辨識率過低的情形，此情形表示或許為訓練時的樣本數不夠，若欲改善此點而提升本系統的廣泛性，學習時需加入差異性大且多人的手勢進行訓練，例如不同的年齡層或是不同的人組成的訓練者；也有可能是角度距離特徵無法很完善的描述手形的原因，因為手部形心至手指輪廓部分通常有阻隔，導致無法完善描述手輪廓變化較大的部分(即較有特徵的部分)，這點或許可以以更好的特徵描述方法替代並改善。

本論文最後設計具有廣泛性的介面，利用手勢影像識別控制游標，並再進一步將此介面用以控制本實驗室設計並組裝完成的機械手臂做為範例，且目前辨識速度與成功率大致符合應用上的需求。未來可將此套系統應用在日常生活各方面上，例如以手勢遠距離控制機械手臂加工，或是行動不便的人將能以手勢控制手臂以完成自身無法完成之動作，如搬運貨物或遞取物品。

## 第五章 結論與未來目標

### 5.1 結論

本論文嘗試將等階集合法應用在手勢辨識中的影像手勢分割上，並且計算手掌角度並完成手掌轉正程序，以及利用傅立葉描述子來消除高頻雜訊，以消除較不重要的特徵，再利用主成份分析大幅降低特徵的維度，保留較重要的特徵，提高辨識率。

本研究提出利用等階集合法的主動式輪廓模型，因為該函數僅利用現有輪廓的周圍影像資訊計算，且利用梯度下降法可以尋找本論文定義的隱函數的區域最小值，使得該函數值收斂時，對應的輪廓會在色差明顯的區域，所以本方法適合用在尋找物體的外部輪廓，且會有強健的分割效果。將此方法應用在手掌影像分割上，將能切割出完整而易於辨識的手部外形，且不會延伸到手掌內部圖形特徵。這項特點將有利於本論文基以手部形狀的辨識。而在第四章採用等階集合法所發展出來的影像切割技術實驗結果，可以看出此方法的確能切割出較完整且易於辨識的手部輪廓。

在影像辨識時，本研究利用主成份分析法降低特徵維度的數目，並將此特徵矩陣與每張訓練影像的特徵矩陣逐一比較，尋找距離最相

近的訓練影像。為了測量特徵大小對辨識率的影響，第四章中也針對特徵數目作探討。實驗中可發現原本針對每張手掌影像蒐集 360 個特徵，但利用奇異值分解，本研究可僅選出前 50 個變異數最大的特徵變數，將維度從 360 降到 50，除了辨識速度可以提高外，發現辨識率比原本不降維度時還要高，每個手勢的辨識率至少都有 85% 以上，每張的辨識速度約在 0.3 秒內。辨識率無法百分之百的可能原因或許為手勢分割的不完整：例如手的晃動、膚色問題、背光環境等，都會造成系統在手勢分割階段切割出不利辨識的手形；也可能為辨識階段的誤判，例如本系統蒐集的訓練影像中沒有與受試者相似的手形、或是取手掌的角度距離特徵並無法完整的區隔各手勢間的差異。

最後本研究將本系統應用於鼠標的控制上，利用手勢即可控制滑鼠動作，且速度能符合需求，將此介面進一步連接其他機電系統，如本論文控制本實驗室提供的機械手臂，達到非接觸式控制的目的。本系統將可幫助一般人更方便的操控儀器，例如電視或遊樂器；若對行動不便的人，也可提供更多的協助，如幫助移動物品或復健等。

## 5.2 未來展望

本研究完成的辨識系統目前使用上仍有下列限制：

1. 背景不能太接近膚色
2. 操控者必須穿著長袖衣物，即只能露出手掌部分
3. 由於在本辨識系統中有定義膚色範圍，所以某些場景如背光區域，  
不適用本辨識系統
4. 對未經訓練的測試者，會有少數手勢辨識率不夠高的情形

由於上列問題，本系統可再往下列方向發展，以臻完善：

1. 在以主動式輪廓切割手勢的步驟中，本論文以參考灰階資訊來移動零階集合，這或許可改為參考機率資訊以移動零階集合，避免背景為膚色的影響，也避免操控者皮膚膚色的問題。
2. 找出可完整分離手掌影像與手臂影像的方法。
3. 利用更好的手部形狀特徵描述方法，例如尋找並計算手指的峰點數目。

## 參考文獻

- [1] Ghassem Tofighi, S.Amirhassan Monadjemi, Nasser Ghasem-Aghaee. "Rapid Hand Posture Recognition Using Adaptive Histogram Template of Skin and Hand Edge Contour", 2010.
- [2] 白文榜, 「基於影像之即時手勢辨識系統設計」, 國立交通大學電控工程研究所, 碩士論文, 民國一百年。
- [3] 余玉田, 「基於手形輪廓之智慧型手勢辨識系統設計」, 國立交通大學電控工程研究所, 碩士論文, 民國九十九年。
- [4] Meenakshi Panwar. "Hand Gesture Recognition based on Shape Parameters", 2012.
- [5] M Kass, A Witkin, Drzopoulos, "Snakes: Active Contour Models", 1988.
- [6] Tony F. Chan, Luminita A. Vese, "Active Contours Without Edges", 2001.
- [7] Anthony Yezzi Jr, "A Fully Global Approach to Image Segmentation via Coupled Curve Evolution Equations", 2000.
- [8] Shawn Lankton, Allen Tannenbaum, "Localizing Region-Based Active Contours", 2008.
- [9] Chopp, D, "Computing minimal surfaces via level set curvature flow" Journal Computational Physics, vol.106, pp.77-91, May 1993.
- [10] D. Mumford, J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems" ,Commun. Pure Appl.Math, vol. 42, pp. 577-685, 1989.
- [11] Dusan Heric, Boiidar Potocnik, "Objective assessment of image segmentation algorithms.", 2006.
- [12] Osher, Stanley, Fedkiw, Ronald, "Level Set Methods and Dynamic Implicit Surfaces", 2003.
- [13] Spencer Patty, "Level Set Framework, Signed Distance Function, and Various Tools", 2010.
- [14] Robert Crandall, Fall, "Image Segmentation Using the Chan-Vese Algorithm", 2009.

- [15] Chunming Li, Chiu-Yen Kao, John C. Gore, Zhaohua Ding, "Minimization of Region-Scalable Fitting Energy for Image Segmentation," 2008.
- [16] LUMINITA A. VESE, TONY F. CHAN, "A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model." 2002.
- [17] Chunming Li, Chenyang Xu, Changfeng Gui, Martin D., "Level Set Evolution Without Re-initialization: A New Variational Formulation." 2005.
- [18] Yongsheng Pan, J. Douglas Birdwell, Seddik M. Djouadi. "Efficient Implementation of the Chan-Vese Models Without Solving PDEs." 2006.
- [19] Spencer Patty, "Level Set Framework, Signed Distance Function, and Various Tools." 2010.
- [20] Nassir Salman, "Image Segmentation and Edge Detection Based on Chan-Vese Algorithm." 2006.
- [21] S. Gao, T. Bui, "Image segmentation and selective smoothing by using mumford-shah model." 2005.

