

國立臺灣科技大學
電機工程系

碩士學位論文

學號: M10007422

應用類神經網路之分類及建模於人形機器人

三維運動之視覺學習

**Neural Network Classification and Modeling for
Visual Learning of Humanoid Robot**

研究 生：王兆葵

指 導 教 授：黃志良 博 士

中華民國 102 年 7 月 15 日

barcode
M10007422



碩士學位論文指導教授推薦書

本校 電機工程系 王兆葵(WANG, CHAO-KUEI) 君

所提之論文：

應用類神經網路之分類及建模於人形機器人三維運動之視覺學習

係由本人指導撰述，同意提付審查。

指導教授：黃志良

指導教授 黃志良

102年 6月 20日



碩士學位考試委員審定書



M10007422

指導教授：黃志良

本校 電機工程系 王兆葵 君

所提之論文：

應用類神經網路之分類及建模於人形機器人三維運動之視覺學習

經本委員會審定通過，特此證明。

學校考試委員會

委

員：

洪敏雄

施慶隆

黃志良

指導教授：

黃志良

學程主任：

吳啓華

系(學程)主任、所長：

中華民國 102年 6月 26日

摘要

本論文提出以立體視覺系統實現人形機器人的三維運動之學習(或模仿)的任務。首先將兩隻機器人以面對面平行的方式放置，藉由指定的機器人(稱之為「表演者」)做出一系列的三維動作，再由另一機器人(稱之為「學習者」)以其身上的立體視覺系統(Stereo Vision System)擷取影像並經由相關的影像處理，例如，背景註冊法進行動態偵測(Motion Detection)、形態學濾波(Morphology Filtering)去除高頻雜訊、姿勢辨識等，得到七個特徵點(即頭部、雙手、手肘及雙腳尖)，並經由立體視覺系統得知特徵點的三維世界座標，紀錄成一系列的軌跡。接著分析雙腳尖及頭部的數據，設計適當的特徵向量(Feature Vector)並以事先學習的多層類神經網路來分類下半身動作，將代表動作(Representative Action, RA)的手尖及手肘座標以事先學習的多層類神經網路來轉換上半身手的驅動馬達與雙手尖三維座標(即所謂的反運動學(Inverse Kinematic))，結合上下半身的動作以實現人形機器人的三維運動之學習(或模仿)的任務。最後以相關實驗證明所建議方法之有效性及可行性。

關鍵詞：視覺學習(或模仿)、背景註冊法進行動態偵測、姿勢辨識、人形機器人、立體視覺系統、多層類神經網路。

Abstract

This paper proposes the visual learning (imitation) of humanoid robot by stereo vision system. At beginning, the sequence of 3-D motion of “the Performer”, which is face to face with “the Learner”, is captured by the stereo vision system (SVS) installed at “the Learner.” The proposed image processing for each sampled image includes motion detector via background registration, morphology filtering of high frequency noise, and estimation of seven feature points (i.e., head, elbows, four tips of two arms and legs). We analyzed the data of head and tips of legs to designing appropriate feature vector and classified lower body by the pre-trained multilayer neural networks (MLNN). Then the two arm tips and elbows with RA is also approximated by a pre-trained MLNN to get upper body. Combined with the RA of lower body and two arm tips of upper body is employed to achieve the visual learning (or imitation) of 3D motion of an human robot. Finally, the corresponding experiments confirm the effectiveness and feasibility of the proposed methodology.

Keywords: *Visual Learning (Imitation), Humanoid Robot, Stereo Vision System, Multilayer Neural Network.*

目錄

摘要.....	I
Abstract.....	II
目錄.....	III
圖目錄.....	V
表目錄.....	VIII
第一章 序論	1
1.1 研究動機.....	1
1.2 實驗平台與場景規劃	2
1.3 實驗系統架構.....	6
1.4 論文架構.....	10
第二章 動態影像偵測與姿態辨識	11
2.1 背景註冊與動態閥值	11
2.2 特徵點辨識.....	17
2.3 手肘位置辨識.....	21
2.4 座標轉換.....	22
第三章 以類神經網路分類下半身動作和關鍵姿勢判斷	23
3.1 下半身動作分類	23
3.2 利用類神經網路分類	26

3.3 代表動作判斷.....	32
第四章 多層類神經網路為基礎的上半身動作 之反運動學 ...	34
4.1 右手神經建模.....	38
4.2 左手神經建模.....	40
第五章 實驗結果	41
第六章 結論及未來展望	47
參考文獻.....	48



圖目錄

圖 1.1 (a)表演者機器人外觀，(b)表演者機器人所具有的自由度	2
圖 1.2 表演者使用的伺服馬達，由左至右為 AX-12、RX-28、RX-64	2
圖 1.3 表演者機器人系統架構圖	3
圖 1.4 人形機器人嵌入式核心系統 RoBoard RB-110.....	3
圖 1.5 (a)學習者機器人外觀，(b)學習者機器人所具有的自由度	3
圖 1.6 學習者所使用的馬達型號，由左至右為 KRS-4034、KRS-2350	4
圖 1.7 學習者所採用的立體視覺系統	4
圖 1.8 學習者控制系統方塊圖	5
圖 1.9 立體視覺系統視野範圍限制與人形機器人擺放位置場景規畫	6
圖 1.10 機器人學習(模仿)實驗系統架構圖	9
圖 1.11 立體視覺系統之視覺學習(模仿)實驗流程.....	9
圖 2.1 兩相鄰灰階影像相素值相除數值之累增直方圖及二值化閾值	12
圖 2.2 不同 k 值之動態偵測結果：(a)、(b) 相鄰影像、(c)-(p)二值化 結果影像.....	14
圖 2.3 不同 k' 值之動態偵測結果：(a) 已註冊之背景、(b) 即時影像、	

(c)-(p) 二值化結果影像	15
圖 2.4 (a)-(c)不同燈光條件下之原始灰階影像，(d)-(f)為(a)-(c)之二值化結果，(g)-(i)為(a)-(c)之長條圖增強結果，(j)-(l)為(g)-(i)之二值化結果，(m)為已註冊之背景	16
圖 2.5 五點特徵點搜尋示意圖	20
圖 2.6 五點特徵點辨識之結果	20
圖 2.7 手肘搜尋示意圖	21
圖 2.8 手肘辨識結果	21
圖 2.9 立體視覺與表演者之世界座標系統	22
圖 3.1 以 Y 軸方向進行偏移狀態判斷	25
圖 3.2 以 Z 軸方向進行抬腳狀態判斷.....	25
圖 3.3 以 X 軸方向進行腳部前後位移方向判斷	25
圖 3.4 十一動作分佈圖	28
圖 3.5 單層 30 神經元類神經架構圖	29
圖 3.6 類神經訓練結果	30
圖 3.7 將腳部座標軌跡帶入類神經分類後結果	31
圖 3.8 濾波之後的結果	31
圖 3.9 群集點搜尋演算法示意圖	33
圖 3.10 將腳部座標軌跡帶入類神經網路分類且濾波後並挑選代表動	

作的結果.....	33
圖 4.1 軌跡座標轉換馬達控制角度	34
圖 4.2 兩個參考點的座標 $[x_1 \quad y_1 \quad z_1]$ 和 $[x_2 \quad y_2 \quad z_2]$	35
圖 4.3 右手類神經建模工作區域示意圖	35
圖 4.4 雙層 30 神經元類神經架構圖	37
圖 4.5 單層 60 神經元類神經架構圖	37
圖 4.6 類神經內插結果(a)1st sub-work space , (b)2nd sub-work space , (c) 3rd sub-work space , (d) 4th sub-work space.....	39
圖 4.7 類神經內插結果(a)1st sub-work space , (b)2nd sub-work space , (c) 3rd sub-work space , (d) 4th sub-work space	40
圖 5.1 表演者動作影像	42
圖 5.2 類神經分類及濾波後結果	43
圖 5.3 挑選代表動作後結果	43
圖 5.4 表演者代表動作之正面(立體視覺拍攝)及側面圖(攝影機)拍攝	45
圖 5.5 學習者動作之正面圖	46

表目錄

表 1.1 立體視覺系統在不同距離之視野範圍	6
表 2.1 特徵點選定法則	19
表 3.1 十一動作分類軌跡判斷結果對照表	23
表 4.1 1 st sub-work space 3 輸入和 6 輸入類神經訓練結果.....	36
表 4.2 右手四個工作區域類神經訓練結果	38
表 4.3 左手各工作區域 6 輸入類神經訓練結果	40



第一章 序論

1.1 研究動機

最近幾年機器人相關的研究的十分熱門，不只在學術上，許多企業界也紛紛成立機器人相關部門，並逐漸推廣到各種領域上，例如工業、教育、娛樂[1]及家庭服務等等，然而現階段機器人有些關鍵瓶頸尚需突破，如與人的互動方式[2]、高準確的控制方式等等。然而，對於一般的使用者而言，這些過程往往必須具備高門檻的相關背景知識，若不能輕易的進行動作編輯，則機器人就算具有智慧型判斷能力，則僅能作為其預設領域用途上之機械。因此，若能通過視覺[3-5]、動作[6]、感測器[7-8]等途徑來控制機器人之行動，將減少一般使用者在使用上的困難，本研究希望藉由影像感測，建立可作為機器人動作學習的判斷演算法則，當擷取的動作影像經過適當之影像處理、軌跡分析與反運動學計算之後，將可獲得機器人行動所需之各項資訊。

近年來關於機器人模仿的研究很多，許多學者提出利用視覺辨識並分析人類動作[9-10]，來使機器人學習重複的任務。有用人類身上的標記作為機器人模仿人的依據[11-12]，也有將收集後的資訊與學習者間建立反運動學的轉換關係[13-14]。因此，訓練機器人學習表演者動作是一個不錯的研究題目。

1.2 實驗平台與場景規劃

本實驗使用兩款自行設計的機器人(學習者和表演者)，針對其硬體特性描述如下所示。首先是表演者，外型為圖 1.1(a)，高度約 55 公分，重量約 3.8 公斤，具有 23 個自由度，分別是頭部 2 個、左右手分別各 4 個、左右腳分別各 6 個自由度以及腰部一個自由度，如圖 1.1(b)所示，圖 1.2 為表演者使用的伺服馬達，圖 1.3 為系統架構圖，圖 1.4 為表演者嵌入式核心系統。

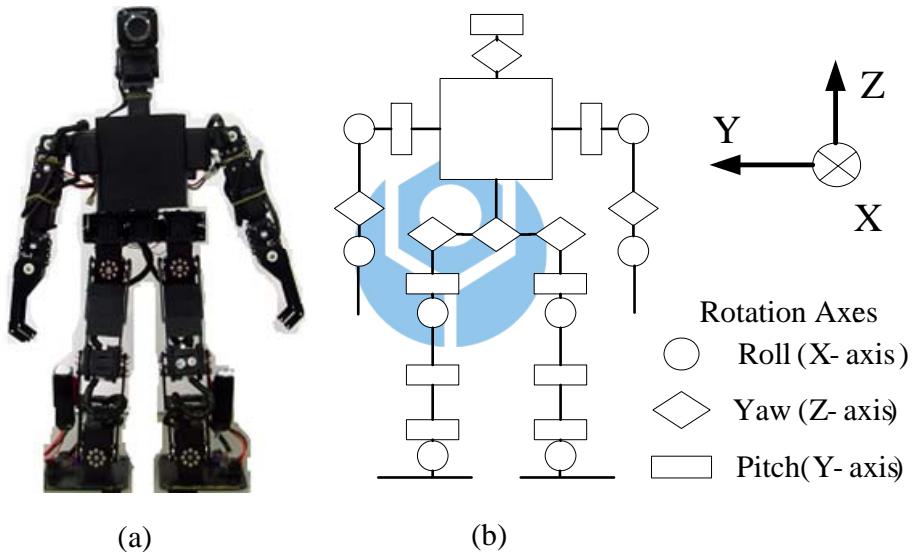


圖 1.1 (a) 表演者機器人外觀，(b)其具有的自由度



圖 1.2 表演者使用的伺服馬達，由左至右為 AX-12、RX-28、RX-64

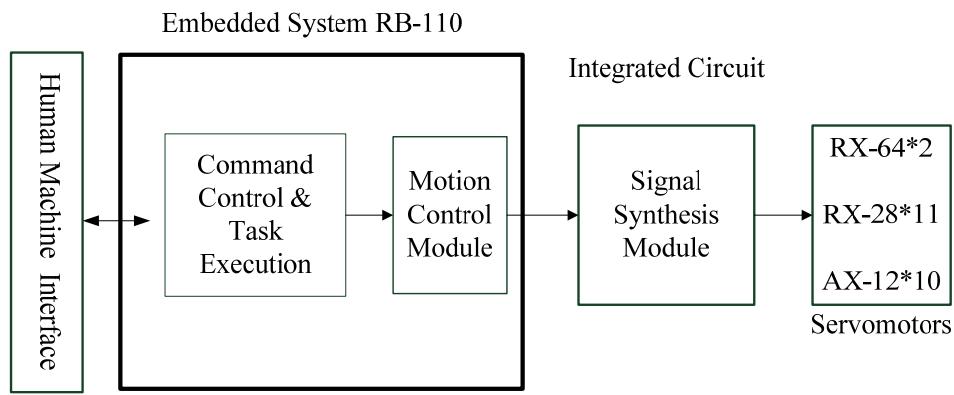


圖 1.3 表演者機器人系統架構圖



圖 1.4 人形機器人嵌入式核心系統 RoBoard RB-110

接下來是學習者外型為圖 1.5(a)，高度約 49 公分，重量約 3.4 公斤，也具有 23 個自由度，分別是頭部 2 個、左右手分別各 4 個、左右腳分別各 6 個自由度以及腰部一個自由度，如圖 1.5(b)所示，圖 1.6 為學習者所使用的馬達型號。

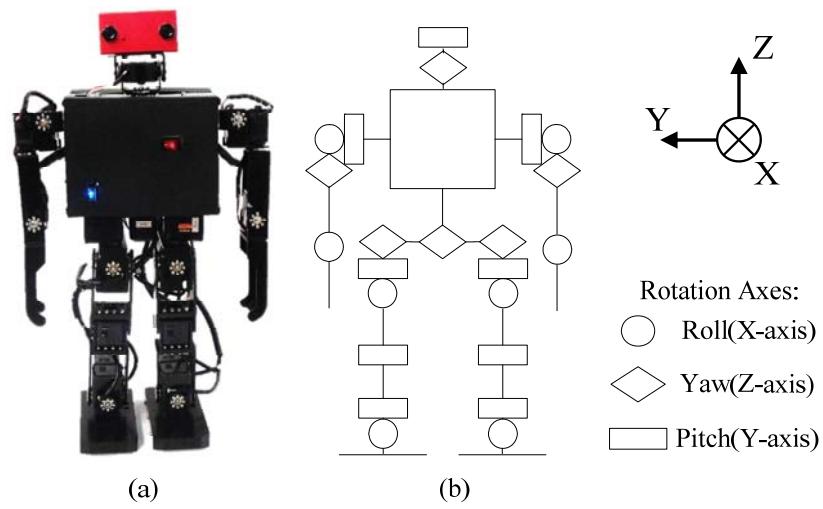


圖 1.5 (a)學習者機器人外觀，(b)其具有的自由度



圖 1.6 學習者所使用的馬達型號，由左至右為 KRS-4034、KRS-2350

在學習者所使用立體視覺系統微 VIDERE 公司所生產的 STOC 立體視覺鏡頭，如圖 1.7 所示，此鏡頭具有 640x480 像素的解析度。鏡頭的大小約為 4.3x8.1x4.3 公分，重量約為 180 公克，因此非常適合我們安裝在小型機器人上。且 STOC 具有內建處理器，可利用其內建之處理器來計算影像的立體訊息，透過 IEEE1394a 介面，可直接提供左、右鏡頭的原始影像及各像素點所對應的三維座標資訊，



圖 1.7 學習者所採用的立體視覺系統

本研究採用個人電腦(Personal Computer, PC)與立體視覺系統連結來進行影像處理及記錄特徵點的軌跡，並將軌跡進行機器人的反運

動學轉換，將轉換後的馬達控制命令以無線方式傳送至學習者的嵌入式系統中，並透過人機介面對馬達進行操作，如圖 1.8 所示。

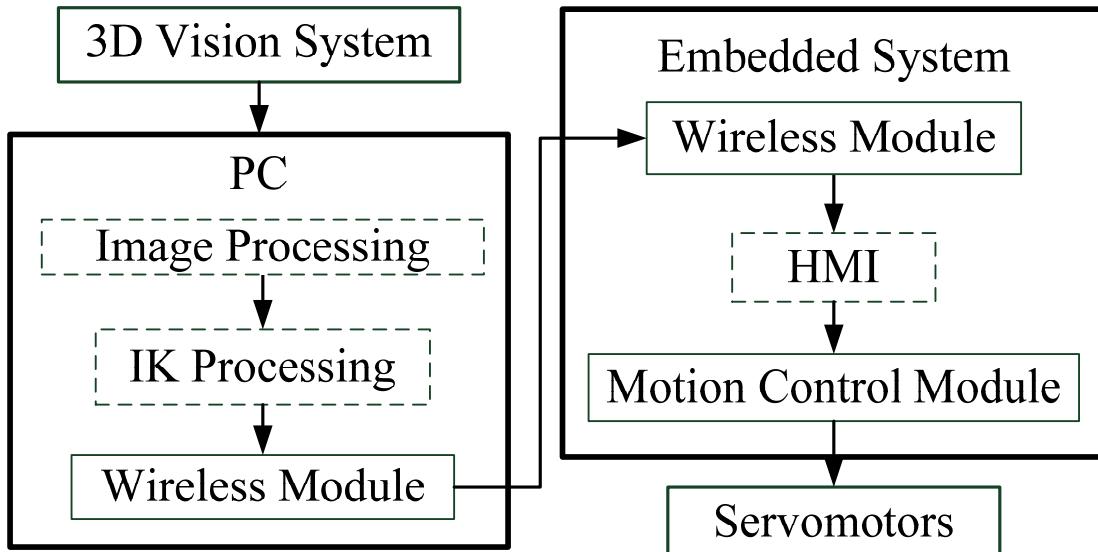


圖 1.8 學習者控制系統方塊圖

在實驗所使用的開發軟體方面上，本研究採用 MicroSoft Visual Studio C++ 2008 編撰所需的影像處理程式、嵌入式系統中動作控制程式和人機介面等等。另外在影像處理程式的開發上，本實驗亦使用 OpenCV 2.0 開放源碼作為影像程式之輔助工具以減少開發的時間。在場景規劃方面，為了配合學習者高度，立體視覺系統 STOC 裝置在學習者身上高度約為 47 公分的位置，且平行於學習者人形機器人座標所定義的 X 平面及 Y 平面。而 STOC 在水平與垂直方向具有不同的視野範圍(Field of View, FOV)，舉例來說，距離鏡頭 40 公分處，STOC 的視野寬（水平）及高（垂直）分別為 43 公分及 39 公分；距離鏡頭 80 公分處，STOC 的視野寬及高分別為 92 公分及 72.5 公分；

距離鏡頭 120 公分處，STOC 的視野寬及高分別為 141 公分及 106 公分，以此類推，如表 1.1 所示。為了讓整個實驗中表演者的活動範圍皆可在視野範圍內，本實驗將表演者放置於距離學習者距離 120cm 處，如圖 1.9 所示。

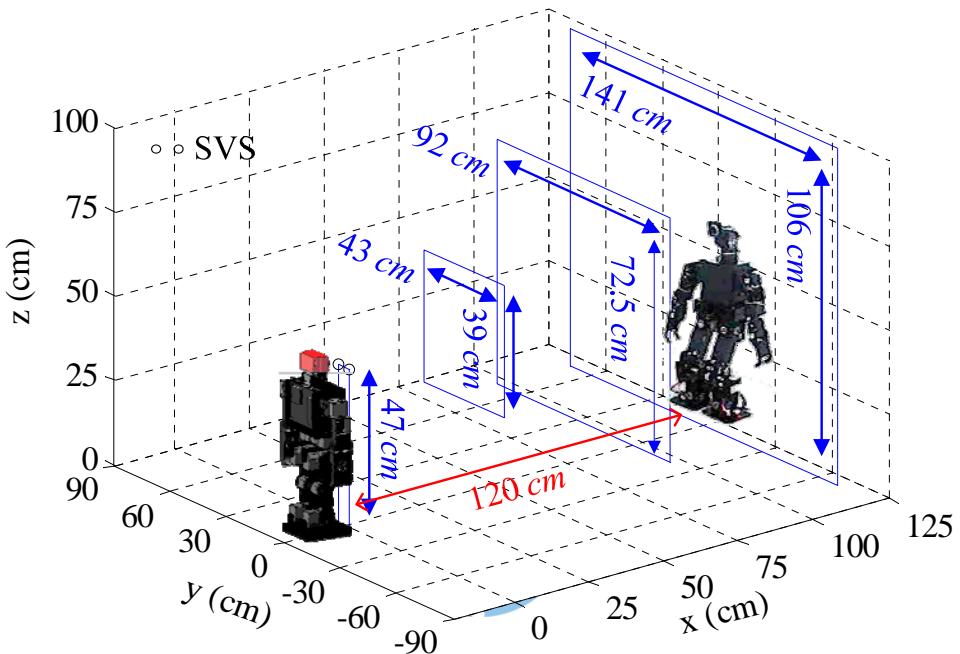


圖 1.9 立體視覺系統視野範圍限制與人形機器人擺放位置場景規畫

表 1.1 立體視覺系統在不同距離之視野範圍

No.	Distance	H-FOV	V-FOV
1	40 cm	43 cm	39 cm
2	80 cm	92 cm	72.5 cm
3	120 cm	141 cm	106 cm

1.3 實驗系統架構

本論文之研究為利用立體視覺系統實現人形機器人於三維座標上的學習。先進行背景註冊法[15]，註冊完後，學習者與表演者面對

面站著，距離大約 120 公分，先由表演者表演一套動作，學習者經由立體攝影機得到表演者的影像。接著透過動態偵測(Motion Detection)及可變閥值方法(Variable Threshold Method)[16]對影像擷取表演者剪影(二值化影像)，並利用形態學濾波及標籤化留下面積最大的部分，並提取得到表演者輪廓。然後進行姿態估算(Posture Estimation)[16]並估算得到表演者頭部、雙手和雙腳末端及手肘 7 個特徵點。接著由立體視覺系統估算特徵點相應的世界座標，並記錄動作過程的軌跡與時間，接著透過分析軌跡判斷平衡狀態，可將動作序列區分為不同的動作分類，並依軌跡的空間頻率分佈決定各分類動作區間之代表動作。各代表動作與學習者動作的反運動學關係可區分為上半身及下半身。藉由將學習者不同平衡狀態下的馬達命令模組化，即可配合軌跡以類神經網路分類後得到的平衡狀態進行轉換，以作為下半身的反運動學關係。上半身的反運動學關係，則是透過事前訓練的類神經模型來進行轉換[13-14]。最後將轉換後的動作命令與時間結合後，操作學習者進行播放，以實現相異人形機器人學習之目的，圖 1.10 為實驗系統架構圖，圖 1.11 為實驗流程。

各流程與所搭配之章節內容如下：

步驟一：以第二章所提出的背景註冊法註冊背景，並由立體視覺系統持續提供即時影像。

步驟二：當表演者執行動作時，以第二章之動態偵測法來處理即時影像與已註冊之背景，之後將獲取表演者的物件資訊。

步驟三：透過第二章提出的姿態辨識法估測表演者的特徵點(頭、手、腳、手肘等)。

步驟四：透過立體視覺系統估測特徵點的世界座標資訊(x, y, z)。

步驟五：持續紀錄各個特徵點之世界座標以作為軌跡資訊。

步驟六：當表演者結束動作後，以第三章之方法得到座標轉換後之軌跡座標($\bar{x}, \bar{y}, \bar{z}$)。

步驟七：將軌跡座標($\bar{x}, \bar{y}, \bar{z}$)利用第三章下半身動作之判斷方法判斷，並依十一種下半身動作分類區分為不同的動作區塊。

步驟八：以第三章的方法分析各分類動作區間的軌跡空間頻率(找出高頻群集點位置)，以挑選該動作分類之代表影格。

步驟九：(1)將代表影格所屬之上半身軌跡資訊，以本論文第四章之類神經反運動學模型轉換為學習者上半身的馬達控制命令。
(2)將代表影格所屬之動作分類，以本論文第三章對應之模組化平衡動作轉換為學習者下半身的馬達控制命令。

步驟十：將步驟九獲得的上半身及下半身馬達控制命令結合成完整學習者動作控制命令。

步驟十一：操作學習者執行控制命令完成學習流程。

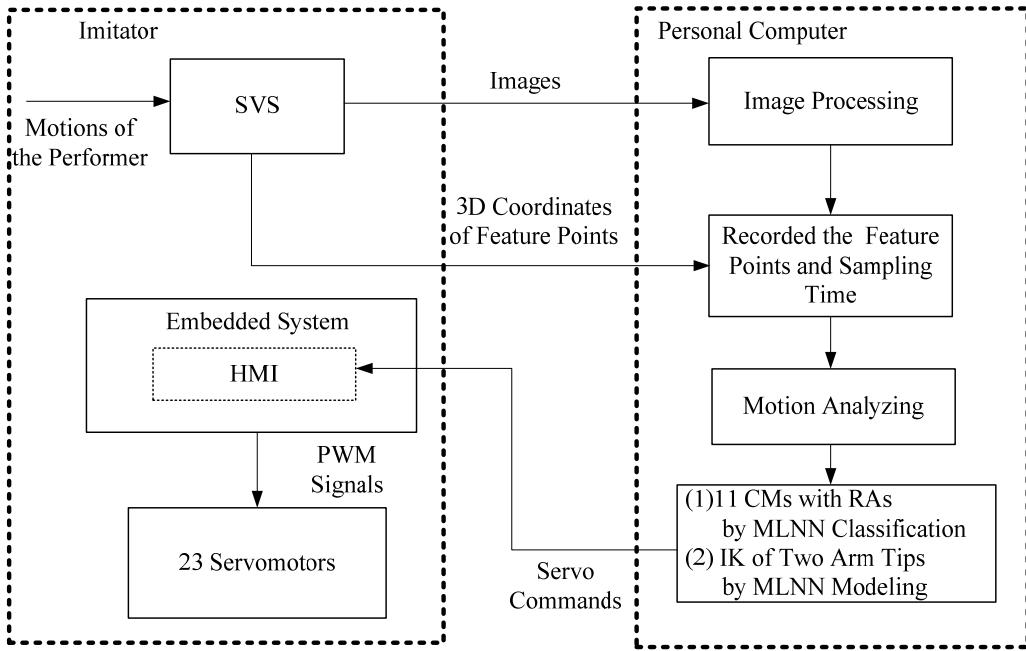


圖 1.10 機器人學習實驗系統架構圖

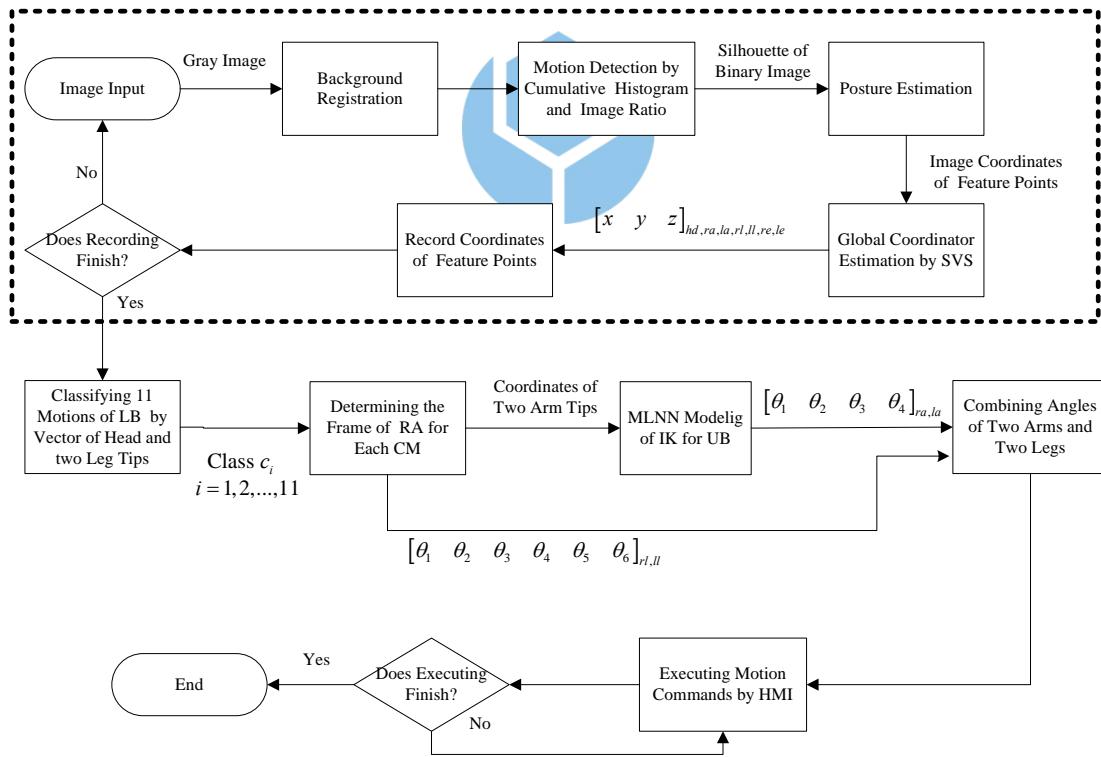


圖 1.11 立體視覺系統之視覺學習實驗流程

1.4 論文架構

本論文後續章節安排架構分述如下：第二章將會介紹學習者於蒐集表演者動作資訊時所使用的影像處理技術，第三章介紹代表動作的擷取及下半身動作分類判斷，第三章介紹人形機器人學習在上半身所使用的反運動學模型，整合影像技術及完整反運動學之實驗結果將於第五章介紹，最後在第六章進行總結與討論未來工作。



第二章 動態影像偵測與姿態辨識

2.1 背景註冊與動態閾值

為了從影像中將表演者與背景分割，得到表演者的剪影，本研究採用累增直方轉折點(Changing Point)的動態閾值演算法來進行影像二值化轉換。首先，假設 $r_n(x, y)$ 為兩個相鄰連續灰階影像 $f_n(x, y)$ 和 $f_{n-1}(x, y)$ 各像素的比值，其中 n 為時間序列， $r_n(x, y)$ 與 $f_n(x, y)$ 及 $f_{n-1}(x, y)$ 的關係可由以下(2.1)式表示：

$$r_n(x, y) = \begin{cases} (f_n(x, y) + k) / (f_{n-1}(x, y) + k), & \text{if } f_n(x, y) > f_{n-1}(x, y) \\ (f_{n-1}(x, y) + k) / (f_n(x, y) + k), & \text{otherwise} \end{cases} \quad (2.1)$$

其中 $1 \leq x \leq M$ ， $1 \leq y \leq N$ ， M 及 N 為螢幕畫面的解析度，在本研究中採用 $M = 320$ 、 $N = 240$ ， k 為一個大於等於 1 的常數值。當計算完整張畫面的像素比值後，接下來要取得二值化的閾值，我們首先計算 $r_n(x, y)$ 像素數目於各像素值的直方圖 $h_n(i)$ ， $i \in [0, 255]$ ，然後我們利用(2.2)式求得 $h_n(i)$ 的累增直方圖 $a_n(i)$ 。

$$a_n(i) = h_n(i) + a_n(i-1), \text{with } a_n(-1) = 0 \quad (2.2)$$

假設 $L_n(i)$ 為在 $a_n(i)$ 上連結 $a_n(i_{\min})$ 與 $a_n(i_{\max})$ 的一條截直線， $D_n(i)$ 則為 $a_n(i)$ 上任意一點至 $L_n(i)$ 的距離，若 $i = i_T$ 時可得到 $D_n(i)$ 的最大值，亦即累增直方圖曲線上最小曲率半徑的位置，則我們即可將 i_T 設定為將 $r_n(x, y)$ 數值二值化的閾值，如圖 2.1 所示。

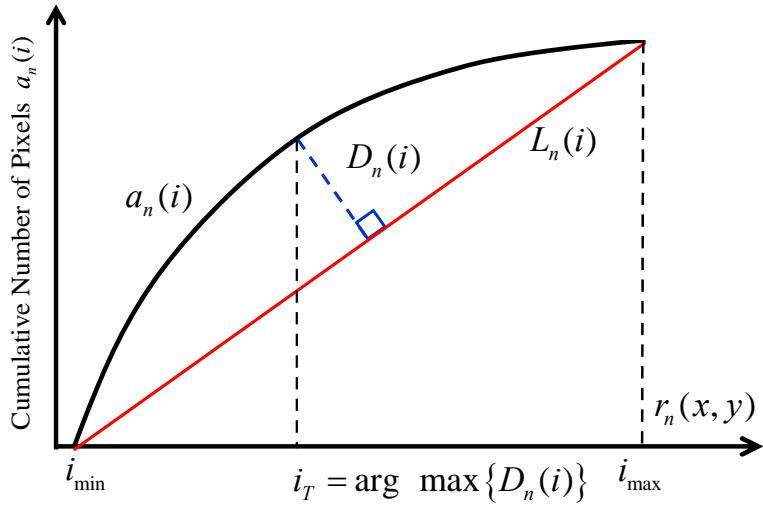


圖 2.1 兩相鄰灰階影像相素值相除數值之累增直方圖及二值化閥值

接著在(2.3)式中令 i_T 為閥值對原始比值直方圖 $r_n(x, y)$ 進行二分法運算，運算之後的二值化結果為 $b_n(x, y)$ 。

$$b_n(x, y) = \begin{cases} 1, & \text{if } r_n(x, y) > i_T \\ 0, & \text{if } r_n(x, y) \leq i_T \end{cases} \quad (2.3)$$

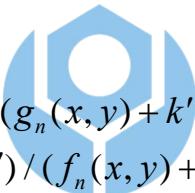
當取得相鄰影像的變動資訊後，即可辨別影像中不常發生變化的區域，即 $b_n(x, y) = 0$ 之像素點，為了將此區塊註冊為背景，我們對連續影像二值化後各像素連續未發生變動的次數進行紀錄；亦即針對 $b_n(x, y) = 0$ 連續發生的次數進行紀錄。當 $b_n(x, y) \neq 0$ 時，該像素的累計次數即歸零，並重新累加。當 $b_n(x, y) = 0$ 連續發生的次數超過預設閥值時，該像素點即被註冊為背景。假設 $s_m(x, y)$ 為 $b_m(x, y)$ 在第 m^{th} 影格的累加次數，則 $s_m(x, y)$ 累加的條件如(2.4)式所示：

$$s_m(x, y) = \begin{cases} s_m(x, y) + 1, & \text{if } b_m(x, y) = 0 \\ 0, & \text{if } b_m(x, y) = 1 \end{cases} \quad (2.4)$$

其中 $n - n_f \leq m \leq n$ 。當其超過預設的閥值 n_f ，本實驗 n_f 取 30，則該點像素值 $f_n(x, y)$ 將會被註冊為背景 $g_n(x, y)$

$$g_n(x, y) = \begin{cases} f_n(x, y), & \text{if } s_n(x, y) > n_f \\ g_{n-1}(x, y), & \text{otherwise} \end{cases} \quad (2.5)$$

當 $g_n(x, y)$ 完成註冊後，即可藉由計算即時影像 $f_n(x, y)$ 與 $g_n(x, y)$ 的差異性來獲得所需之變動物件。與相鄰影格差異性比較時所用的演算法類似，假設 $r'_n(x, y)$ 為 $f_n(x, y)$ 與 $g_n(x, y)$ 差異性比較後所得到的比值，怎可定義(2.6)式如下：



$$r'_n(x, y) = \begin{cases} (f_n(x, y) + k') / (g_n(x, y) + k'), & \text{if } f_n(x, y) > g_n(x, y) \\ (g_n(x, y) + k') / (f_n(x, y) + k'), & \text{otherwise} \end{cases} \quad (2.6)$$

k' 值同樣為一個大於等於 1 的常數值。同樣藉由計算 $r'_n(x, y)$ 的累增直方圖、並求得該曲線上最佳二分法的閥值 j_T 後，則 $r'_n(x, y)$ 藉由 j_T 進行二值化的結果可由(2.7)式來表示：

$$b'_n(x, y) = \begin{cases} 1, & \text{if } r'_n(x, y) > j_T \\ 0, & \text{if } r'_n(x, y) \leq j_T \end{cases} \quad (2.7)$$

其中 $b'_n(x, y) = 1$ 的像素點即為第 n^{th} 影格中發生變動之像素，同一影格中所有 $b'_n(x, y) = 1$ 的像素點即代表與已註冊背景不同的變動前景物件，之後再進行形態學與取利用面積大小來篩選所需要的物件資訊之

影像處裡，可得到表演者的剪影圖。

由於當分母值較小時，像素相除法所得到的結果（商值）容易受到分子的變化而影響，因此在(2.1)式與(2.6)式中所使用的 k 值與 k' 值，可有效降低小分母值的影響。然而 k 值(k' 值)將會影響動態偵測的敏感性，過大會降低敏感性，過小則無法有效降低小分母值的影響，如圖 2.2 和 2.3 所示。依據實驗結果，本研究採用[50,100]與[40,80]等區間分別作為 k 及 k' 值的值域。

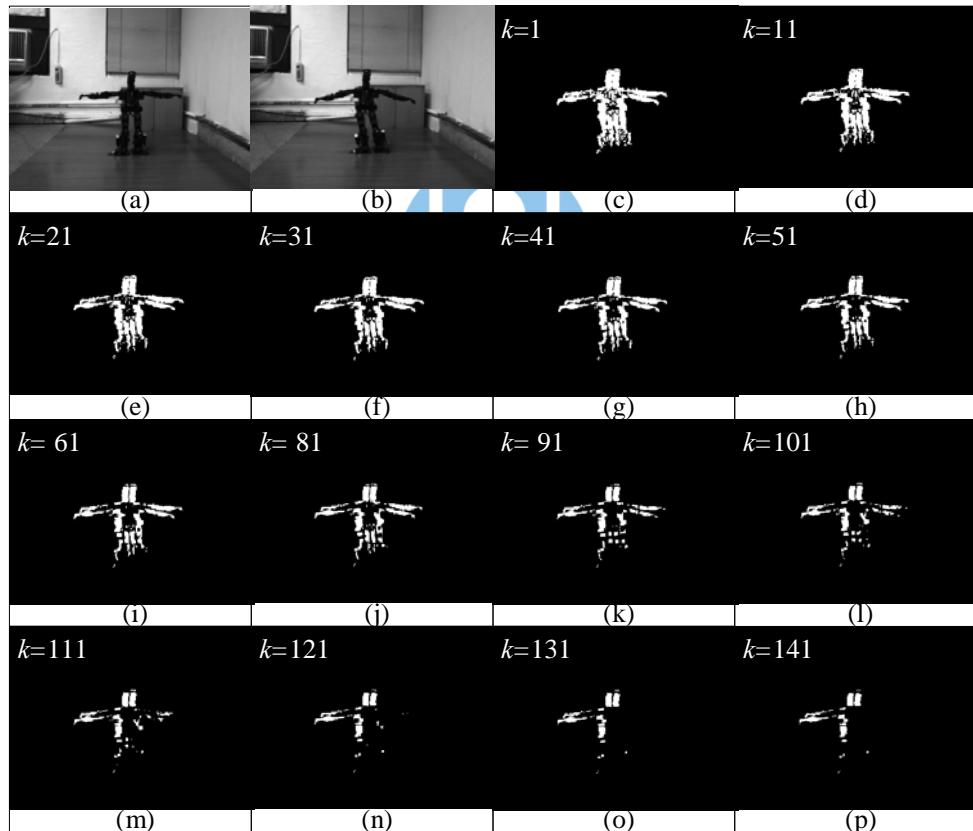


圖 2.2 不同 k 值之動態偵測結果：(a)、(b) 相鄰影像、(c)-(p)二值化
結果影像



圖 2.3 不同 k' 值之動態偵測結果：(a) 已註冊之背景、(b) 即時影像、
 (c)-(p) 二值化結果影像

雖然透過上述的動態閾值演算法可有效地自註冊背景中區並且分出前景之變動物件，也可克服小範圍之燈光變化，但當環境燈源出現大幅度變化時，無法正確判斷。因此，本研究利用長條圖強化法(Histogram Equalization)修正灰階值過於集中至某區域的輸入影像，計算公式如以下所示：

$$f'_n(x, y) = \frac{255}{M \times N} \times \sum_{i=0}^{f_n(x, y)} \hat{h}_n(i) \quad (2.8)$$

其中 M 與 N 同樣為影像平面長及寬的解析度， $i \in [0, 255]$ ， $\hat{h}_n(i)$ 為同

一影格中所有像素灰階值的直方圖， $f'_n(x,y)$ 則為轉換後的灰階像素值。透過此公式，即可將當前像素灰階值配合對應之累增直方圖與全畫面像素數目的比例關係，均勻的展延至[0,255]的值域範圍，可克服燈光大幅度變化的情況，如圖 2.4 所示。

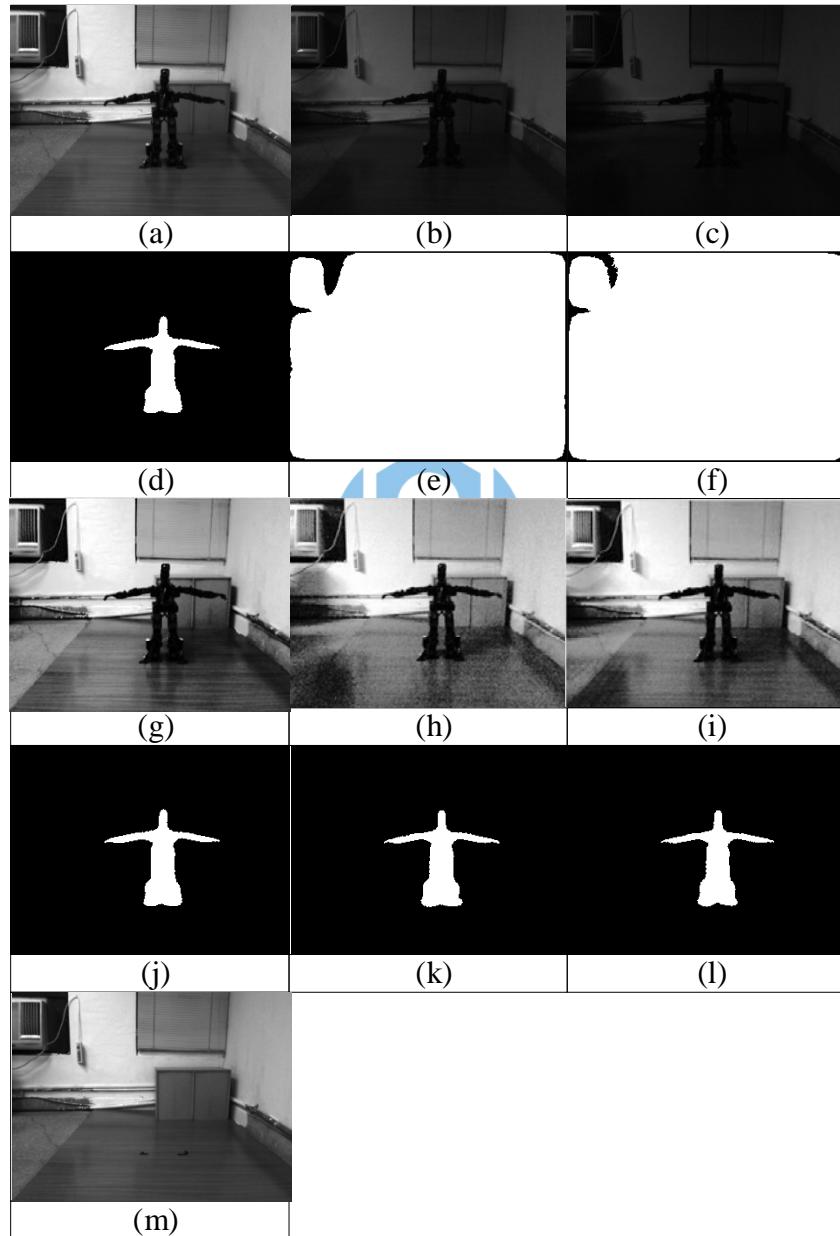


圖 2.4 (a)-(c)不同燈光條件下之原始灰階影像，(d)-(f)為(a)-(c)之二值化結果，(g)-(i)為(a)-(c)之長條圖增強結果，(j)-(l)為(g)-(i)之二值化結果，(m)為已註冊之背景

2.2 特徵點辨識

透過上一節動態閥值對移動物件的偵測與變動背景註冊法的處理後，我們可獲得欲處理的前景物件，此物件為表演者在影像平面空間的投影資訊。本節的研究重點為獲得表演者的頭、雙手及雙腳特徵點資訊。首先，先找到表演者剪影重心(Center of Gravity, COG)，重心在影像座標上的位置為 (x_c, y_c) ，可藉由(2.9)式求得：

$$x_c = \sum x/A, y_c = \sum y/A \quad (2.9)$$

其中 x 及 y 為影像座標表演者的剪影各像素座標值， A 為影像平面上之面積（像素數量）。接下來判斷表演者影像的方向性，我們利用中央動差法(Central Moment)，此方法可計算空間座標中橢圓物體最小慣量的旋轉軸，該旋轉軸方向近似於橢圓的長軸方向，且亦近似於身體軀幹的方向，可藉此方向與輪廓之交點作為選定特徵點之依據。因此本研究採用中央動差法來判斷物件的方向性。中央動量 μ 可透過(2.10)式求得：

$$\mu_{p,q} = \sum \sum (x - x_c)^p (y - y_c)^q \quad (2.10)$$

其中 p 與 q 值代表計算所需的指數，分別為介於 0 到 2 (含) 之間的整數。最後，利用(2.11)式，即可利用中央動量 μ 值計算出最小旋轉慣量軸的角度 θ_p ，亦即表演者於影像座標上身體方向的傾斜角度：

$$\theta_p = \tan^{-1} \left\{ 2\mu_{1,1} / [\mu_{2,0} - 2\mu_{0,2}] \right\} / 2 \quad (2.11)$$

計算出表演者影像之方向後，長軸與表演者影像輪廓可得到兩個交點 \mathbf{p}_a 與 \mathbf{p}_b （相對位置位於影像平面上方者定義為 \mathbf{p}_a ，下方者為 \mathbf{p}_b ），如圖 2.5(a) 所示，接著，便可藉由此兩點作為參考基準點，挑選及定義適合之頭、手及腳之特徵點。由於頭、手及腳等特徵點均位於身體輪廓之凸點(Convex Point)上，所以我們要先找身體輪廓之凸點候選點。首先，我們利用輪廓的曲率變化來判斷頂點的位置，利用(2.12)式計算其向量夾角

$$\theta_c = \cos^{-1} \left(\mathbf{v}_1 \cdot \mathbf{v}_2 / (\|\mathbf{v}_1\| \|\mathbf{v}_2\|) \right) \quad (2.12)$$

其中， \mathbf{v}_1 和 \mathbf{v}_2 是在剪影輪廓上凸輪廓點的兩個向量。接著我們沿著輪廓點順時針及逆時針方向取得最接近 \mathbf{p}_a 之凸點，且定義頭部之向量夾角限制範圍介於 $[30^\circ, 150^\circ]$ 之間，最後選取一個符合角度範圍以及最靠近 \mathbf{p}_a 之凸輪廓點為頭部特徵點，如圖 2.5(b) 所示。

當頭部的特徵點定後，接下來尋找腳部的特徵點，我們以長軸與輪廓交點 \mathbf{p}_b 作為搜尋起點，沿著輪廓以順時針及逆時針方向搜索，並定義腳部特徵點的限制角度為小於 155° 。當角度介於限制範圍內，且為距離頭部距離最遠之點時，即定義為右腳或左腳之特徵點，如圖 2.5(c) 所示。找完了頭與腳的特徵點後，由於手的位置通常位於兩者之間，因此即可利用頭與腳之間的輪廓來搜尋候選點作為手部特徵

點。我們定義手部特徵點須為距離 \mathbf{p}_o 最遠之點，其中 \mathbf{p}_o 為 COG 與 \mathbf{p}_{hd} 兩點間之中點。同樣的，我們定義角度小於 120° 為手部特徵點的角度限制範圍。因此，當候選點位於頭與腳之間、角度介於限制範圍內、且為候選點中距離 \mathbf{p}_o 最遠之點，即為右（左）手特徵點，如圖 2.5(d) 所示。透過上述尋找的方法，我們即可獲得表演者的頭、手及腳 5 點的位置，最後將收尋方法整理如表 2.1：

表 2.1 特徵點選定法則

特徵點	判斷參考點	角度限制	其他判斷條件
頭部 (\mathbf{p}_{hd})	\mathbf{p}_a	[$30^\circ, 150^\circ$]	當候選點超過兩點時，選定最靠近 \mathbf{p}_a 之點作為特徵點。
左腳 (\mathbf{p}_{ll})	\mathbf{p}_b	< 155°	位於 \mathbf{p}_b 之逆時針方向輪廓上，且距離 \mathbf{p}_{hd} 直線距離最遠之候選點。
右腳 (\mathbf{p}_{rl})			位於 \mathbf{p}_b 之順時針方向輪廓上，且距離 \mathbf{p}_{hd} 直線距離最遠之候選點。
左手 (\mathbf{p}_{la})	$\mathbf{p}_{hd}, \mathbf{p}_{ll}, \mathbf{p}_o$	< 120°	介於 \mathbf{p}_{hd} 與 \mathbf{p}_{ll} 之間，且距離 \mathbf{p}_o 直線距離最遠之候選點。
右手 (\mathbf{p}_{ra})	$\mathbf{p}_{hd}, \mathbf{p}_{rl}, \mathbf{p}_o$		介於 \mathbf{p}_{hd} 與 \mathbf{p}_{rl} 之間，且距離 \mathbf{p}_o 直線距離最遠之候選點。

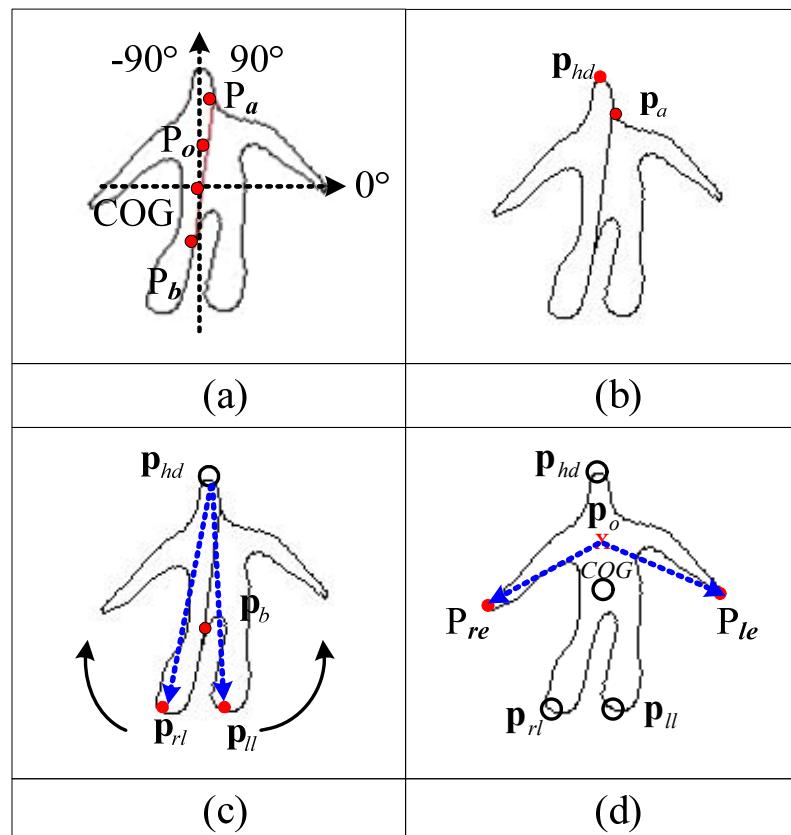


圖 2.5 五點特徵點搜尋示意圖

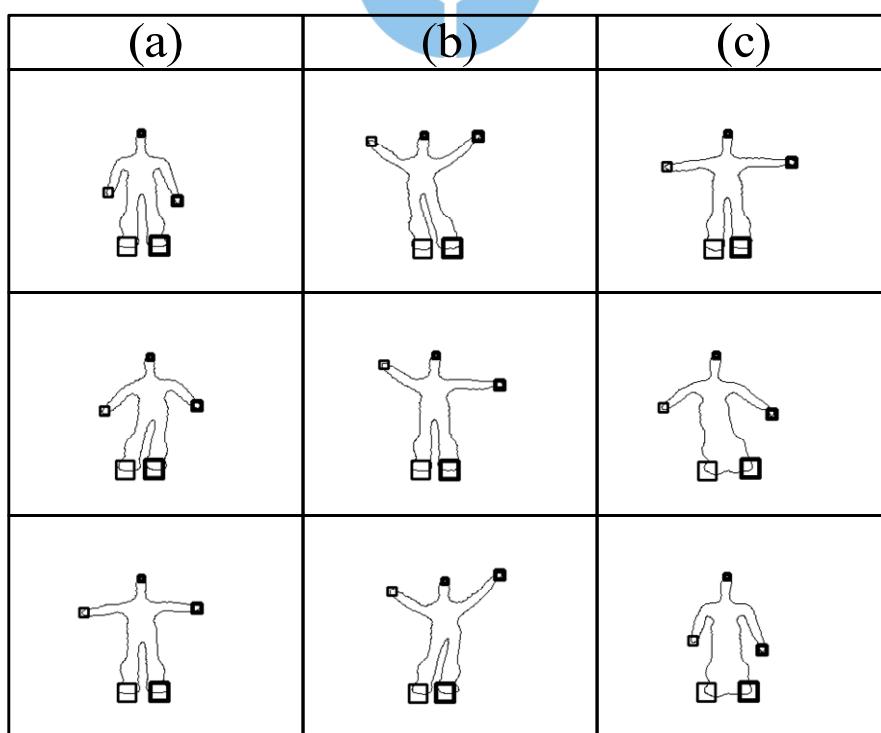


圖 2.6 五點特徵點辨識之結果

2.3 手肘位置辨識

由於第四章的手部類神經直接將左右手的三維座標建立馬達角度估算的模型訓練結果並不理想，因此我們多加了手肘對手指的向量座標，這時會需要手肘的位置。我們以手部特徵點位置為起點，沿著輪廓以順時針及逆時針找約全輪廓長度 3% 的長度，將這兩點取中間位置為手肘，如圖 2.7 和圖 2.8。

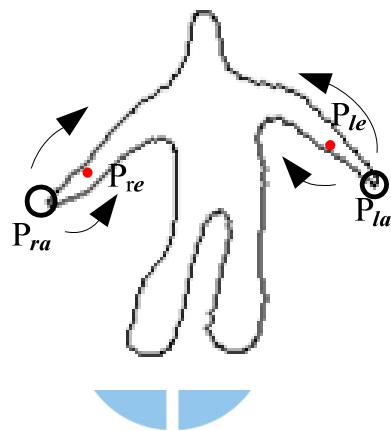


圖 2.7 手肘搜尋示意圖

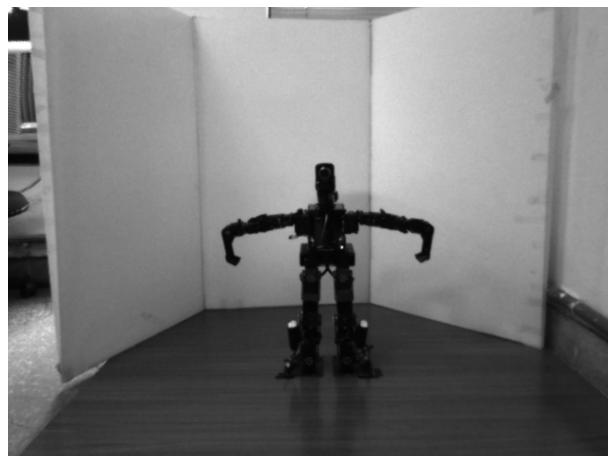


圖 2.8 手肘辨識結果

2.4 座標轉換

由於立體視覺系統所提供之三維世界座標，是以鏡頭本身為座標原點所得到的三維座標，因此當機器人與鏡頭的相對位置發生變化時，紀錄的軌跡也會隨之變化。所以為了要保持軌跡數值的穩定與手部類神經的應用，我們必須先將軌跡進行座標轉換。

假設 $(\tilde{x}, \tilde{y}, \tilde{z})$ 為表演者影像的 \mathbf{p}_o 以鏡頭為座標原點的三維座標， (x, y, z) 為特徵點的三維座標，若定義 \mathbf{p}_o 為新座標的參考原點，透過(2.13)式轉換成新的世界座標：

$$\begin{cases} \bar{x} = z - \tilde{z} \\ \bar{y} = -(x - \tilde{x}) \\ \bar{z} = -(y - \tilde{y}) \end{cases} \quad (2.13)$$



其中 $(\bar{x}, \bar{y}, \bar{z})$ 為轉換後表演者各特徵點且以 \mathbf{p}_o 基準之世界座標，此座標即可作為類神經模型於上半身反運動學轉換之輸入資訊，如圖 2.9。

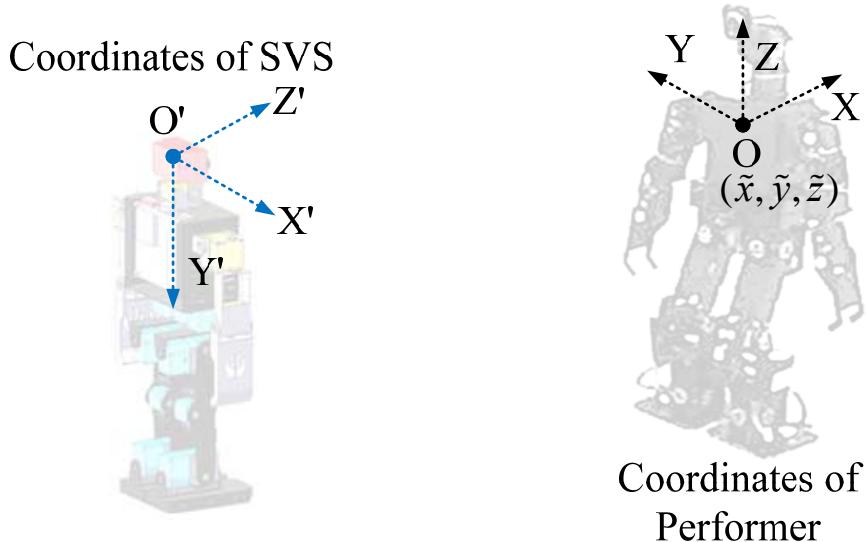


圖 2.9 立體視覺與表演者之世界座標系統

第三章 以類神經網路分類下半身動作 和關鍵姿勢判斷

3.1 下半身動作分類

由於本實驗的人形機器人並未擁有主動平衡的機制，因此我們在這邊提出了模組化的動作，將機器人下半身動作建立分類，透過將三個方向進行分析，將動作分成 11 種類，假設座標轉換後表演者的頭部座標 $(\bar{x}, \bar{y}, \bar{z})_{hd}$ 、左腳與右腳的世界座標分別為 $(\bar{x}, \bar{y}, \bar{z})_{ll}$ 與 $(\bar{x}, \bar{y}, \bar{z})_{rl}$ ，先透過 Y 軸方向的左右偏移判斷，再來是 Z 軸方向的高度判斷抬腳及支撐腳類型，最後是 X 軸方向的前後位置判斷，透過這三部份的判斷後得到下半身動作的分類，如表 3.1 所示。

表 3.1 十一動作分類軌跡判斷結果對照表

動作分類	偏移方向判斷 Y 軸	支撐腳類型 Z 軸	前後位移判斷 X 軸	動作類型
7	LW		NEU	DSP-L
11	NEU		R-FW	DSP-L-F
6	NEU		R-FW	
1	RW		NEU	DSP-N
5	RW		L-FW	DSP-R-F
8	LW		L-FW	
9	LW		NEU	DSP-R
10			NEU	SSP-L
4			BW	SSP-L-B
3	RW		FW	SSP-L-F
2	RW		NEU	SSP-R
			BW	SSP-R-B
			FW	SSP-R-F

透過三個方向的判斷後，如圖 3.1、圖 3.2、圖 3.3、，我們可以利用這些結果來確認表演者下半身動作的狀態，透過 Y 軸方向確認是否發生左偏移(Leftward, LW)、右偏移(Rightward, RW)或是未偏移(Neurtal, NEU)，再透過 Z 軸方向確定是否有單腳支撐(Single Support Phase, SSP)或者雙腳支撐(Double Support Phase, DSP)，最後透過 X 軸方向確認是否有前踢(Forward Motion, FW)及後踢(Backward Motion, BW)或是未踢(Neurtal, NEU)，組成 11 動作狀態。

我們使用的第一個代號為 DSP 或 SSP 代表是雙腳或單腳支撐，第二個代號為 R、L 及 N 代表右偏移、左偏移及未發生偏移，第三個代號為 F、B 或沒有代表非重心腳(支撐腳)的前踢、後踢或未踢。分成的 11 種動作分類為：(1) DSP-R-F: 左跨步(2) SSP-R-F: 右偏移抬左腳前踢(3) SSP-R-B: 右偏移抬左腳後踢(4) SSP-R: 右偏移抬左腳(5) DSP-R: 右偏移(6) DSP-N: 站立(7) DSP-L: 左偏移(8) SSP-L: 左偏移抬右腳(9) SSP-L-B: 左偏移抬右腳後踢(10) SSP-L-F: 左偏移抬右腳前踢(11) DSP-L-F: 右跨步。

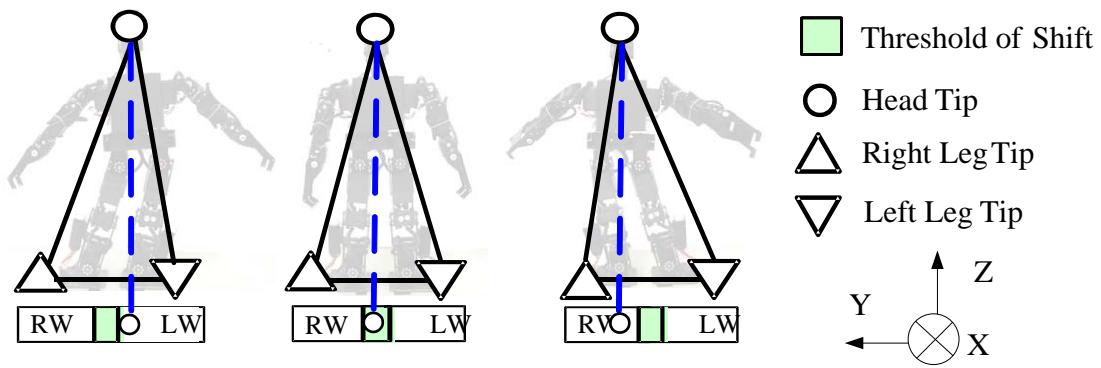


圖 3.1 以 Y 軸方向進行偏移狀態判斷

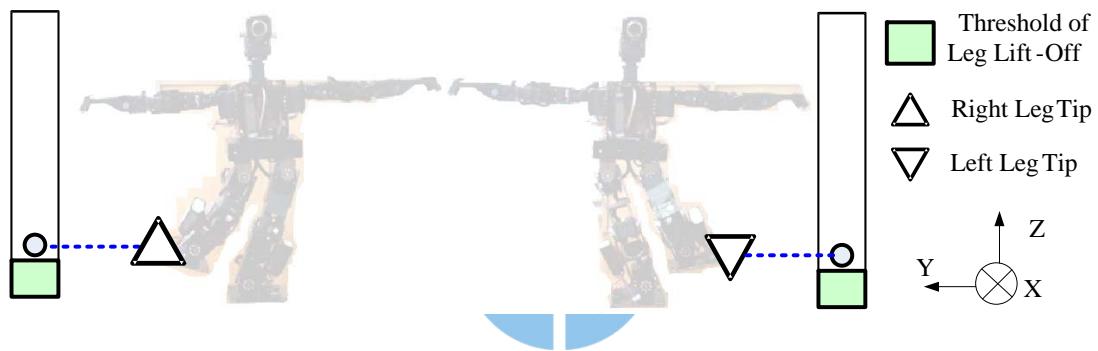


圖 3.2 以 Z 軸方向進行抬腳狀態判斷

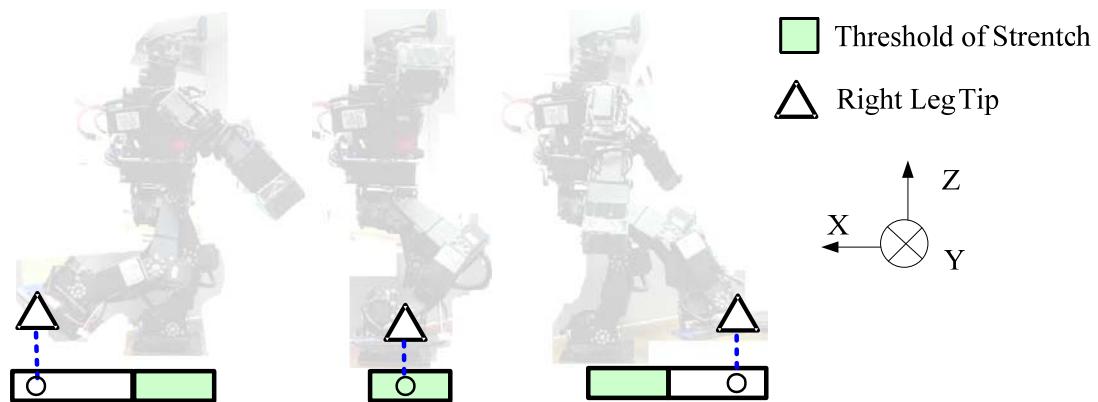


圖 3.3 以 X 軸方向進行腳部前後位移方向判斷

3.2 利用類神經網路分類

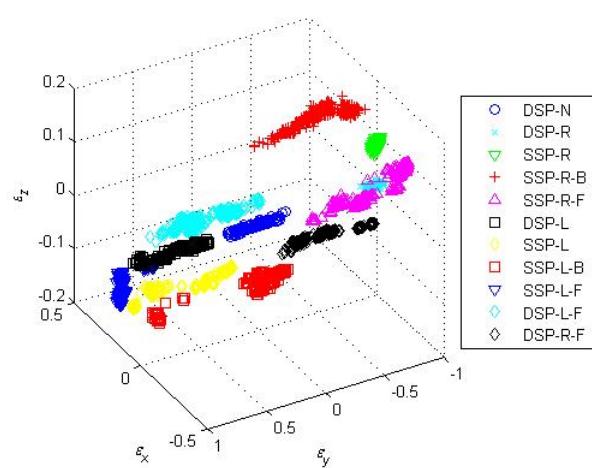
本研究希望有個通用法則在機器人與人類的大小不同下都能適用，但在判斷閥值的定義上很難有個適用值，所以之前的方法我們是採用將所有軌跡錄製後取同方向的平均值來當作閥值，但此方法無法實現即時學習的實驗。因此希望利用類神經網路來實現我們的需求，使之後實驗能達到即時學習的效果。首先將之前獲得的頭部座標 (x_{hd}, y_{hd}, z_{hd}) 、右腳座標 (x_{rl}, y_{rl}, z_{rl}) 及左腳座標 (x_{ll}, y_{ll}, z_{ll}) 代入以下公式(3.1)、公式(3.2)及公式(3.3)

$$\varepsilon_x = \frac{c_{x1}(x_{ll} - x_{rl})}{c_{x2}|z_{ll} - z_{hd}| + c_{x3}|z_{rl} - z_{hd}|} \quad (3.1)$$

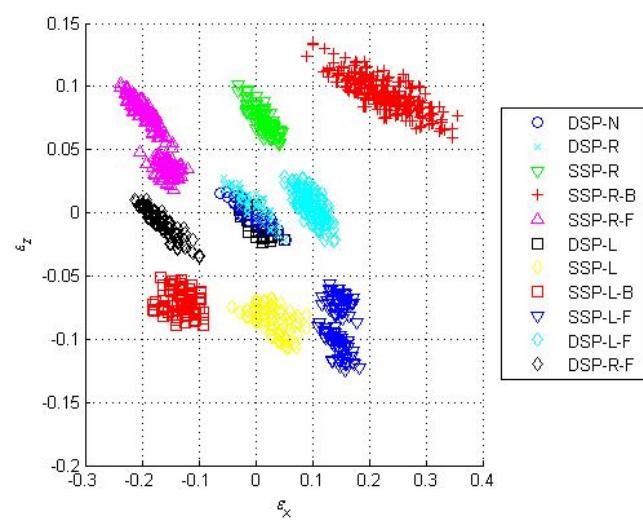
$$\varepsilon_y = \frac{c_{y1}|y_{hd} - y_{rl}| - c_{y2}|y_{hd} - y_{ll}|}{c_{y3}(y_{rl} - y_{ll})} \quad (3.2)$$

$$\varepsilon_z = \frac{c_{z1}(z_{ll} - z_{rl})}{c_{z2}|z_{ll} - z_{hd}| + c_{z3}|z_{rl} - z_{hd}|} \quad (3.3)$$

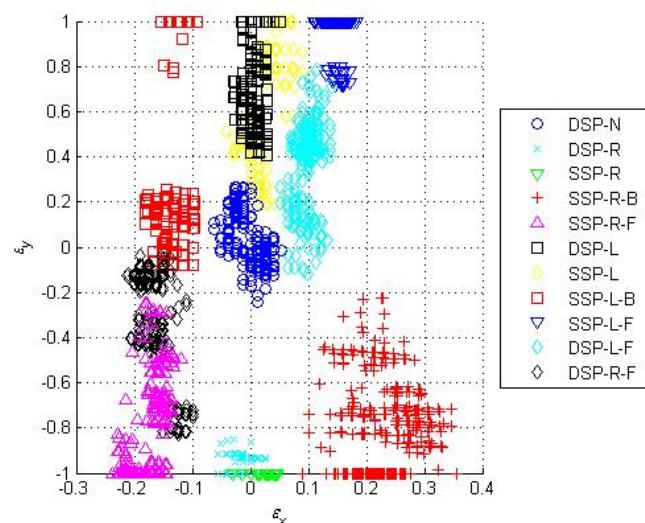
其中 ε_x 為stretch因子， ε_y 為shift因子， ε_z 為lift-off因子， c_{xi}, c_{yi}, c_{zi} , $i = 1, 2, 3$ 是用來提高每一種分類差別的係數，使每一類分類的重疊區域減少，藉以使分類結果更好， $\varepsilon_x, \varepsilon_y, \varepsilon_z$ 皆介於-1和1之間，將獲得的特徵因子 $\mathbf{E} = [\varepsilon_x \quad \varepsilon_y \quad \varepsilon_z]^T$ 當作輸入，帶入類神經網路做分類，輸出為動作分類的結果，圖3.4為十一動作的特徵向量分佈圖。



(a) $[\epsilon_x \quad \epsilon_y \quad \epsilon_z]$.



(b) Project to $[\epsilon_x \quad \epsilon_z]$.



(c) Project to $[\epsilon_x \quad \epsilon_y]$.

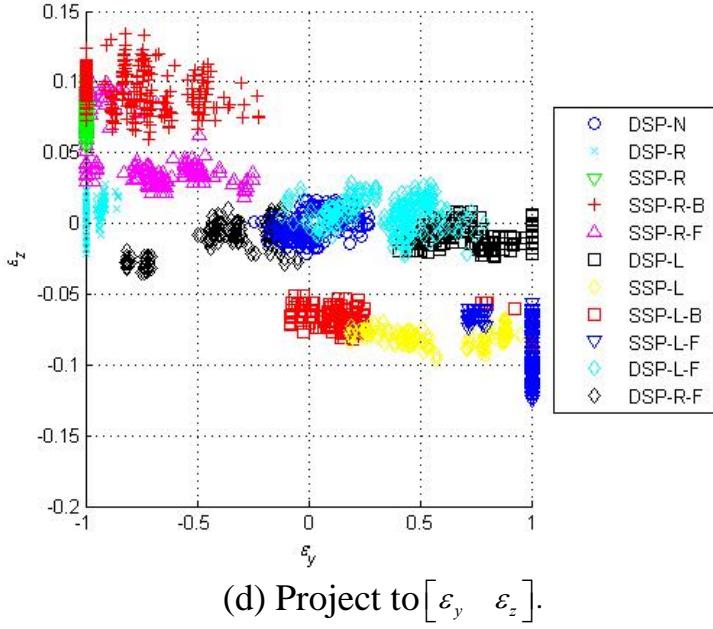


圖 3.4 十一動作分佈圖

圖 3.5 為我們所運用的多層感知器類神經網路架構圖，包含輸入層、1 個隱藏層和輸出層，基本類神經網路的描述如下：

$$\mathbf{a}^{j+1} = \mathbf{f}^{j+1}(\mathbf{W}^{j+1}\mathbf{a}^j + \mathbf{b}^{j+1}), j = 0, 1, \dots, M-1 \quad (3.4)$$

其中 $\mathbf{a}^0 = \mathbf{P} \in \Re^m, m=3$, 輸入層的輸入為 $\mathbf{E} = [\varepsilon_x \quad \varepsilon_y \quad \varepsilon_z]$ ，輸出為

$\mathbf{a}^M = [c_1 \quad c_2 \quad \cdots \quad c_{11}]^T$ 。此外，第 $l-th$ 的隱藏層權重為 $s^l, l = 1, 2, \dots, M-1$ ，轉

移函數為 $f_i^l(n_i^l) = \left(1 - e^{-n_i^l}\right) / \left(1 + e^{-n_i^l}\right), i = 1, 2, \dots, s^l, l = 1, 2, \dots, M-1$,

$f_j^M(n_j^M) = n_j^M, j = 1, 2, \dots, 5$ ，加權值 W^i 以及偏權值 $\mathbf{b}, i=1, 2, \dots, M$

本研究採用 Levenberg Marquardt 的倒傳遞(Back Propagation)學習演算法來作為類神經架構演算法，此演算法在函數的逼近上有較快的處理速度並且有更小的平方誤差。

$$F(\mathbf{x}) = \sum_{i=1}^Q [\mathbf{T}_i - \mathbf{a}_i^M]^T [\mathbf{T}_i - \mathbf{a}_i^M] = \sum_{i=1}^Q \sum_{j=1}^{s^M} [e_{j,i}]^2 = \sum_{i=1}^N [v_i]^2 \quad (3.5)$$

其中 Q 代表輸入的數目， \mathbf{T}_i 和 \mathbf{a}_i^M 代表第 i 個輸出， $s^M = 5$ 為輸出層的數目，而 $N = Q \times s^M$ ，LMBP 的學習法的描述如下：

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left[\mathbf{J}^T(\mathbf{x}_k) \mathbf{J}(\mathbf{x}_k) + \mu_k \mathbf{I}_n \right]^{-1} \mathbf{J}^T(\mathbf{x}_k) \mathbf{v}_k \quad (3.6)$$

k 為時間序列，學習權重為 $\mathbf{x}_k \in \Re^n$ ， $n = s^1(m+1) + s^2(s^1+1) + \dots + s^M(s^{M-1}+1)$ ，誤差向量 $\mathbf{J}_k \in \Re^{N \times n}$ 為 Jacobian 矩陣， $\mu_k > 0$ 確保有使得可逆矩陣， \mathbf{I}_n 為 n 維的方陣。此外，存在以下的關係式：

$$\mathbf{x}_k^T = \begin{bmatrix} w_{1,1}^1 & w_{1,2}^1 & \cdots & w_{s^1,m}^1 & b_1^1 & \cdots & b_{s^1}^1 & w_{1,1}^2 & w_{1,2}^2 & \cdots & b_{s^M}^M \end{bmatrix} \quad (3.7)$$

$$\mathbf{v}_k^T = \begin{bmatrix} e_{1,1} & e_{2,1} & \cdots & e_{s^M,1} & e_{1,2} & \cdots & e_{s^M,Q} \end{bmatrix} \quad (3.8)$$

$$[\mathbf{J}_k]_{pq} = \partial v_p / \partial x_q, p = 1, 2, \dots, N, q = 1, 2, \dots, n. \quad (3.9)$$

利用這個方法我們反覆調整類神經網路的加權值和偏權值，使得類神經網路的輸出值與我們的目標輸出值越來越接近，使誤差越來越小。

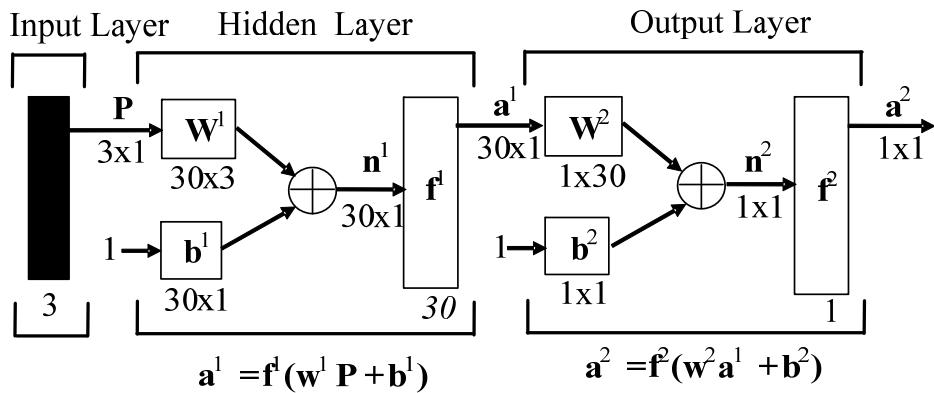


圖 3.5 單層 30 神經元類神經架構圖

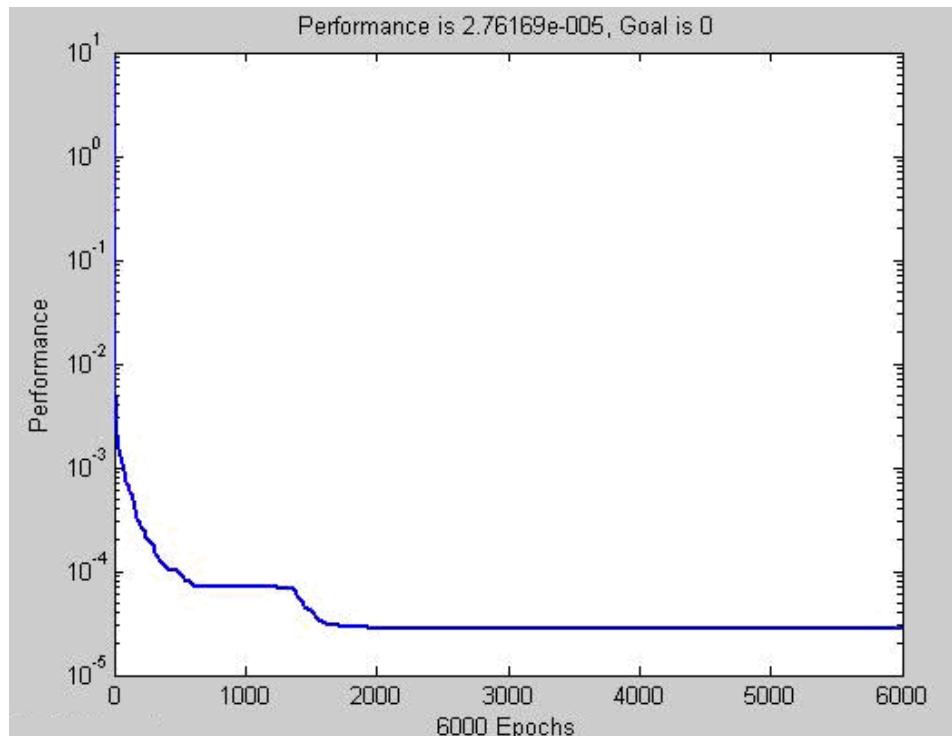


圖 3.6 類神經訓練結果

我們將一連串動作的腳步軌跡透過類神經分類並將結果做整數的 4 捨 5 入，希望能藉此判斷腳部動作，但由於動作與動作之間抓取特徵點時會出現誤判等原因，導致雜訊的產生，如圖 3.7。因此我們會將類神經分類後的結果，經由濾波來確定腳步的動作，我們的濾波方法是同一分類動作的影格數必須連續出現 20 次以上，要不就視為雜訊，如圖 3.8。

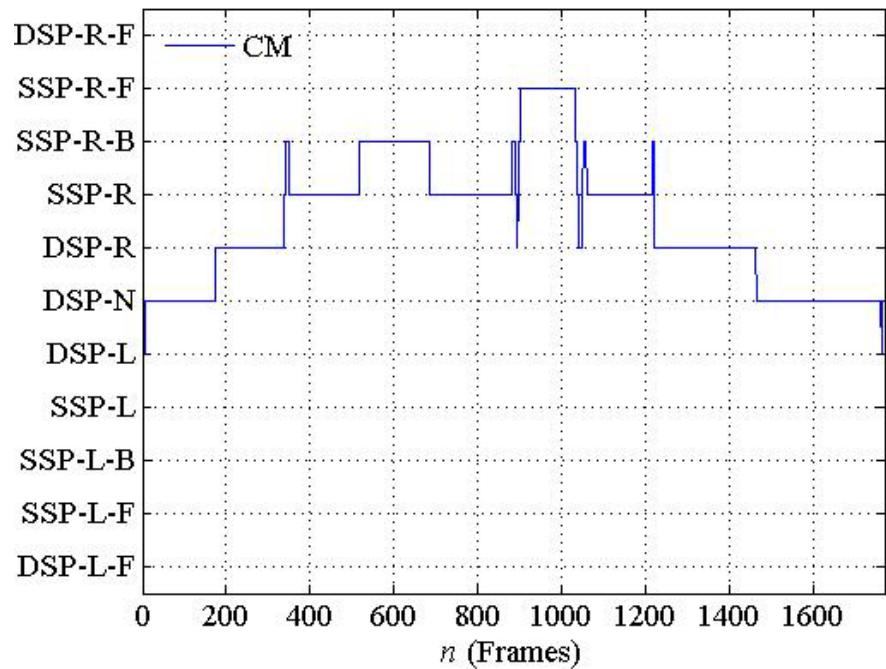


圖 3.7 將腳部座標軌跡帶入類神經分類後結果

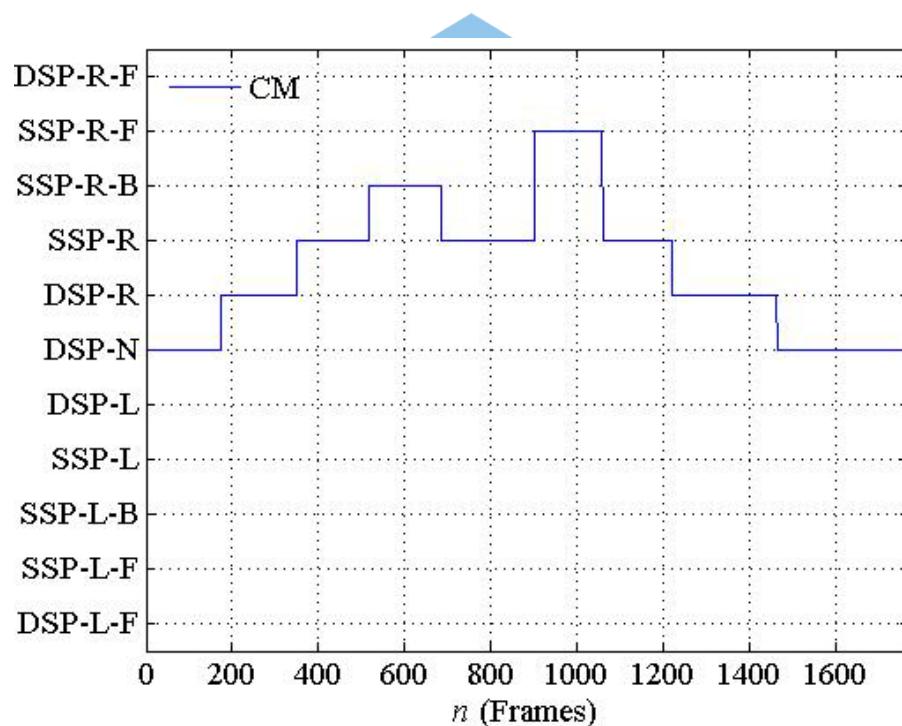


圖 3.8 濾波之後的結果

3.3 代表動作判斷

在透過上一節的下半身動作類神經網路分類後，我們可以成功的將表演者軌跡依十一種分類切割成不同的動作區間，接下來我們要找尋各動作區間中的關鍵動作來完成學習的實驗，因為我們紀錄了所有表演者的軌跡，若將全部動作資料皆去做的話非常浪費資源，會大大降低系統整體的運算效能。因此，找尋關鍵動作是我們所希望的，本節中我們提出利用空間頻率的概念，希望對各動作區間進行分析，找尋某座標點最具有高的空間頻率，並定義為該區間的代表動作(Representative Action, RA)。

我們定義 $\mathbf{D} = [\bar{x} \pm L, \bar{y} \pm L, \bar{z} \pm L]$ 為一個 3 維空間的搜尋方塊 (Searching Cube)，本研究採用 10 mm 作為 L 的長度。假設 $\mathbf{t}_{\bar{n}}$ 為各動作區間第 \bar{n} 個影格的軌跡座標，該空間頻率 $fs_{\bar{n}}$ 可表示成：

$$fs_{\bar{n}} = \begin{cases} fs_{\bar{n}} + 1, & \text{if } \mathbf{t}_{\bar{m}} \in L, \bar{m} \in [\bar{n}_{\min}, \bar{n}_{\max}] \\ fs_{\bar{n}}, & \text{otherwise.} \end{cases} \quad (3.4)$$

當某個 $fs_{\bar{n}}$ 為最大值時，我們認定這個 $\mathbf{t}_{\bar{n}}$ 為當區間之代表動作，如圖 3.9 所示，圖 3.10 為挑選代表動作的結果。

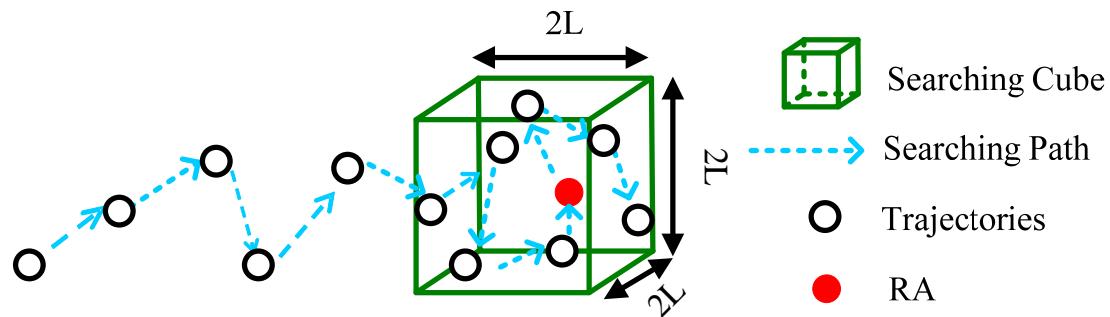


圖 3.9 群集點搜尋演算法示意圖

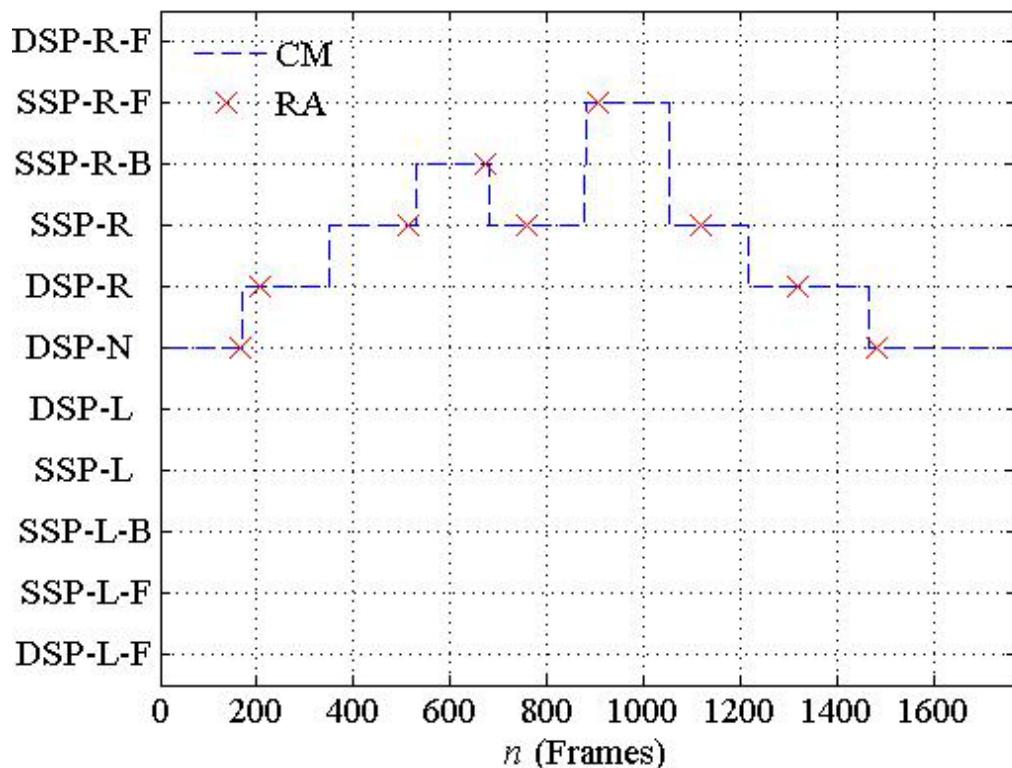


圖 3.10 將腳部座標軌跡帶入類神經網路分類且濾波後並挑選代表動作的結果

第四章 多層類神經網路為基礎的上半身動作 之反運動學

本論文運用類神經網路來建立手部馬達角度的反運動學，利用類神經網路建立手部三維空間座標資訊，由先前得到的關鍵姿勢手部座標當作輸入，表演者機器人左右手之各 4 顆馬達的角度作為輸出，經過類神經網路的訓練與學習，建立出其估算的模型，如圖 4.1。採用類神經網路來做為反運動學之模型，其優點為可以克服複雜的演算過程與節省計算時間。

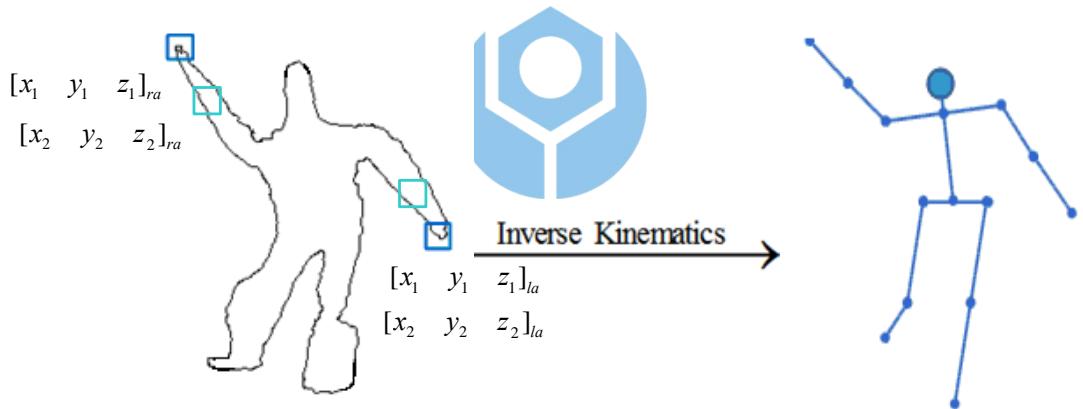


圖 4.1 軌跡座標轉換馬達控制角度

在本研究中，原本是直接將左右手的三維座標建立馬達角度估算的模型訓練，但其結果並不理想。因此，我們提出 2 個改進方法，第一個改進方法是多加了手肘對手尖的向量座標，將輸入點從 3 個輸入改成 6 個輸入，增加訓練的空間資料，如圖 4.2 及表 4.1 所示，從表 4.1 可得知 6 輸入不論在單層 60 神經元或雙層 30 神經元的均方誤差

(Mean Square Error, MSE) MSE 結果比 3 輸入的結果佳，但花費的時間較多，最後我們選用 6 輸入的方式去做類神經網路訓練。另一個改進方法是將其手部空間範圍利用 x 和 z 區分成左右手共 8 個區域分別去做類神經模型，如圖 4.3 所示。

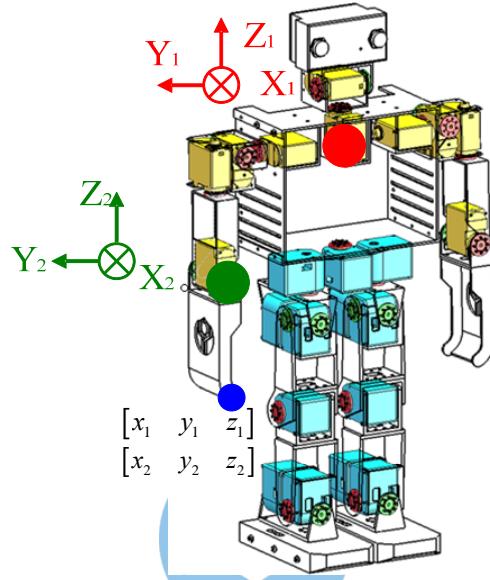
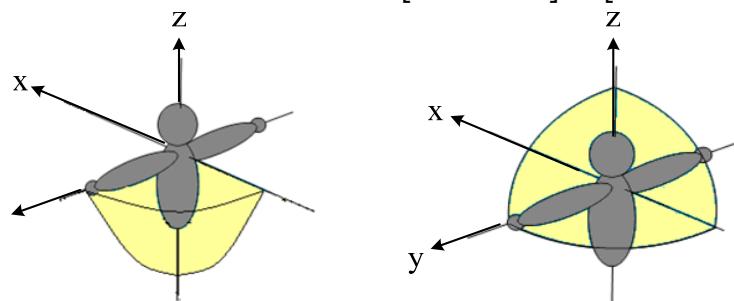
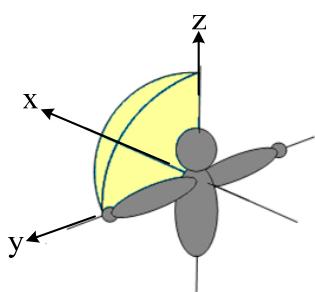


圖 4.2 兩個參考點的座標 $[x_1 \ y_1 \ z_1]$ 和 $[x_2 \ y_2 \ z_2]$

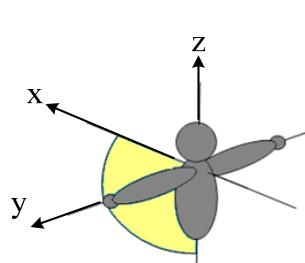


1st sub-work space $x \leq 0, z \leq 0$

2nd sub-work space $x \leq 0, z \geq 0$



3rd sub-work space $x \geq 0, z \geq 0$



4th sub-work space $x \geq 0, z \leq 0$

圖 4.3 右手類神經建模工作區域示意圖

表 4.1 1st sub-work space 3 和 6 輸入(訓練數目 2908)類神經訓練結果

隱藏層數	神經元數	輸入數	時間(分)	MSE
One	60	3	57	9.72 E-03
	60	6	98	1.25E-03
Two	30	3	237	8.10 E-03
	30	6	296	3.73E-04

圖 4.4 和圖 4.5 為我們所運用的多層感知器類神經網路架構圖，圖 4.4 用在第一工作區域和第二工作區域，圖 4.5 則用在第三工作區域和第四工作區域。圖 4.4 中包含輸入層、2 個隱藏層和輸出層，輸入層為左(右)手的三維座標 $[x_1 \ y_1 \ z_1 \ x_2 \ y_2 \ z_2]_{la}$ ($[x_1 \ y_1 \ z_1 \ x_2 \ y_2 \ z_2]_{ra}$)，第一層的加權值 W^1 設定為 30 組而轉移函數 f^1 為使用雙正切 S 型函數 $a^1 = (e^{n^1} - e^{-n^1}) / (e^{n^1} + e^{-n^1})$ ，第二層轉移函數 f^2 為使用雙正切 S 型函數 $a^2 = (e^{n^2} - e^{-n^2}) / (e^{n^2} + e^{-n^2})$ ，第三層轉移函數 f^3 為使用線性函數 $a = n$ ，輸出層的輸出為 $a^3 = [\hat{\theta}_1 \ \hat{\theta}_2 \ \hat{\theta}_3 \ \hat{\theta}_4]_{la}$ (or $a^3 = [\hat{\theta}_1 \ \hat{\theta}_2 \ \hat{\theta}_3 \ \hat{\theta}_4]_{ra}$)， $[\hat{\theta}_1 \ \hat{\theta}_2 \ \hat{\theta}_3 \ \hat{\theta}_4]$ 為估算出馬達角度命令。

圖 4.5 中包含輸入層、1 個隱藏層和輸出層，輸入層為左(右)手的三維座標 $[x_1 \ y_1 \ z_1 \ x_2 \ y_2 \ z_2]_{la}$ ($[x_1 \ y_1 \ z_1 \ x_2 \ y_2 \ z_2]_{ra}$)，第一層的加權值 W^1 設定為 60 組，轉移函數 f^1 為使用雙正切 S 型函數

$a^1 = \frac{(e^{n^1} - e^{-n^1})}{(e^{n^1} + e^{-n^1})}$ ，第二層轉移函數 f^2 為使用線性函數

$a = n$ ，輸出層的輸出為 $a^2 = [\hat{\theta}_1 \ \hat{\theta}_2 \ \hat{\theta}_3 \ \hat{\theta}_4]_{la}$ ($a^2 = [\hat{\theta}_1 \ \hat{\theta}_2 \ \hat{\theta}_3 \ \hat{\theta}_4]_{ra}$)，

$[\hat{\theta}_1 \ \hat{\theta}_2 \ \hat{\theta}_3 \ \hat{\theta}_4]$ 為估算出馬達角度命令。

本研究採用 Levenberg Marquardt 的倒傳遞(Back Propagation)學習演算法作為類神經學習方法，此演算法在函數的逼近上有較快的處理速度並且有更小的平方誤差，訓練結果後面會說明。

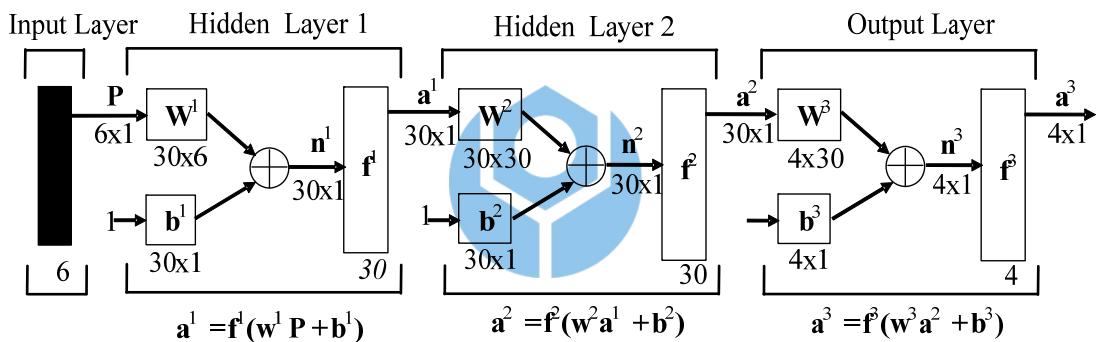


圖 4.4 雙層 30 神經元類神經架構圖

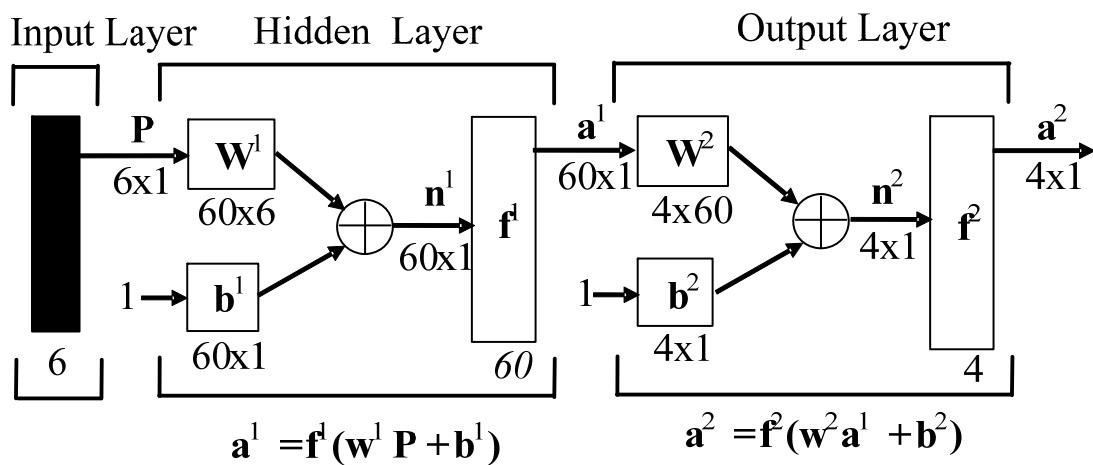


圖 4.5 單層 60 神經元類神經架構圖

4.1 右手神經建模

本章中我們手部類神經訓練資料範圍為第一顆馬達：0~180(每 10 度)、第二顆馬達：0~110(每 10 度)、在第一及第二區域中第三與第四顆馬達：-90~90(每 22.5 度)、在第三及第四區域中第三與第四顆馬達：-90~90(每 15 度)，並根據不同的工作區域各自做類神經建模。

表 4.2 右手四個工作區域類神經訓練結果

工作區域	No. of Data	隱藏層數	神經元數	時間(分)	MSE
1 st	2908	One	30	35	2.43E-03
			40	52	1.80E-03
			50	72	1.55E-03
			60	98	1.25E-03
		Two	30	296	3.73E-04
2 nd	1259	One	30	33	2.38E-03
			40	51	1.79E-03
			50	73	1.55E-03
			60	97	1.42E-03
		Two	30	276	8.97E-04
3 rd	1906	One	30	22	1.7E-03
			40	33	1.23E-03
			50	46	1.01E-03
			60	66	8.27E-04
		Two	30	191	3.78E-04
4 th	1906	One	30	21	2.75E-03
			40	32	2.11E-03
			50	46	1.48E-03
			60	63	1.35E-03
		Two	30	200	1.14E-03

表 4.2 是機器人手臂 4 軸依據工作區域的不同去做類神經網路所

得的時間及 MSE，可以從以下結果得知輸入數和神經元數越多，可以使得 MSE 越好，但花費的時間越多，而雙層跟單層的類神經模擬相比，雙層的訓練結果也會有比較好的 MSE 結果，但花費的時間也是越多。

本節中我們手部內插測試的資料範圍為第一顆馬達：0~180(每 10 度)、第二顆馬達：0~110(每 10 度)、在第一及第二區域中第三與第四顆馬達：-90~90(每 -60,30,30,60 度)、在第二及第三區域中第第三與第四顆馬達：-90~90(每 22.5 度)。比較各內插過後，我們選用在第一及第二區域以雙層 30 神經元類神經網路，在第三及第四區域以單層 60 神經元類神經網路，如圖 4.6 所示。

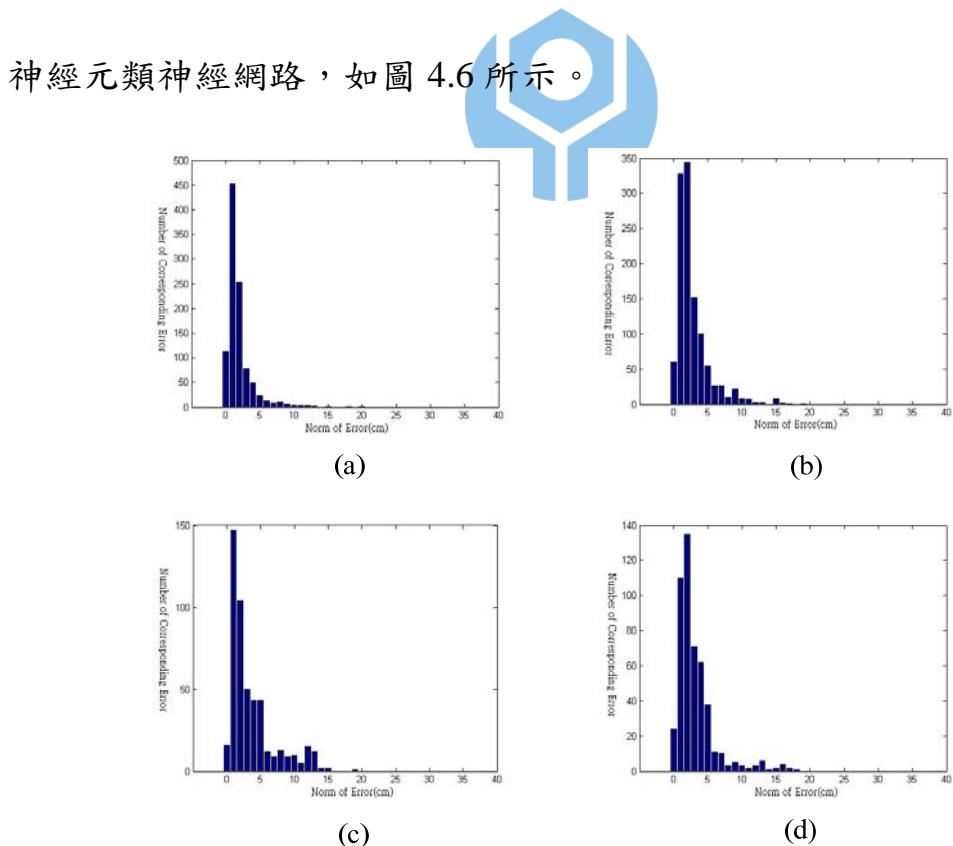


圖 4.6 類神經內插結果(a)1st sub-work space , (b)2nd sub-work space , (c) 3rd sub-work space , (d) 4th sub-work space

4.2 左手神經建模

根據上面右手所獲得的結果，我們左手也使用相同架構，在第一及第二區域以雙層 30 神經元類神經網路，在第三及第四區域以單層 60 神經元類神經網路，結果如表 4.3 與圖 4.7。

表 4.3 左手各工作區域 6 輸入類神經訓練結果

工作區域	No. of Data	隱藏層數	神經元數	時間(分)	MSE
1 st	2908	One	60	95	1.24E-03
		Two	30	283	4.31E-04
2 nd	2908	One	60	96	1.55E-03
		Two	30	276	7.26E-04
3 rd	1906	One	60	69	9.01E-04
		Two	30	196	3.93E-04
4 th	1906	One	60	68	1.33E-03
		Two	30	201	7.83E-04

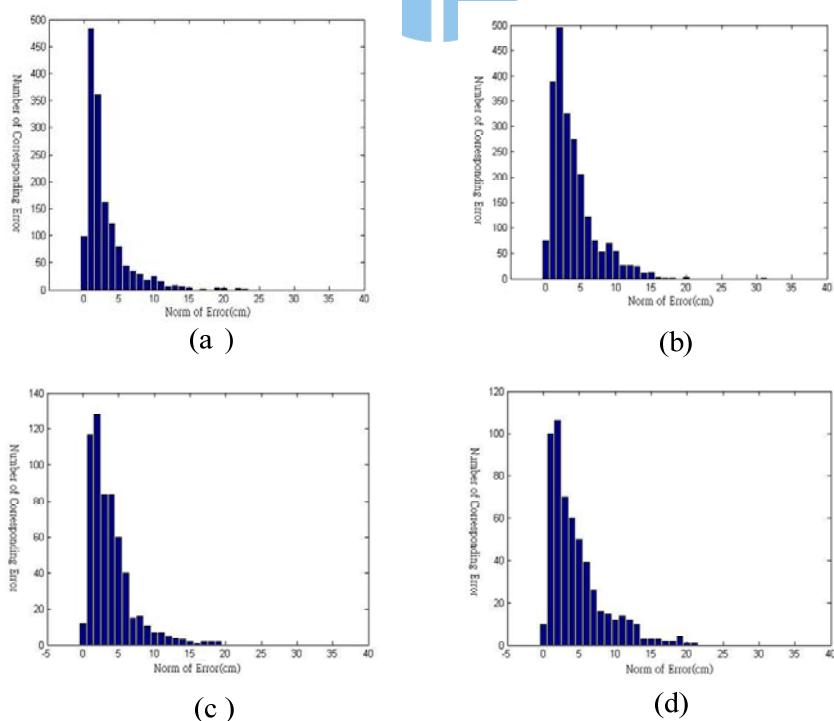


圖 4.7 類神經內插結果(a)1st sub-work space , (b)2nd sub-work space , (c) 3rd sub-work space , (d) 4th sub-work space

第五章 實驗結果

本論文之研究為利用立體視覺系統實現人形機器人於三維座標上的學習。學習者與表演者面對面站著，距離大約 120 公分，藉由立體視覺鏡頭透過 IEEE1394 將資料傳輸到電腦，經由影像處理擷取到表演者的 5 個特徵點，將 5 個特徵點的軌跡透過下半身類神經網路分類及上半身類神經網路模型轉成馬達的角度，再透過無線網路將角度傳給嵌入式系統，並透過人機介面讓學習者完成動作。表演者的動作如圖 5.1。下面的圖 5.2 是類神經分類及濾波後結果，圖 5.3 為擷取的代表動作，圖 5.4 和圖 5.5 為學習者的學習結果。



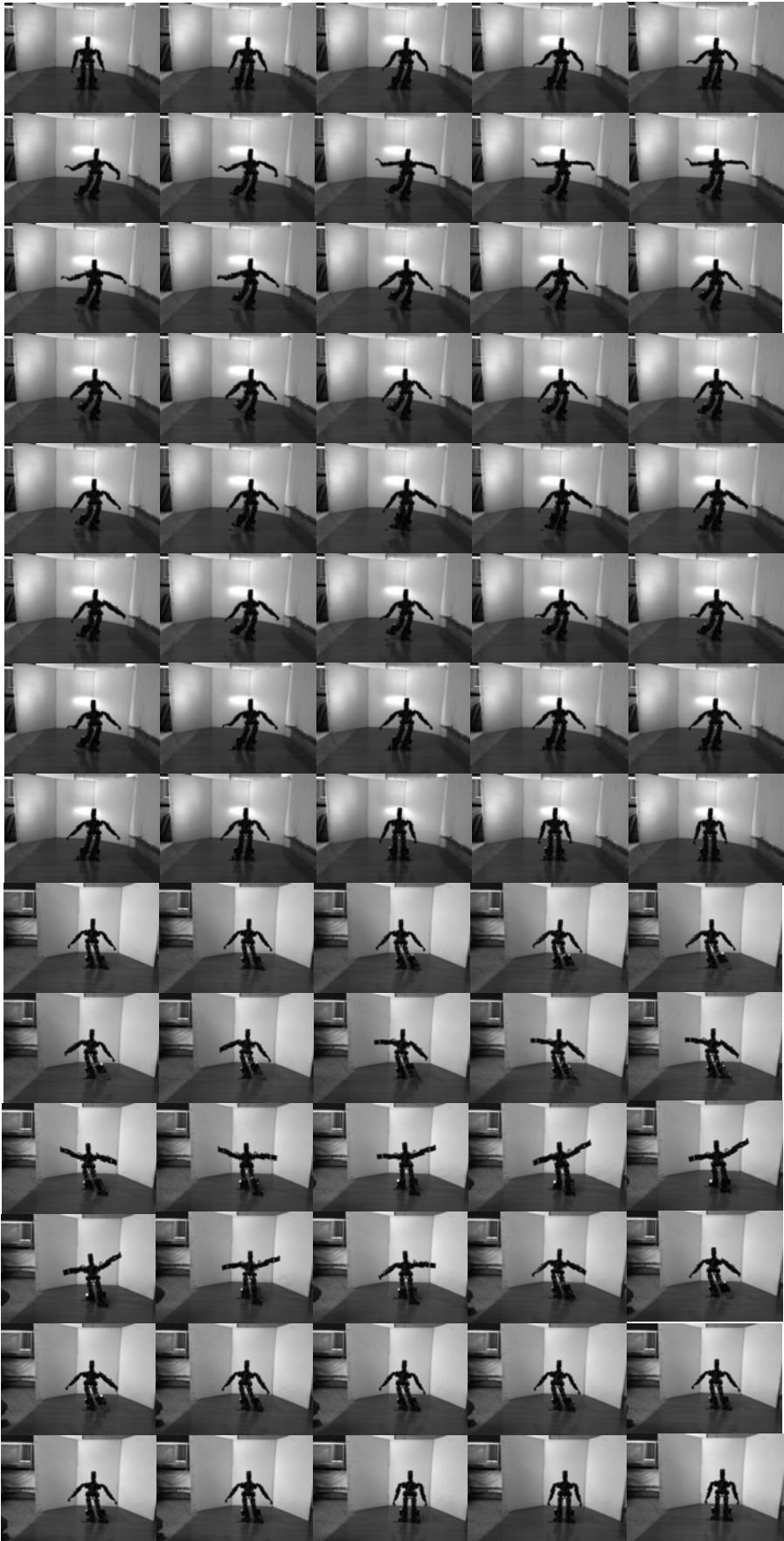


圖 5.1 表演者動作影像

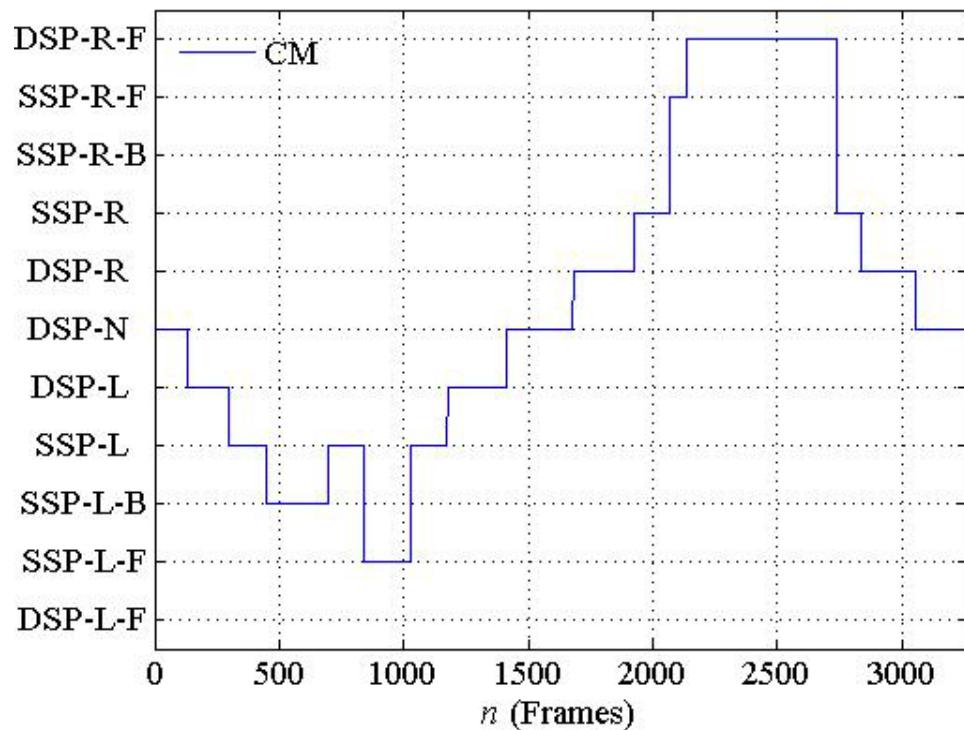


圖 5.2 類神經分類及濾波後結果

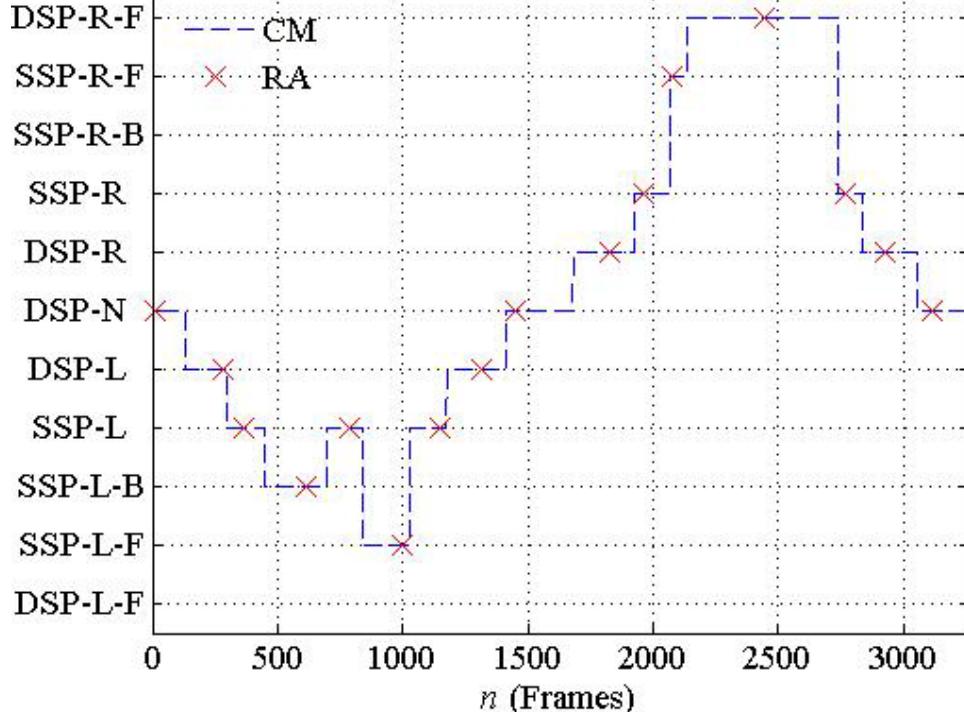


圖 5.3 挑選代表動作後結果

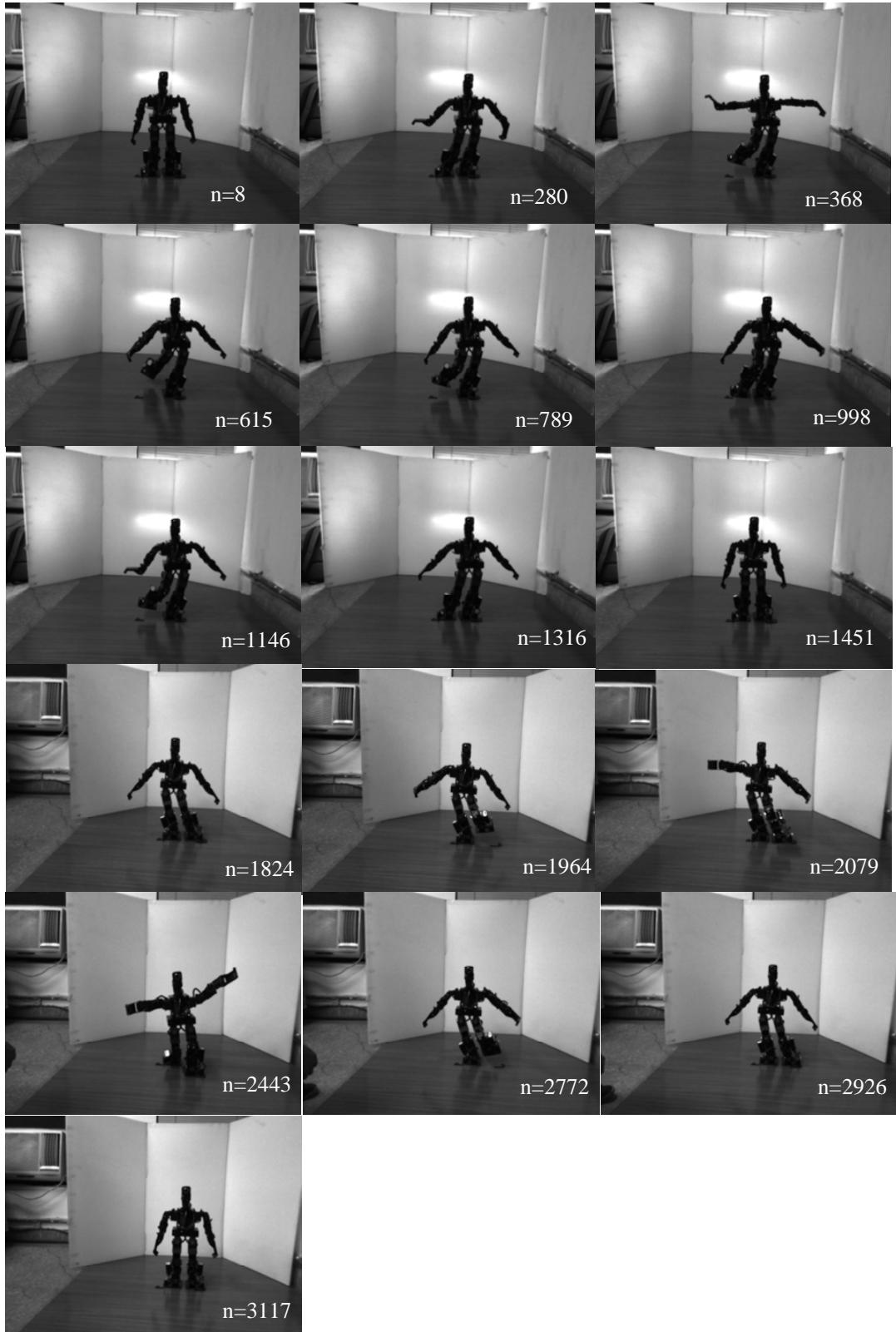




圖 5.4 表演者代表動作之正面(立體視覺拍攝)及側面圖(攝影機)拍攝

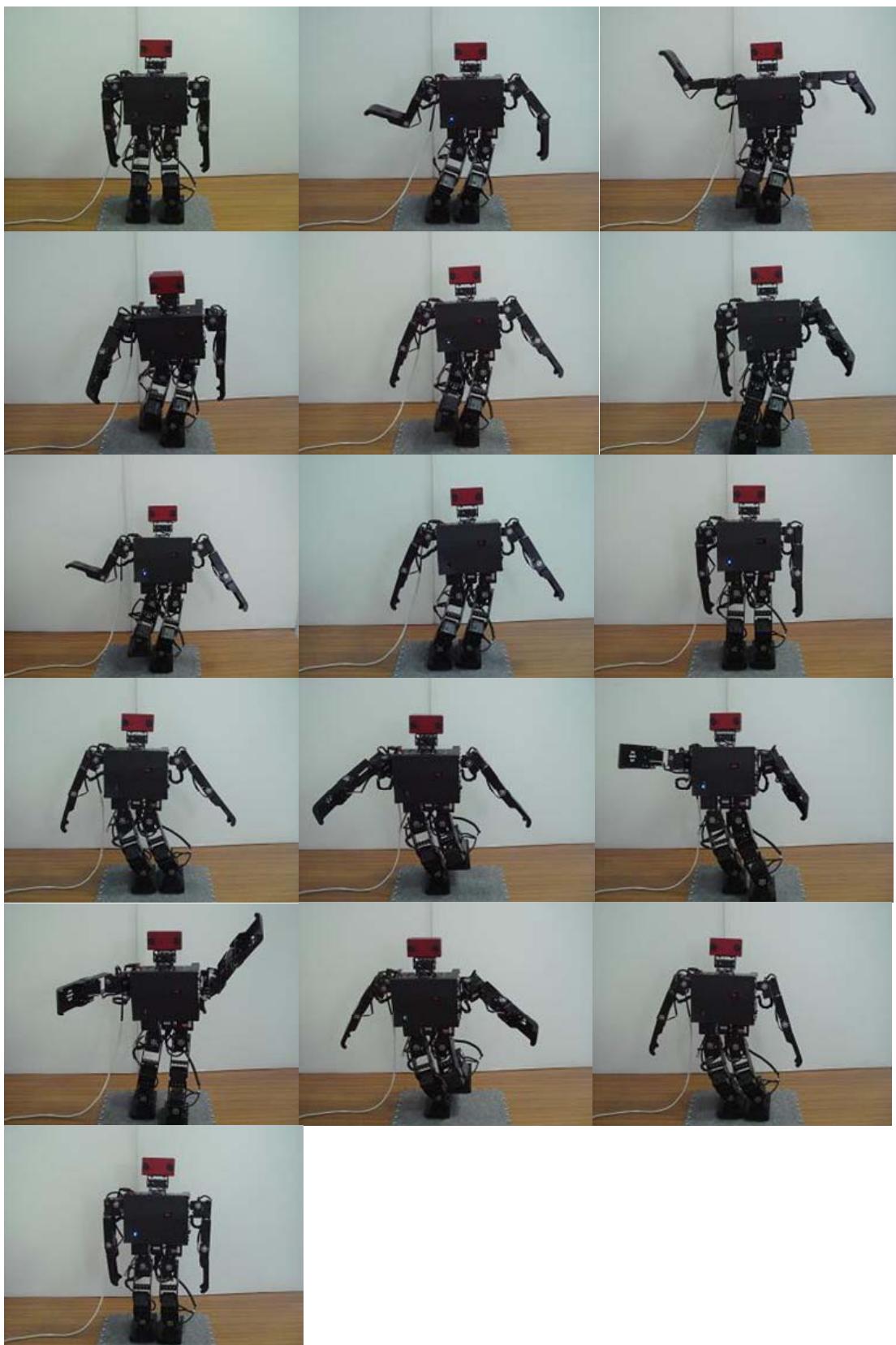


圖 5.5 學習者動作之正面圖

第六章 結論及未來展望

本論文利用立體視覺鏡頭(STOC)將影像資料透過 IEEE1394 傳送到電腦，並在電腦上做相關影像處理，包含動態閾值、背景註冊及特徵點辨識，並擷取學習者的特徵點，將特徵點的三維座標透過類神經網路來分類下半身動作並區分出動作區間，透過所建立的模組化動作，即可實現各對應平衡狀態的反運動學目標。再藉由分析軌跡空間頻率的分佈狀況，可將分類出的動作區間各自挑選出適合學習的代表動作，然後將挑出代表動作的特徵點三維座標以類神經網路計算出所對應的馬達角度。最後，相關實驗證明了方法的可行性，成功地實現人形機器人視覺學習之目標。



在未來展望中，希望能將學習實驗即時化，不用在等表演者做完動作經處理後才能做出學習的動作，而下半身模組化動作也限制了下半身學習的動作，若能結合壓力感測器實現平衡功能，如此一來將可呈現更好的學習動作。由於我們利用輪廓去判斷特徵點，若能提升影像處理技術，在動作重疊時也能判斷出特徵點，以及增加膝蓋的特徵點使動作能更精細。另外更換新的立體視覺系統設備也能實現本研究所提出之學習理論，如華碩公司推出的 XtionPro 等立體視覺系統，透過紅外線偵測技術以及 openNI 的演算法，即使重疊也能精準地掌握到肢體的特徵資訊。

參考文獻

- [1] Schaal, S. and Atkeson, C. G., “Learning control in robotics,” *IEEE Robotics & Automation Magazine*, vol.17, no.2, pp. 20-29, 2010.
- [2] Kober, J. and Peters, J., “Imitation and reinforcement learning,” *IEEE Robotics & Automation Magazine*, vol. 17, no. 2, pp. 55-62, 2010.
- [3] Kruger, V., Herzog, D., Baby, S., Ude, A., and Kragic, D., “Learning actions from observations,” *IEEE Robotics & Automation Magazine*, vol. 17, no. 2, pp. 30-43, 2010.
- [4] Chersi, F., “Learning through Imitation: a biological approach to robotics,” *IEEE Transactions on Autonomous Mental Development*, vol. 4, no. 3, pp. 204-214, 2012.
- [5] Hüser, M. and Zhang, J., “Visual programming by demonstration of grasping skills in the context of a mobile service robot using 1D-topology based self-organizing-maps,” *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 463-472, 2012.
- [6] Calinon, S., D'halluin, F., Sauser, E. L., Caldwell, D. G., and Billard, A. G., “Learning and reproduction of gestures by imitation,” *IEEE Robotics & Automation Magazine*, vol.17, no.2, pp. 44-54, 2010.
- [7] Khansari-Zadeh, S. M. and Billard, A., “Learning stable nonlinear dynamical systems with Gaussian mixture models,” *IEEE Transactions on Robotics*, Vol. 27, No. 5, pp. 943-957, 2011.
- [8] Ude, A., Gams, A., Asfour, T., and Morimoto, J., “Task-specific generalization of discrete and periodic dynamic movement primitives,” *IEEE Transactions on Robotics*, Vol. 26, No. 5, pp. 800-815, 2010.

- [9] Kim, S., Kim, C. H., You, B. J., and Oh, S. R., "Stable whole-body motion generation for humanoid robots to imitate human motions," *IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS), St. Louis, MO, USA, pp. 2518-2524, 2009.
- [10] Chen, G., Xie, M., Xia, Z., Sun, L., Ji, J., Du, Z., and Wang, L., "Fast and accurate humanoid robot navigation guided by stereovision," *International Conference on Mechatronics and Automation* (ICMA), Jilin, China, pp. 1910-1915, 2009.
- [11] Thobbi, A. and Sheng, W., "Imitation learning of hand gestures and its evaluation for humanoid robots," *Proceedings of the 2010 IEEE Int. Conf. on Information and Automation (ICIA)*, Harbin, China, pp.60-65, 2010.
- [12] Liu, H. Y., Wang, W. J., Wang, R. J., Tung, C. W., Wang, P. J. and Chang, I. P., "Image recognition and force measurement application in the humanoid robot imitation," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 1, pp. 149-161, Jan. 2012.
- [13] Tarokh, M. and Mikyung, K., "Inverse kinematics of 7-DOF robots and limbs by decomposition and approximation," *IEEE Transactions on Robotics*, vol. 23, no. 3, pp. 595-600, 2007.
- [14] Hwang, C. L. and Huang, J. Y., "Neural-network-based 3-D localization and inverse kinematics for target grasping of a humanoid robot by an active stereo vision system," *The 2012 International Joint Conference on Neural Networks (IJCNN)*, Brisbane, Australia, pp. 1-8, 2012.
- [15] Juang, C. F., Chang, C. M., Wu, J. R., and Lee, D., "Computer vision-based human body segmentation and posture estimation," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 39, no. 1, pp. 119-133, 2009.
- [16] Wu, Q. Z., Cheng, H. Y., and Jeng, B. S., "Motion detection via

change-point detection for cumulative histograms of ratio images,"
Pattern Recognition Letters, vol. 26, no. 5, pp. 555-563, 2005.



國立臺灣科技大學博碩士論文授權書

(本授權書裝訂於紙本論文內)

本授權書所授權之論文為王兆葵〔M10007422〕在國立臺灣科技大學電機工程系 101 學年
度第 2 學期取得碩士學位之論文。

論文題目： 應用類神經網路之分類及建模於人形機器人三維運動之視覺學習
指導教授： 黃志良

茲同意將授權人擁有著作權之上列論文全文〔含摘要〕，依下述授權範圍，以非專屬、無
償授權本校圖書館及國家圖書館，不限地域、時間與次數，以紙本、微縮、光碟或其他數
位化方式將上列論文重製典藏，並提供讀者基於個人非營利性質之線上檢索書目、館內閱
覽、或複印。

授權人 王兆葵

簽章
(請親筆正楷簽名) 

備註：

1. 兹確認本論文之著作權係王兆葵單獨擁有。
2. 授權人不因本授權而喪失上述著作之著作權。
3. 本授權書請授權人簽章後，裝訂於紙本論文內。

指導教授簽章：

中 華 民 國 102 年 7 月 15 日