

教育部100年度「學海築夢」獎助大專校院
海外專業實習計畫心得報告

原就讀學校科系：國立臺北科技大學電腦與通訊研究所

姓名：陳玠霖同學

實習國家：德國

實習機構：德國明斯特大學與德國科技中心

獲獎年度：100 學年度

一、前言

立體視訊系統的發展已經有百餘年的歷史，它的起源可以追溯至 19 世紀歐洲的一位科學家根據左右眼成像的差異而發明的立體眼鏡，使得配戴眼鏡的左眼與右眼接收到不同的畫面，而產生立體的效果。這樣的原理發展至今也沒有突破性的變化，雖然人們使用各種不同的技術來實現，卻皆是依循相同的原理。德國一直以來被人們視為工業強國，除了眾所皆知的重工業與汽車工業等，他們在科技產業上的發展也不容小覷，而立體電視相關的技術也是發展甚早，並且從立體影像的拍攝開始一直到最終的顯示端均有發展，擁有的完整技術讓我們這趟實習獲益良多。

這次的實習，我們的規劃是以「三維立體視訊系統」為主軸，因此相關的規劃皆是圍繞著此主軸為實習的目標。將「三維立體視訊系統」從理論端、影像來源端、內容端、演算法端直到最後的顯示端，整個設計流程做個深入的研究探討，在此的同時，我們必須於實習期間，依據所見所學，完成實習實作，並且將實習成果在國外發表。為了進一步了解德國人引以為傲的「德國精神」，我們同時安排了工業參觀的行程。

二、行前規劃

在出發之前，我們依照實習的城市及日期，事先預訂好每一個住宿的地點，以確保實習的順利與安全。而交通方面，我們的規劃以鐵路作為交通工具，我們在出發前事先準備了當地的鐵路通行證(Rail Pass)，讓我們可以在工業實習與參訪時，利用此通行證到達實習的地點。

我們所參與的「德國三維立體視訊系統實習計劃」，透過實務學習，進一步了解三維立體電視的各個流程。

我們的實習獲得來自於德國明斯特大學的協助，在三維立體視訊系統演算法與技術上，習得深入的技術，我們以德國明斯特大學為主要學習的對象，透過這一次的海外實習，深入了解三維立體視訊技術。

接著為了能夠具備實務實習的經驗，DeutschesTechnikmuseum、3D Exploramuseum、Potsdam Filmpark 與 Deutsches Museum 提供我們實習操作的機會，讓我們除了習得系統演算法之外，能夠接觸到工業級的三維立體視訊系統。我們整個實習包含了「三維立體視訊系統演算法與技術實習」、「三維相機攝影機的光學系統前端設計實習」、「三維立體視訊系統成像端設計實習」、「三維立體視訊互動系統實習」、「實習實作：三維立體電視之自然使用者介面設計」、「涵養德國工

藝精神-參訪參訪魯爾工業區(Ruhr Area)」與「實習評量考核-成果發表」。

在實習訓練中，我必須完成實習的重點項目「**實習實作**」，並且須上台報告成果，與三維立體視訊專家，做進一步的意見交流，作為實習評量的考核，我的實習題目訂為「三維立體電視之自然使用者介面設計」，需設計完整的演算法實現三維立體電視之自然使用者介面。

在實習的過程之中，為了瞭解德國精益求精的科技精神，老師安排了一天的工業參訪，我到德國工業發展的發源地--魯爾工業區(Ruhr Area)，進一步體會學習德國工業的精神。

三、啟程

實習的起點是由台灣桃園國際機場搭乘中華航空 CI-61 的班機直飛往德國的法蘭克福機場。我們預計抵達德國後直接利用事先準備好的鐵路通行證搭乘當地的城際快車(Inter City Express)前往實習地點。



圖 1 華航 CI-61 台北往法蘭克福登機證



圖 2 歐洲鐵路通行證

四、Westfälische Wilhelms-Universität Münster

我們實習的地點位於德國的 Universität Münster，Universität Münster 位於北萊因-西發里亞邦的 Münster，創立於西元 1771 年，校內學生約有 37,200，Universität Münster 是德國最大和最著名的大學之一，Universität Münster 目前包含 7 所學院及 15 個科系。學校最早的歷史可以追溯到 1588 年建立的教徒學校。

Universität Münster 學術成就十分卓越，先後有 1 位校友和 1 名教授各獲得諾貝爾獎，10 位教授獲得萊普尼茲獎，5 位教授獲得馬克思-普朗克獎，1 位教授獲得洪堡獎。此外，Universität Münster 為德國培養出了聯邦德國總統 Gustav Walter Heinemann 以及許多部長、州長和大企業家。Universität Münster 具有良好的讀書氣氛，處處充滿書香，因此變成北萊因-西發里亞邦之著名大學城。

此外 Münster 向來有歐洲腳踏車之都的美譽，到了這個城市之後，會覺得四處皆可以看到人們騎著腳踏車。

這次協助我們「德國三維立體視訊系統實習計劃」的便是 Universität Münster 數學與計算機科學系(Mathematics and Computer Sciences)的蔣教授實驗室。



圖 3 指導教授與蔣教授合影於 Universität Münster

蔣教授的實驗室在三維立體視訊系統演算法設計上十分知名，博班學生在學就有 11 位博士學生，研究成果十分卓越。

實習的歲月裡，蔣教授實驗室教導我們許多關於的三維立體視訊的實務研究，包含三維視訊影像處理演算法、三維醫療影像處理、DIBR 技術，其中，蔣教授的一位博士班學生目前正在從事三維顯示之 DIBR 相關研究，這位博士班學生教導我們三維立體影像處理的訣竅，並辨別三維影像的優劣。在實習的過程當中，我們和 11 位博士班學生學習到三維立體視訊處理與影像處理演算法設計的經驗，讓我們獲益良多，經歷這一段實習的過程，相信對於我未來無論是在工作上或是學業上，將會有莫大的幫助。



圖 4 三維立體視訊模擬環境



(a)



(b)



(c)



(d)

圖 5 Universität Münster 與 Münster 城市一隅，(a) Universität Münster，(b) Universität Münster 之假日跳蚤市場，(c) Aasee Lake, Münster，(d) Münster 城市一隅。

五、三維相機攝影機的光學系統前端設計實習

在實習過程中，我們同時必須熟悉三維立體視訊系統的核心技術，為了瞭解德國在立體影像處理的各個層級所使用的器材與方式，Film Museum 與 3D Explora Museum 協助我們操作三維相機攝影機的光學系統前端設計，透過許多互動的方式讓我們操作電影或立體電影拍攝的設備、方式與手法。圖 6 即為不同時代所使用的各式錄影設備與不同的鏡頭。此外，我們也操作了一些立體影像拍攝與觀看的設備，這兩端的光學設計通常都是透過兩個鏡頭來拍，抑或是提供使用者兩個不同視角的方式來呈現立體的效果



圖 6 拍攝電影的錄影器材



圖 7 拍攝立體影像的相機

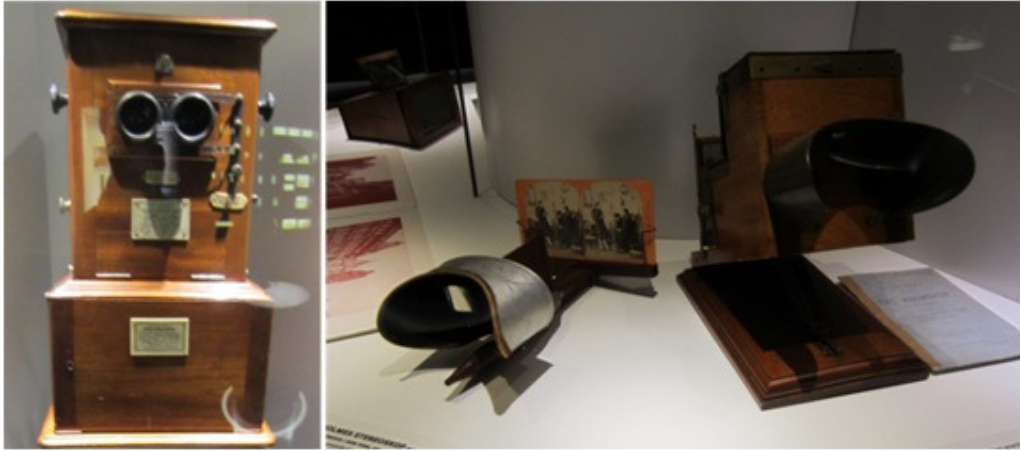


圖 8 立體影像觀看設備

六、三維立體視訊系統成像端設計實習

緊接著是「三維立體視訊系統成像端設計實習」，我們學習到許多立體影像的成像方法。在實習的過程中，我們先學習歷史上早期的立體影像相關技術，利用多層次圖像創造出的立體效果，雖然在三維顯示還未成熟的時代，就已經有立體影像呈現的概念，早期的技術將場景中不同深度的景象分別畫出，再將它們前後排列成一直線，最後觀賞的人從正面開的小洞觀看場景，這個方法所呈現出來的立體效果相當顯著，因為當中的物件是真的放在不同深度的位置，因此在視覺上有良好的效果，可惜的是每個物件本身並沒有立體感。



圖 9 利用多層次圖像創造出的立體效果

接著我們學習近代較常見的立體影像成像方式，如圖 9 中的立體影像與紅青雙色的立體影像。這兩種方式雖然呈現的方式不同，但是它們都讓人們的雙眼看到不同視角的影像所產生的立體效果。這兩種方式分別是利用柱狀透鏡的折射以及利用紅色與青色的濾光鏡(通常以對應顏色的玻璃紙實作)使得對應的眼睛可以看到正確視角的影像因而產生立體效果。

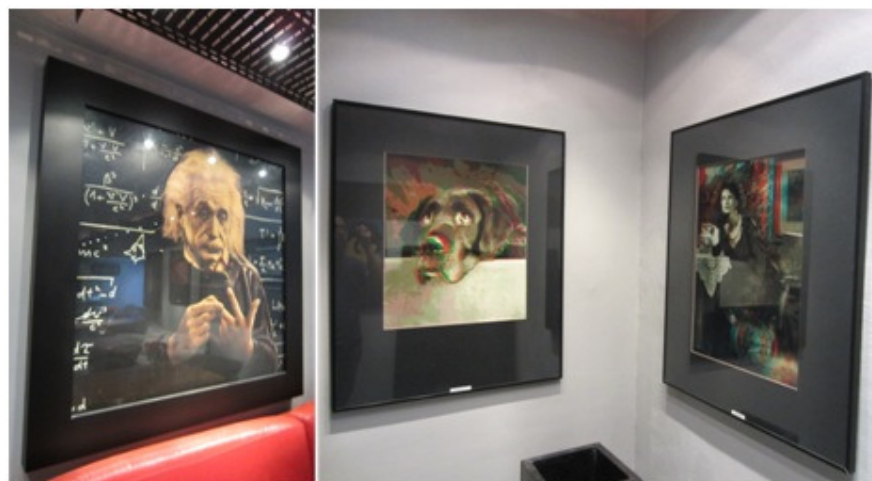


圖 10 利用光學方式呈現的立體效果

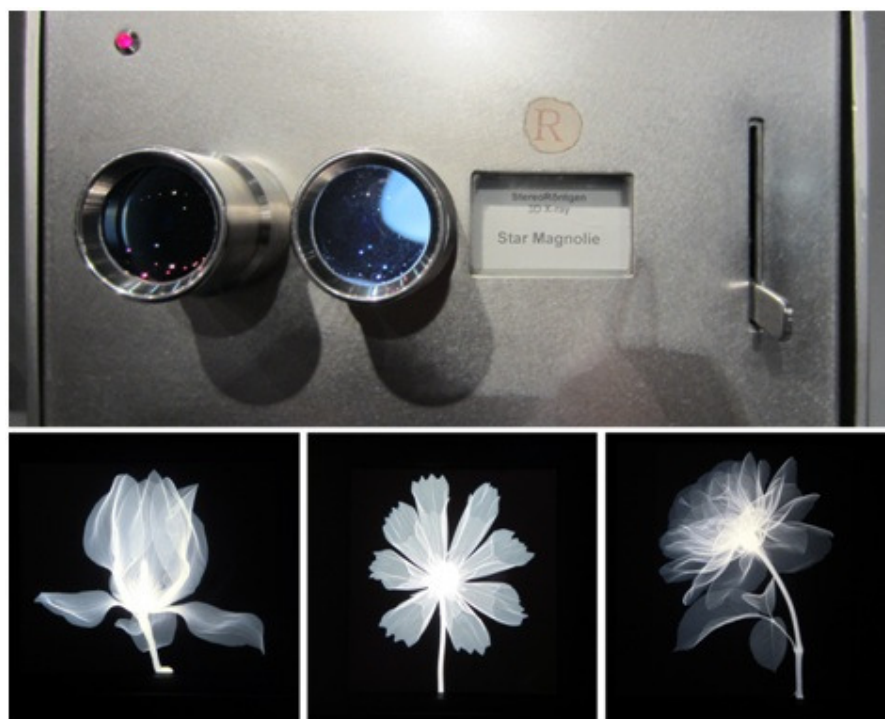


圖 11 將兩個帶有視差的圖片透過兩視角的呈現產生立體效果

七、三維立體視訊互動系統實習

完成「三維立體視訊系統成像端設計實習」之後，接著是「三維立體視訊互動系統實習」，我們實際操作三維立體視訊互動系統，實際透過三維立體互動系統編輯體驗，剪接知名電影的片段，排列不同角度所拍攝的畫面，讓我們習得三維電影剪輯的技巧，使我留下深刻的印象。

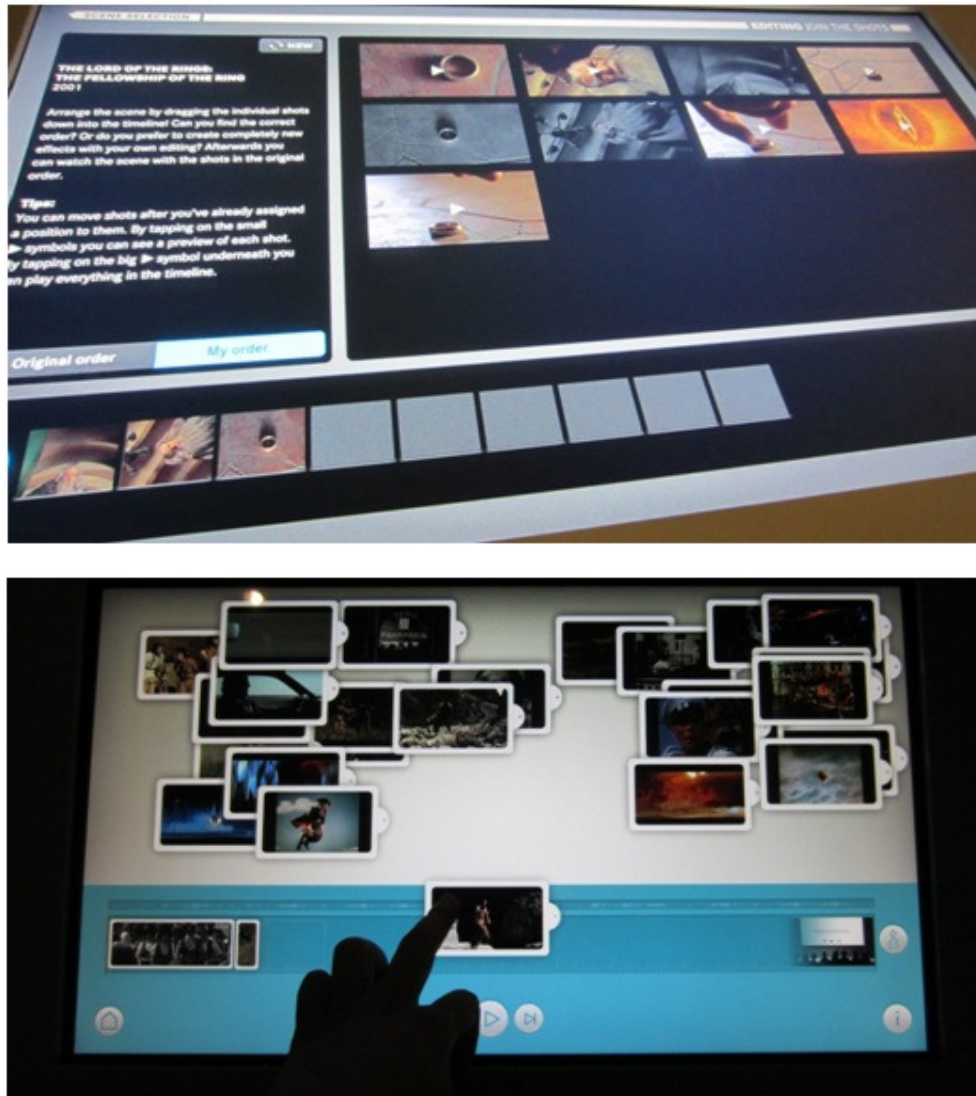


圖 12 三維立體互動系統編輯體驗

接著我們進入三維立體即時特效場景的環境中實習，如下圖，只要使用者站在一個綠色的布幕當中，就可以看到自己出現在眼前的電視中，而綠色的場景備置換成動態的真實影像，如同真實拍攝電影一般，進一步拍攝出電影的內容。



圖 13 三維立體即時特效場景的環境

八. 實習實作：三維立體電視之自然使用者介面設計

完成了「三維立體視訊互動系統實習」之後，接著進入到實習最重要的部分「實習實作」，實務實習的概念源自於德國務實的工匠精神，我們必須於實習期間，完成所訂定的實習實作，並且要通過實習評量考核，我的實習題目訂為「三維立體電視之自然使用者介面設計」，以下簡述我所完成的實習實作內容。

目前市面上的電視逐漸朝著智慧型電視的方向發展，然而使用者的介面與輸入裝置，大多停留在遙控器、鍵盤、滑鼠或是觸控等實體的介面。遙控器的介面簡單方便，但是對於要做出複雜控制的智慧型電視的使用上，可能會使人感到力有未逮。而上述除了遙控器外的介面大都不是為了電視本身而設計的輸入裝置，因此在使用上也有一些限制與不便。觸控介面雖然就貼近於直覺的方式，但是在家用的場合使用電視時，一般使用者與電視會保持一段距離，因此使用觸控做為電視介面也不甚理想。

為了解決上述的問題，本實習實作致力於設計一個自然使用者介面，做為智慧型電視的輸入介面。不同於市面上的體感輸入裝置，大多利用三維的偵測技術，建立環境的深度做為體感量測的基礎。本實習實作只使用單一攝影機做為影像來源，雖然在使用者的判別上較費工夫，卻也擁有重量輕巧、體積輕薄、成本低廉以及硬體需求較低等優點。

本實習實作在讀取到攝影機的影像後，經由影像處理過濾出膚色資訊，並且透過我所提出的演算法進行強健的手部追蹤，再從手掌的四周尋找出手指出現的區域，並且利用兩個類神經網路的搭配，將膚色的區域利用空間上的特徵分類為多隻候選手指，最後再經過後續的最佳化處理，過濾掉誤判的手指，便可偵測出畫面中每隻手指的座標與長度。

系統架構

本實習實作提出一個透過三個步驟實現手指追蹤的系統，其架構如圖 14 所示。自影像從視訊攝影機擷取至系統開始，經過三個階段的處理，分別為前處理(Pre-processing)、手部追蹤(Hand Tracking)以及手指追蹤(Finger Tracking)。可以從畫面中判斷出使用者的手掌中心位置、每隻手指的資訊(包含手指頭尾座標、手指長度與旋轉角度等)，並且透過連續畫面間的判斷，可以辨識出使用者手指點擊(Click)的動作。

前處理 (Pre-processing)

前處理主要由兩個功能所組成，首先系統會利用一個事先設定好的動作來判斷使用者的手是否進入畫面。本實習實作判斷使用者是否在位於畫面右上角的關注區域(Region of Interest, ROI)做出「推」(Push)的動作做為初始化的依據，當使用者的手部做出「推」的動作後，系統便會透過影像處理的方式過濾畫面中不需要的資訊，並且於第一次依照使用者的手部膚色建立膚色模型，做為往

後追蹤時過濾畫面的依據。

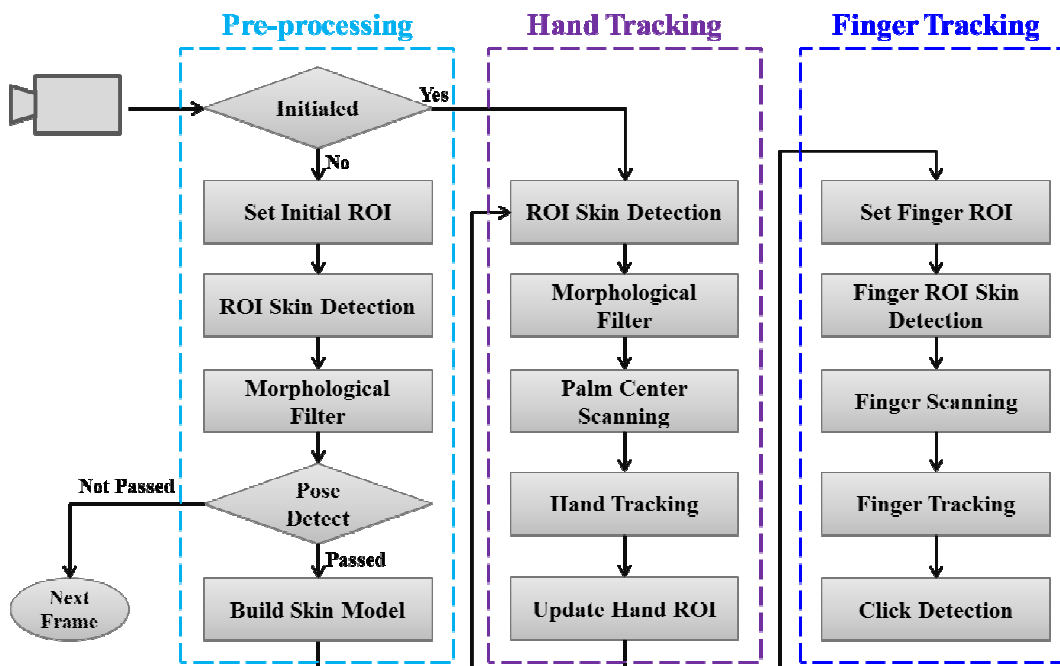


圖 14 系統架構圖

關注區域 (Region of Interest, ROI)

本實習實作採用單一攝影機做為影像的來源擷取使用者的畫面。然而若是單純地採用顏色做為判斷或辨識使用者手部的依據，雖然不無可能，卻由於日常生活的環境中，充斥著許多與膚色顏色相近的物品，因此容易造成誤判的情況，造成準確度的降低，進而影響使用者體驗。因此本實習實作採用關注區域的方式，與使用者約定好初始關注區域的位置以及初始的確認動作，如圖 15 右上角的矩形區域，不僅可以大大的減少誤判的機率。更由於系統大部分時間只需處理關注區域當中的影像而非完整的畫面，故可加快系統處理與分析的速度，可謂一舉兩得。

初始動作偵測 (Initial Pose Detection)

在本實習實作當中的初始化動作以 Push 做為判斷的依據，由於這個動作相當容易做到，並且透過適當的演算法，相當的容易判斷，因此我們採用這樣的動作做為初始化的動作。然而在判斷之前，為了要適當的濾除不必要的雜訊，我們以 YCbCr 色彩空間當中的 CbCr 分量來初步的判斷膚色，減少雜訊。接著，由於手部在 ROI 當中佔有許多的比例，可視為相對較大的物件，因此我們利用影像處理中型態濾波的「斷開」(Opening)技巧，濾除較為細小並且被誤判為膚色的部分後，才進行動作的偵測。



圖 15 初始畫面的關注區域

觀察使用者在攝影機前做出的 Push 動作可發現，使用者的手部會從距離攝影機較遠處漸漸的移近攝影機，接著再漸漸地遠離。利用分析這樣的特性我們可得到，在使用者做出 Push 的動作時，不論 Push 幅度的大小，都必定會經過一次的漸近轉變為漸遠的轉折，因此我們採用此轉折做為判斷使用者做出 Push 動作的依據。

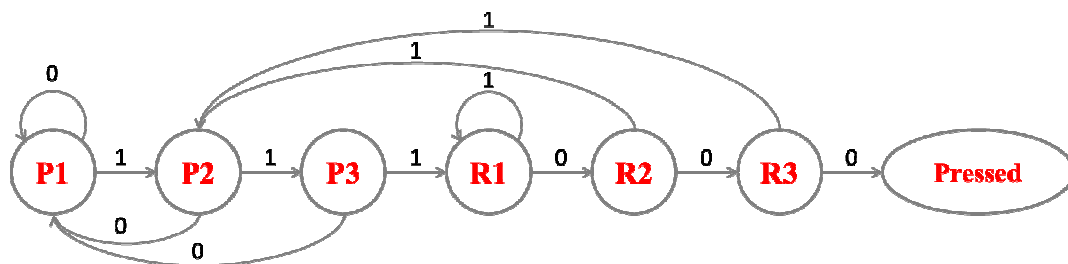


圖 16 偵測使用者 Push 之狀態圖

偵測的方式如圖 16 的狀態所示，當使用者的手部與攝影機距離越近，手部的大小就會隨之變大，反之則會縮小。因此我們將放大的狀態設為 1 而縮小設為 0，則可由圖中所示狀態判斷出使用者的動作。在本實習實作中我們採用連續三次的放大，接著連續三次的縮小做為判斷的標準，其中在放大的過程當中，若未滿三次手部便開始縮小，則系統會判定為雜訊而忽略此部分，若放大超過三次，則代表使用者 Push 的幅度大於系統預設的數值，因此狀態會停留在放大的最後一個狀態，直到使用者的手部縮小，則進入下一個狀態。

連續放大與縮小的次數，乃是經過多次實驗的結果，由於視訊攝影機在每個畫面拍攝時，常有雜訊與誤差產生，因此若單採用一次的放大縮小就判定為

動作成立，會非常容易發生誤判的情形；而若是判斷的標準設定得太嚴苛，則會容易受到雜訊的干擾，讓使用者在使用時不易達到系統的標準，因此我們在此兩者間取得平衡，使得系統不易誤判又不致門檻過高而難以達成。

建立膚色模型 (Build Skin Model)

在通過初始動作的偵測後，系統假設在關注區域當中的物件，即為使用者的手部。因此，為了改善膚色偵測的品質與降低誤判的情形，我們採用動態膚色偵測(Adaptive Skin Color Detection)的方法。它會取樣使用者的手部的顏色，並且搭配 Normalized RGB(r, g, b)以高斯分佈建立膚色模型，讓系統可以更精確的過濾出與使用者手部相近的膚色。

膚色模型的建立是透過使用者膚色區域的均值與標準差的計算。如下列公式所示，(1)式計算出使用者手部正規化紅色(r)、正規化綠色(g)與 RGB 色彩空間紅色(R)的均值(Mean)，以 μ_i 表示之。式中 hand 代表手部的膚色區域而 n 則為 hand 所包含的 Pixel 個數。(2)式計算的則是上述三個色彩分量的標準差(Standard Deviation)，並以 σ_i 表示之。

$$\mu_i = \frac{1}{n} \sum_{(x,y) \in \text{hand}} \text{Image}(x,y), \quad i = r, g, R \quad (1)$$

$$\sigma_i = \sqrt{\frac{1}{n} \sum_{(x,y) \in \text{hand}} (\text{Image}(x,y) - \mu_i)^2}, \quad i = r, g, R \quad (2)$$

有了上述三個色彩空間分量的均值與標準差後，我們透過膚色偵測方法，利用 Higher Bound 以及 Lower Bound 來判斷影像中的顏色，是否屬於所建立使用者膚色模型的範圍。而 Higher Bound 與 Lower Bound 的計算，如(3)、(4)與(5)式所示，透過判斷 r、g 與 R 三個空間的數值，若三者皆介於 Higher Bound 與 Lower Bound 之間者，即可判定為膚色。其中(3)與(4)當中的 C 是一個常數，用來決定顏色範圍的大小。而 μ_i 及 σ_i 則是(1)與(2)式的結果，(5)式中的 Skin 則是透過判斷三個色彩分量是否皆介於高低邊界之間，若同時成立則視為膚色。

$$\text{Higher Bound}_i = \mu_i + C \times \sigma_i, \quad i = r, g, R \quad (3)$$

$$\text{Lower Bound}_i = \mu_i - C \times \sigma_i, \quad i = r, g, R \quad (4)$$

$$\begin{aligned} \text{Skin}(x,y) &= \text{Higher Bound}_r > r > \text{Lower Bound}_r \ \& \\ \text{Skin}(x,y) &= \text{Higher Bound}_g > g > \text{Lower Bound}_g \ \& \\ \text{Skin}(x,y) &= \text{Higher Bound}_R > R > \text{Lower Bound}_R \end{aligned} \quad (5)$$

手部追蹤 (Hand Tracking)

第二個階段的手部追蹤則是建立於前一步驟所建立之膚色模型的基礎，對關注區域做膚色的過濾後，會產生一個膚色的遮罩(Mask)。再對此遮罩做雜訊的濾除後，系統則會計算使用者手掌的中心，並依據此中心持續的追蹤使用者的手掌位置，以提供給下一個階段，即手指追蹤的部分做為搜尋手指的依據。

手掌中心掃描 (Palm Center Scanning)

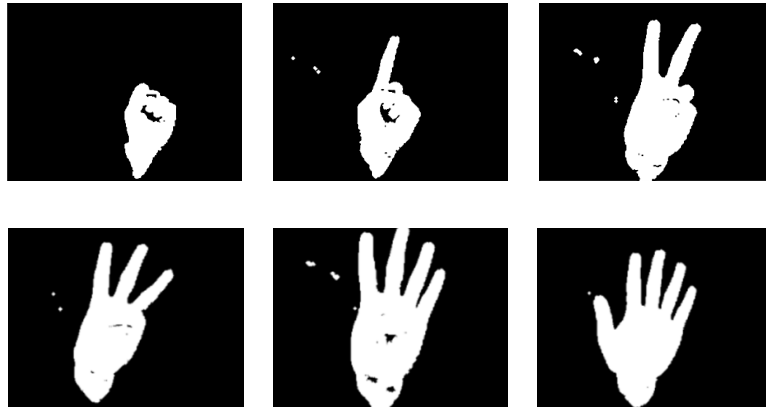
由於手指在視訊的畫面當中，屬於相對較小的物件，較容易受到雜訊的干擾而誤判。反之手掌相對的比手指大許多，因此為了提高系統追蹤的可靠度，我們將手掌與手指的追蹤分成兩個部分，分別進行追蹤，讓系統不至於在手指偵測受到雜訊干擾時，也一併遺失了移動中的手掌圖形。

手掌中心的掃描，我們採用重心的運算，如(6)式所示，重心 C 是來自於手掌點座標的總和除以手掌點座標的個數，其中 $I(x,y)$ 是在 (x,y) 座標時膚色遮罩的數值，1 代表膚色，0 則代表不是膚色。然而使用重心做為計算手掌中心的方法固然有效，但卻會受到使用者手指的方向與數目影響而不穩定，例如使用者將手掌完全張開時的重心相較於握拳時的重心會有偏高的情形，此情況是因為手指也被加入了重心的計算，導致其位置不穩定。

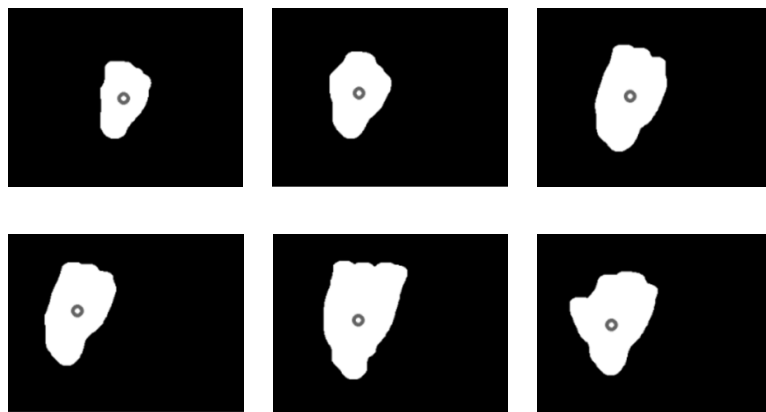
$$C = \frac{\sum_{(x,y) \text{ chand}} I(x,y) \cdot (x,y)}{\sum_{(x,y) \text{ chand}} I(x,y)} \quad (6)$$

為了解決上述的問題，我們使用影像處理中的型態斷開(Morphological Opening)濾波器，並且設定其結構元素的大小大於手指而小於手掌，如此便可使畫面中的手指自關注區域中濾除而留下手掌的部分。因此，不論使用者的手部握拳亦或是開掌，對於重心座標的影響都大大的減小，使得系統的重心不致於受到手指的干擾而產生不穩定的現象。

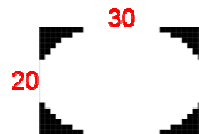
其效果如圖 17 所示，在濾波前的手部截圖如圖 17(a)，圖中的手會因為使用者的手勢變化而影響到重心計算的結果。此外，由於手部握拳時，容易在手指與手掌的交界處(手心附近)產生陰影，因此在做型態斷開的運算之前，我們先使用了一個型態閉合(Morphological Closing)濾波器，將拳頭當中的縫隙填滿，再以型態斷開將手指濾除，其結果便如圖 17(b)所示，雖然手指濾除後的手掌不盡相同，但此運算仍可將大部分的手指濾除，因此也已經達到穩定手掌重心的目的。



(a) 型態濾波前之手部截圖



(b) 型態濾波後之手部截圖



(c) 用於型態濾波的橢圓形結構元素

圖 17 型態濾波前後比較圖

在此實驗中，如圖 17(c)所示，我們使用大小為 30×20 的結構元素做為型態濾波的基礎，因此可輕易的濾除寬度較窄的手指。然而經過多次的實驗後，由於考量到使用者的手指可能轉向不同的角度，因此我們將型態濾波器所使用的結構元素修正為 30×30 的圓形結構元素，如此一來便可接受使用者各個方向的手指，而不影響其效果。

手部追蹤 (Hand Tracking)

在關注區域當中，手掌是相對相當大的物件，常佔有關注區域中百分之五十以上的區域。此外由於視訊攝影機在捕捉畫面時會有每秒鐘30張的更新率，因此手掌在每一張畫面中的移動距離不致過大，因此我們簡化追蹤的方法，直

接假設上一張畫面的手掌中心在此張畫面還會落在手掌當中。基於這個假設，我們可以直接重新計算此物件的中心點，而不需要重新計算整個關注區域的手部重心，因此可省去上萬次的累加與龐大的除法運算。

取而代之的方法，我們透過手掌的寬度與高度的計算，不斷的從上一張畫面的手掌中心尋找更新的手掌中心。方法如圖 18 所示，圖中灰色圓形代表上一張畫面的手掌中心點，由於在此張畫面手部位置已經更新，因此手掌中心也必須更新，以達到追蹤的效果。

圖 18(b)即為中心計算的第一步，自前一張畫面的中心點出發，計算手掌的左右邊界位置，並找到其中心點。接著如圖 18(c)，尋找手掌的上下邊界，並且計算上下兩點的中心點。如此重複計算，直到中心點幾乎不再移動，就代表已經相當趨近於手心，運算便可停止。用此方法雖然無法像重心的計算可以精準地找到手掌的形狀中心，卻可以十分快速的讓手掌中心維持在接近手掌重心的

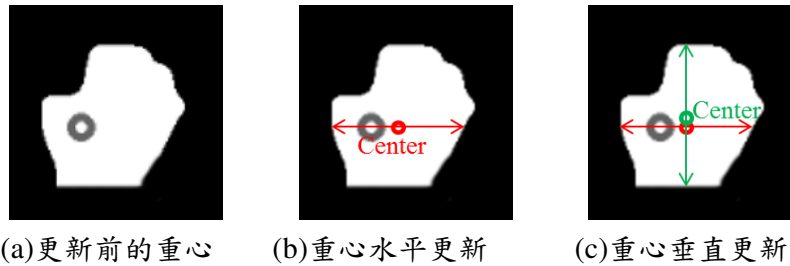


圖 18 手掌中心重新計算

關注區域更新 (ROI Updating)

由於手掌中心會隨著每張畫面捕捉的位置不同而重新計算，因此關注區域也必須隨著手部的移動而更新，否則在使用者的手部移開原本的關注區域後，便無法再捕捉到其手部畫面。因此我們在得到穩定的手掌中心點後，也會隨之修改關注區域的位置，讓它保持在可以捕捉到手掌的位置，如此便可持續的重複追蹤使用者的手部，以提供手指追蹤的更進一步資訊。

手勢方位偵測

在每個畫面當中偵測到手掌的資訊後，為了提供使用者更多操作方式，我們加入了手勢的方位辨識。而方位辨識的啟動方法是將手部的五指併攏，系統便會開始偵測手部移動的方向，並計算出移動的角度以及做出方向的判斷。而系統判斷的方式是透過濾除手指前的膚色遮罩與濾除手指後的膚色遮罩相減，得到了一張去除掌面的手指遮罩。當使用者手指沒有併攏時，會被系統辨識為手指，而進行手指追蹤；反之，當手指併攏時，系統會將整個手部辨識為手掌，而上述的手指遮罩數值為 1 的數目會大幅的下降，因為系統偵測不到手指。

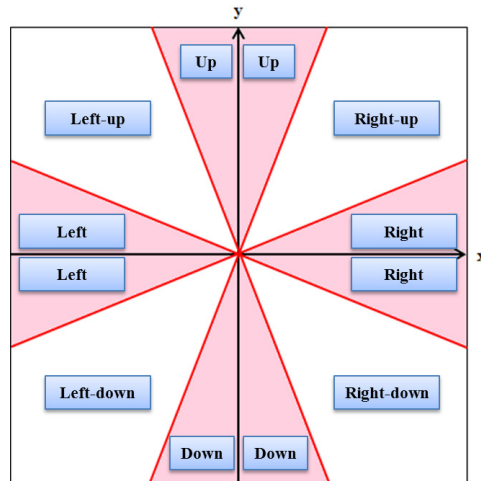


圖 19 手部移動方位劃分

當上述條件成立時，系統會連續的紀錄每張畫面所偵測到的手掌中心，若是持續成立超過 5 個畫面，系統便會計算在此 5 張畫面當中的平均移動角度。並且透過如圖 19 當中的方位分割圖，決定使用者手部所移動的方位屬於何者。

手指追蹤 (Finger Tracking)

手指的追蹤則是分為手指的偵測、追蹤以及點擊動作偵測三個部分。手指的偵測建立於手部追蹤的基礎上，由於手部追蹤結果可以提供我們使用者手掌的位置，而手指必然出現在手掌的周圍，因此我們從手掌的周圍搜尋符合膚色的區域，再利用類神經演算法做手指的區分、辨別與追蹤。在取得使用者各個手指的詳細資訊後，我們將其結果儲存起來，並且透過連續畫面間的手指變化來判斷使用者是否做出點擊的動作。

手指關注區域設定 (Finger ROI Setting)

本實習實作當中的手指追蹤是建立於手部追蹤的基礎，透過手掌的掃描搭配手部的追蹤，我們可以得知使用者的手部位置，有了手部位置便可確定使用者的手指必然會出現在手掌的周圍，如圖 20 所示。然而使用者手指出現的位置並非為一個固定區域，其出現區域可能因為使用者的距離遠近、本身手部大小、旋轉的角度等原因有所變動，可能只出現於中心的手部 ROI 當中，也可能出現在其周圍八個相鄰區塊之中。

要解決這個問題有不同的解決方法，一種方法是盡量的偵測所有可能出現的區域，即將周圍八個區塊與原本中心區塊的關注區塊組合成一塊較大的關注區域，並且在整個區域當中掃描，尋找使用者的手指。此方法幾乎可確保不會漏掉任何的手指，但是卻由於影像的大小在拼湊九塊區域後變為原本的九倍大，造成偵測運算時的運算量較為龐大。

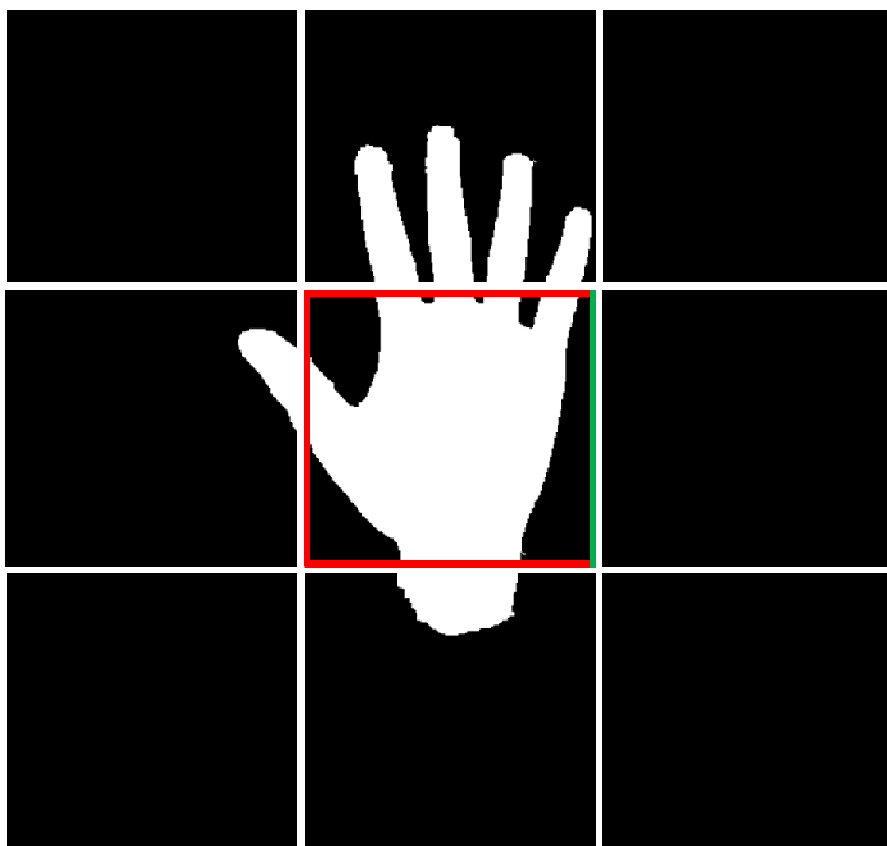


圖 20 手指關注區域偵測

因此本實習實作利用人類手部的特性來試圖減少偵測區域，以達到系統效率的提升。此方法是建立於人類的手指必定連接於手掌的基礎上，因此當使用者的手不只出現於中央的關注區域(即手掌關注區域)時，也必定穿越其邊緣。因此我們在做手指偵測前，先掃描手掌關注區域的四個邊緣是否有連續的膚色。只要該邊緣出現大於門檻值的連續膚色，我們就將該邊緣所對應的區塊列入手指可能出現的候選區域。

如圖 20 當中使用者的手部有跨越中央手掌關注區域的邊緣(圖中紅色邊緣)

，因此在偵測手掌關注區域的四周時，便會得到有三個邊緣(上、左與下)出現符合條件的連續膚色，此外，當使用者的手指同時出現在上方與左方時，我們假設他的手指也可能會出現在左上方的區塊，左下角亦是如此。因此我們將這些區域相接合成一塊手指關注區域以供手指的偵測與掃描。

透過上述方法，我們可以在此情況下將手指掃描的關注區塊縮小為原本的 66%。若是使用者的手部恰巧只出現在中央手掌關注區域，則手指關注區塊等於原本的手掌關注區塊，相對於完全不偵測手指候選區域，直接掃描所有手指可能出現的區塊可以減少約 88%的掃描面積。對於接下來手指掃描的效率有

一定的幫助與改善。

手指掃描 (Finger Scanning)

適應性共振理論(Adaptive Resonance Theory, ART)是屬於類神經網路當中的非監督式類神經網路，亦即此網路在學習或訓練的階段當中，沒有輸出的期望值可回授到網路中做為修正的依據，而是直接依照輸入的狀態來決定輸出的結果。使用者的手在影像當中是變化多端的，因此我們很難用一個公式或是一個標準來告訴電腦「什麼是手指」，所以當然也不會有標準答案的期望值。

一般要建構一個類神經網路有幾個階段，分別是決定網路架構、設定神經元(Neuron)數量再對網路做充足的訓練與學習。當訓練的結果收斂後，便可以將之應用於實際的系統當中，此應用類神經網路的過程稱之為回想(Recall)。因此在一個類神經網路當中，當決定網路的架構後，最重要的部分就是學習。在學習的階段，類神經網路會隨著訓練樣本以及重複的疊代而調整其權重(Weight)值或群聚(Cluster)的中心。

在本實習實作當中，我們不以傳統的方式使用類神經網路，由於手指的影像千變萬化，因此網路的學習很難會達到收斂的狀態。因此我們讓類神經網路持續的維持在學習的情況，使之可以隨著手部的變化而隨時調整，以因應變化多端的手指，而不使網路停止學習，導致過於生硬而難以辨識各種情況。

適應性共振理論的網路有個與其它類神經網路不同的特點，它在初期的時候不需設定好固定的神經元與群聚個數。反之，它會隨著網路的訓練依據輸入的特徵即時的歸類，並且將差異過大的特徵歸納成為一個新的群組。本實習實作之所以採用此種類神經網路便是由於這項重要的特性，因為使用者的手指在畫面中，並不一定有著固定的數量，可能隨著情況的不同，出現一到五隻甚至更多的手指。因此我們利用適應性共振理論會隨著特徵的變化，自動歸納與產生群聚的特性來偵測使用者的手指。

從網路的結構來看，如圖 21，適應性共振網路可分為輸入與輸出兩層，輸入層神經元所儲存的即為系統所給予的輸入向量，網路會依照此向量對影像做分類。因此我們必須慎選做為輸入向量的特徵，才能確保網路有足夠的資訊可以利用所提供的特徵分辨出每隻手指的差異。我們的輸入向量選擇了兩個特徵，可以用來分辨每隻手指最大的差異，就在於它的空間特性。而我們的掃描，是建立於連續膚色基礎上，因此我們將符合預設條件的連續膚色中心座標(x,y)值做為輸入向量。

輸出層在我們的系統中直接代表為手指的資訊。在本實習實作中，我們對原本的適應性共振理論做了一個修改，即初始輸出節點的設定。由適應性共振理論的網路架構來看，初始的輸出層應由一個權重為 1 的結點組成。但是本實習實作的情況下，手指偵測的狀況中，經常會出現偵測不到手指的情況。因此我們在對網路初始化時，將輸出節點設為 0 個，直到出現第一個符合設定條件的特徵，才會產生輸出節點。

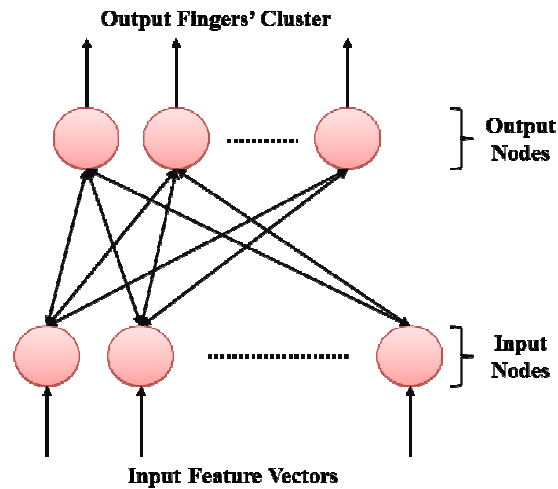


圖 21 自適應共振理論網路架構

在做手指的掃描時，我們自手指關注區域的左上角開始沿著水平掃描線尋找連續膚色。當一段連續膚色結束時，必須先經過寬度的過濾才會將其特徵做為類神經網路的輸入向量，不符合條件的特徵就會被忽略。由於手指在畫面上會有一定的寬度範圍，這個範圍我們用一個上下限的門檻值做為判斷的依據。但是由於使用者每次使用的位置與其手部大小皆不固定，因此門檻若設為定值不免失去其準確度。所以我們的門檻值，使用的是相對的數值。由於手指與手掌間存在一個比例關係，因為大部分的人都擁有 5 隻手指，因此手指可能的寬度範圍在略小於 1/5 倍的手掌寬度與略大於 1/5 倍的手掌寬度之間。故我們用此設定來做為過濾手指的依據，可以使系統在使用者的遠近或手部大小不同的情況下，依然能偵測到他們的手指。

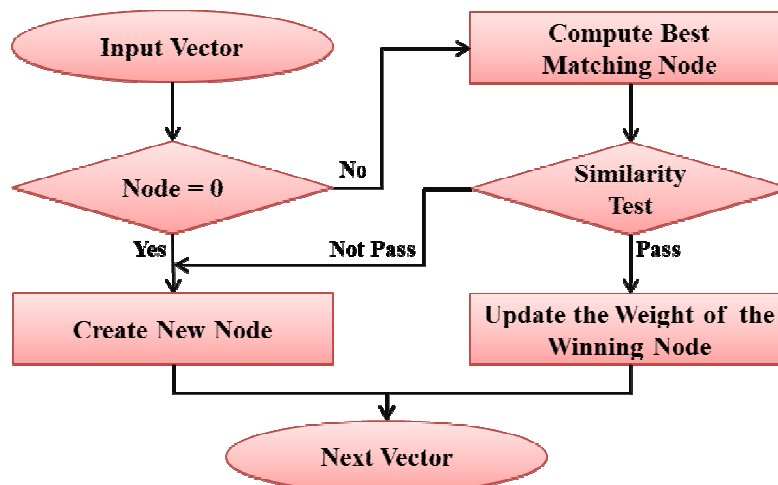


圖 22 適應性共振網路運算流程

通過條件判斷的連續膚色，就會被做為特徵向量送入類神經網路當中歸類。類神經網路會經過一連串的處理程序，將系統偵測到可能為手指的部分自動依照手指在空間上的特性分類成數隻手指。其步驟如圖 22 說明當符合手指寬度的特徵向量輸入後，系統會判斷目前是否已存在節點，若無節點存在則創造一個新的節點(手指)。若已有節點存在則會依照(7)式比較目前輸入的向量與現存所有節點的差異，其中 x 代表輸入向量、 ω_j 是第 j 個節點的權重值，而 N_{out} 則是輸出節點的個數，差異最小者則視為勝利節點(Winning Node)。接著相似度檢查會以(8)式比較此輸入向量與勝利節點間的差異大小， j^* 代表勝利的節點，若是差異小於門檻值 ρ 則會依照(9)式更新節點權重 ω_j ，反之則會因差異過大而成立一個新的節點。其中 $Cluster_{j^*}^{old}$ 代表勝利節點在更新前類聚的叢集向量。而新節點產生時，由於只有一筆特徵歸納到此節點，因此我們直接將權重設為該筆輸入向量，如(10)式所示。

$$j^* = \arg \min(\|x - \omega_j\|), \quad \text{where } j = 0 \text{ to } N_{out} - 1 \quad (7)$$

$$\|x - \omega_{j^*}\| < \rho \quad (8)$$

$$\omega_{j^*} = \frac{x + \omega_{j^*}^{old} \times \|Cluster_{j^*}^{old}\|}{1 + \|Cluster_{j^*}^{old}\|} \quad (9)$$

$$\omega = x \quad (10)$$

經過上述方法的處理，我們可以掃描出使用者在正常情況下未經旋轉的手指。但是為了確保使用者可以自由的旋轉手部，保證使用時的方便性，我們擷取手部的影像，並且每隔 30 度旋轉一次，針對各個角度做實驗，得到圖 23 的結果。從圖中我們可以看到，當角度為 0° 與 30° 時可以正確地偵測出每一隻手指。但是當旋轉角度接近 90° 時，由於手指逐漸轉成橫向，導致水平的連續膚色變大，以致於超出門檻的範圍而偵測失敗如圖中紅色圓圈所包含的手指。

為了解決這個問題，我們對相同的圖片再做一次基於垂直掃描的手指偵測。如圖 24 所示，結果恰好與水平掃描互補，旋轉角度越接近 90° 偵測的效果越好，於是我們決定合併這兩個結果，並且以不同顏色表示不同隻手指如圖 25 所示，可以達到相對完整的手指。此外，我們可以由過程中觀察到，接近手掌底端的部分有些許的誤判，藉由觀察這些誤判的區域，我們可以發現這是由手掌轉折處寬度較短的地方所造成的，因此我們直接判斷手指長度，移除不合理的部分，就可以留下正確的手指。

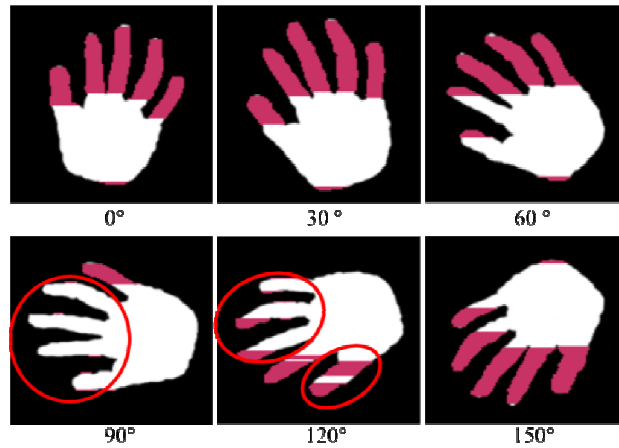


圖 23 利用 ART 執行水平手指偵測

此外，為了分辨手指的頭尾，我們以手心的座標作為起點向外掃描，因此每隻手指的第一個偵測到的點即為手指尾端，反之則為指尖。如圖 26 我們將通過檢查的每隻手指以最先與最後偵測到的兩個點與一條直線來表示，可以看到手指底下誤判的部分已經被忽略，只留下正確的五隻手指。

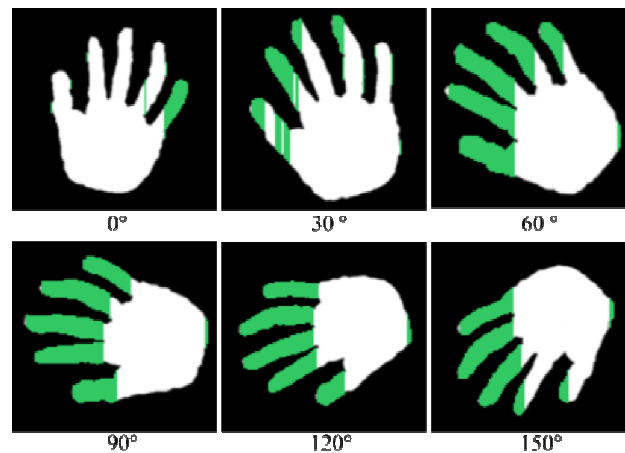


圖 24 利用 ART 執行垂直手指偵測

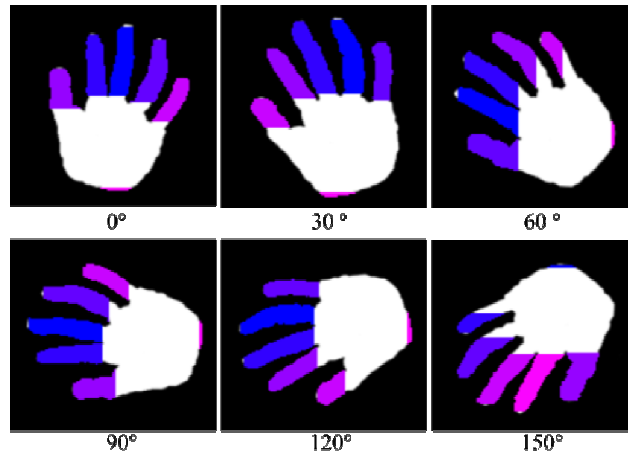


圖 25 合併水平與垂直手指偵測

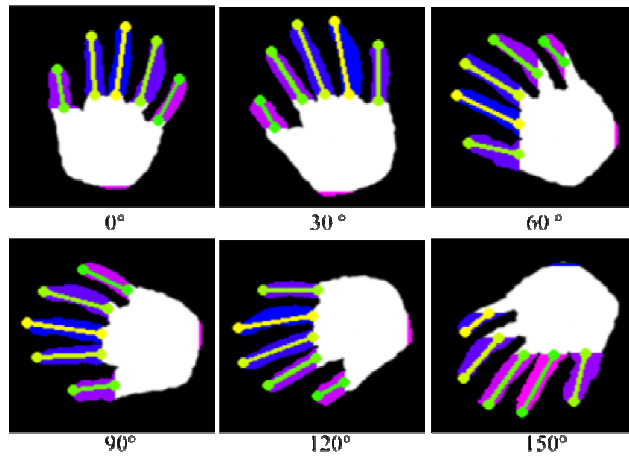


圖 26 合併手指偵測並移除誤判與標示出手指資訊

手指追蹤 (Finger Tracking)

常見的追蹤方法，如卡爾曼濾波器(Kalman Filter)常利用物件在不同時間的線性變化來預測它的走勢。此類的方法容易在物體轉折的時候發生誤差，尤其像手指這種細微距離又近的物件。本實習實作提出另一個方法，利用全域(Global)的變化來預測區域(Local)，即手指的走向，做為追蹤演算法的根據。這個方法是針對人類的手部來設計的一個方法，利用人類手指本身的特性，做為追蹤的方式。由於人的手指關節，大部分是往垂直方向的移動，僅有少部分可以少量的往水平方向移動，而且在移動時通常會帶動其他手指一起移動，很難單獨的橫向移動。因此我們利用這樣的特性來預測手指的位置，並加以追蹤。

詳細的追蹤方法是在每次的手指掃描後，分別計算所有手指頭與其末端的平均座標值，並且將此數值減去上一張畫面的平均座標值，如(11)式與(12)式，其中目前畫面使用者手指的平均移動量為 $AvgMV(t)$ ， $Coordinate_i$ 代表第 i 隻手指的座標，Finger Number 為偵測到的手指數目。此數值便是針對手指通常在

移動時會一起移動來設計，因此我們計算所有手指的平均移動量來預估它們在這個畫面時可能的位置，用這樣的方式來追蹤手指，不會受到轉折的影響而發生錯誤，因為手指的平均移動是即時的，與要追蹤的手指是同一張畫面的資訊，並非線性的由前兩張畫面估測而來。

有了平均的移動資訊，我們便可將之累加至上一張畫面的手指座標，做為手指座標的預測值，如(13)式所示。接下來，我們將預測出的手指座標與新偵測到的手指座標一一比對，找出最鄰近的手指即為目前該隻手指的實際位置，便完成追蹤的程序。

$$\text{AvgCoordinate}(t) = \frac{\sum_i \text{Coordinate}_i(x, y)}{\text{Finger Number}} \quad (11)$$

$$\text{AvgMV}(t) = \text{AvgCoordinate}(t) - \text{AvgCoordinate}(t - 1) \quad (12)$$

$$\text{Finger}_i(t) = \text{Finger}_i(t-1) + \text{AvgMV}(t) \quad (13)$$

手指點擊偵測 (Click Detection)

本實習實作所提出的手指點擊偵測，是透過在一般室內環境中實際的觀察，手指在點擊時與非點擊時的差異，並且把這些差異利用我們所偵測出的手部、手指資訊來判斷與實現。在觀察的過程當中，我們發現有幾項特徵是手指在點擊按下的時候會發生的變化，我們將之整理成表格如表 1 所示：

表 1 手指點擊時的特徵與其辨識難易度

項次	特徵	辨識的難易度
1	指尖與手心距離變短	低
2	手指長度變短	中
3	手指亮度變暗	中
4	指節紋路變暗(清晰)	高

表 1 當中所列之難易度的標準，是建立於手指經由網路攝影機拍攝後，經由影像處理或直接計算的方法，辨識的難易度。其辨識的難度通常來自於視訊攝影機的雜訊，如手指的長度與亮度。由於視訊攝影機在拍攝每一張畫面時皆會有些許的偏移與差異，因此手指的參數資訊也會隨著畫面的偏移而改變，而手指在畫面上是個相對較小的物件，因此其長度值容易受到偏移的影響而跳動，因而影響判斷點擊的準確度。手指的亮度亦然，此一現象在現實的觀察當中由於光線的陰影，相當明顯、容易觀察，但是在畫面當中手指的亮度也是隨著畫

面而不斷變動，因此雖可做為判斷的參考，但不足以完全當作點擊的判斷依據。

指節的紋路也是一個可以做為判斷的線索，但是卻受限於視訊攝影機解析度的影響在畫面當中不甚清楚，因此在我們的實驗中並未使用這項參數。綜合以上觀察的結論，第 2 項的手指長度由於較容易受到攝影機的偏移所影響，且其性質與第 1 項相近，因此我們使用第 1 項與第 3 項做為判斷手指點擊的基準。

結論

我們描述了本實習實作所提出之「三維立體電視之自然使用者介面設計」的主要演算法與系統架構。並詳細介紹了各個功能區塊的原理與其運作方式。我們採用了單一鏡頭的視訊攝影機做為影像擷取的來源，經過了前處理、手部追蹤與手指追蹤三大區塊的處理後，可以得到手部的掌心資訊、使用者手指資訊以及可以判斷出手指點擊的狀態。若將此介面應用於智慧型電視當中，想必可以成為一個合適的輸入介面，並且提供使用者良好的使用經驗。在德國實習的過程當中，接觸到前瞻的技術與演算法，啟發了我對於實習實作的靈感，在這實習的過程之中，我所做的「三維立體電視之自然使用者介面設計」為我的實習過程留下一個完美的註記。

九. 涵養德國工藝精神-參訪參訪魯爾工業區(Ruhr Area)

除了實習的學習、訓練與實習實作之外，老師為了讓我們能夠涵養德國工藝精神，特別安排了一天做工業的參訪，到德國工業發展時期的重鎮--魯爾工業區，學習德國工藝精神。

由於德國不只是立體影像技術發展的早，在工業方面的成就也是眾所皆知，處事態度一向給人細心謹慎的印象。因此我們安排了德國魯爾工業區的參訪，除了了解德國工業的發展以外，更是為了學習他們的處事精神與態度，希望未來能夠對我們的研究與人生等各方面都發揮正向的影響。

盧爾工業區位於科隆的北方的埃森(Essen)，因此我們搭乘短程的火車前往埃森，再透過轉乘地面上的電車就可以抵達我們要參訪的主要地點。盧爾工業區以發展重工業最為著名，如：採礦、鋼鐵、機械或化工等相關的產業都有相當程度的發展。我們這次前往參訪的，是當地的一個主要的博物館，此博物館是由早期的礦坑所改建，如圖 27 所示，其建築外觀保留了礦坑的原貌，在設計時也保存了許多的礦坑設備如圖 28，當中的展覽也留下了當地工業發展的軌跡。



圖 27 魯爾工業區礦坑博物館



圖 28 博物館中保留的礦坑設備

十、實習評量考核-成果發表

為了考核實習的成效，在德國實習的期間，我們參與了一場實習發表會與研討會，並且用英文發表了實習的成果，介紹我們實習實作中「三維立體電視之自然使用者介面設計-立體影像自動化深度量測與產生」的技術。在參加發表會的過程當中學習到許多知識，並且第一次在國外發表實習實作的過程當中，獲得了許多難以言喻的寶貴經驗，同時與會的專家學者給予我許多寶貴的建議，對於進一步改善實習實作的成果，有很大的幫助。



圖 29 實習發表會與研討會現場



圖 30 實習發表會與研討會的課程



圖 31 實習發表會發表實習實作成果

十一、結論

在這次的實習的過程中，我們接觸到許多立體視訊的相關系統設備與技術，也在過程當中見識到許多互動的設備。這些設施不僅可以增加了實習過程中的研習深度，也讓我留下深刻的印象。我的實習實作題目是要設計一個三維立體電視之自然使用者介面設計，因此在使用這些設施的時候也讓我得到了一些實作的啟發。

在我的實習實作當中，要實現一個基於手勢控制的使用者介面，我的構想源自於實習訓練過程之中，所操作的互動設備與觸控介面。但是由於三維立體電視通常與使用者有段距離，因此觸控要重新設計成手部的偵測與追蹤，並且偵測出使用者的每一隻手指，再判斷出每隻手指點擊的動作以達到控制的目的。

我所設計的實習實作，自影像從視訊攝影機擷取至系統開始，經過三個階段的處理，分別為—前處理(Pre- processing)、手部追蹤(Hand Tracking)以及手指追蹤(Finger Tracking)。可以從畫面中判斷出使用者的手掌中心位置、每隻手指的資訊(包含手指頭尾座標、手指長度與旋轉角度等)，並且透過連續畫面間的判斷，可以辨識出使用者手指點擊(Click)的動作。這一個系統的實現，讓我可以像在遠端的觸控電視一般，不需要遙控器直接用手就可輕鬆的操作電視。

綜觀這次來德國進行學海築夢實習計畫，收穫十分豐富。在德國實習期間，讓我對於三維立體電視之自然使用者介面有著全新的設計理念，同時啟發了許多新穎的想法，同時，對於研究工作也有著十分大的幫助。

完成實習之後，我將實習過程之中，相關的成果 Natural User Interface for Interactive Televisionaper，寫成國際研討會論文，並且投稿至 2013 IEEE International Symposium on Circuits and Systems - ISCAS 2013，由此可以看出，這一次的海外實習過程中，確實是收穫滿載。

最後，非常謝謝我的指導教授范育成老師讓我前往德國實習一個月，這個過程中，雖然忙碌，但卻十分充實，學習到許多新的事物，包括德國的工藝精神、外語能力、專業知識與多元的文化，相信這一切對我的人生來說，是一個重要的里程碑，並且在一生中受用不盡。在此的同時，非常感謝明斯特大學的蔣教授及其學生，在我們實習的過程中，給予我們許多的幫助，同時詳細的教導我們相關的實習研究主題。此外，特別感謝教育部提供我們參與學海築夢海外實習計畫的機會，透過這次的實習計畫，開啟我們不一樣的人生與視野，這對於我未來的人生，將會有深遠的影響與幫助。