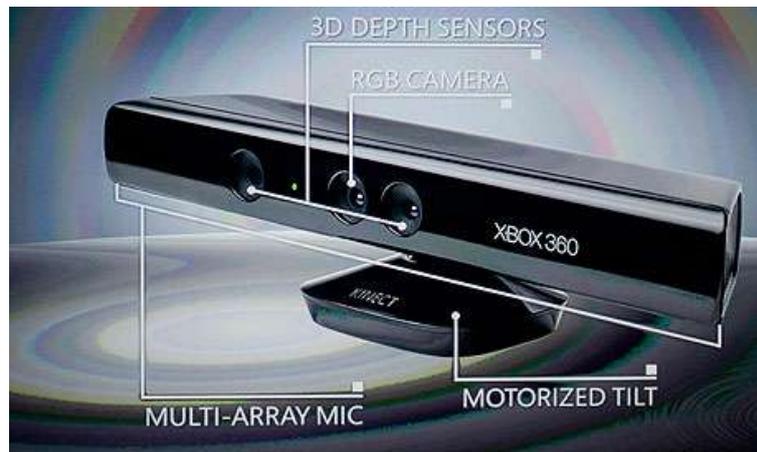


中華大學資訊工程學系  
100 學年度專題製作期末報告

以視覺為基礎的簡報系統



組員：劉禹廷

黃冠瑜

蔡仁儀

指導教授：黃雅軒 教授

# 目錄

摘要.....	1
第一章、緒論.....	2
1.1 研究動機.....	2
1.2 論文架構.....	4
1.3 相關研究.....	5
1.4 立體視覺.....	6
1.5 時差測距.....	7
1.6 連續光編碼.....	8
第二章、擷取分析感測器相對應參數.....	10
2.1 簡介 Microsoft Kinect 感測器.....	10
2.2 擷取串流影像.....	11
2.3 擷取深度影像.....	12
2.4 骨架追蹤及座標轉換.....	13
第三章、手部追蹤及手勢辨識.....	15
3.1 流程圖.....	15
3.2 架構流程說明.....	15

第四章、成果展示及延伸應用 .....	21
4.1 系統實驗環境.....	21
4.2 透過手勢辨識，目前已實作應用在.....	22
第五章、結論與未來展望.....	26
甘梯圖.....	27
致謝.....	28
參考文獻.....	29

## 摘要

這篇研究是在提出一個即時辨識手勢互動的系統，使用 Kinect 三維感測器，一種新穎且價格合理的範圍傳感器。

此系統是利用準確的三維深度資訊及準確標定關節去實踐局部手部的擷取。然則會根據手部座標範圍內的灰階值，經過二值化…等處理既可得完整的手部資訊。最後再以半徑做掃描來得到手指數資訊。

而透過左、右手的標定，所以當手部跟其他身體部位重疊，並不易發生誤判。基於得到手勢的動作及精確的手指數後，使用 FSM 有限狀態機的模式來讓使用者下達正確的指令，因為使用這台可提供深度及骨架資訊的感測器，所以縱使在背景複雜地方操作，或光線不一強弱的情況下，皆可達到即時 performance at 30 fps 跟高度的準確性。

如此一來，使用者可藉由此台機器，將現實中的自然直覺式操作，轉換成電腦可判讀指令資訊。

**關鍵字:**Kinect、二值化、深度、骨架資訊、手部擷取、手指數

# 第一章、緒論

## 1.1 研究動機

在 2010 年六月份的 E3 電子遊戲大展中，三大遊戲主機中的 Microsoft 和 Sony 也推出了體感裝置，但最令人期待的莫過於是，不需要在手上拿著遙控器，由微軟發表的體感裝置「Kinect」。該產品不但實現了在電影「關鍵任務」中，主角使用手勢自由的操作介面跟處理資料，也將「使用者介面沿革：從 CLI、GUI 到 NUI」中的 NUI 概念正式的帶入到整個電子產業中。

”你就是控制器”（You are the controller.）—這句令人印象深刻的廣告詞，透過 Kinect 人們不用再學習新的操控方式，用本身最直覺式的動作，來將現實跟虛擬世界連結再一起。為了讓虛擬世界中能完整呈現肢體自然地揮動的資訊，Kinect 使用多個感測器同時擷取、過濾、分群、辨識的技術，及隱藏在這些背後感測器的關鍵「3D 深度測量」，都是在本研究中所利用的資訊。

目前的動態手勢識別方法分為感測器和基於電腦視覺的方法。而基於感測器具的方法要求用戶配戴如數據手套的穿戴式感測器，以提

供系統的可用性和辨識準確性，如“SixthSense” [1]中。然而額外的穿戴式感應器通常也帶來使用者不便及在視覺上看起來不太具有美感。基於視覺的系統上，另外一方面，只需要一個可接收影像的感應器，以視覺處理方式來辨識我們的手勢。但是當遇到背景複雜或光線強弱光源不一的條件下，或手部及其他相似物體互相遮蔽情況下，驗證不同程度的背景雜波，會將是十分困難的。那在這麼多的嚴苛條件下使用三維深度資訊是個很好解決辦法。目前已有越來越多研究的趨勢建議使用三維的範圍攝影機。如雙眼視覺、time-of-flight (ToF) range cameras。

這個專題是研究利用這套三維感測器來達到即時主動式追蹤手部手勢辨識系統，此靈感來自於目前大多用於智慧型手機的多點觸控技術。我們利用的是一款新穎且價格合理的三維距離感測器，且能即時應用在動態的手勢追蹤辨識。

這套硬體提供顯卓的效能超過傳統攝影機，並且非常穩定提供每秒刷新三十張頁面，且可使用深度資訊來表示物理上的距離，也因此簡化了在解決手部局部定位的複雜度。而也幫助在計算手指數及各種手勢時有較高的穩定性，且能有效解決在手部重疊遮蔽情況下做判斷。在手指和手於三維空間中所形成的姿態，用有限狀態機(FSM)方法來

變成各種的操作手勢。未來此套系統發展可用於各種識別領域：手語識別、人身行為操控機器人、製作3D模組動畫、身臨其境的遊戲、娛樂系統、家庭安全、人行偵測、車輛碰撞偵測、多物體空間距離檢測、體態訓練、功夫運動、矯正姿勢及中風病人進行肢體及腦部認知的復健...等。

本專題研究是藉著 Kinect 對自然世界中深度的感測及人體骨架的追蹤技術，來開發可透過直覺式的肢體動作，來達成一個能夠辨識手勢模擬滑鼠及互動式簡報的系統。

## 1.2 論文架構

本論文共分為五章，各章節內容安排如下：

第一章說明本論文的研究動機及相關研究，本論文主要目的是利用一台可測深度的感測器來達成多項應用。

第二章主要說明由微軟開發的感測器，透過其函式庫我們可以取出那些有用的資訊。

第三章利用感測器提供的資訊，來針對局部的追蹤及取樣有很大的幫助。

第四章的成果展示了透過此感測器，目前可以做到的應用。

第五章為結論與未來展望，以及對未來可繼續改進的部分，以及繼續研究的方向做一個說明。

### 1.3 相關研究

近年來，電腦視覺發展越來越迅速，許多的應用在廣泛層面中效果顯卓。而直覺手勢操控系統也因應多點觸控流行，近來類似的研究蓬勃發展。然則擷取手勢特徵轉換編碼成為非常重要的資訊，使用者手勢的準確性和性能，在不同的硬體辨識系統和其所使用的演算法有不同的精確穩定性。

基於視覺系統通常提取的方式是從二維數字圖像的數據，這些數據反應了真實世界的三維場景。此外，在第三維的幾何關係(即深度)之間的雙手和背景，也因擁有不同的深度值，所以可以有效地分割出欲得到的資訊。除此之外，有相當多的研究在從三維場景轉換成二維圖像的形式，透過不同的硬體及不一樣的設備，其儲存格式及演算法，來獲得有用的特徵。

其餘本文會將主要探討三維深度的使用方式：立體的視覺和飛行時間，使用 Kinect 利用光編碼技術。本文採用創新三維範圍感測器。

## 1.4 立體視覺

電腦立體視覺參考所採集的圖像採用兩個或更多的攝影機，水平的放置或用其他的擺放方式來產生模擬生物的視覺。透過左右邊的攝影機分別表示左眼及右眼。而使用這種方式來描繪一個場景，攝影機將提供兩種投射場景的差異視野，那將得到的差異影像做比較後產出額外的資訊，如場景中物體的深度等。因此，從立體影像做配對提取場景中三維的結構稱之為「立體視覺」。

目前，立體視覺所面臨的問題是在於兩個影像的匹配，而通常此兩張影像皆有不一致的現象，顏色及強度各自不相同。另外也有個問題是決定兩個像素點中唯一匹配點，因產生多個候選點導致可能無法做正確的選擇。然則立體視覺匹配面臨最大的問題是，當左邊的影像像素點，跟右邊影像像素點要做匹配時，發生相對應像素點不存在，這時立體成像可能需要額外的機制去針對此事件做處理。

儘管有這些問題的產生，立體視覺仍是一個標記三維深度感測的好方法，也在各種應用中看到了它的用途。為了呈現立體視覺，常見的方法為使用兩個攝影機和一個膚色模型，來對使用者的頭部和手部位位置進行標定。

## 1.5 時差測距(TOF, "Time-of-flight" sensing)

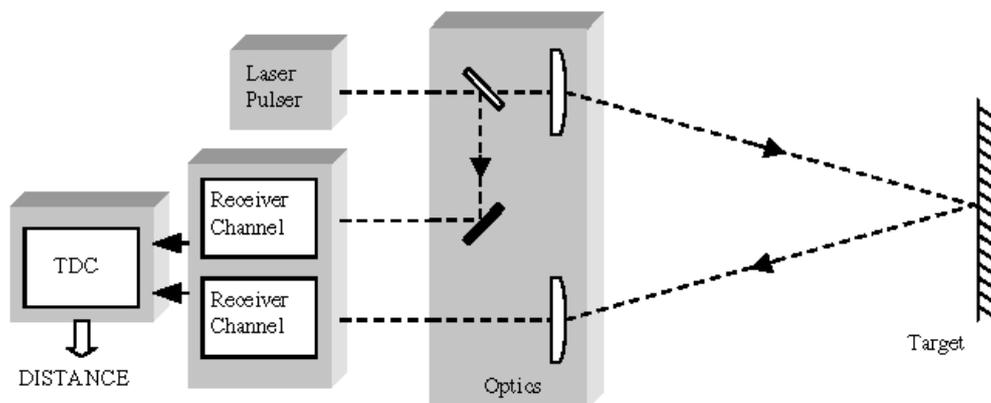


圖 1-1、TOF 示意圖[2]

使用時差測距” ToF”[3]的方法來做手勢辨識，其基於使用紅外線光 IR 的方式，然後計算光線飛行的時間。先發射一道光(雷射或是紅外光)，並計算該道光從發出後撞到物體返回的時間差。由於光速是已知，透過時間差即可計算出距離。

而利用此方法來測量時間差的關鍵在於如何感測反射回來的光？較常採用的方法有「被動計數」與「主動開關」兩種。「被動計數」是透過砷化鎵晶片 (Indium Gallium Arsenide) 來紀錄每個像素感光的時間，「主動開關」則是透過開關鏡頭或開關感光元件 (如 CCD 或 CMOS) 來取得一道光的兩段折返時間。

目前較常使用為「主動開關」，因為它能與光源發射器同步，以便測量特定區域、光源的折返。時差測距的好處為大幅降低演算法的複雜度，雖說其也有不同的困難點，但單看解決光線問題就已經提供一個不錯的效能，所以其演算法簡單且測量速度非常快。後者也是目前許多研究著眼的重點，因為只有測量速度夠快才能應用至即時追蹤辨識中。然而，時差測距面臨的最大挑戰是環境光源及待測物體的表面特性，這些都會影響測量機感應光的準確度，進一步影響時間差的計算而錯估距離。

## 1.6 連續光編碼 (Light Coding)

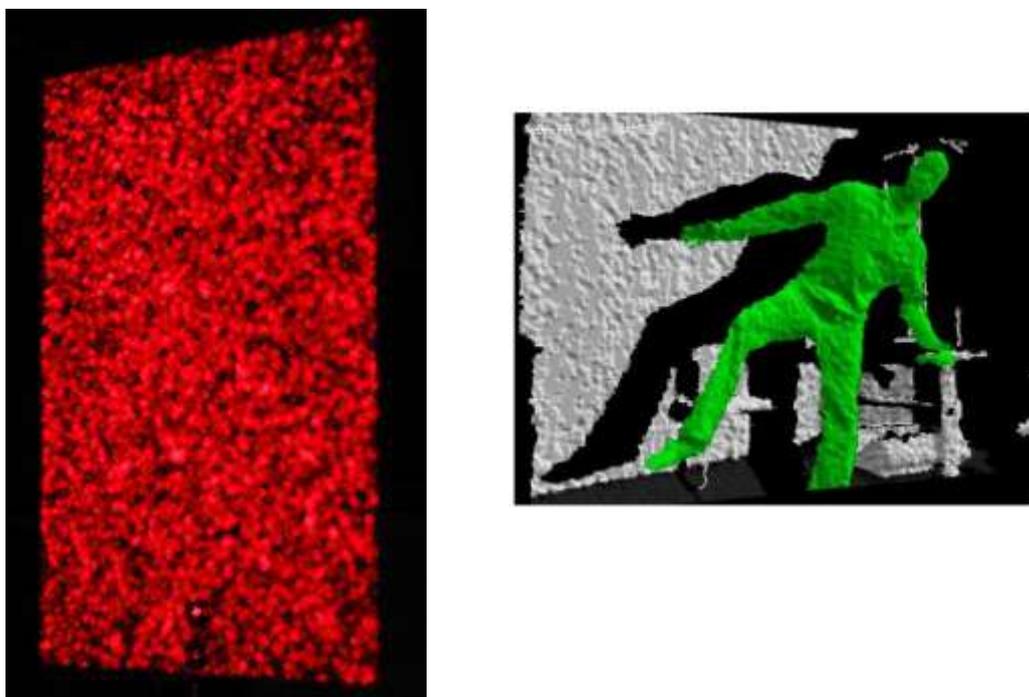


圖 1-2、雷射光斑[4]

連續光編碼（Light Coding）技術理論是利用連續光對欲測量空間進行編碼，來對特定空間進行「標定」（register）的方法完成三維深度辨識。藉由感測器讀取編碼後的光線，交由晶片運算進行解碼後，產生一張具有深度圖的影像。

而其中關鍵的技術為雷射光散斑（Laser Speckle），意即當發出雷射光後照射到物體或是穿透毛玻璃後，會形成反射斑點，稱之為散斑。這些散斑不但在同一平面上形狀完全不同，隨著距離變換也會有所不同。空間中任何兩處的散斑都會是不同的圖案，等於是將空間中每一個點都加上標記，換句話說，投射到空間上的散斑全部都不相同，就像人類的指紋一般。

我們只要針對待測空間按照一定距離取出參考平面，並紀錄每一個平面上所有的散斑，如此一來之後只要有物體在該空間中活動，我們就能透過比對散斑來定位出該物體在空間中的位置。此一方法在概念上與結構光頗為類似，也同樣享有掃描時間快速、掃描面積廣大的優點，透過隨機形狀的雷射散斑，我們還能夠進一步得到整個空間的定位資訊。

## 第二章、擷取分析感測器相對應參數

### 2.1 簡介 Microsoft Kinect 感測器

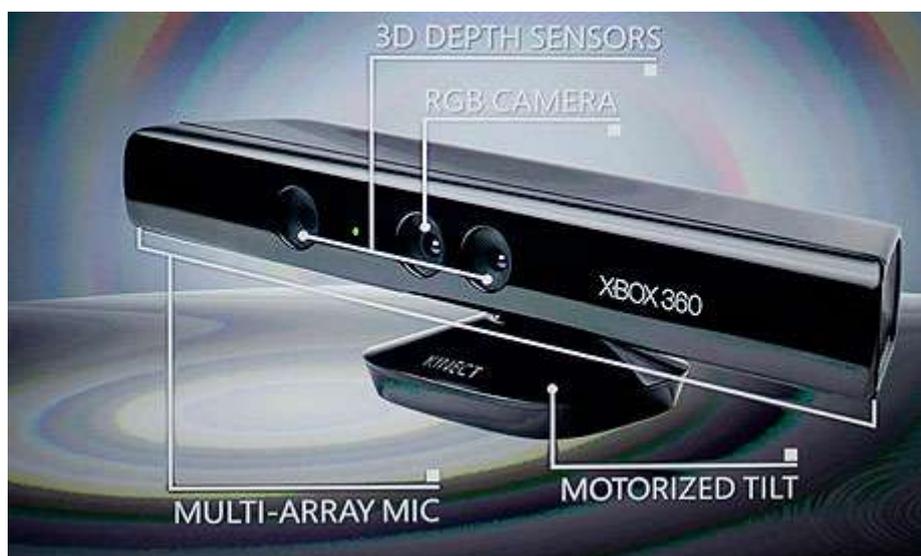


圖 2-1、Kinect 示意圖[5]

本文所使用的這台由微軟所開發的 Kinect 感測器，其不僅能提供每秒平均 30 頁面刷新率，還能提供三維深度資訊，如此的確可為手勢辨識帶來極佳的效能。雖然目前已有相關演算法來解決手部追蹤及手勢分類的相關問題，但系統的整體性可能會因為使用硬體不同，產生差異的效能。而一般在該領域是依靠標準的 RGB 攝影機，來實現手勢辨識系統數據採集的主要方法。本文使用由微軟提供的” SDK”[6]及硬體來蒐集 Light Coding、3D Depth Data、Skeleton 等資訊實作手勢追蹤辨識。此外，最近研究趨勢顯示，有越來越多手勢識別的相關研究，已開始使用三維深度感測技術來提高整體效能及穩定性。

在本文中，設計即時的手勢辨識系統，其概念類似目前新的直覺操控方式”多點觸控系統”。但其有先天上的限制，就是要以實體碰觸時，才能夠產生一連串操控的訊號。但本文提出的辦法是希望能以隔空的方式，對該平台作操控。

## 2.2 擷取串流影像

Kinect 提供平均每秒 30 刷新率的頁面，而也因其提供的是串流影像，並非一般傳統直接可拿來使用的 Bitmap 或可提供給 OpenCV 可用的 MplImage。所以首先在得到影像之後，將鎖定其記憶體存放的串流讀出來，轉換成以下影像兩種格式

1. Bitmap (在不使用 OpenCV 情況下，即可做影像處理)
2. MplImage(可供 OpenCV 來使用的影像格式)。



圖 2-2、由 Kinect 擷取出的影像

## 2.3 擷取深度影像

Kinect 紅外線感測避免在做骨架追蹤定位時，受到周圍環境光照條件影響。而該 CMOS 的紅外線感測器，透過黑白光譜的方式來感測環境。純黑代表無窮遠，純白代表無窮近。而黑白之間的灰色地帶則對應物體到感測器的物理距離。概念類似圖[2-3]。



圖 2-3、深度示意圖[7]

紅外線感測器以平均每秒三十頁面的速度，蒐集視線範圍內的每一點，然後生成周圍環境的三維深度圖。如圖[2-4]。

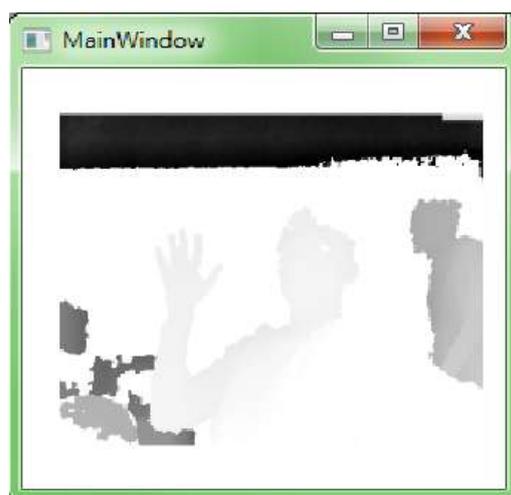


圖 2-4、擷取出由 Kinect 產生的深度圖

## 2.4 骨架追蹤及座標轉換

利用感測器蒐集到的資訊整合，其會根據追蹤到的二十個關節點去生成骨架系統。Kinect 也會評估硬體輸出的每一個可能像素來確定關節點。基於可信度高的信息，Kinect 可以精確獲取人體實際在空間中的所在位置。除此之外 Kinect 還在模型匹配階段時，就可產生輸出濾鏡來平滑化影像及關節點重疊等例外事件。

而開發者也可以藉由任意的關節事件組合，來開發各種遊戲體驗及應用程式，這也是體感中骨架追蹤系統[8]的目標之一。



圖 2-5、骨架標定

透過骨架標定後每個支點，將其轉換相對應螢幕中的位置，即可得到座標。得到這些座標後，就變成極有意義的事情。因為這些座標

可反映出目前使用者的肢體。如圖 2-6 中左方是將座標顯示在支點部位旁，右方是將得到的座標資訊並列顯示。



圖 2-6、上半身骨架標定，右邊顯示為骨架標定座標



圖 2-7、標定身體座標，轉成整數型態

## 第三章、手部追蹤及手勢辨識

### 3.1 流程圖

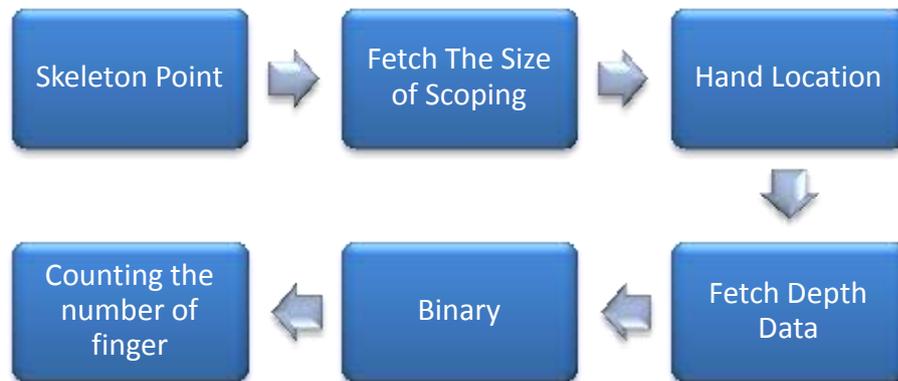


圖 3-1、流程圖

### 3.2 架構流程說明

#### Step 1. 針對全身做骨架標記

先對全身各關節點進行標記，如此可利用骨架的間距，來得知使用者距離攝影機的遠近。



圖 3-2、全身骨架標定

Step 2. 得到手部範圍框選的動態參數  $D$

1. 擷取右手掌心位置得到  $RHC_{position}(x1,y1)$
2. 擷取右手腕位置得到  $RHW_{position}(x2,y2)$

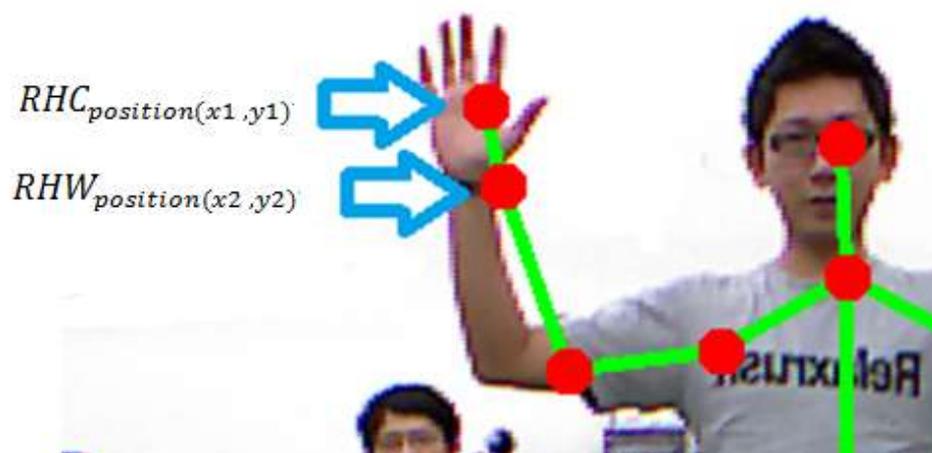


圖 3-3、擷取掌心及手腕座標

3. 再對空間中的兩點計算公式求解

$$Distance = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$

其中  $x1,y1$  來自於  $RHC_{position}(x1,y1)$

$x2,y2$  來自於  $RHW_{position}(x2,y2)$

Distance 為掌心節點到手腕節點的距離

得到動態參數為

$$D = Distance \times C$$

( $C = 1.3$ ，透過三十種手獲得的最佳參數)

Step 3. 根據 D 來決定框選手部的範圍

1. 利用在 Step1 中所得的右手掌心及左手掌心的座標

$$RHC_{position(x1,y1)} \setminus LHC_{position(x1,y1)}$$

兩座標中的  $Position(x,y)$  做決定範圍計算。得到

a. 右手掌的範圍

$$Upper\ Left_{(x,y)} = RHC_{position(x1-D,y1-D)}$$

$$Lower\ Right_{(x,y)} = RHC_{position(x1+D,y1+D)}$$

b. 左手掌的範圍

$$Upper\ Left_{(x,y)} = LHC_{position(x1-D,y1-D)}$$

$$Lower\ Right_{(x,y)} = LHC_{position(x1+D,y1+D)}$$

2. 以得到的左上角及右下角的座標，再搭配掌心座標。如圖 3-3。

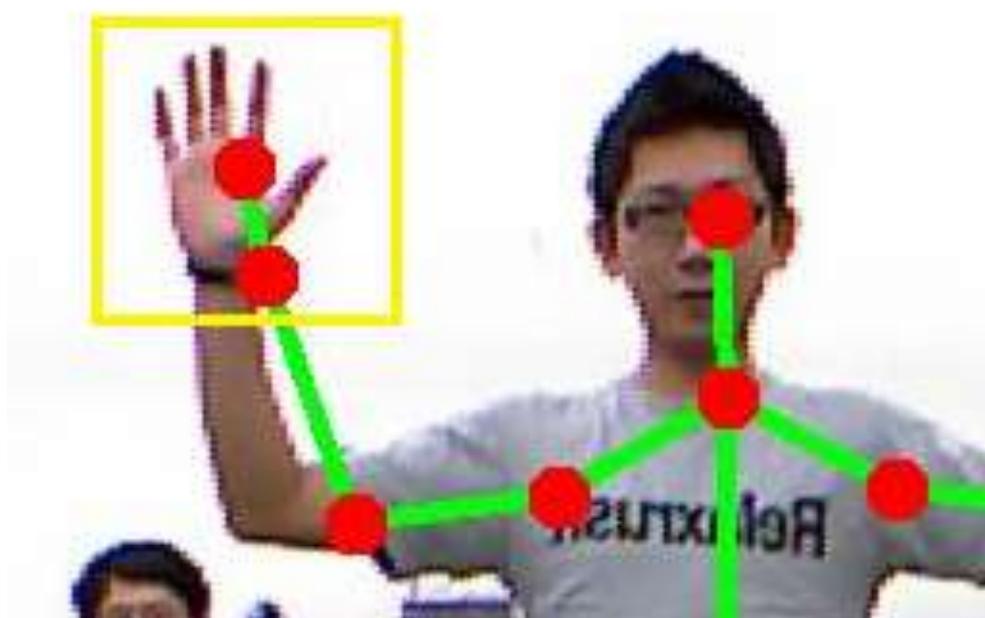


圖 3-4、框選適中範圍，以右手掌為例

#### Step 4. 得到完整的手掌

使用在 Step 1. 中，獲得

$$RHC_{position(x1,y1)}、LHC_{position(x1,y1)}$$

雙手掌心的空間座標，將其座標對應到相對位置的深度圖上，抓取在深度圖中該點座標的灰階值為 $G_{el}$ [9]。

感測器所產生出的深度圖，會依照距離遠到近表示暗到亮。不同的距離會有不一樣的灰階，因此根據 $G_{el}$ ，即可能定位出在空間中手部的資訊及其位置。



圖 3-4、深度影像

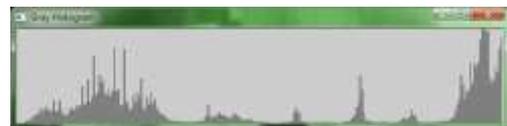


圖 3-5、左圖的灰階分布直方圖

1. 得到  $G_{el}$  的值後，對框選後範圍後的深度圖做二值化。只將亮度值等於  $G_{el}$  的像素質設為 1，否則設為 0。

$$G(x, y) = \begin{cases} 1, & |f(x, y) - G_{el}| \leq 5 \\ 0, & \text{Otherwise} \end{cases}$$

$f(x, y)$ 、 $G(x, y)$  分別為處理前、處理後的影像中位於  $(x, y)$  上的某個像素的亮度值， $G_{el}$  為臨界值。

2. 即可得到做完二值化後的手部資訊。

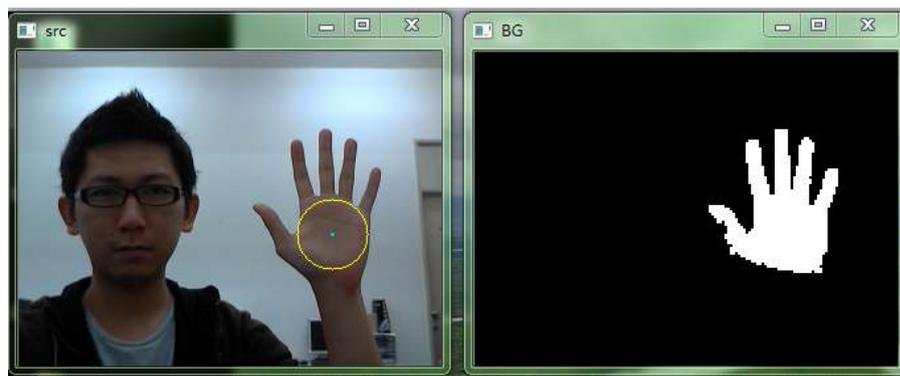


圖 3-6、得到手部區塊

#### Step 5. 手部辨識

1. 因手掌會隨著使用者的移動，在空間中會有遠近距離的問題，所以利用 Step3. 中獲得的動態參數  $D$  來求得圓半徑。

$$R = D * 0.7857$$

(0.7857 為統計三十種手，所得最佳參數)

2. 以  $R$  為圓半徑長度，去環繞左、右手掌心。因為是對二值化後影像做掃描，所以我認為當繞圈走訪像素值時遇到  $0 \rightarrow 1$  後再遇到  $1 \rightarrow 0$  的情況時就可視為遇到一根手指頭。如圖 3-7，此張會產生五次的  $0 \rightarrow 1$ ，所以代表手部目前狀態是伸出五根手指頭。可簡單加入手指寬度和高度資訊來排除手臂，如此即可使用更自然的手勢。

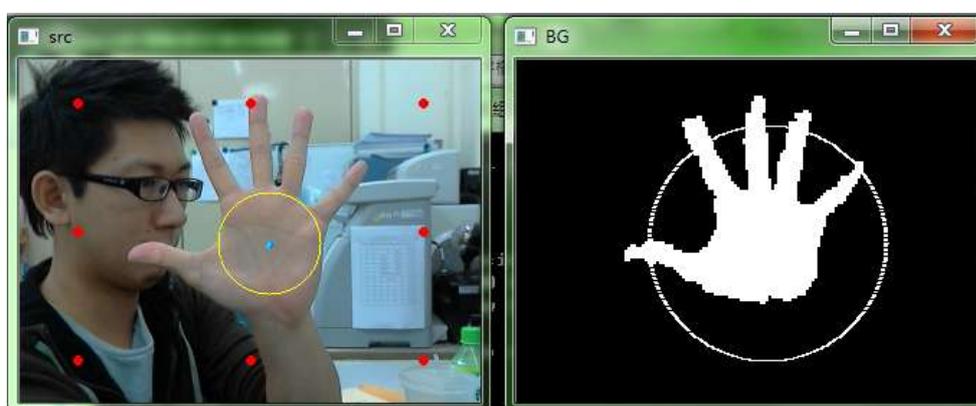


圖 3-7、左邊顯示內圓，右圖顯示外圓

## 第四章、成果展示及延伸應用

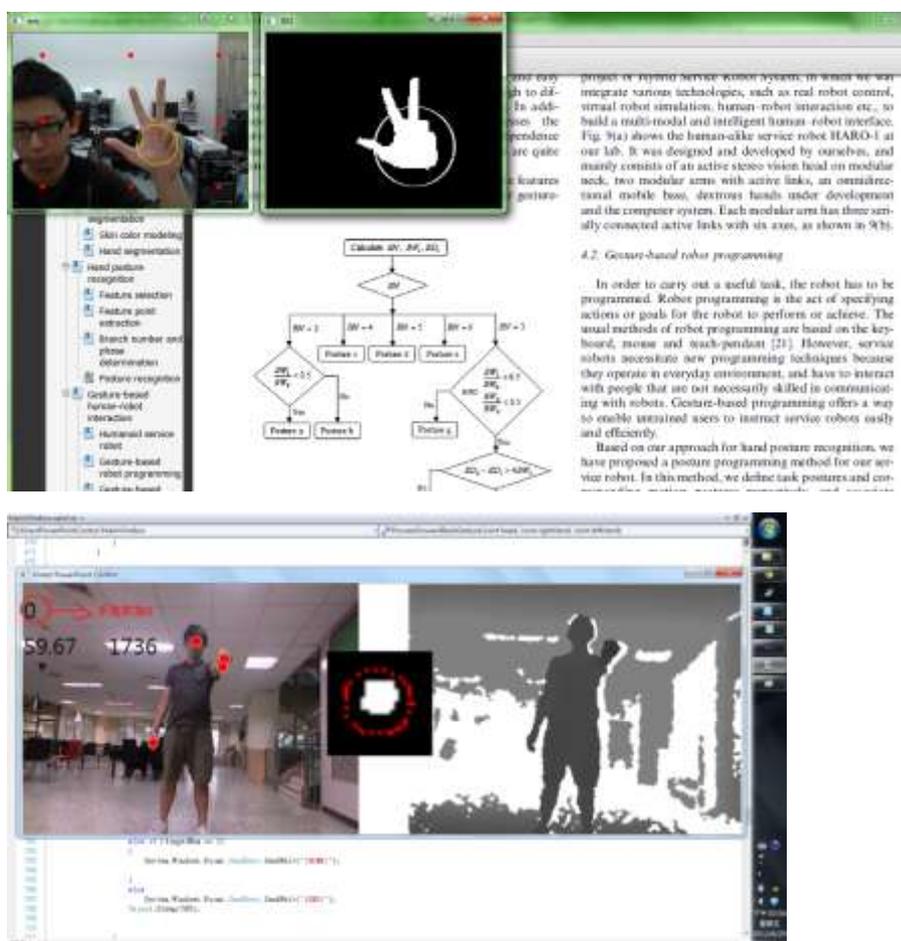
### 4.1 系統實驗環境

系統的實驗硬體設備	
CPU	Intel(R) Core(TM) i5 2.67GHz
RAM	6.0GB
系統開發環境	
作業系統	Microsoft Windows 7
開發環境	Visual Studio.NET 2010
使用函式庫	Kinect SDK、OpenCV Library

表 4-1、系統實驗環境

## 4.2 透過手勢辨識，目前已實作應用在

我們將手指數辨識及手勢辨識整合至 Microsoft PowerPoint 中。其中定義手指數為 0 時，會結束投影片放映；手指數為 1 時，會回到第一張投影片；而手指數為 5 時，會直接到最後一張投影片。為了避免與手勢指令衝突造成誤判，在左手靠近頭部時，右手比出的 0、1 或 5 的指令才會獲得系統接受。



上圖是手指數為 0。



上圖是手指數為 1。



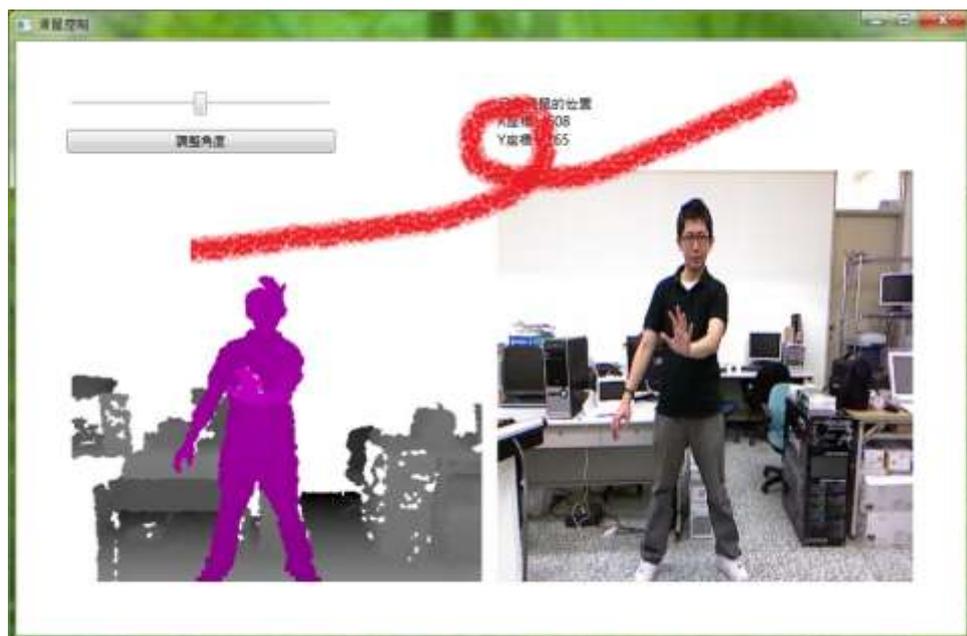
上圖是手指數為 5。

利用深度及骨架資訊，使用手勢操作 PowerPoint。當放映投影片時，可利用左手手勢與右手手勢進行上一頁與下一頁的控制。當右手伸直由頭上向右邊畫四分之一圓時，投影片會往下跳一頁；而左手伸直由頭上向左邊畫四分之一圓時，投

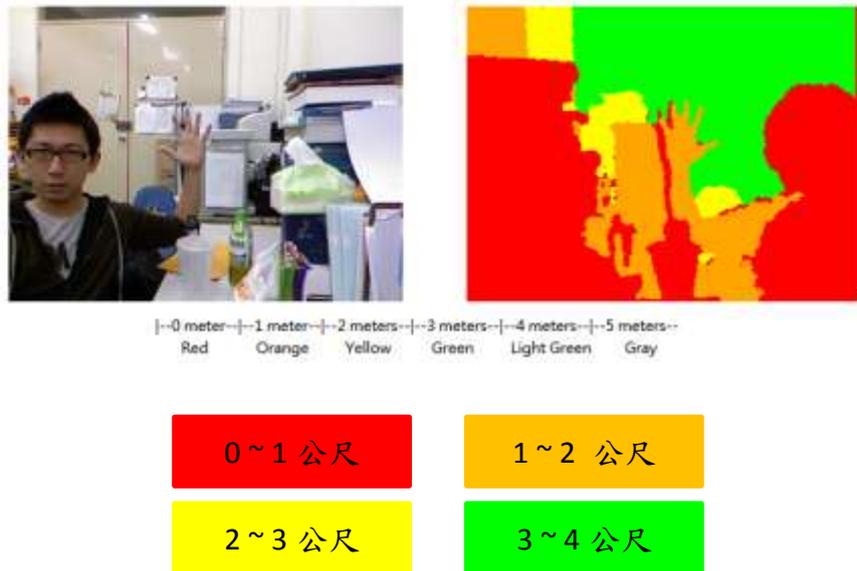
影片會往回跳一頁。



滑鼠操控:使用掌心座標，做滑鼠的移動。目前用預測路徑法來大幅降低操控時滑鼠抖動情況。下圖左邊是使用顏色來區分不同使用者，圖中紅色筆畫代表滑鼠移動的軌跡。



對空間中物體做距離量測:使用深度資訊、影像處理。對空間中距離感測器的遠近,來使用顏色區分。誤差值在±5公分內。



語音控制:利用 Kinect 內附的音訊資料,能讓系統接收使用者發出的語音。當使用者發出” Open PowerPoint”時,系統便會開啟使用者預設的投影片;當使用者發出” Close PowerPoint”時,投影片會結束放映。

## 第五章、結論與未來展望

由於現今科技進步日異月新，人機互動介面的重視已不可同日而語。好的人機介面將有助於拉近人與機器的距離，而本文提出一個藉由使用者的自然肢體語言，去操控其目標平台，利用 Kinect 來實現互動式手勢操控系統。如 Kinect 所提供的骨架標定資訊，利用如向右畫圓或向左畫圓等手勢，用來操作投影片的下一或上一頁。而深度資訊與影像資訊可定位出手掌在空間中的位置及形狀，再根據手掌的位置或形狀來算出手指數，進行投影片操作像是第一頁、最後一頁或結束投影片放映。音訊資訊則可用於控制投影片的開啟或結束，同時也可用於增加使用者操作上的便利。

結合深度資訊及骨架標定，將可對電腦視覺產生極大的穩定性。在未來，可以利用這台穩定的感測器，進行人臉追蹤及表情辨識。目前想法是依照框選的範圍內，搭配 SVM 分類學習器，將不同特徵、紋理來進行統計後，針對不同表情的來做判斷。

未來更是可以增加使用 Kinect 所提供的聲音辨識，嘗試不同的組合，或找出具有良好特性的特徵及演算、將可使人性化操作介面(NUI)的效能大幅提升。其中特徵演算也是一項相當重要的技術，目前本文尚還在前處理步驟，而這部分在未來也是可以改進的主要方向。

## 甘梯圖

	%	%	%	%	%	%	%	%	%	%	%
專題規劃											
收集資料											
研究書籍											
熟悉軟體操作											
程式撰寫											
除錯測試											
預計累積進度(%)	10	20	30	40	50	60	70	80	90	100	

## 致謝

首先感謝指導老師黃雅軒，在學業與研究上給予很多的指導與建議，在百忙之中還不時的與小組討論研究相關的內容，且能精確發現在開發過程中所忽略的細節並給予修改的方針，也讓大家學習到如何靠自己來發現問題、找出問題與解決問題的能力，這不只在專題中有所幫助，在未來研究生涯上或工作上更是受用無窮。

感謝實驗室的學長陳禹仲，在提供專題方面的演算法上，給予非常多的建議與改善，使得在專題製作上在遇到困難時，能得到最好的研究方向。感謝學長莊順旭，提供在程式撰寫上的指導，感謝學長林啟賢提供予實驗資料庫…等多方面的幫忙協助，也感謝實驗室的學長歐志鴻、彭國達、張偉偉、陳宣有對於專題開發過程有所疑問時能盡其所能的回答與協助。

## 參考文獻

- [1] P. Mistry and P. Maes, "SixthSense - a wearable gestural interface," Proceedings of SIGGRAPH Asia 2009, 2009.
- [2] ToF Image.[Online].Available:  
<http://herkules.oulu.fi/isbn9514269667/html/c305.html> [Data accessed]
- [3] Tof.[Online].Available:  
<http://herkules.oulu.fi/isbn9514269667/html/c305.html> [Data accessed]
- [4] 激光散斑圖.[Online].Available:  
<http://www.bb.ustc.edu.cn/jpkc/guojia/dxwlsy/kj/part2/grade3/LaserSpeckle.html> [Data accessed]
- [5] Kinect示意圖.[Online].Available:  
<http://static.guim.co.uk/sys-images/Technology/Pix/pictures/2010/6/15/1276594054160/Kinect-002.jpg> [Data accessed]
- [6] Microsoft Research , "Kinect for Windows SDK beta" , 2010
- [7] 深度示意圖.[Online].Available:  
<http://edit.jgospel.net/media/28095/.35358.jpg> [Data accessed]
- [8] Ko-Hsin Cheng , Chaur-Heh Hsieh and Chang-Chieh Wang , " HUMAN ACTION RECOGNITION USING 3D BODY JOINTS" , Ming-Chuan University , 2010
- [9] Yu-Yuan Chen , Tzyh-Jong Shang and Hsu-Chun Yen , "REAL TIME DYNAMIC HAND GESTURE RECOGNITION WITH 3D DEPTH SENSING" , National Taiwan University , 2010
- [10] Xiaoming Yin and Ming Xie , "Finger identification and hand posture recognition for human–robot interaction" , In IEEE Conference on Advanced Image and Vision Computing , 2007