

# 國立交通大學

## 資訊科學與工程研究所

### 碩士論文

基於影像掃描之  
即時單攝影機 3D 人體姿態估測

Real-time 3D Human Posture Estimation Based on Image  
Scanning with Monocular Camera

研究生：楊朝吉

指導教授：莊仁輝 教授

指導教授：彭文孝 教授

中華民國九十八年十一月

基於影像掃描之即時單攝影機 3D 人體姿態估測  
Real-time 3D Human Gesture Estimation Based on Image  
Scanning with Monocular Camera

研究生：楊朝吉

Student : Chao-Chi Yang

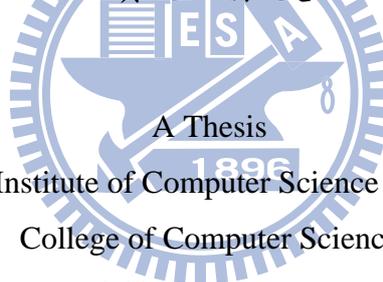
指導教授：莊仁輝

Advisor : Jen-Hui Chuang

指導教授：彭文孝

Advisor : Wen-Hsiao Peng

國立交通大學  
資訊科學與工程研究所  
碩士論文



Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

November 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年十一月

# 基於影像掃描之即時單攝影機 3D 人體姿態估測

學生：楊朝吉

指導教授：莊仁輝 教授

彭文孝 教授

國立交通大學

資訊科學與工程研究所碩士班

## 摘 要

近年來，電腦輔助之人類行為認知的應用越來越廣，這些應用包括事件偵測、行為分析、姿態辨識和人機互動等方面。當應用在人機互動時，我們可以利用不同的姿勢或姿態來表示不同的命令或指令，如作遊樂器的控制。為了能在三度空間中做姿態的辨識，通常會加入控制感應器，增添人機互動應用的臨場感跟多元性。但是多數 3D 遊戲控制需使用昂貴的設備，或是要求使用者使用特殊的控制感應器，容易造成局部運動傷害，且方便性不足。

為了改善上述情形，我們希望以電腦視覺的方法，在單一攝影機下，不使用控制感應器，即能自動的辨識人體 3D 姿態，讓使用者只需擺動肢體，即可在三度空間中進行人機互動。如此可降低局部運動傷害，也使人機互動應用的空間更廣、更真實。本研究期望在架設有一般網路攝影機的簡單平台下，提供使用者人機互動的介面。

# Real-time 3D Human Posture Estimation Based on Image Scanning with Monocular Camera

Student : Chao-Chi Yang

Advisors : Prof. Jen-Hui Chuang

Prof. Wen-Hsiao Peng

Institute of Computer Science and Engineering  
National Chiao Tung University

## ABSTRACT

Recently, the computer-aided applications in human behavior recognition are becoming more and more popular, which include event detection, behavior analysis, posture recognition, and human-machine interaction. For human-machine interaction, e.g., for game control, various gestures or postures can be used to indicate different orders or commands. In order to distinguish behaviors, e.g., for game control, remote controllers are often used to make users feel as if they were in a real-world environment, like the sports field. Compared with the traditional game controllers, the above approach adds more immediate reactions and varieties. However, most 3D game controllers require expensive equipments or ask user put on sensors, which are often inconvenient, and may result in exercise injuries.

To improve the above situations, we hope to develop computer vision method by using monocular camera to recognize 3D human postures. The user can then move the limbs in the 3D space to perform the human-machine interaction. In this way, the exercise injuries can be reduced while enabling more realistic application. This research hopes to employ computer vision to make the human-machine interaction system more affordable on ordinary webcams.

# 致 謝

首先必須要感謝我的指導教授，莊仁輝老師與彭文孝老師。兩位老師不僅給予我許多研究方面的提點，也在處事態度及人生經驗上讓我受益良多。能順利的拿到碩士學位，實在是因為有兩位老師的用心指導與包容。另外也要感謝我的口試委員，黃正能教授及白宏益博士，對這篇論文所提出的指導和建議，讓本論文的內容更加的完整及充實。

接著要感謝實驗室的所有夥伴，尤其是永昌、聖中、尚一、宣良、維辰、邦展、世旻與芝穎。由於大家的熱心協助及扶持，使得我的研究可以更有效率的躍進。也謝謝學弟妹及隔壁實驗室的同學們，由於你們帶給大家的歡樂，使得實驗室的氣氛一直很和樂，也因此使得研究的生活並不會孤單及煩悶。在此要特別感謝星陸、泓宏及國華學長，在我程式與實驗上有任何問題時，你們總是不吝惜的指導，讓我在研究上可以更上一層樓。

最後感謝我的父母、家人對我的關懷、支持及鼓勵，以及所有幫助過我的朋友、同學們，你們都是我心靈上最大的支柱。沒有你們，我或許無法堅持到最後，僅以此篇表達我最誠摯的謝意。

# 目 錄

中文摘要.....	I
英文摘要.....	II
致謝.....	III
目錄.....	IV
圖目錄.....	V
表目錄.....	VII
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 相關研究.....	2
1.3 論文架構.....	6
第二章 人體區域偵測.....	8
2.1 前景物擷取與影子濾除.....	8
2.2 前景物區域的膚色擷取.....	13
第三章 基於掃描線演算法之肢體近似中心軸線分析.....	16
3.1 掃描線演算法.....	16
3.2 掃描線截面線段之近似中軸線段建立.....	19
第四章 2D 人體模型建立.....	25
4.1 人體區域的外接矩形.....	25
4.2 頭部選取與軀幹位置估測.....	26
4.3 腳部位置估測.....	29
4.4 手部位置估測.....	32
第五章 3D 人體模型建立.....	37
5.1 關節景深資訊的估算.....	37
5.2 虛擬視角畫面建立.....	39
第六章 實驗結果與加速方法.....	42
6.1 關節對應的結果討論.....	42
6.2 加速方法.....	50
第七章 結論及未來工作.....	53
參考文獻.....	54

# 圖目錄

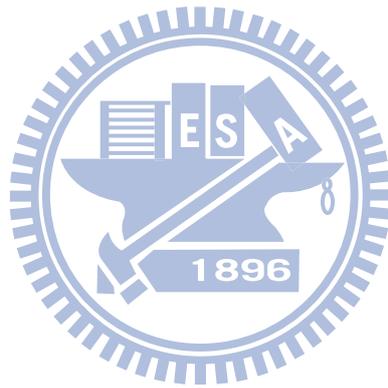
圖 1.1 骨架計算[1].....	3
圖 1.2 在圖中標示關節位置，以重建 3D 人體模型[4].....	4
圖 1.3 依輪肢體廓來調整姿態的人體模型[7].....	4
圖 1.4 調整 3D 人體模型的方法[8].....	5
圖 1.5 戶外環境下建立的 2D 人體模型結果[9].....	5
圖 1.6 系統流程圖.....	7
圖 1.7 人體正面的關節對應圖.....	7
圖 2.1 人體區域偵測流程圖.....	8
圖 2.2 以 RGB 絕對值差的和，偵測前、背景的結果.....	10
圖 2.3 對圖 2.2 做影子濾除的結果.....	10
圖 2.4 以 RGB 色彩空間表示輸入像素與背景像素的亮度與色度扭曲[11].....	11
圖 2.5 以文獻[11]偵測前、背景與影子區域的結果.....	12
圖 2.6 膚色在不同色彩空間中的分布[12].....	13
圖 2.7 以文獻[12]對前景物區域偵測膚色的結果.....	15
圖 3.1 建立肢體近似中軸線段流程圖.....	16
圖 3.2 以兩個不同方向的掃描線掃描人體區域後所得之截面線段.....	17
圖 3.3 將圖 3.2 中寬度過短之截面線段濾除的結果: (a) $0^\circ$ , (b) $90^\circ$ .....	18
圖 3.4 彌補截面線段的破損.....	19
圖 3.5 水平方向( $0^\circ$ )掃描線的擷取結果.....	21
圖 3.6 垂直方向( $90^\circ$ )掃描線的擷取結果.....	22
圖 3.7 移除過寬或過短的近似中軸線段.....	23
圖 3.8 以近似中軸線段與掃描線夾角推算肢體寬度的示意圖.....	23
圖 3.9 擷取較正確的近似中軸線段端點位置.....	24
圖 3.10 兩個方向掃描線得到的前景物與膚色近似中軸線段.....	24
圖 4.1 以近似中軸線段的端點，定義人體外接矩形.....	26
圖 4.2 以近似中軸線段的外接矩形，定義人體外接矩形.....	26
圖 4.3 頭部選取的結果.....	27
圖 4.4 改良頭部選取方法的結果.....	28
圖 4.5 肩膀與髖骨的連線.....	28
圖 4.6 腳部的偵測範圍.....	29
圖 4.7 調整髖骨的位置.....	30
圖 4.8 以橢圓表示腳部位置.....	31
圖 4.9 手部的偵測範圍.....	33
圖 4.10 擴增手部的偵測範圍.....	33
圖 4.11 以橢圓表示手部的初始位置.....	34

圖 4.12 以膚色近似中軸線段調整手部位置，短袖的例子.....	35
圖 4.13 以膚色近似中軸線段調整手部位置，長袖的例子.....	36
圖 5.1 人體側面的姿勢圖[18].....	38
圖 5.2 手部關節的相對深度值計算.....	39
圖 5.3 虛擬攝影機示意圖.....	40
圖 5.4 以不同視角建立人體 3D 模型.....	41
圖 6.1 短袖的偵測結果.....	43
圖 6.2 腳末端顏色接近背景顏色的狀況.....	44
圖 6.3 手部偵測錯誤的狀況.....	44
圖 6.4 膚色區域、衣著接近背景顏色的偵測結果.....	45
圖 6.5 動態調整肢體長度.....	46
圖 6.6 腳部部分重疊的偵測結果.....	46
圖 6.7 複雜場景中的偵測結果.....	47
圖 6.8 背面的偵測結果.....	48
圖 6.9 受測者距鏡頭較遠的偵測結果.....	49
圖 6.10 受測者蹲下的偵測結果.....	50
圖 6.11 以降低解析度的概念，減少偵測的像素資料量.....	50



# 表目錄

表 5-1 肢體相對長度估測值.....	39
表 6-1 影像來源、掃描線方向、肢體表示及辨識比較.....	52



# 第一章 緒論

使用單一攝影機來辨識3D人體姿態，在電腦視覺與影像處理的領域，一直是個有趣且有挑戰性的議題，可應用在事件偵測、行為分析、姿態辨識和人機互動等方面。由單一攝影機擷取到的畫面(frame)，只提供我們像素的2D資訊，缺乏像素的景深值(depth of field)。為了偵測物體在空間中的位置，一般我們會使用兩台以上的攝影機，並且需要知道攝影機在空間中的位置關係。但除非物體及攝影機皆靜止不動，否則我們還需要解決不同攝影機影像畫面間的同步問題，才可以從兩張以上的畫面，正確的計算物體在空間的位置，相關的計算方法複雜且需使用較多軟、硬體設備。

為了改善上述情形，本論文以電腦視覺的方法，在單一攝影機下，自動的辨識的受測者姿態，再以運動學(kinematics)的活動限制，即時重建對應的3D人體模型。本章將依序介紹我們的研究動機、相關研究及論文架構。

## 1.1 研究動機

由於以多台攝影機偵測物體的空間位置，需要使用較多的軟、硬體設備，所以本論文旨在探討如何以單一攝影機的影像，來估測人體3D姿態。我們可經由影像處理的技術，在2D影像上擷取人體區域。然而人體區域會因為肢體自我遮蔽(self-occlusion)，使得肢體在畫面中消失(例如，手在軀幹後方)或是在人體區域內部形成不易偵測的輪廓(例如，手在軀幹前方)。在第二種自我遮蔽的情況下，本論文希望利用影像畫面的彩色資訊，分析人體區域內部的輪廓，來推測肢體的位置。

肢體中以頭部選取最為關鍵，因頭部自我遮蔽的情況較少。若以膚色或人臉

特徵來辨識頭部，會因為頭飾、眼鏡或口罩等因素，使得頭部選取不易。所以本論文利用人體區域輪廓寬窄變化的特徵，來選取頭部位置，亦可衍生此方法做其他肢體位置的辨識。然而，如果我們對畫面中所有的像素做輪廓(邊界)檢測(Canny)，其運算量較大。為了可以即時的運算，我們參考[15]、[16]、[17]的研究，以改良的掃描線演算法，來擷取肢體的近似中軸線段，再辨識關節位置。本論文提出類似降低畫面解析度的概念，適當的調整掃描線間距，來減少像素的前景物偵測運算量，使系統盡可能達到即時運算。

在得到2D人體模型後，若要重建3D模型，還需要計算關節在畫面上的景深資訊。由於人體是由固定長度的肢體所組成，且肢體間以關節連接，所以人體的肢體有運動學上的活動範圍限制，像是肩膀與手連接、髖骨與腳連接；手肘一般狀況往身體後方曲折、膝蓋往身體前方曲折等。依人體比例的關係[7]，如果可以得知畫面上頭頂到肩膀的長度，以此長度為參考長度，即可自動計算其它肢體約略的相對長度。而後，我們即可依2D人體模型在畫面上的肢體長度與它的相對長度，根據畢氏定理(Pythagorean theorem)，來計算關節的相對景深值。本論文以關節在畫面中的位置及景深值資訊，來重建3D人體模型。

## 1.2 相關研究

在電腦視覺與影像處理的相關領域中，分析人體姿態的研究，可應用在事件偵測、行為分析、姿態辨識和人機互動等方面。在不可控制的監控環境中，會因為背景的凌亂或光線的變化，使擷取到的人體區域產生破損。即使完美擷取人體區域，也必須解決自我遮蔽的問題，才能得到較正確的肢體位置。

有關分析人體姿態的研究，通常會建立 2D 人體模型，來做人體姿態的比對。建立 2D 人體模型的方法，必須在畫面中的人體區域找到關節的位置。文獻[1]所提的方法是在人體區域依形狀標示肢體末端，當人體成大字形張開時，可以得到圖 1.1 的結果。此方法會因為自我遮蔽的情況，像是手在身體前方或後方，雙

腳靠近，使肢體末端遺失。

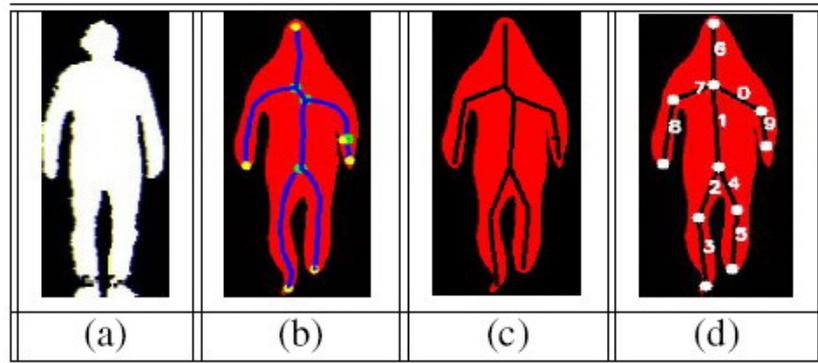
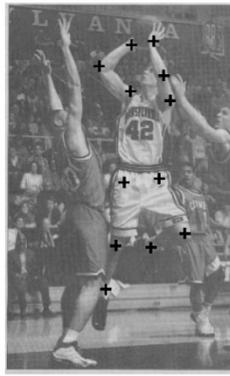


圖 1.1 骨架計算[1]，(a) 擷取人體區域，(b) 骨架計算，(c) 取得骨架線段，(d) 將骨架線段編號來辨識肢體。

為了偵測人體在三度空間中的姿態，一般會架設多台攝影機[2][3]。相關的研究需要知道攝影機在空間上的相對位置；亦或在不同攝影機的畫面中，找對應的地標，以計算攝影機的相對位置。此方法需要處理不同攝影機影像間的同步關係，也需使用較多的軟、硬體資源。由於每台攝影機在分析 2D 影像時，需對不同的肢體做標記，並在不同攝影機畫面中做標記的對應，但是這些標記所標示的肢體可能在畫面上遺失或不易偵測。

在 2D 畫面上重建 3D 人體模型，有一種方法是藉由使用者以手動方式在 2D 畫面上標示關節的位置[4]，如圖 1.2，再由使用者給定肢體的長度，來計算關節的相對景深值。此方法提供我們在找到影像中的關節座標時，估算關節間相對景深值的概念。然而在一般應用中，系統不能在即時的影像中，以手動標示的方法，來給予關節位置及肢體間的長度。

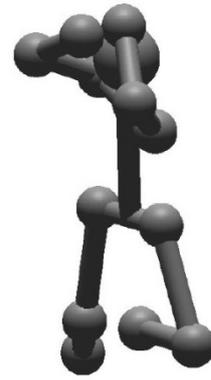
基於上述問題，近年來有針對單一影像畫面重建人體 3D 模型的相關研究，主要探討取得關節位置的方法。常見的作法兩種，第一種為偵測人體區域輪廓的特徵，個別建立肢體，再還原重建人體模型[5]、[6]、[7]；第二種為預先建立 3D 人體模型物件，再局部依肢體分布找尋對應的關節位置[8]、[9]。



(a)



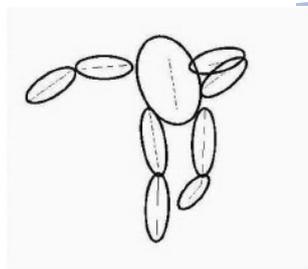
(b)



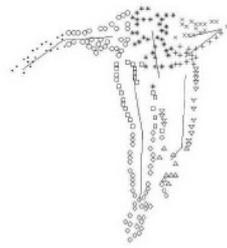
(c)

圖 1.2 在圖中標示關節位置，以重建 3D 人體模型[4]：(a) 以十字標示關節的位置，(b)、(c) 以文獻[3]重建不同角度的對應 3D 人體模型。

為了解決自我遮蔽的問題，相關研究除了偵測外部輪廓，也進一步偵測內部輪廓[5]、[6]、[7]，如圖 1.3(a)。此方法由於必須擷取人體內部的輪廓、分析對應的樣本點，再調整樣本點以改變模型姿態，使得此方法的判斷相對複雜。



(a)



(b)



(c)

圖 1.3 依肢體輪廓來調整人體模型[7]：(a) 人體模型，虛線為骨架位置，(b) 在肢體輪廓自動給定樣本點，(c) 調整樣本點以調整對應的姿勢。

文獻[8]中，提到預先儲存 3D 人體模型樣本的做法，首先對輸入影像來分析頭部、手及腳的位置分布，再調整 3D 人體模型的肢體到對應肢體的位置，如圖 1.4。圖中，部分自我遮蔽的腳部，係以重疊方式在畫面中顯示。

由於人類的脖子寬度在一般情況下窄於頭部與軀幹，所以文獻[9]以前景物截面的寬度判定脖子位置，在得知脖子位置後，依序選取頭部、軀幹及四肢。

圖 1.5 中，畫面中的人體為側身，左右肩膀與髖骨重疊顯示在畫面上。此方法以

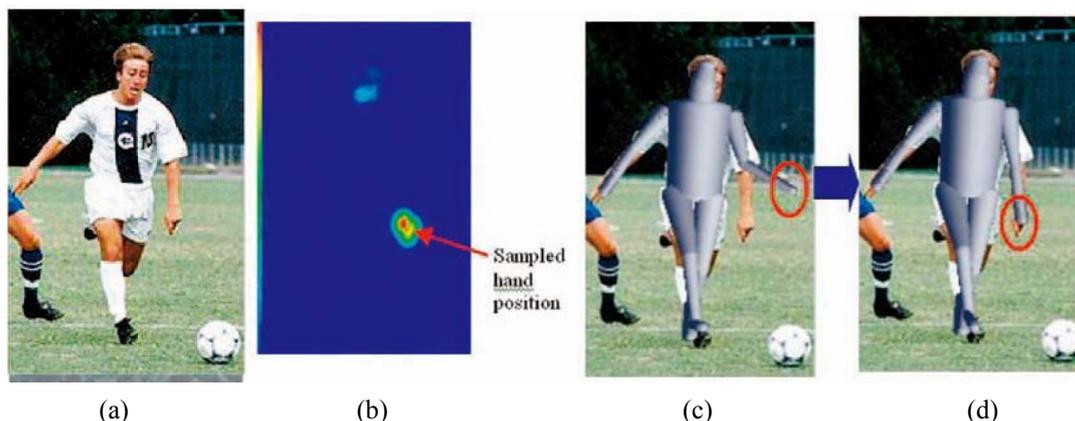


圖 1.4 調整 3D 人體模型的方法[8]：(a) 原始影像，(b) 擷取左手(紅點)將對應到模型上，(c) 正面姿態對應，(c) 調整模型的左手到紅點位置。

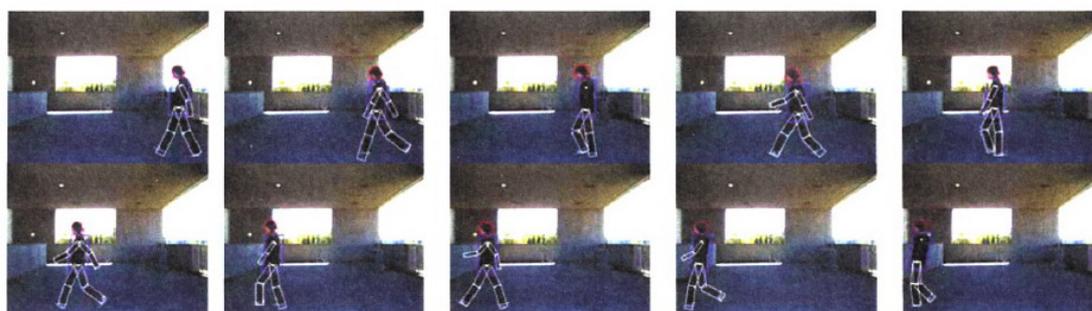


圖 1.5 戶外環境下建立的 2D 人體模型結果[9]。

機率及運動學的限制，估測手部與腳部位置，來自動建立 2D 人體模型，並可進一步重建 3D 人體模型[10]。

為了得到不同肢體的位置，[15]、[16]、[17]的研究中以四種不同角度( $0^\circ$ 、 $90^\circ$ 、 $45^\circ$ 、 $135^\circ$ )的掃描線，在人體區域擷取不同方向的肢體近似中軸線段。系統會對多段近似中軸線段進行整合，並標示肢體在畫面上的位置。研究[17]更進一步分析不同近似中軸線段所代表的肢體是為頭部、手部或腳部。由於上述研究是以灰階影像來作分析，且當自我遮蔽時，亦會產生部分肢體消失的情況。如果此方法應用於醫療看護上，則系統只需要偵測受測者是否發生跌倒或長時間躺下，部分肢體的消失，不至於影響系統的偵測結果。

### 1.3 論文架構

針對前述研究，我們發現在自我遮蔽的情況下，必須額外在人體區域的內部擷取輪廓資訊，才可能偵測到人體區域內部的肢體。本系統將對前景物區域作膚色的偵測，藉此可省略其他色彩輪廓的擷取與辨識。本論文首先將[15]、[16]、[17]研究中，對灰階影像的分析衍生到彩色影像上，並在擷取前景物的肢體近似中軸線段時，同時建立前景物上膚色區域的近似中軸線段。本論文亦改良[15]、[16]、[17]中擷取近似中軸線段的演算法，再延伸前述3D人體模型相關研究的方法，在單攝影機下以人體區域外接矩形資訊，以及肢體活動範圍內建立可動態調整的3D人體模型，期望藉由對於人體不同姿勢的即時分析，來發展基於3D姿態辨識的人機互動介面。

圖1.6為本論文的系統流程圖，以綠色虛線矩形分為三個部分。第一部份是利用前背景分離的技術擷取人體區域，並在人體區域上作膚色的偵測，這部份將在第二章做詳細的介紹。接著第二部份以[14]的近似中心軸線概念為基礎，但改良原先以四個方向( $0^\circ$ 、 $90^\circ$ 、 $45^\circ$ 、 $135^\circ$ )的掃描線演算法，而以更具彈性的演算法，以水平( $0^\circ$ )及垂直( $90^\circ$ )兩種掃描線擷取近似中軸線段，詳細的作法會在第三章做介紹。第四章將利用擷取到的近似中軸線段，建立2D人體模型。第三部份計算畫面上人體關節的相對景深值，以重建3D人體模型，經座標轉換後，建立不同視角的虛擬畫面來做人機互動的介面，將在第五章做介紹。第六章是本論文的實驗結果討論及加速方法。第七章探討本論文的結論及未來工作。

本系統藉由近似中軸線段的端點位置，來估測2D人體模型的16個關節(joints)，如圖1.7。論文將依序在人體區域做頭部的選取(圖1.7中紅點連線)、軀幹位置的估測(圖1.7中棕色點內部區域)、腳部(圖1.7中藍點連線)與手部(圖1.7中綠點連線)位置的估測。

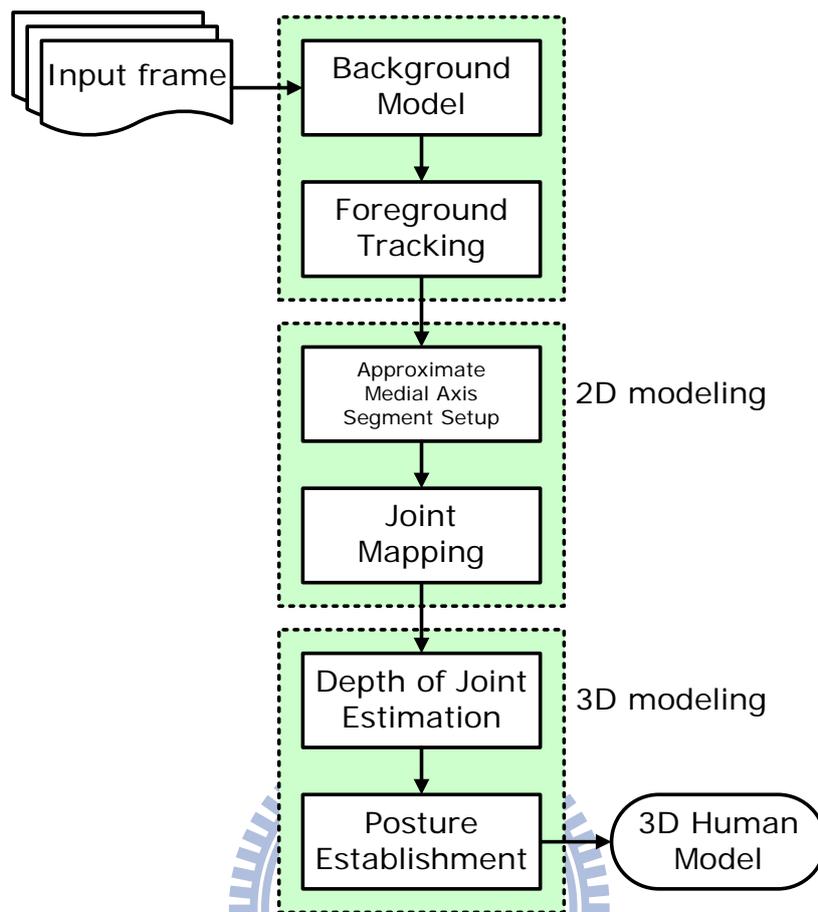


圖 1.6 系統流程圖。

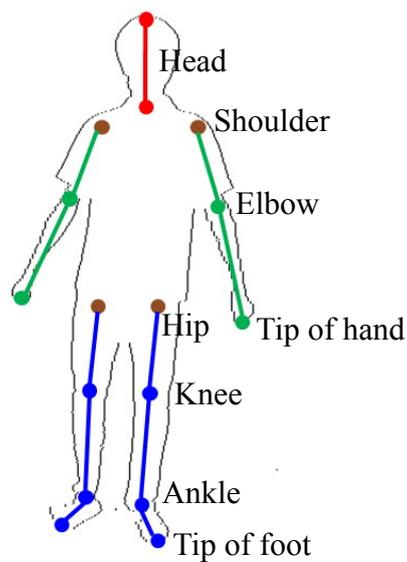


圖 1.7 人體正面的關節對應圖。

## 第二章 人體區域偵測

為了建立人體模型，我們首先擷取影像中的人體區域。由於彩色影像提供的特徵資訊較灰階影像多，故相關的研究多已從灰階影像轉成使用彩色影像來進行偵測及辨識。儘管如此，人體區域的偵測與辨識還是存在許多問題需要解決，如：人體側身的確切角度、部分自我遮蔽的問題、不同衣著、膚色造成的顏色辨識問題、不可預期的髮型及體型、複雜的背景及光源的影響等。

本章節將介紹整個系統的前端部分，也就是人體區域偵測。首先本系統對輸入的影像做前景物及影子偵測，再於前景物區域中偵測膚色區域。圖 2.1 為流程圖，接下來我們將會在各個小節逐一介紹每個步驟。

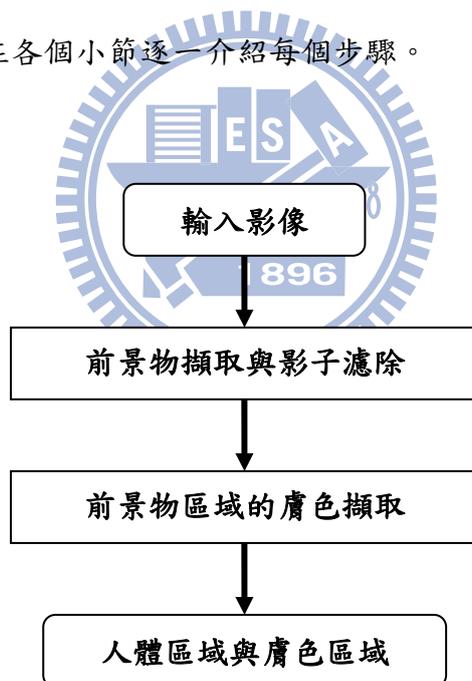


圖 2.1 人體區域偵測流程圖。

### 2.1 前景物擷取與影子濾除

若直接對影像作人體偵測，只能以畫面中的輪廓特徵做分析，此方法不但耗時且可能因複雜的背景而影響正確性。有鑑於此，我們預先建立監控環境的背景

模型(background model)，再對每張新進的畫面作前景物偵測，如此才能濾掉大部份的背景資訊，減少人體區域擷取的錯誤。

常見的前景物偵測法有兩種，分別為高斯混合模型(Gaussian mixture model)與背景減去法(background subtraction)。第一種方法，會學習光影變化，動態的更新背景模型，若 GMM 學習速度過快，在畫面中有前景物移動量極小時，會將該前景物學習成為背景的一部分。因此系統運作時，需適當的調整 GMM 學習速度，此方法較適用於光影變化較大、前景物移動量較大的監控環境。第二種方法，背景減去法需要事先建立背景模型，隨著監控環境的變更，需要更換背景模型，適用於光影變化不大、前景物移動量較小的監控環境。依上述分析，在模擬人體姿態模型的應用下，適合使用背景減去法來偵測人體區域。

背景減去法直覺上的作法，為對目前輸入影像與背景模型對應座標的像素顏色差異的判斷，表示式如下。

$$\sum_{C=R,G,B} |I_C - B_C| \geq T_{abs} \quad (2.1)$$

其中， $I$  為目前輸入影像的像素 RGB 值， $B$  為其對應背景模型的像素 RGB 值。若  $I$  與  $B$  絕對值差的和大於門檻值( $T_{abs}$ )，表示顏色差異夠大，則判定該像素為前景物(foreground)；否則，判定該像素為背景(background)，此判斷式在前景物與背景 RGB 顏色相近時，會造成前景物區域的破損，如圖 2.2 中的臉部、手部區域(紅圈)。

即使  $T_{abs}$  適當的調整，以較完整的擷取前景物區域，也要再對前景物區域作影子的偵測。本論文參考文獻[13]，將 RGB 色彩空間轉至 HSV 色彩空間，以下列條件式是否為影子，如果下列條件式均成立就判定該像素為影子，而不作後續前景物的分析。

$$(a) \alpha \leq \frac{P_V(x,y)}{B_V(x,y)} \leq \beta$$

$$(b) |P_S(x,y) - B_S(x,y)| \leq T_S$$

$$(c) |P_H(x, y) - B_H(x, y)| \leq T_H$$

其中， $P_H(x, y)$ 、 $P_S(x, y)$ 、 $P_V(x, y)$ 分別為目前影像座標 $(x, y)$ 的 HSV 強度值， $B_H(x, y)$ 、 $B_S(x, y)$ 、 $B_V(x, y)$ 則分別為背景模型座標 $(x, y)$ 的 HSV 強度值， $\alpha$ 、 $\beta$ 、 $T_H$ 、 $T_S$ 是手動指定之門檻值，經由參考文獻[17]實驗，設定其值分別為 0.45、1.0、32 和 32。

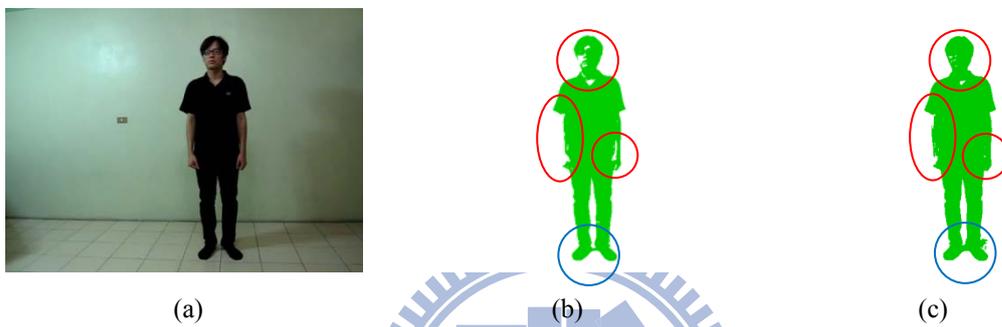


圖 2.2 以 RGB 絕對值差的和，偵測前、背景的結果：(a) 輸入影像，(b)、(c) 以  $T_{abs}=145$ 、90 代入(2.1)式得到前景物區域(綠色)，我們發現當  $T_{abs}$  較大時，會因為前景物像素與背景像素 RGB 顏色差異不夠大，而產生人體區域破損(紅圈)及影子區域(藍圈)。

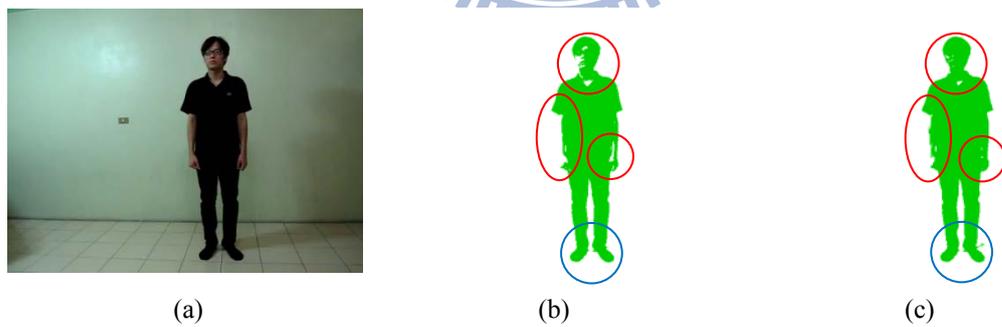


圖 2.3 對圖 2.2 做影子濾除的結果：(a) 輸入影像，(b)、(c) 為圖 2.1(b)、(c)濾除影子後留下的前景物區域(藍圈)。

由圖 2.3 我們可以看到圖 2.2 腳附近，因為影子造成腳部相連的區域(藍圈)，得以正確地被分開。為了加快運算速度，此方法只針對背景減去法後，被判定是前景物的像素做影子偵測。但是以(2.1)式判別前景物，仍容易造成前景物的擷取

遺漏，使得前景物破損嚴重，及後續對前景物區域做顏色偵測的像素來源不足。我們以另一個較穩定的方式，擷取前景物。

我們參考文獻[11]，以統計的方式，較穩定的偵測前景物及影子。首先在未出現前景物的影像，建立背景模型。對新進輸入像素與對應背景模型像素，分別定義兩個的元素：亮度扭曲值  $\alpha$  (brightness distortion)與色度扭曲值  $CD$ (chromaticity distortion)，如圖 2.4。圖中  $I$  為輸入像素色彩向量(R,G,B)， $E$  為期望的對應背景模型像素色彩向量。

$$\alpha = \frac{\bar{I} \cdot \bar{E}}{|\bar{E}|} \quad (2.2)$$

$$CD = \|\bar{I} - \alpha \bar{E}\| \quad (2.3)$$

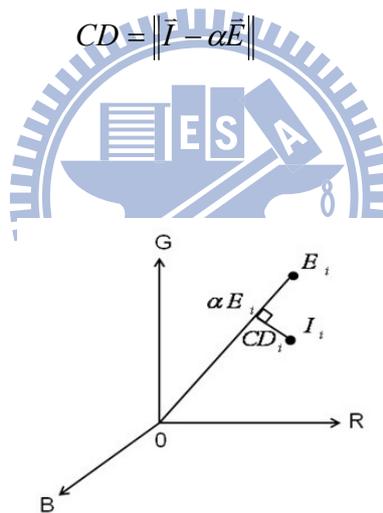


圖 2.4 以 RGB 色彩空間表示輸入像素  $I$  與背景像素  $E$  的亮度與色度扭曲[11]。

根據  $\alpha$  和  $CD$ ，可以將目前影像的像素  $P$  推類成以下四類：

1. 原始背景(*Original Background*,  $B$ )：當  $P$  與背景像素的亮度與色度都相似
2. 影子(*Shadow*,  $S$ )：當  $P$  與背景像素的色度相似，但亮度較暗
3. 明亮背景(*Highlight Background*,  $B$ )：當  $P$  與背景像素的色度相似，但亮度較亮
4. 移動前景物(*Moving Foreground*,  $F$ )：當  $P$  與背景像素色度有明顯差異

我們設定色度扭曲與亮度扭曲的門檻值，來辨別像素的屬性。

$$\begin{aligned} F: & \text{CD} > \tau_{CD} \text{ or } \alpha < \tau_{\alpha}, \text{ else} \\ S: & \alpha < 1, \text{ else} \\ B: & \text{otherwise} \end{aligned} \quad (2.4)$$

其中，以  $\tau_{CD}$  及  $\tau_{\alpha}$  為色度扭曲及亮度扭曲的門檻值，本系統分別取  $\tau_{\alpha}$  為 0.45，取  $\tau_{CD}$  為 10 至 25(若環境光較亮，適度地調低  $\tau_{CD}$  值，減少前景物破損；若環境光較暗，適度地調高  $\tau_{CD}$  值，減少誤判成前景物的雜訊及影子區域)。上述得到的前、背景與影子區域如圖 2.5，相較於圖 2.3，影子區域(藍色)較靈敏地被偵測到，前景物也較完整地被偵測出來，後續我們只對前景物區域(綠色)做分析。

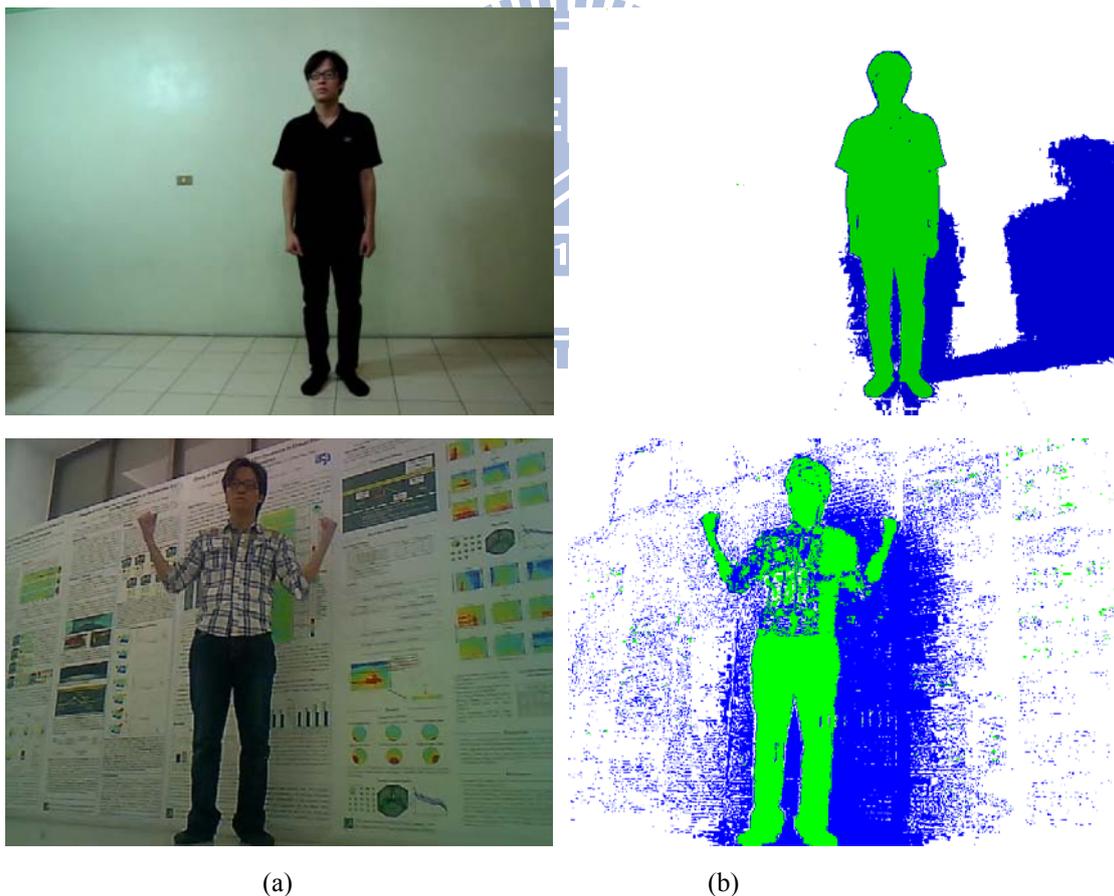


圖 2.5 以文獻[11]偵測前、背景與影子區域的結果：(a) 輸入影像，(b) 綠色像素表示人體(前景物)區域，藍色像素表示影子區域，白色像素表示背景區域(上方是影像品質較好攝影機的影像，背景較單純， $\tau_{CD}=15$ ，下方是網路攝影機的影像，背景稍複雜且衣著接近背景顏色， $\tau_{CD}=12$ )。

## 2.2 前景物區域的膚色擷取

為了在前景物區域偵測膚色，必須定義皮膚(skin)顏色。我們首先選擇一種合適的色彩空間來具體描述皮膚顏色。過去的研究，曾使用過的色彩空間有 RGB、HSV、YCbCr、XYZ、YIQ 與 YES 等，但環境光線變化影響膚色的判定甚大，所以本論文選擇能夠從色彩中分離出亮度元素的 YCbCr 色彩空間。修改 YUV 色彩空間而得的 YCbCr 色彩空間，分為亮度元素(Luminance, Y)與色度元素(Blueness, Cb; Redness, Cr)，將亮度與色度分開操作，較適用於此時的影像處理。

本論文採用 YCbCr 色彩空間判斷膚色區域，因為 YCbCr 色彩空間對於光線變化的反應比 HSV 色彩空間不敏感，對於不同人的膚色也有較好的收斂性。文獻[12]中針對 YCbCr 及 HSV 的色彩空間，詳盡探討膚色分割，圖 2.6 為膚色在 YCbCr 及 HSV 色彩空間中的分布。

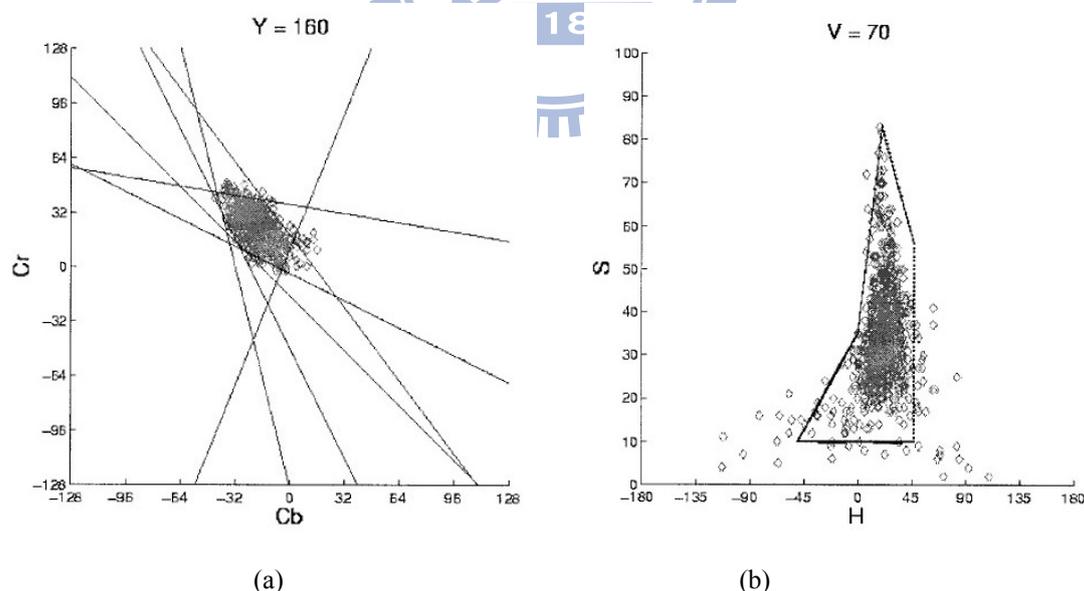


圖 2.6 膚色在不同色彩空間中的分布[12]:(a) 膚色在 YCbCr 色彩空間中的分布,(b) 膚色在 HSV 色彩空間中的分布。

由圖 2.6 可以看出，膚色分布在 YCbCr 色彩空間較 HSV 色彩空間集中，容易將

膚色部分與非膚色部分區分開來。

研究[12]中提出另一種 YCbCr 色彩空間轉換公式(2.5)及膚色判斷公式(2.6)、(2.7)，本系統即採用此種膚色判斷式：

$$\begin{cases} Y = 0.2989 \times R + 0.5866 \times G + 0.1145 \times B \\ Cb = 0.5647 \times (B - Y) \\ Cr = 0.7132 \times (R - Y) \end{cases} \quad (2.5)$$

$$\begin{cases} \text{if}(Y > 128) \\ \theta_1 = -2 + \frac{256 - Y}{16} \\ \theta_2 = 20 - \frac{256 - Y}{16} \\ \theta_3 = 6 \\ \theta_4 = -8 \\ \text{if}(Y \leq 128) \\ \theta_1 = 6 \\ \theta_2 = 12 \\ \theta_3 = 2 + \frac{Y}{32} \\ \theta_4 = -16 + \frac{Y}{16} \end{cases} \quad (2.6)$$

$$\begin{cases} C_r \geq -2(C_b + 24); & C_r \geq -(C_b + 17); \\ C_r \geq -4(C_b + 32); & C_r \geq 2.5(C_b + \theta_1); \\ C_r \geq \theta_3; & C_r \geq 0.5(\theta_4 - C_b); \\ C_r \leq \frac{220 - C_b}{6}; & C_r \leq \frac{4}{3}(\theta_2 - C_b); \end{cases} \quad (2.7)$$

(2.6)及(2.7)中的膚色判斷式為圖 2.6 中七條包含住膚色區域的邊界線，由於膚色分布區域在 Y 值大於或小於等於 128 時有顯著不同，因此以 128 為分界做不同的判斷處理。圖 2.7 為膚色區域分割測試的例子，由圖中可以看到，在肢體部分自我遮蔽的情況，若在前景物上加以分析膚色區域，其結果有助於判別肢體的位置。

實驗的輸入影像品質，會影響系統擷取前景物的結果。如果影像色彩飽和度較差或色調不對，會造成偵測到的膚色區域與實際上有落差。系統必須依環境光的變化，調整偵測前景物色度扭曲的門檻值( $\tau_{CD}$ )，以擷取較完整的前景物，並作膚色區域偵測。但調低  $\tau_{CD}$ ，會使得部分雜訊及影子誤判成前景物，如圖

2.7(b)，我們將在下一章介紹濾除雜訊的方法。

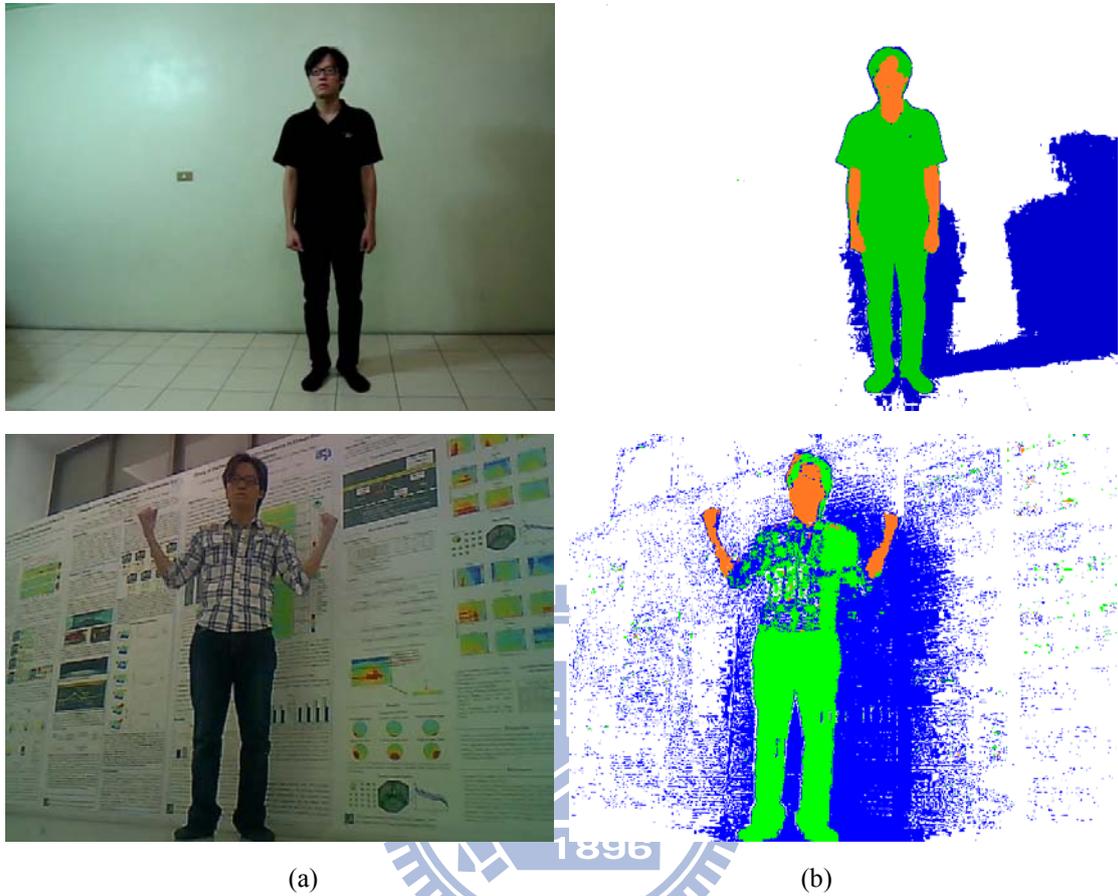


圖 2.7 以文獻[12]對前景物區域偵測膚色的結果，以膚色像素表示膚色區域：(a) 輸入影像，  
(b) 在前景物區域偵測膚色(上方是影像品質較好攝影機的影像，背景較單純， $\tau_{CD}=15$ ，  
下方是網路攝影機的影像，背景稍複雜且衣著接近背景顏色， $\tau_{CD}=12$ )。

# 第三章 基於掃描線演算法之肢體近似 中心軸線分析

本章節是利用[15]、[16]、[17]研究中的概念，來擷取肢體的近似中心軸線。首先我們使用兩個方向( $0^\circ$ 、 $90^\circ$ )及固定間隔的掃描線，使用第二章提到的偵測方法來掃描輸入的影像畫面，得到前景物的截面線段。掃描過程中，系統將濾除過短的截面線段，再利用 3.2 節的規則來建立近似中心軸線段，並濾除過寬或過短的近似中軸線段，留下適合代表肢體的近似中軸線段。圖 3.1 為本章的流程圖，接下來我們會在各個小節介紹每個步驟。



圖 3.1 建立肢體近似中軸線段流程圖。

## 3.1 掃描線演算法

在建立不同肢體近似中軸線段的過程中，我們首先參考[15]、[16]、[17]的研究，以四種不同方向( $0^\circ$ 、 $90^\circ$ 、 $45^\circ$ 、 $135^\circ$ )掃描線掃瞄人體區域的作法，改良以更具彈性的建立近似中心線演算法，放寬截面線段中心點的相連規則，只使用水

平(0°)及垂直(90°)掃描線來擷取人體區域之截面線段。

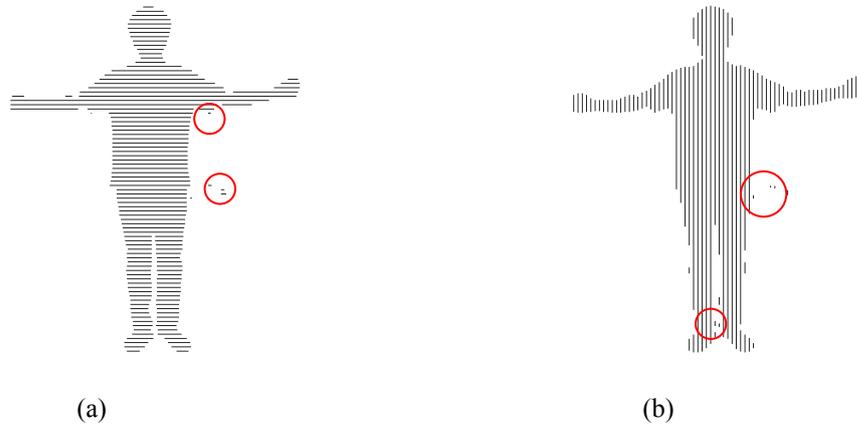


圖 3.2 以兩個不同方向的掃描線掃描人體區域後所得之截面線段，圖中每條線段各代表一個截面線段，紅圈出標示出寬度過短之截面線段：(a) 0°(橫截面線段)，(b) 90°(縱截面線段)。

圖 3.2 中，本系統使用類似降低畫面解析度的概念，以  $\delta$  (單位為像素) 為掃描線的間距，適當地調整  $\delta$  值，可減少偵測的像素，以達到較好的效能，其改變的資料量關係式如下。

$$d = \frac{1}{\delta + 1} xy \quad (3.1)$$

其中， $d$  為偵測的像素個數， $\delta$  為掃描線的間距， $x$ 、 $y$  分別為畫面寬與高的像素個數。當  $\delta$  越小，分析資訊較完整，產生的肢體線段會越接近影像中的人體區域，但是計算量較高，執行速度較慢。如果系統運算能力較差，可以調寬  $\delta$  值，省略偵測的像素個數，當  $\delta$  越大，掃描到像素及肢體截面越少，速度較快。但是若  $\delta$  過大，會因擷取到的近似中心軸線末端位置較不準確，使得在連續的畫面中肢體末端跳動嚴重，我們將在下一節探討如何擷取較正確的近似中軸線段端點。有鑑於此，本系統取  $\delta$  為 5，則資料量減少為原來的 1/6，使系統在偵測像素的階段，速度增快了 6 倍，本論文將在第六章節探討加速方法。

為了解決前景物破損與雜訊的問題，傳統做法利用型態學的技巧，以侵蝕 (erosion) 做雜訊的濾除；擴張 (dilation) 做前景物破損區域的彌補。但是對畫面全

域的侵蝕，會使前景物的面積變小，使得過細的前景物甚至會因此而產生碎損；全域的擴張亦會使雜訊及前景物面積擴大，因此我們需要更有彈性地利用這些概念。

前述的目的旨在濾除雜訊及彌補前景物破損的地方，本論文參考[15]、[16]、[17]的研究，以改良掃描線演算法取代畫面中全域的對前景物做侵蝕與擴散。為了判斷截面線段是否為雜訊，本系統對截面線段做以下的條件判斷式

$$W_{section} < T_{noise} \quad (3.2)$$

其中，若截面寬度( $W_{section}$ )小於門檻值( $T_{noise}$ )，則將此截面線段視為雜訊，予以濾除，本論文以 640x480 的影像測試後，取 7 (單位為像素) 為  $T_{noise}$ ，若畫面解析度較小，可將  $T_{noise}$  調低。

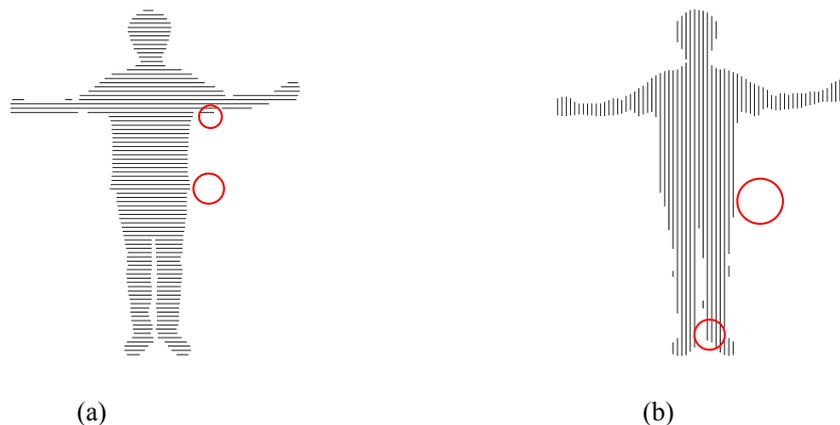


圖 3.3 將圖 3.2 中寬度過短之截面線段濾除的結果：(a)  $0^\circ$ ，(b)  $90^\circ$

另一方面，為了彌補前景物的破損，如圖 3.4，本系統以旗標(*flag*，初始值為 0)作為輔助，以下列三個條件做判斷。

- (a) 當目前像素( $x, y$ )為前景物像素，且 *flag* 為 0。紀錄( $x, y$ )為截面起始座標( $x_s, y_s$ )，並將 *flag* 設為 1。
- (b) 當目前像素( $x, y$ )為前景物像素，且 *flag* 為 1。將( $x, y$ )暫存為的截面終點( $x_{tmp}, y_{tmp}$ )。
- (c) 當目前像素( $x, y$ )為背景像素，與截面終點暫存值( $x_{tmp}, y_{tmp}$ )距離為  $T_{interval}$  (單



建立近似中心軸線，其端點連線即為肢體近似中軸線段(Approximate Medial Axis Segment)。在[14]、[15]、[16]、[17]研究中，使用四個方向(0°、45°、90°、135°)的掃描線掃描影像畫面，來擷取的近似中軸線段。由於四個方向的掃描需要平行或個別處理畫面四次，掃描的像素資料量及運算量是處理一張的四倍，故較耗時且得到近似中軸線段的需整合與剔除，過程較繁複。

本論文改良判斷法則，放寬近似中心點的相連條件，以水平(0°)及垂直(90°)掃描線，即可擷取到適用的肢體近似中軸線段，如此可降低整合的複雜度，並減少像素偵測的運算時間，提升系統效能。將前一節掃描線演算法擷取到截面線段的中心點座標( $x_c, y_c$ )相連成近似中心軸線，在[15]、[16]、[17]研究中，決定兩個截面線段的中心點是否相連，所考慮的判斷條件式如下。

(a) 兩中心點( $x_c, y_c$ )連線與截面線段垂直方向，其間夾角  $\theta$  是否小於等於 45°

(b) 兩相鄰橫截面寬度的差值，是否小於等於 5 個像素。

條件(a)確保兩中心點間距夠近；條件(b)確保兩截面線段寬度差小，符合上述兩個條件，即將此兩個截面線段視為同肢體的一部分，並將中心點( $x_c, y_c$ )相連。

由於上述條件判斷式較嚴謹，以水平(0°)掃描線為例，當肢體與地面夾角接近 45° 或 135° 時容易遺漏，需要加入 45° 及 135° 的掃描線才能擷取到此肢體，再對所有近似中軸線段重複的區域，作剔除或合併的整合步驟。基於上述，本論文根據兩截面線段的寬度，改良水平(0°)掃描線建立的近似中心軸線，條件如下

$$(A) \quad |y_c - M_i \cdot y_{tmp}| = \delta + 1$$

$$(B) \quad |x_c - M_i \cdot x_{tmp}| < T_d$$

$$(C) \quad |W_c - M_i \cdot W| < T_d, \quad T_d = \alpha \max(W_c, M_i \cdot W)$$

其中，( $x_c, y_c$ )為目前的截面中心點座標，( $M_i \cdot x_{tmp}, M_i \cdot y_{tmp}$ )為近似中心軸線( $M_i$ )的暫存末端座標， $M_i \cdot W$ 為近似中心軸線目前擷取到的截面線段平均寬度， $T_d$ 為判別是否為同一近似中心軸線的變異門檻值。

當肢體越接近 45° 或 135°， $W_c$  或  $M_i \cdot W$  亦隨之變寬，根據兩截面線段寬度的

最大值( $\max(W_c, W_{tmp})$ )的  $\alpha$  倍，做為動態調整變異門檻值的依據。若滿足上面的三個條件，則表示  $(x_c, y_c)$  為近似中心軸線( $M_i$ )的延伸，將  $(M_i.x_{tmp}, M_i.y_{tmp})$  修改為  $(x_c, y_c)$ ；若不滿足其中一個條件式，則建立新的近似中心軸線( $M_{i+1}$ )，起點及  $(M_i.x_{tmp}, M_i.y_{tmp})$  皆為目前的中心點  $(x_c, y_c)$ ，在對下一個中心點做上面條件的判斷。

上述條件(A)中，當目前中心點與近似中軸線段( $M_i$ )暫存末端的  $y$  座標間距為掃描線間距  $\delta$ ，表示目前得到的截面中心點  $(x_c, y_c)$ ，可能為近似中軸線段( $M_i$ )的延伸。條件(B)對兩中心點座標  $(x_c$  與  $M_i.x_{tmp})$  判斷是否夠接近。條件(C)中，對兩截面寬度  $(W_c$  與  $M_i.W)$  判斷寬度是否差距不大。本論文經實驗，在擷取前景物與膚色前景物近似中軸線段時，本系統對(B)分別取  $\alpha$  為 0.5 及 0.8，對(C)取  $\alpha$  為 0.4。將條件(A)、(B)、(C)中的  $x$  與  $y$  互相對調，可以得到垂直( $90^\circ$ )掃描線的改良判斷。

上述改良的近似中心軸線的相連判斷式，擷取到的近似中心軸線如圖 3.5(a)。圖 3.5(b)中，紅色橢圓的長軸為近似中心軸線的兩端點連線，即為近似中軸線段；短軸為擷取到的  $N$  個截面線段平均長度值。

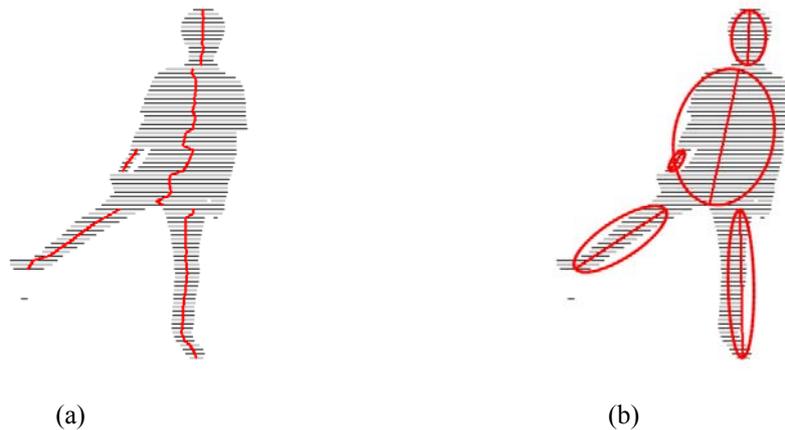


圖 3.5 水平方向( $0^\circ$ )掃描線的擷取結果，(a) 近似中心軸線，(b) 橢圓的長軸為近似中軸線段，短軸為截面線段的平均寬度值。

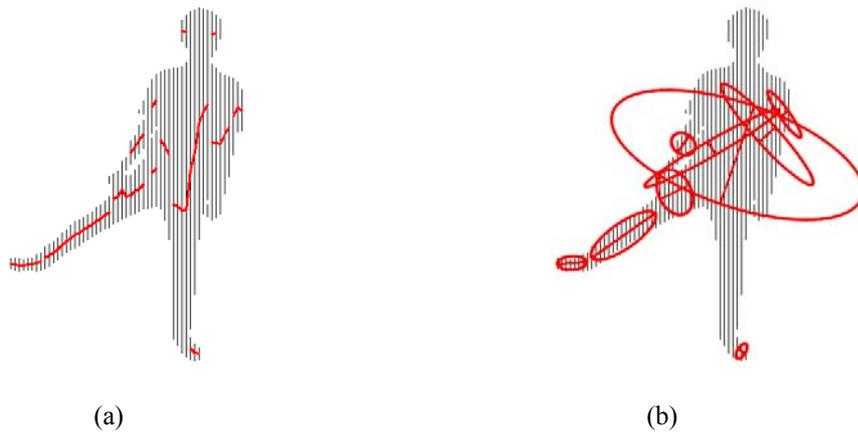


圖 3.6 垂直方向(90°)掃描線的擷取結果，(a) 近似中心軸線，(b) 橢圓的長軸為近似中軸線段，短軸為截面線段的平均寬度值。

[15]、[16]的研究提出，由於人的肢體具有一定長度，且不太可能呈現過寬或過短的形狀，如圖 3.6(a)所示，粗黑線表示近似中軸線段，以紅色橢圓標出了過寬或過短的情況，我們可以清楚地看到它們並不適合代表人的肢體。因此 [15]、[16]進一步地利用下列條件，來濾除較不合適的中心軸線段：

- (a) 該中心軸線段是否由三個以上的截面線段所構成
  - (b) 中心軸線段的平均截面線段寬度是否小於等於該中軸線段之長度
- 由以上兩個條件得到的結果如圖 3.6(b)。

我們希望再進一步得知，肢體的約略寬度。由於以我們改良的演算法，亦可擷取到較不水平或垂直的肢體，使得肢體截面線段的平均寬度值大於實際的肢體寬度值，如圖 3.8。圖 3.8 為肢體方向與水平掃描線呈現夾角  $\theta$  的示意圖，黑色線段為掃描線演算法擷取到的截面線段；藍色線段為截面線段的平均寬度值；紅色線段為近似中軸線段；綠色線段為近似中軸線的寬度值。

$$W_{axis} = \left( \frac{\sum W_i}{N} \right) \times \sin \theta \quad (3.3)$$

其中， $W_{axis}$  為近似中軸線段的寬度值， $W_i$  為第  $i$  個截面線段的寬度，假設此近似中軸線段有  $N$  個截面線段。 $\theta$  為近似中軸線段與掃描線的夾角，經由(3.3)的計算得到近似中軸線段的寬度值。

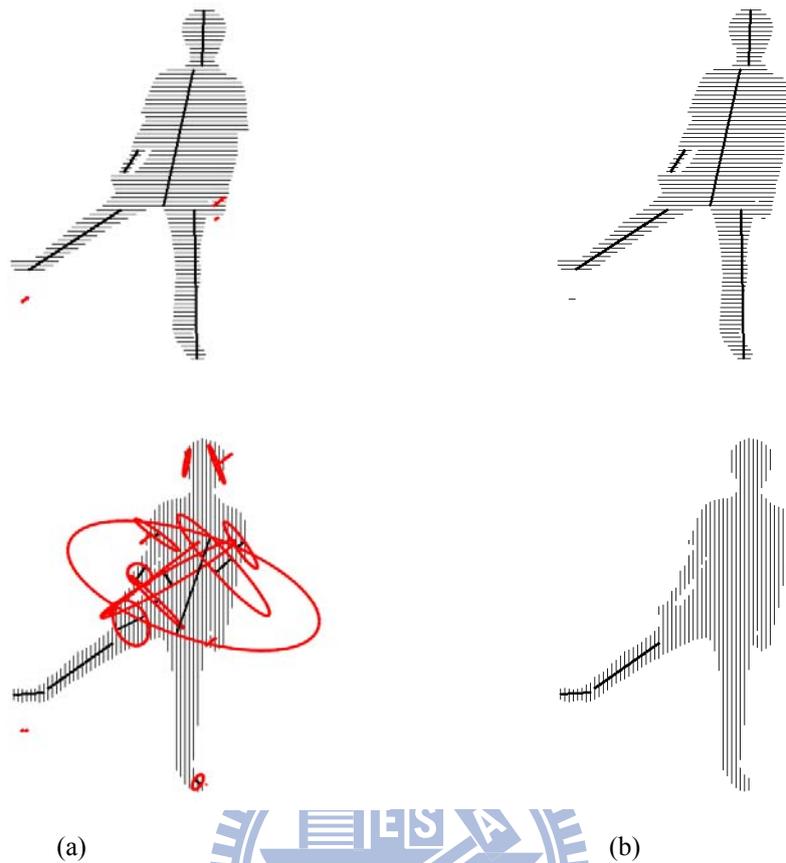


圖 3.7 移除過寬或過短的近似中軸線段：圖上方為  $0^\circ$ ，下方為  $90^\circ$ ，(a)紅色橢圓標示出不適合表示肢體的近似中軸線段，(b)濾除(a)中過寬或過短的近似中軸線段。

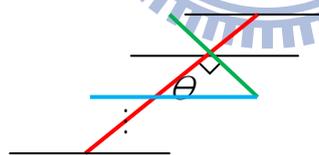


圖 3.8 以近似中軸線段與掃描線夾角推算肢體寬度的示意圖。

當系統運算速度較慢時，可以調整掃描線的間距，以減少像素的偵測判斷，增快執行速度。但是系統掃描線間距變寬，會使得擷取到的近似中軸線段端點誤差變大，使得連續畫面中肢體的末端位置跳動嚴重。我們可依近似中軸線段端點(起點 $(x_S, y_S)$ ，終點 $(x_E, y_E)$ )連線的延伸方向找尋較確切的肢體末端位置，如圖 3.9。

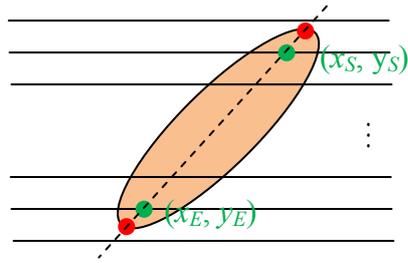


圖 3.9 擷取較正確的近似中軸線段端點位置，往端點(綠點)連線延伸，擷取較接近實際的端點(紅點)位置。

本系統在建立前景物近似中心軸線時，同時建立前景物上膚色區域的近似中心軸線，藉此得到前景物的近似中軸線段與膚色的近似中軸線段，如圖 3.10 所示。圖 3.10 中，我們可以發現，以改良的近似中軸線段演算法，使用水平( $0^\circ$ )及垂直( $90^\circ$ )的掃描線，亦可擷取到接近  $45^\circ$  或  $135^\circ$  的肢體近似中軸線段，如圖 3.10 中左腳，則可省略  $45^\circ$  及  $135^\circ$  掃描線擷取步驟。當肢體發生自我遮蔽的情況時，我們可藉由膚色中軸線段端點位置，來推測肢體的位置。本論文將在第四章整合這些近似中軸線段，並建立 2D 人體模型。

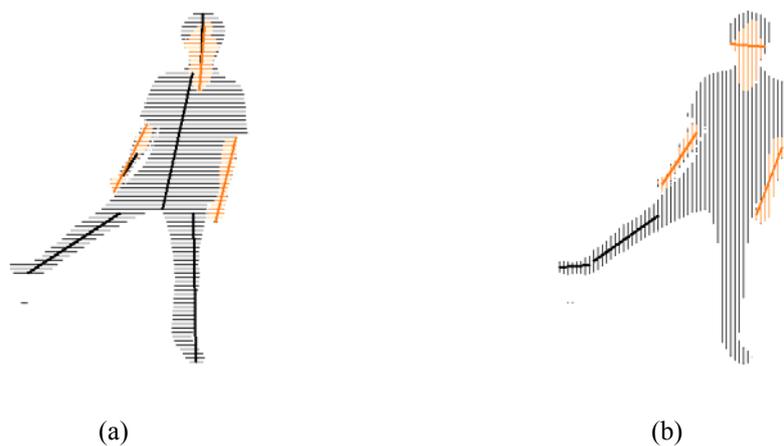


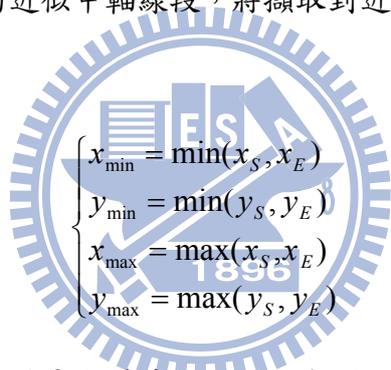
圖 3.10 兩個方向掃描線得到的前景物與膚色近似中軸線段，(a)  $0^\circ$ ，(b)  $90^\circ$ 。

## 第四章 2D 人體模型建立

我們藉由第三章的方法，在 2D 影像中擷取到數段前景物與膚色區域的近似中軸線段。本章將整合這些近似中軸線段，並辨識人體關節位置，來建立 2D 人體模型。本論文參考文獻[9]的概念，依序辨識頭部、軀幹、腳部及手部。

### 4.1 人體區域的外接矩形

本系統在這個階段，首先藉由前景物近似中軸線段的位置資訊，來定義人體區域的外接矩形(bounding box)。在前一節濾除寬度過短的前景物截面線段(視為雜訊)或不適合表示肢體的近似中軸線段，將擷取到近似中軸線段的端點座標( $x_S, y_S$ )及( $x_E, y_E$ )代入下式，


$$\begin{cases} x_{\min} = \min(x_S, x_E) \\ y_{\min} = \min(y_S, y_E) \\ x_{\max} = \max(x_S, x_E) \\ y_{\max} = \max(y_S, y_E) \end{cases} \quad (4.1)$$

式中，( $x_{\min}, y_{\min}$ )及( $x_{\max}, y_{\max}$ )是人體外接矩形的左下、右上座標，假設影像畫面的左下方像素座標為(0, 0)，右上方像素座標為( $W-1, H-1$ )，其中  $W$  與  $H$  分別為畫面的寬與高(單位為像素)，系統可由(4.1)得到由近似中軸線段端點定義的人體區域外接矩形，如圖 4.1。

由圖 4.1 中，我們發現單純以近似中軸線段的端點來判定外接矩形的邊界，並不能得到較準確的人體區域外接矩形。有鑑於此，本系統在擷取前景物截面線段端點( $x_s, y_s$ )、( $x_e, y_e$ )的過程中，個別依(4.1)定義近似中軸線段的外接矩形(bounding box)，再以這些外接矩形得到較正確的外接矩形，如圖 4.2 所示。

本論文經由外接矩形的資訊，定義何時開始分析人體區域。當判定外接矩形不碰觸畫面邊界，即外接矩形滿足條件(4.2)，表示人體區域已全部進入畫面中，才開始做分析；否則，表示人體區域未完全進入畫面，不做分析。

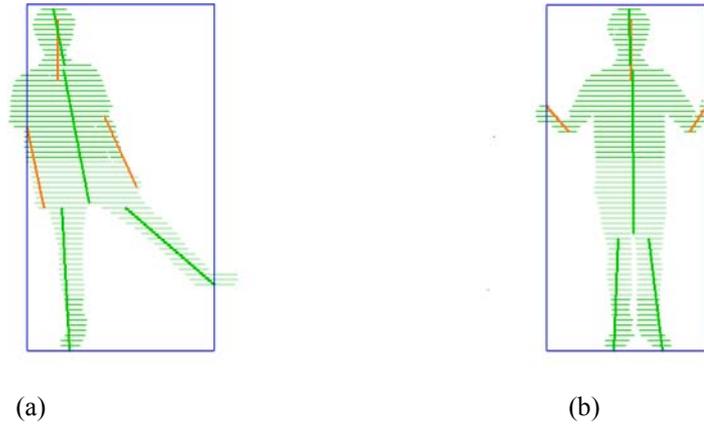


圖 4.1 以近似中軸線段的端點，定義的人體區域外接矩形(藍色矩形)

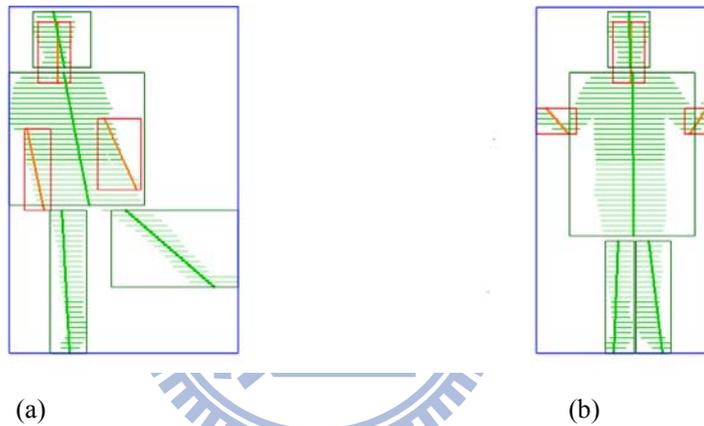


圖 4.2 以近似中軸線段的外接矩形，定義的人體區域外接矩形(藍色矩形)

$$\begin{cases} x_{\min} > 0 \\ y_{\min} > 0 \\ x_{\max} < W - 1 \\ y_{\max} < H - 1 \end{cases} \quad (4.2)$$

## 4.2 頭部選取與軀幹位置估測

我們首先在人體區域選取頭部位置。由於脖子的寬度在人體區域的外部輪廓與頭部及軀幹有明顯的差異，依第三章的方法可擷取到頭部的近似中軸線段。由近似中軸線段的端點連線為橢圓的長軸，截面的平均寬度為橢圓的短軸，可在畫

面上以橢圓表示不同肢體的分布位置及形狀，如圖 4.2。當人站立且手部不高於頭部時，以水平掃描線由上而下擷取到的第一個至少跨越三條掃描線的近似中軸線段為選取的頭部，如圖 4.3 中紅色橢圓。其他姿態的頭部選取，在[17]的研究中有詳細論述。

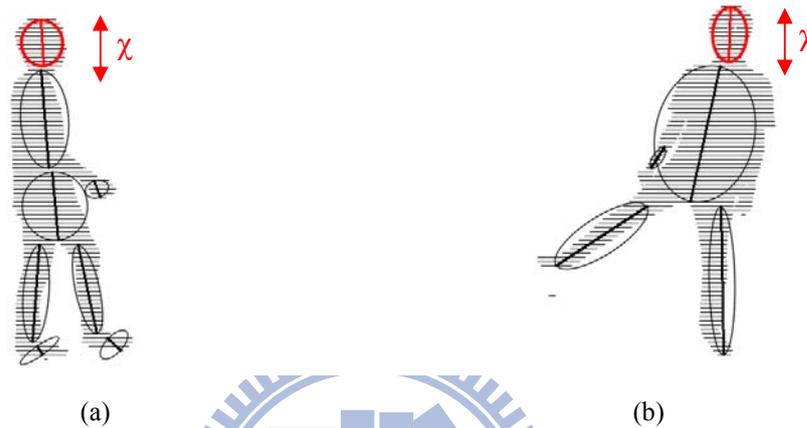


圖 4.3 頭部選取的結果，(a) 側面例子，(b) 正面例子。

如果身體與頭部的寬度接近或截面線段內破損較嚴重而未能彌補，會使得頭部無法以擷取到的近似中軸線段表示，如圖 4.4。在圖 4.4 中，我們發現頭部長度明顯與身高不成人體比例(依人體比例[18]，假定頭頂到肩膀的長度  $\chi$  約為身高的 1/6 至 1/7)，此時我們在肩膀上方，選取合適的頭部長度及寬度，本系統設為  $0.9\chi$  及  $0.8\chi$ 。系統在得到人體身高(以像素為單位)後，若頭部選取不合理，可依上述來調整選頭部的大小。

我們在畫面上選取頭部位置後，依人體比例及外接矩形，估測軀幹在頭部的下方位置，本系統設定軀幹的長度約為  $2\chi$ 、肩膀寬度約為  $1.2\chi$ 、髖骨寬度約為  $0.9\chi$ ，得到估測的位置，如圖 4.5。論文[9]中提到當人體為側面時，肩膀與髖骨位置重疊於畫面中，由於平行於鏡頭側身步行的情況，必須追蹤偵測左右腳，才能正確得知左右腳的前後關係。在真實世界中，軀幹位置可能因衣服顏色與背景接近，使得正面與側面難以判定。由於受測者較常以正面朝螢幕來操控人

機介面，且攝影機常會架設在螢幕上或內嵌於筆記型電腦上，故本論文主要在探討人體正面姿態的偵測。當腳部向左或右伸展時，初始的髖骨位置會產生明顯誤差，如圖 4.5(b)。我們將在下一節，作腳部位置估測，並進一步調整髖骨位置。

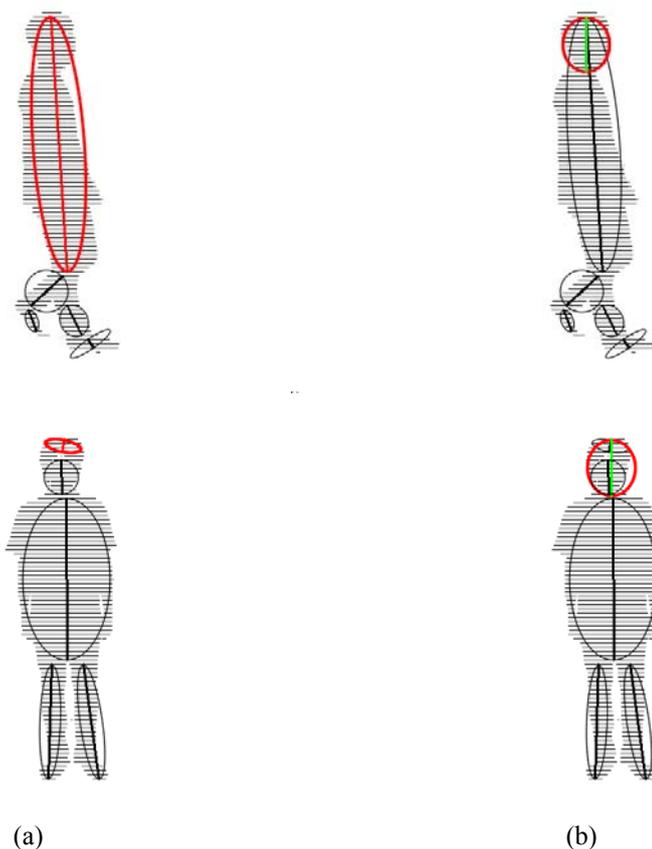


圖 4.4 改良頭部選取的結果，(a) 頭部選取錯誤的狀況，(b) 改良後的頭部選取。



圖 4.5 肩膀與髖骨的連線，(a) 站立例子，(b) 腳部往外伸展的例子。

### 4.3 腳部位置估測

延續前一節，本系統得到的頭部、肩膀與髖骨在影像畫面上的位置，藉此我們可以定義左、右腳的活動範圍。由頭頂到肩膀的長度 $\chi$ 為參考長度，來計算肢體的相對長度。我們依人體比例，假定大腿長度約為 $1.5\chi$ ，小腿長度約為 $1.5\chi$ ，則髖骨到腳底的最大距離為 $3\chi$ ，並設定最大活動距離為 $3.2\chi$ ，給予腳部伸展的彈性活動空間，藉此以髖骨為圓心， $3.2\chi$ 為半徑，在畫面上定義左、右腳可能的活動範圍，如圖 4.6(a)。

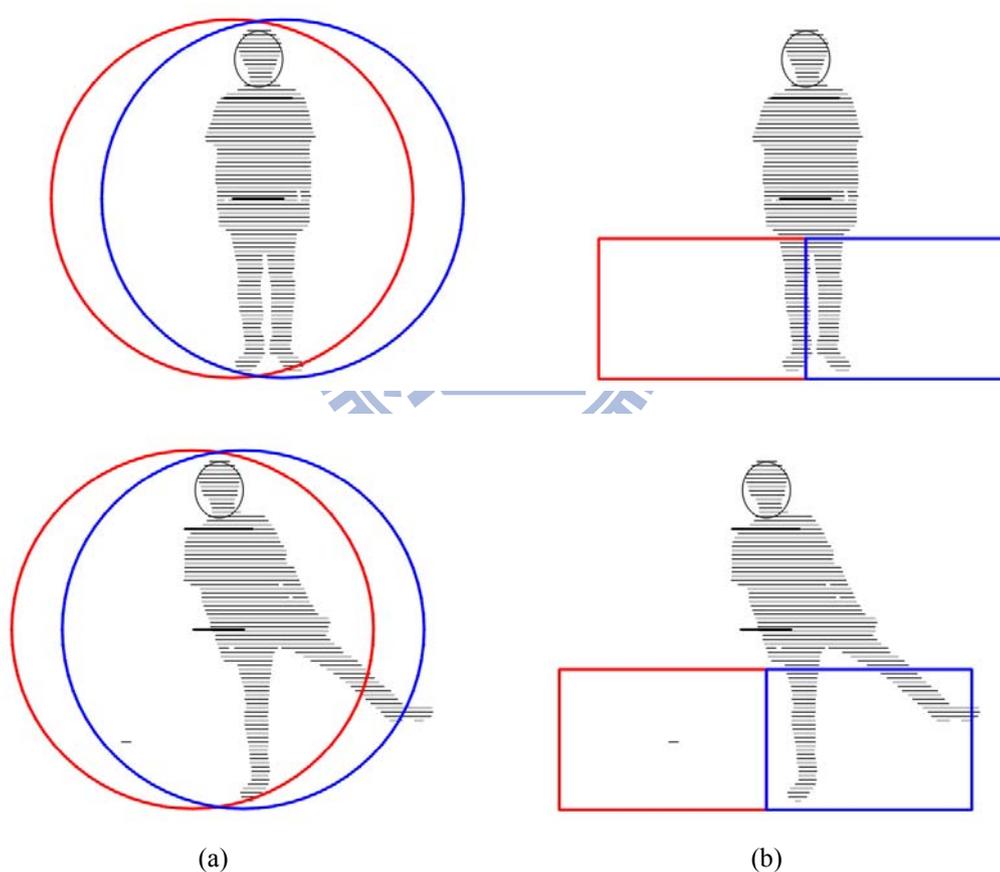


圖 4.6 腳部的偵測範圍，分別以紅色、藍色內部區域，表示左、右腳的活動範圍，(a) 可能的活動範圍，(b) 實際的偵測範圍。

如果腳部與軀幹自我遮蔽的情況較嚴重，會因為褲子與衣服的颜色相近，使得腳與軀幹很難以人體區域內部的輪廓作區分。當畫面中出現前述狀況時，本系

統以預設的腳部姿態取代之。我們以手部活動範圍的下方位置，來定義正面時，腳部實際的偵測範圍，如圖 4.6(b)。在圖 4.6 中可以發現，若在頭部正下方給予腳部的估測位置，當腳往身體左、右方伸展時，會發生兩腳的末端都落在同一腳的活動偵測範圍內，如圖 4.6 下方的狀況。

為了避免上述情況造成的誤判，本論文以腳部區域的前景物近似中軸線段，來定義腳部外接矩形，如圖 4.7 中的紅色矩形。圖 4.7(a)下方的圖中，我們發現即使中心軸線段斷成多節，也可較正確定義腳部區域的邊界。

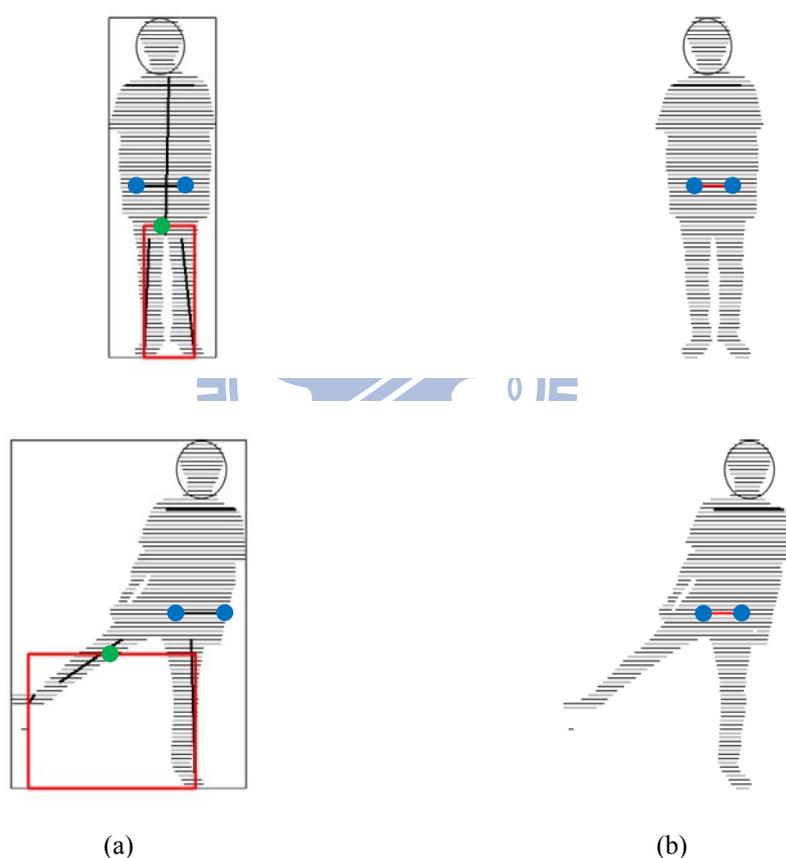


圖 4.7 調整髁骨的位置，(a) 腳部區域的外接矩形，(b) 調整後的髁骨位置連線。

本系統進一步以腳部外接矩形的上部中心點位置(圖 4.7(a)中的綠點)，與髁骨原始的中心點位置，根據下式來調整髁骨的位置。

$$P_1 = \alpha P_2 + (1 - \alpha) P_3 \quad (4.3)$$

其中本系統對髁骨的橫向位置( $x$  座標)做權重值的調整， $P_1$  為調整後的髁骨位置

(圖 4.6(b)中的藍點)， $P_2$  為初始髌骨位置(圖 4.7(a)中的藍點)， $P_3$  為腳部外接矩形的上部中心點位置(圖 4.7(a)中的綠點)，系統設定  $\alpha$  值為 0.7。在腳部外接矩形內，分別計算近似中軸線段端點與中心點位置(圖 4.7(a)中的綠點)左、右距離最長的端點位置，視為左、右腳的末端，其髌骨與腳末端的連線即為腳部，如圖 4.8。

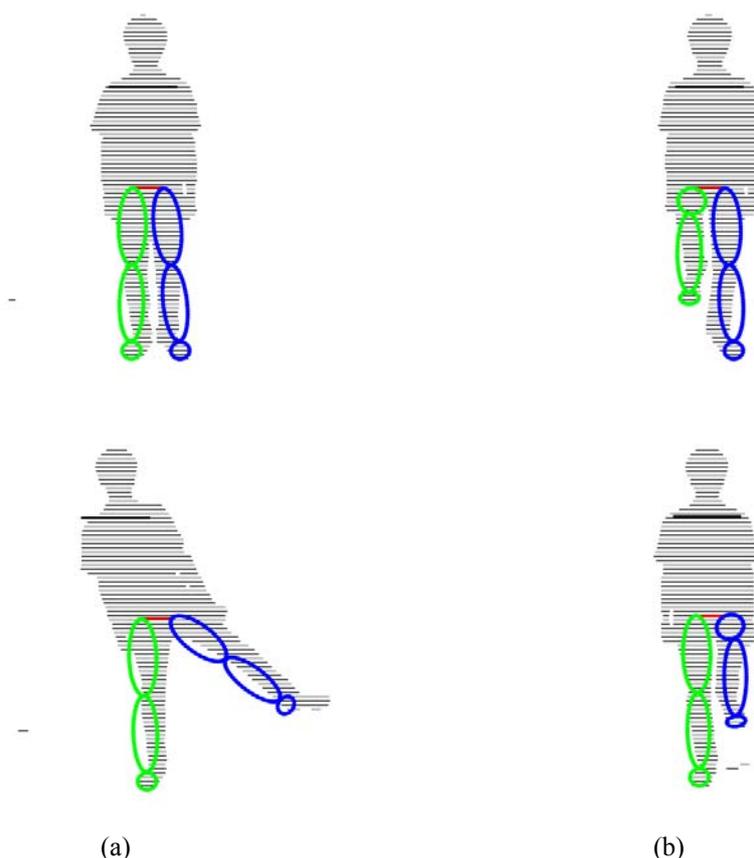


圖 4.8 以橢圓表示腳部位置，(a) 腳部伸直的情況，(b) 腳部有曲折的情況。

圖 4.8(a)下圖中，我們改善了圖 4.6(b)下圖因為兩腳落在同一偵測範圍內，所造成其中一腳偵測不到或偵測錯誤的情況。本系統自髌骨到腳末端的連線位置，依連出三個表示大腿、小腿、腳底板位置的橢圓。本系統依人體比例，定義腳部的大腿、小腿、腳底板的長度占腳長度的比例為 0.45、0.45、0.1，藉此估測膝蓋及腳踝的位置，如圖 4.7 中橢圓的長軸為大腿、小腿及腳底板的端點位置連線，短軸為頭部寬度的 0.6、0.5、0.4 倍。

當腳部的長度小於腳伸直長度( $3\chi$ )，表示腳部有曲折。若以上述定義髌骨

到腳踝的中心位置為膝蓋，表示腳伸直抬高，屬於高難度姿態，一般情況下很難長時間的保持此姿態。我們假定腳發生曲折時，膝蓋的位置是垂直往上舉起，如圖 4.8(b)。我們根據前景物的近似中軸線段端點，來估測腳部位置，所以在腳部位置所形成的膚色近似中軸線，不會影響判別。本系統在腳末端與背景的颜色差異夠大，即使腳部的近似中軸線段斷成多節，也不致影響腳部末端的估測結果。

#### 4.4 手部位置估測

在本節中，我們再依人體比例來定義手部的活動範圍，推測上臂長度約為  $\chi$ ，手肘到手末端(前臂)長度約為  $1.5\chi$ ，則肩膀到手末端的最大距離為  $2.5\chi$ ，並設定最大活動距離為  $2.7\chi$ ，給予手部伸展的彈性活動空間，藉此以肩膀為圓心， $2.7\chi$  為半徑，則可以在畫面上定義左、右手可能的活動範圍，如圖 4.9(a)。

當手部與頭部彼此遮蔽時，由於肢體顏色都為膚色，會造成肢體輪廓邊界難以辨別，以預設姿態表示。我們以頭部的位置定義左、右手的活動範圍，來劃分頭部與手部的分布區域實際的偵測範圍，如圖 4.9(b)。本系統在只在手部活動範圍內，對膚色近似中軸線段做整合，以辨識手部的末端位置。如此不會偵測到分布在頭部區域的膚色近似中軸線段。我們可以由髖骨位置，再擴增手部的偵測範圍，如圖 4.10。我們可以先依前述手部的比例，推測手部的初始位置，以橢圓(以  $\chi$  與  $1.5\chi$  為橢圓的長軸，頭部寬度的 0.5、0.4 倍為橢圓的短軸)表示，如圖 4.11。

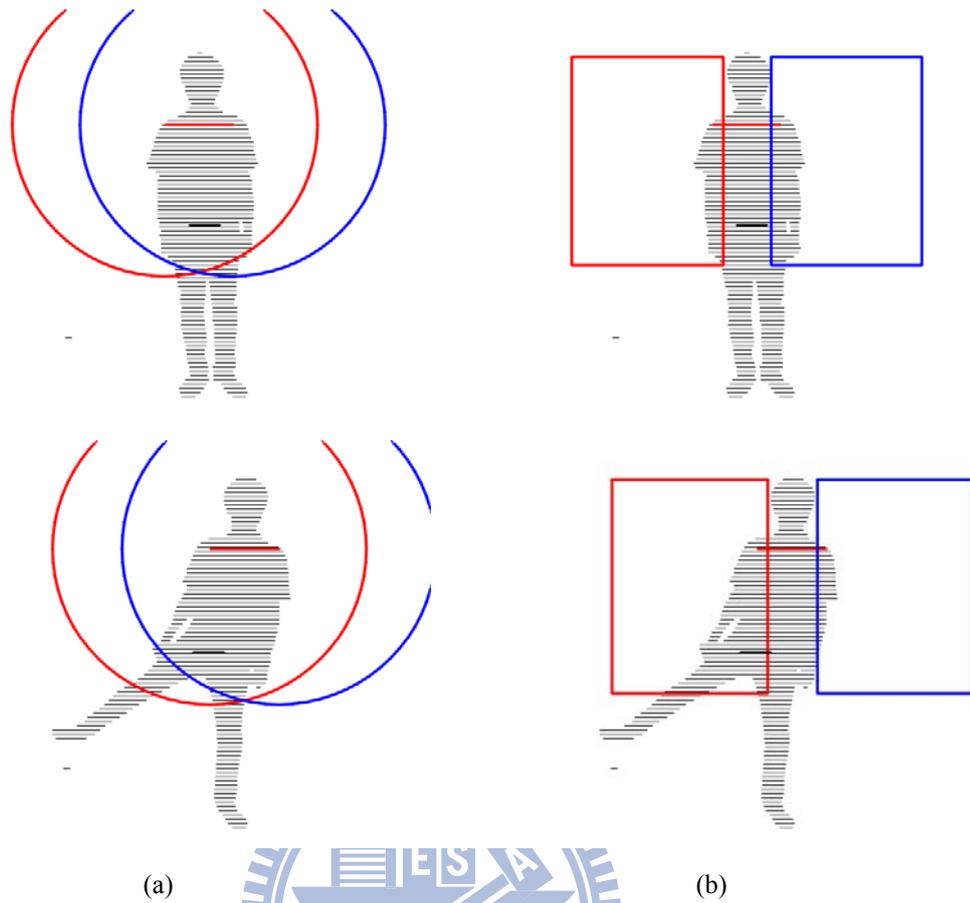


圖 4.9 手部的偵測範圍，分別以紅色、藍色內部區域，表示正面時左、右手的活動範圍，  
 (a) 可能的活動範圍，(b) 實際的偵測範圍。

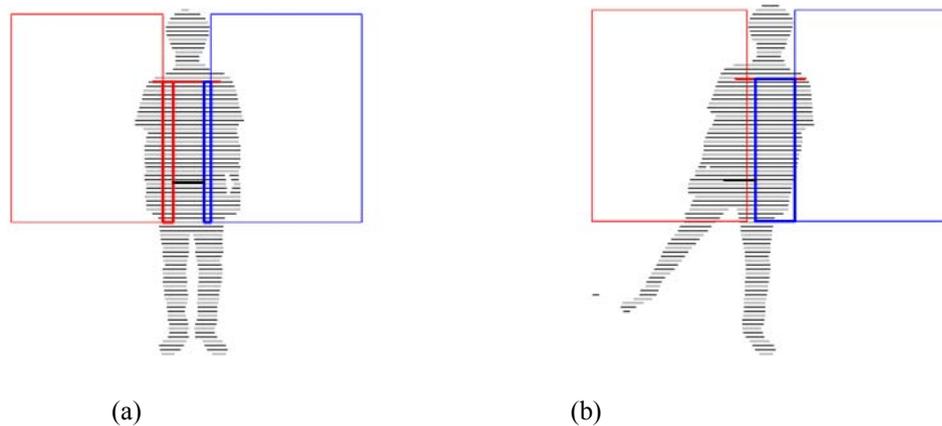


圖 4.10 擴增手部的偵測範圍，分別以紅色、藍色粗框矩形內部的區域表示，(a) 站立的例子，  
 (b) 腳部往外伸展的例子。

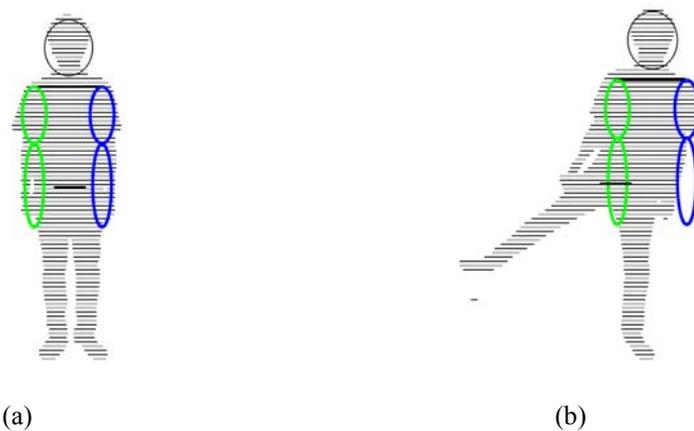


圖 4.11 以橢圓表示手部的初始位置，(a) 站立的例子，(b) 腳部往外伸展的例子。

本論文探討不同衣袖時，膚色近似中軸線段在手部偵測範圍內的分布，如圖 4.12 的短袖例子及圖 4.13 的長袖例子。由圖 4.12(a) 下方，我們發現膚色線段可能斷成多節，為了整合水平掃描線擷取到的膚色近似中軸線段(圖 4.12 及圖 4.13 中的以膚色線段表示)，與垂直掃描線擷取到的膚色近似中軸線段(圖 4.12 及圖 4.13 中的以紅色線段表示)。本系統計算這兩種顏色的近似中軸線段端點的到手肘初始位置的距離長度，我們定義距離最長的端點為手末端，距離最短的端點為手肘，可以得到圖 4.12(b)。

由上述計算，本系統可以得到手肘的位置。但如果手肘與肩膀在畫面上的連線(上臂)長度大於上臂最大長度( $\chi$ )，表示袖子較長，如圖 4.13(b)。我們必須調整手肘的位置，以手末端與肩膀的連線的中心位置為手肘的估測位置，如圖 4.13(b) 下方。由圖 4.13，我們看到赤腳所形成的膚色近似中軸線段，由於不在手部的活動範圍內，所以不影響手部末端的判定。由於我們是以前景物近似中軸線段的端點來辨識腳末端位置，所以膚色近似中軸線段亦不會影響腳部位置的判斷。

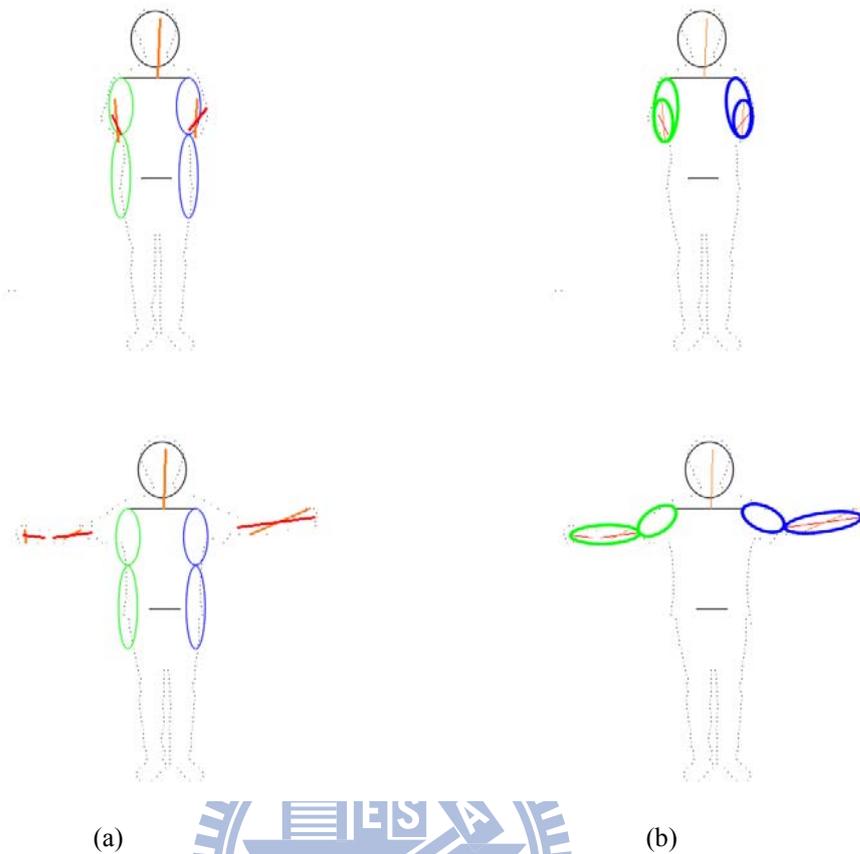


圖 4.12 以膚色近似中軸線段調整手部位置，短袖的例子，(a)膚色近似中軸線段的分布，  
(b)手部位置。

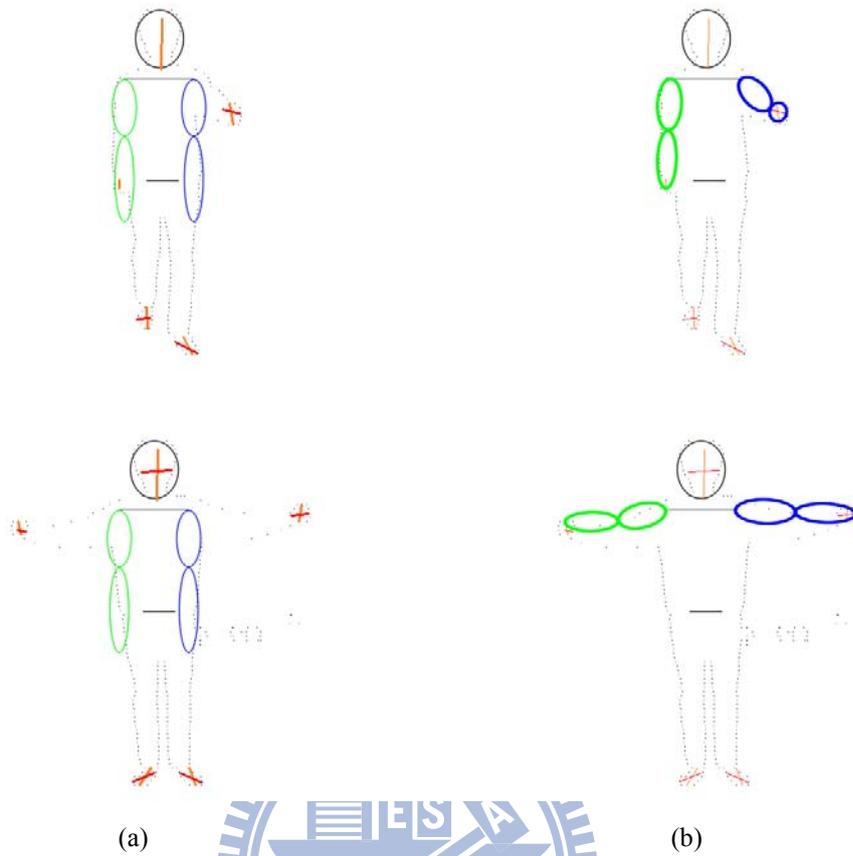


圖 4.13 以膚色近似中軸線段調整手部位置，長袖的例子，(a) 膚色近似中軸線段的分布，  
(b) 手部位置。

# 第五章 3D 人體模型建立

在現實中，人類可在三度空間中自由的活動，攝影機也可以任意角度自由擺設。本實驗主要探討將網路攝影機(webcam)的視線與受測者接近垂直擺設，並調整受測者距離鏡頭的長度為身高的 1.35 倍至 4.5 倍之間，使人體區域較容易全身入鏡不致超出畫面邊界而分析錯誤，也不致使人體區域面積過小而難分析。本系統即時的自動建立互動式 3D 人體模型，以方便在螢幕上觀察自身的 3D 姿態變化。

本章將在第一小節介紹如何計算前一章 2D 人體模型的關節景深值，重建 3D 人體模型；第二小節介紹以人體模型的 3D 資訊，建立虛擬的影像畫面。

## 5.1 關節景深資訊的估算

為了建立 3D 人體模型，需要進一步的取得 2D 人體模型關節在畫面上的景深資訊，我們參考[3]，以畫面上的關節座標來計算景深值。

$$l^2 = (X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2 \quad (5.1)$$

$$dZ = \sqrt{l^2 - ((X_1 - X_2)^2 + (Y_1 - Y_2)^2)} \quad (5.2)$$

其中， $l$  為肢體長度，可由空間中兩關節的 3D 座標計算得到。當得到肢體的長度後，並得知該肢體兩端關節在畫面上的 2D 座標，可由(5.2)式計算關節的景深值( $dZ$ )。由於 2D 影像畫面未能提供我們景深資訊，所以肢體的長度只有在  $Z_1$  及  $Z_2$  恰等於 0 時，才能經由(5.1)式得到，此時肢體恰與鏡頭視線垂直，且此長度在 2D 畫面上可能恰為最大長度值。隨著人體距鏡頭的遠近，肢體長度與景深值亦會有動態的長短變化，所以我們不能以畫面上兩關節間的最大長度值，作為該肢體的長度值。由此得知，當人體與鏡頭的距離變化時，肢體的長度亦要動態的更新。

有鑑於前述，由於一般人體是由固定長度的肢體所組成，且左右肢體長度有對稱的關係。我們參考[18]，推測人類肢體存在著比例關係，以比例關係做為估測肢體長度的依據，其人體側面姿勢如圖 5.1。

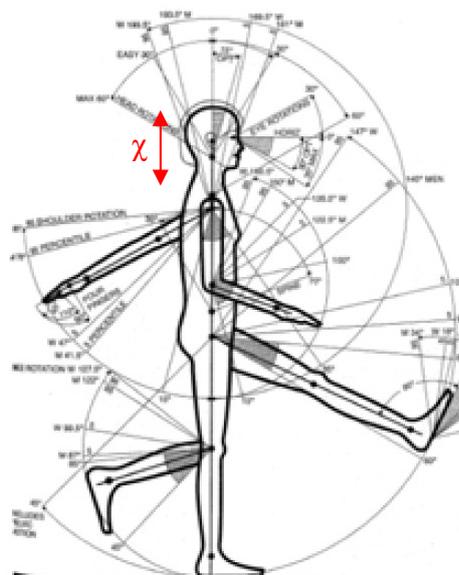


圖 5.1 人體側面的姿勢圖[18]。

由圖 5.1，我們觀察到人體頭頂到肩膀的長度  $\chi$  約為人類身高的 1/6 至 1/7，本論文將在第六章探討以此方法造成的誤差關係。長度  $\chi$  與其他肢體長度的關係如下

$$L_i = \alpha\chi \quad (5.3)$$

其中  $L_i$  為各個肢體的長度，亦是(5.1)式、(5.2)式中的  $l$ ， $\chi$  為頭頂到肩膀的長度。由圖 5.1 的人體側面圖，我們推測頭部長度約為  $0.9\chi$ ，頭部寬度約為  $0.8\chi$ ，軀幹長度約為  $2\chi$ ，肩寬約為  $1.2\chi$ ，髌骨寬約為  $0.9\chi$ ，肩膀到手肘(上臂)的長度及腳底板的長度約為  $\chi$ ，手肘到手末端(前臂)、髌骨到膝蓋(大腿)及膝蓋到腳踝(小腿)約為  $1.5\chi$ 。

表 5-1 肢體相對長度估測值( $L_i$ )

頭頂到肩膀	頭長、寬	肩寬	髖骨寬	上臂	前臂	大腿	小腿
$\chi$	$0.9\chi$ 、 $0.8\chi$	$1.2\chi$	$0.9\chi$	$\chi$	$1.5\chi$	$1.5\chi$	$1.5\chi$

由第四章建立的 2D 人體模型，得知關節在螢幕上的位置後，由前述的肢體相對長度估測值，再依基本的運動學概念(手肘一般情況下往身體後方曲折，腳踝一般狀況下往身體前方曲折)，可以由(5.2)式以下式計算得到當人體正面對鏡頭時的關節相對景深值，如圖 5.2。

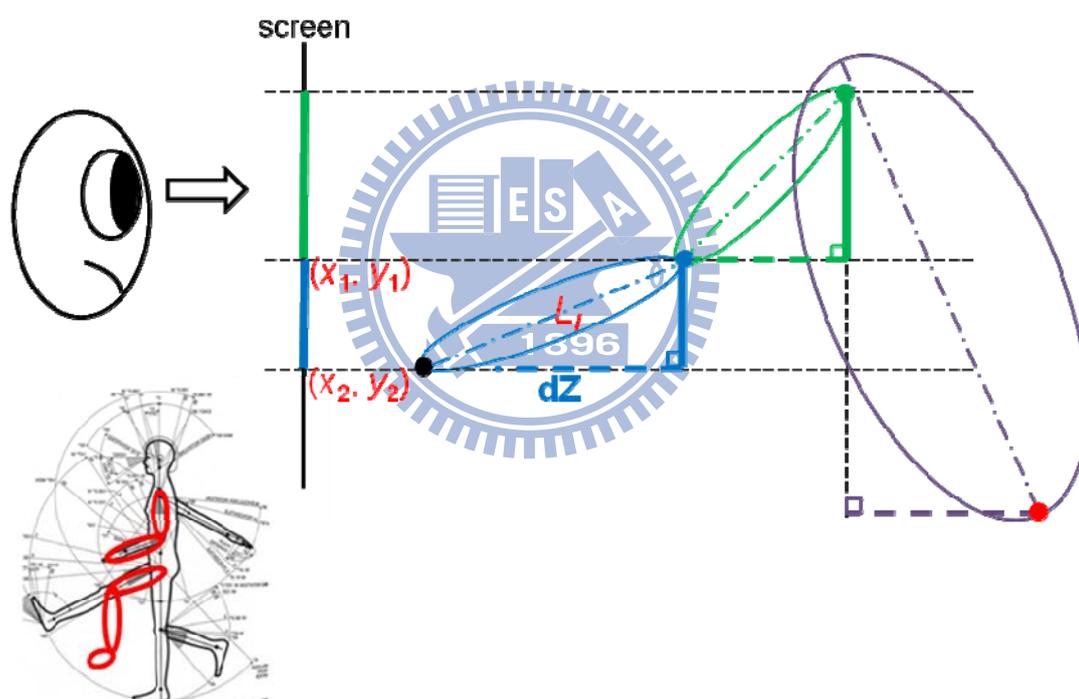


圖 5.2 手部關節的相對景深值計算。

## 5.2 虛擬視角畫面建立

本系統由前一小節計算得到的關節景深值，藉由座標轉換，可以得到虛擬視角的畫面，如圖 5.3。圖 5.3 中，藍色攝影機為實際的擺設位置，紅色攝影機為虛擬的擺設位置，在得知關節的景深資訊後，我們可經由座標轉換，建立俯視與

側視虛擬視角的 3D 人體模型畫面，如圖 5.4。

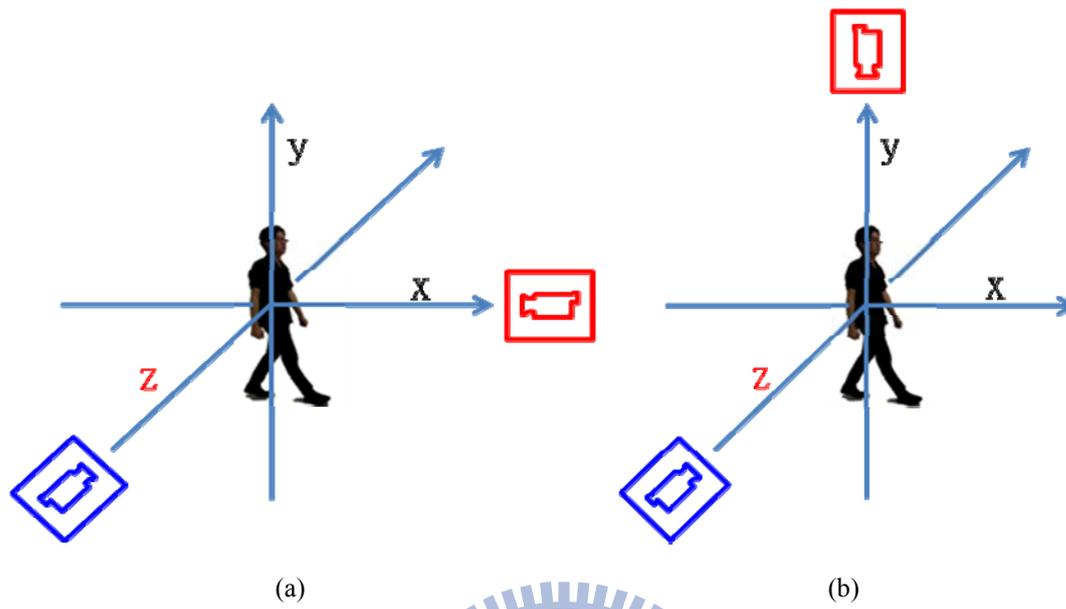


圖 5.3 虛擬攝影機示意圖，藍色為真實攝影機的擺設位置，紅色為虛擬攝影機的擺設位置，  
(a) 側視，(b) 俯視。



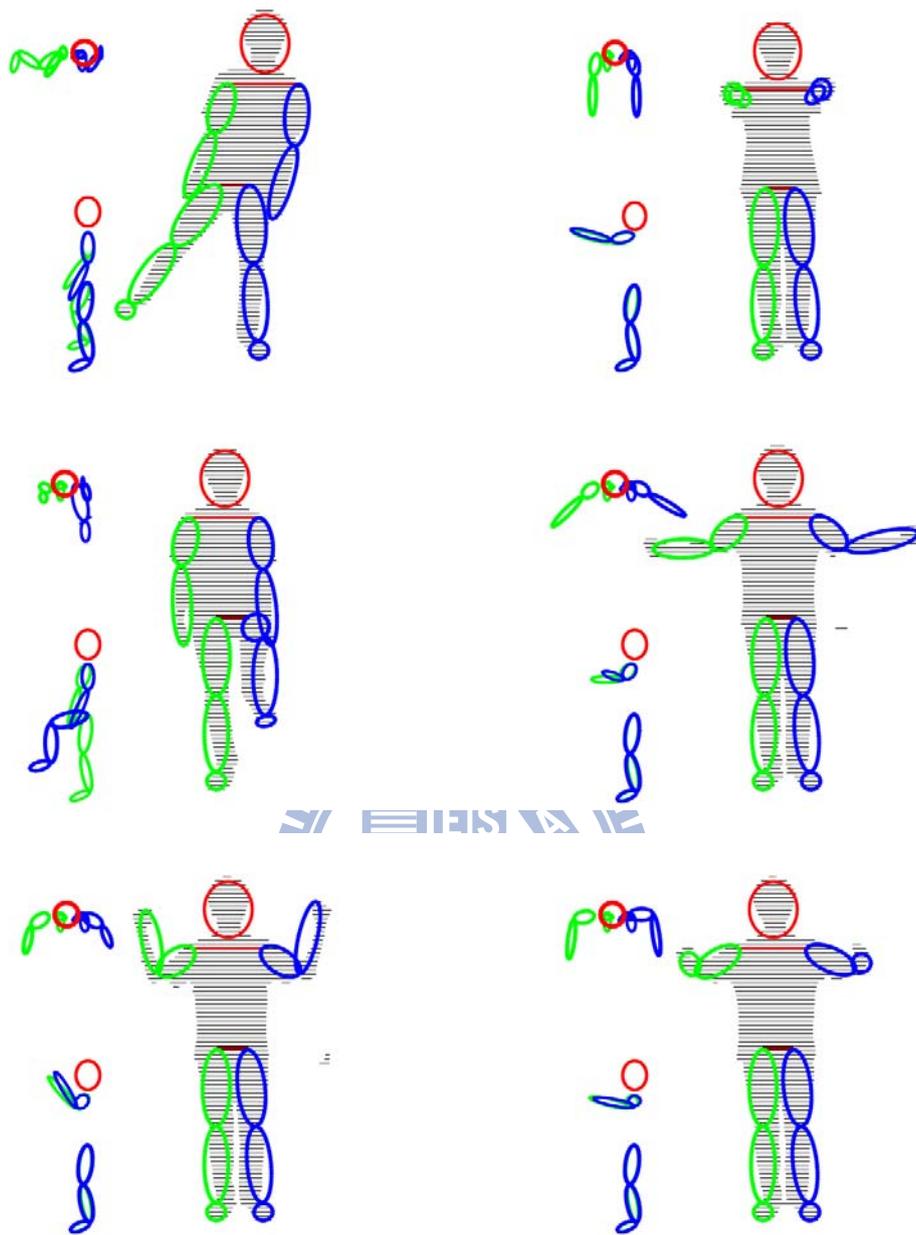


圖 5.4 以不同視角建立人體 3D 模型：(圖中左上)為俯視的 3D 人體模型，人體正面向下，  
 (圖中左下)為側視的 3D 人體模型，人體正面向左。

## 第六章 實驗結果與加速方法

本章節中，將依照前兩章所提到 3D 人體模型建立的方法，利用程式進行模擬實驗並分析實驗結果。本系統使用的程式語言為 C++，程式作業環境為 Microsoft Visual Studio 2005，測試電腦環境為 Pentium 4 3 GHz，2G RAM，作業系統為 XP 及 Vista，分別以影像品質較好的攝影機與筆電的網路攝影機，在單純及複雜的背景環境下，對使用者進行 3D 姿態的偵測，以俯視及側視的虛擬視角重建 3D 人體模型，來顯示系統辨識的正確性並對結果作分析。

### 6.1 關節對應的結果探討

我們在不同環境下對人體姿態做偵測，由於只對前景物上的膚色做偵測，在影像品質穩定的情況下，背景的膚色區域不會影響肢體的判斷。如果背景的雜訊為較大而未能被濾除，且偵測為膚色，亦在手部的偵測範圍內，此狀況則會造成短暫的手部辨識錯誤，如圖 6.3。另一方面，由於本系統是以前景物的近似中軸線段來調整腳部姿態，所以在腳部形成的膚色近似中軸線段，像是短褲，不會影響腳部的判定，如圖 6.1(a)、(b)。

一般成人為 7 到 9 頭身，若以頭部長度為參考長度，來計算肢體的相對長度，其誤差較大。所以本系統以畫面上頭頂到肩膀距離的長度  $\chi$  (約為身高的 1/6 至 1/7) 為參考，來動態調整肢體長度，此方法會因為不同受測者人體比例的不同，而產生誤差約為身高的 0 倍至 1/42 倍。若畫面的高度為  $H$ ，在受測者全身入鏡時，由於距離鏡頭的距離變化，會使在畫面上的人體身高為 0 至  $H$ 。本系統當受測者在畫面上身高為  $0.3H$  (如圖 6.8) 至  $0.9H$  時，若對不同人固定以身高的 1/6 定義  $\chi$ ，肩膀的高度與其它肢體長度的誤差為 0 至  $0.0216H$ ；如果受測者真實身高範圍為 100 至 250 公分內，則實際肩膀的高度與其它肢體長度的誤差為 0 至 5.95 公分。這個誤差值，並不會影響像是擺動手部、拳擊、抬腳、原地走路、原地跑

步等姿態的辨識。我們在系統介面的輸入影像畫面中，即時於左上方重建俯視視角的 3D 人體模型，左下方重建側視視角的 3D 人體模型。

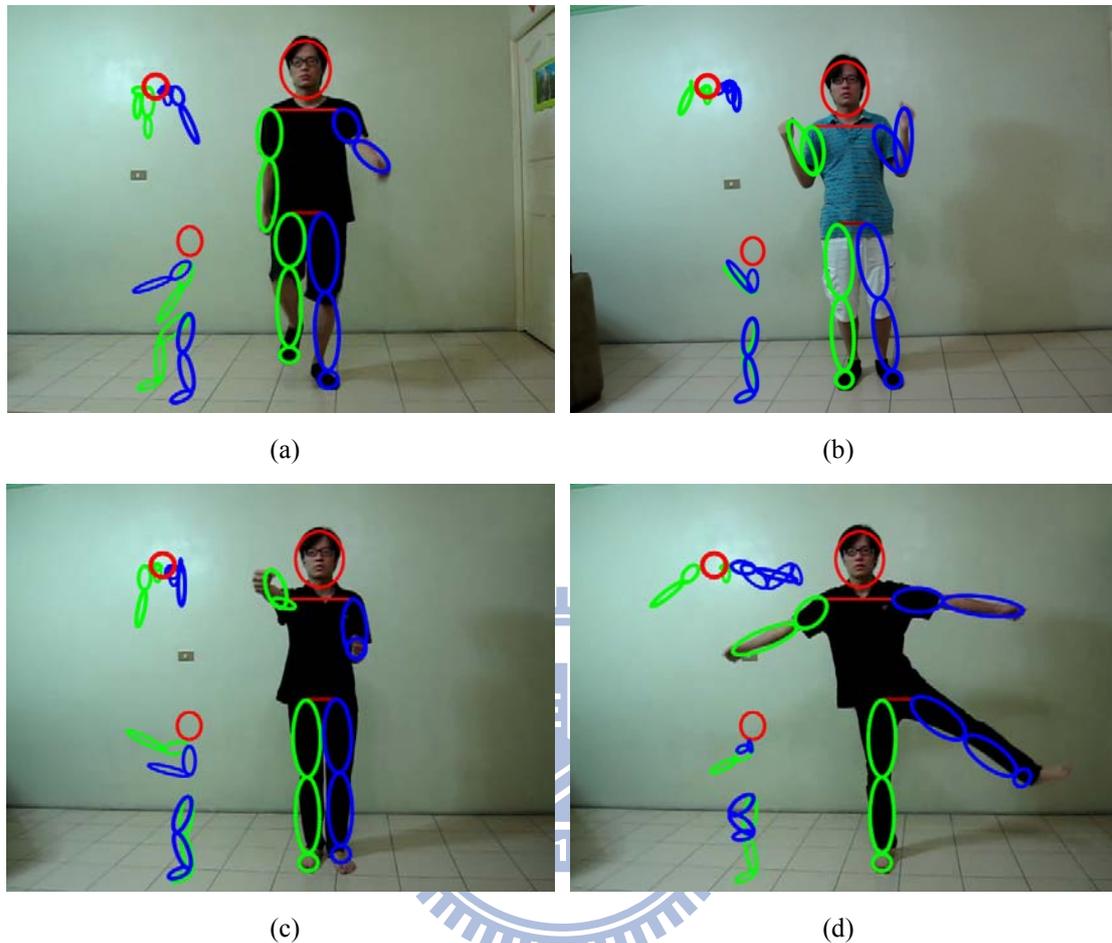


圖 6.1 短袖的偵測結果。

圖 6.1 中，由於腳部位置的偵測是基於前景物的近似中軸線段做判定，因此受測者褲管的長短，所造成的腳部膚色近似中軸線段，不會影響系統對腳部的判定。如果前景物與背景顏色接近，會使得前景物截面線段破損嚴重，造成近似中軸線段形狀過扁或過小而被濾除，此狀況下會發生錯誤的偵測，如圖 6.2 腳部區域。此時我們可適度調低前背景色度扭曲門檻值( $\tau_{CD}$ )，增加擷取的前景物像素面積。

在手部的活動範圍內，以膚色近似中軸線段的端點，來定位手末端及手肘位置。若雜訊造成過大膚色前景物而未能被濾除，且在手部的活動偵測範圍內，會造成手部偵測錯誤的狀況，如圖 6.3(a)。若未能手部的偵測出膚色近似中軸線段，

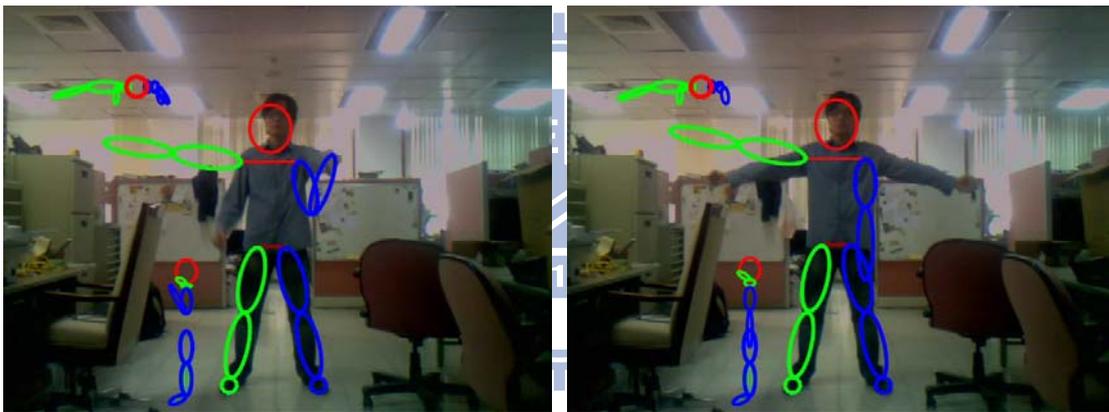
本系統以預設手部姿態表示，如圖 6.3(b)。未來在不同環境下的對膚色在色彩空間中的分布作學習，期望可以使膚色辨識率提升。



(a)

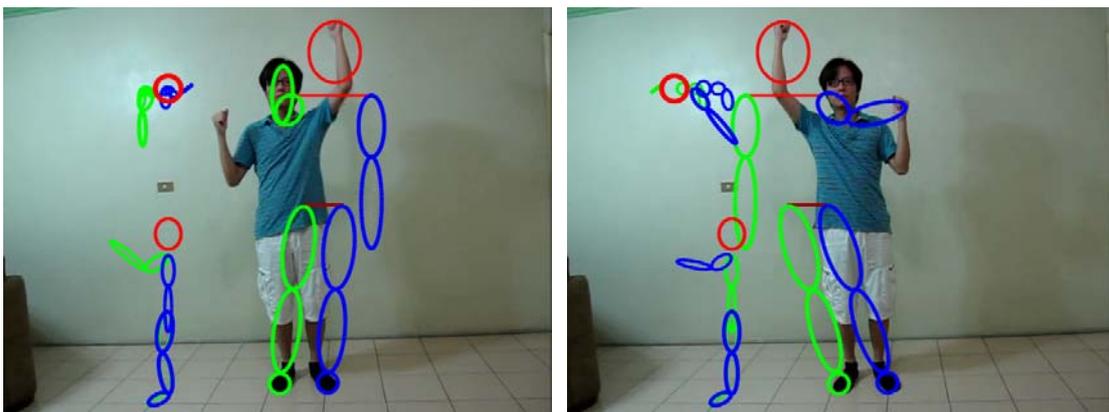
(b)

圖 6.2 腳末端顏色接近背景顏色，造成腳部擷取過短的狀況，(a) 拖鞋與背景顏色接近的例子，(b) 赤腳的例子。



(a)

(b)



(c)

(d)

圖 6.3 手部偵測錯誤的狀況，(a) 雜訊造成過大的膚色前景物，且恰在手部的活動範圍內，使右手的末端偵測錯誤，(b) 右手同(a)，左手因為偵測出膚色，而以預設姿態表示，(c)、(d) 手舉過頭的情況。

假設頭部在最高位置，則以水平掃描線掃描到的第一個跨越 3 條掃描線以上的近似中軸線段的起點，即當作頭頂的位置，如圖 6.3(c)、(d)。未來可以加入時間因素，以前幾張關節位置的變化量，來預測頭部及其它關節的位置。另一種作法，可以依近似中軸線段平均截面線段的寬度，來分析頭部或手部位置。但是由於背景複雜或顏色與前景物近似的偵測環境下，會使得頭部、手部人體區域破損

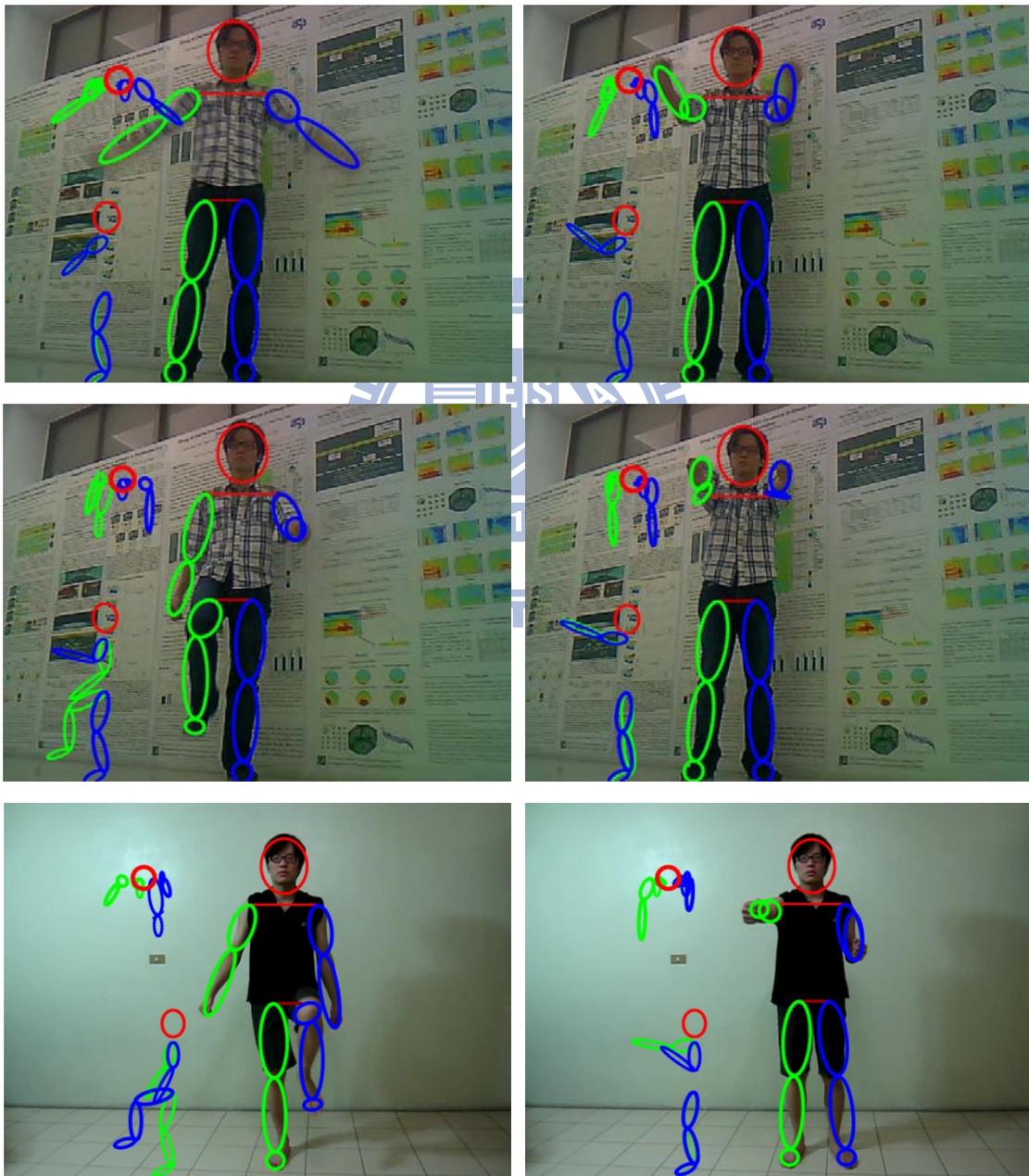


圖 6.4 膚色區域、衣著接近背景顏色的偵測結果。

嚴重，很難以破損的近似中軸線段來辨識頭部或手部，這時必須先整合破損的近似中軸線段後，再以分布區域面積的大小，來分辨頭部或手部。

若單純以輸入影像與背景模型的像素 RGB 絕對值差的和來取得前景物，會使膚色區域或衣著，在與背景顏色接近時，難以分離前背景區域。本系統採[11]的方法，可以得到較佳的前背景分割，以建立較佳的 3D 人體模型，如圖 6.4。

我們進一步以不同顏色的衣著，在背景較凌亂的場景中做測試，其測試結果如圖 6.5、圖 6.6、圖 6.7。由前一章所述，本系統自動依人體在畫面中的身高，動態調整肢體的長度，肢體的景深值亦產生相對的變化，如圖 6.5。當腳部部分重疊時，由於前景物截面線段的寬度變化，只要不為同一近似中軸線段，系統可以建立此兩腳的末端資訊，如圖 6.6。



(a) (b)  
圖 6.5 動態調整肢體長度，(a) 距鏡頭較近，(b) 距鏡頭較遠。



圖 6.6 腳部部分重疊的偵測結果。

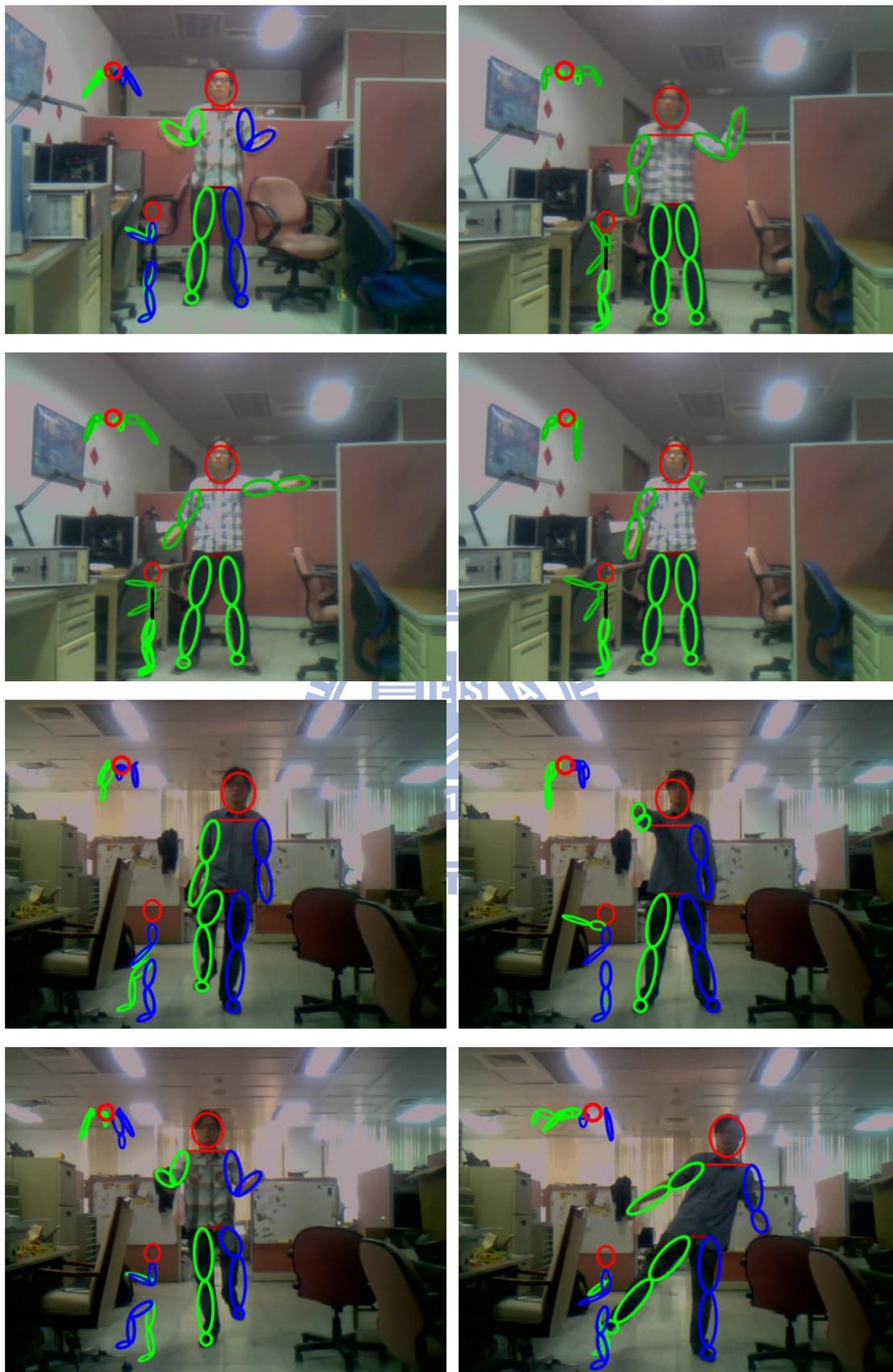


圖 6.7 複雜場景中的偵測結果。

由於本論文是依據人體區域輪廓及人體比例來選取頭部位置，因此在人體背向螢幕時，亦可建立 3D 人體模型，但是人體方向會誤判成是正面，如圖 6.8。未來我們可以進一步的分析膚色在頭部區域分布的面積大小，來辨識受測者為正面或背面，甚至進一步辨識畫面中的人臉方向，但是此方法會因為頭部遮蔽物(口罩、墨鏡等)或光頭(人臉正、背面的膚色面積差不多大)，使得此方法在某些情況下不適用。當受測者距離鏡頭漸遠時，我們很難以人體區域外部輪廓判寬窄的變化，來選取頭部的的位置，此時本系統依人體比例動態調整肢體相對長度，在距離鏡頭甚遠時，亦可計算畫面中人體關節的相對景深值資訊，如圖 6.9。

本論文進一步探討軀幹往前曲折的狀況。當人體區域的立足點不變，表示受測者沒有往鏡頭視線的前或後方移動，因此不調整其它比肢體(手、腳部)與  $\chi$  的比例，此時若外接矩形的高度變小，表示受測者做原地蹲下的動作。如果頭部與

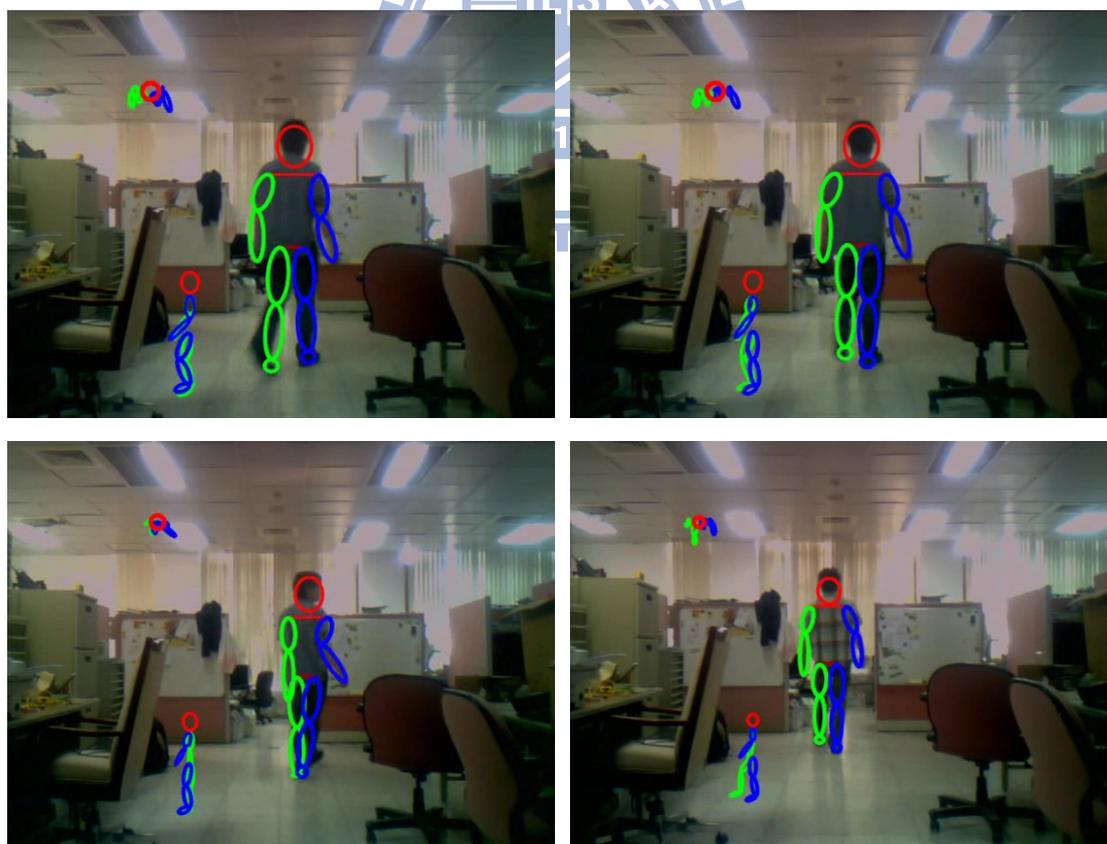


圖 6.8 背面的偵測結果。

身體互相遮蔽，會造成頭頂到畫面上肩膀的距離縮短。我們可以由頭頂到畫面上肩膀的距離長度  $\chi$  的變化，動態調整軀幹在畫面上的長度 ( $2\chi$ )，藉此估測軀幹在螢幕上的長度，來計算軀幹的景深值，如圖 6.10。受測者在往下蹲時，如果身體曲折，可由前述的方法，如圖 6.10 所示；如果身體不曲折，則軀幹在螢幕上的長度不變，只在畫面上往下移動。此方法的限制為衣著顏色明顯與背景有差異，才能在頭部下方找到畫面上肩膀的位置。

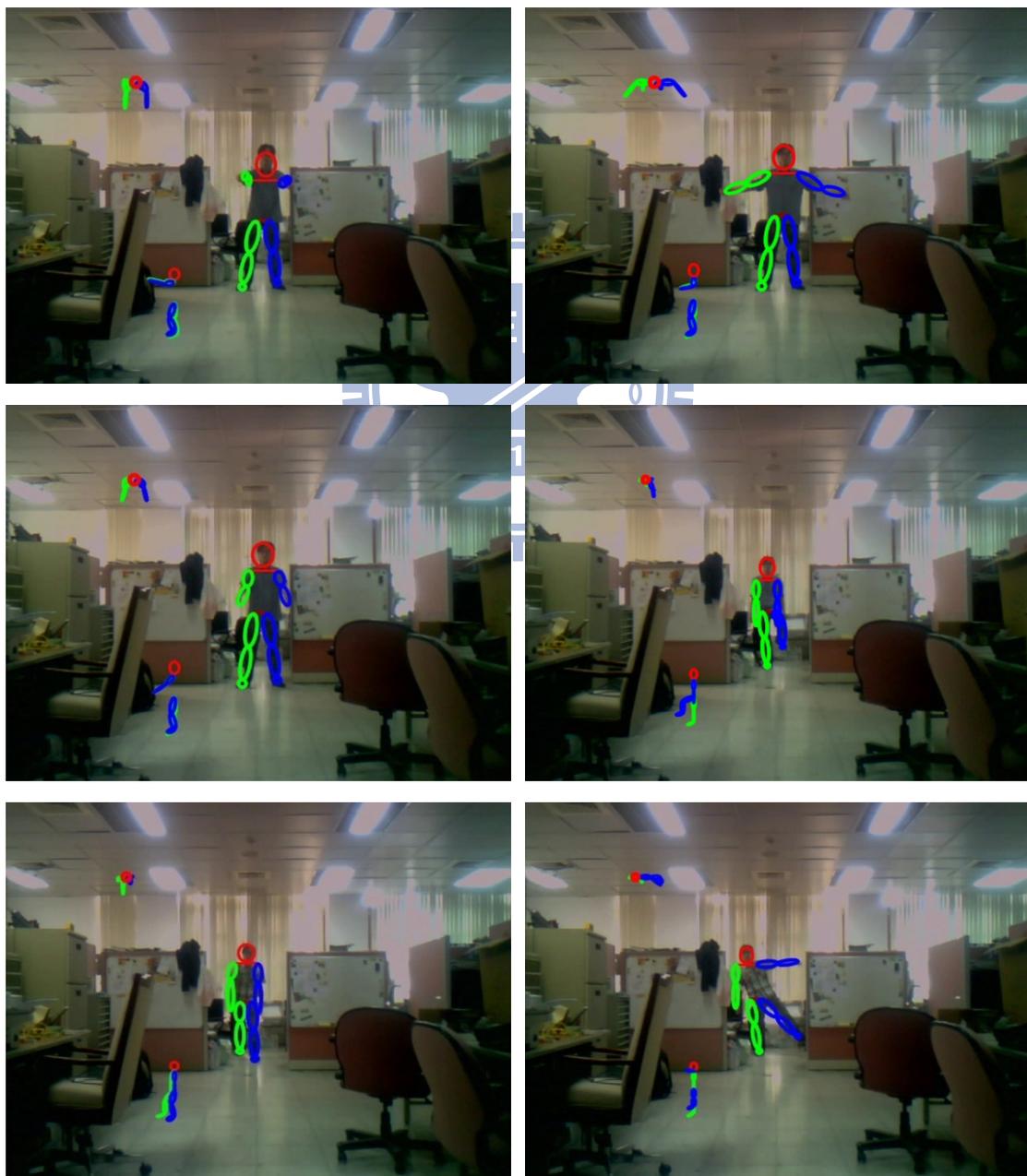


圖 6.9 受測者距鏡頭較遠的偵測結果。

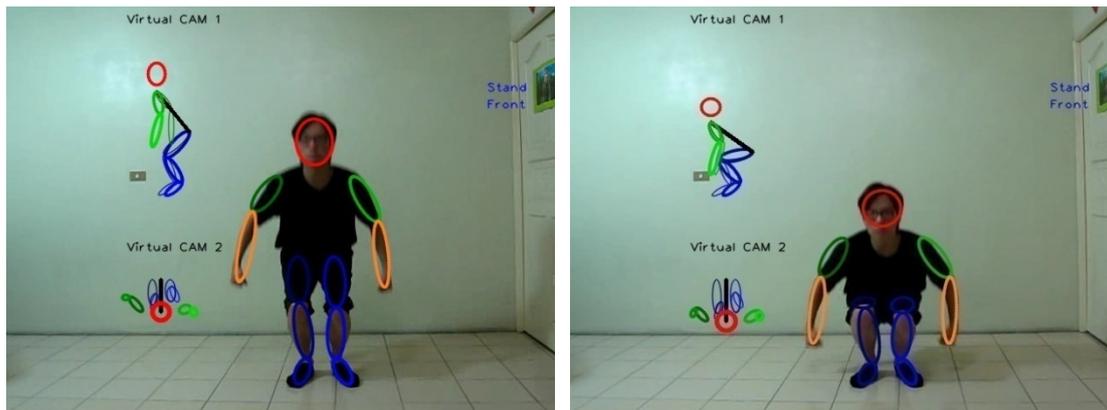


圖 6.10 受測者蹲下的偵測結果。

## 6.2 加速方法

由第三章介紹的掃描線演算法，我們發現對畫面中的每個像素做前景物偵測，其運算量較大。本系統使用類似降低解析度的概念，調整掃描線的間距以減少偵測的資料量，以提升系統運算速度。

為了得知每條掃描線上前景物截面線段的端點位置，除了完整的偵測掃描線上所有像素外，可再衍生降低解析度的概念，來減少偵測的像素資料量。

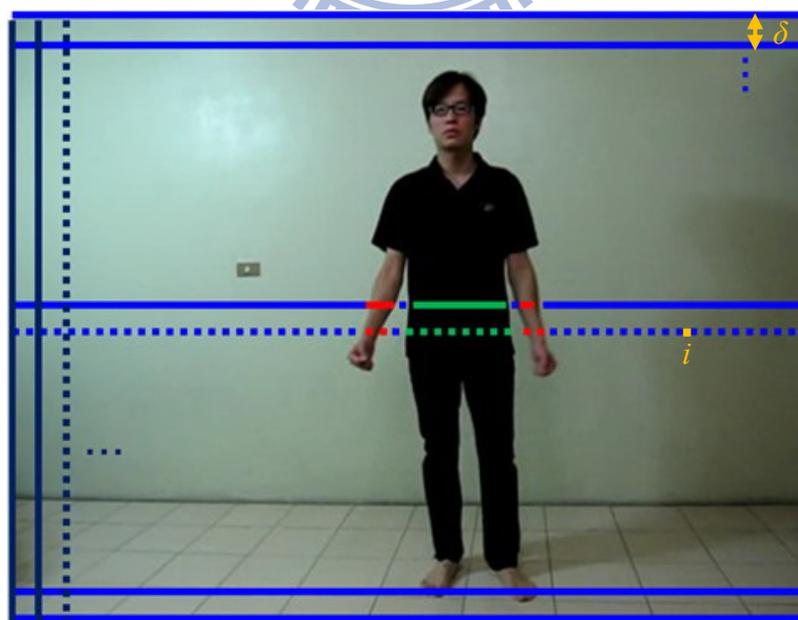


圖 6.11 以降低解析度的概念，減少偵測的像素資料量，圖中淺、深藍色的實線，分別表示一般的水平、垂直掃描線所需偵測的像素；淺、深藍色的虛線，分別表示以  $i$  為掃描線上像素的偵測間距；紅色線段表示偵測到的前景物截面線段。

6.11 圖中，減少後所需偵測的像素資料量與原本需要偵測的像素資料量關係是如下， $d$  為偵測的像素個數， $\delta$  為掃描線的間距， $x$ 、 $y$  分別為畫面寬與高的像素個數， $i$  為掃描線上偵測的像素間距。

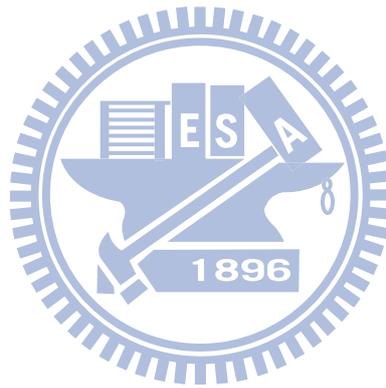
$$d = \frac{1}{(\delta + 1) \times (i + 1)} xy \quad (6.1)$$

若間距  $i$  等於 0，表示對掃描線上所有的像素作前景物偵測，資訊較完整，速度較慢；當  $i$  越大，表示須偵測的像素越少，速度越快，但是在影像品質較差的攝影機畫面下，容易因擷取資料量的不完整，而造成人體區域破損更嚴重。雖然第三章中提到掃描線上前景物截面線段破損的彌補方法，但是在前景物破損越嚴重時，需要調大門檻值才能予以彌補其破損區域，這會造成其它應被濾除的雜訊殘留下來，甚至形成多餘的近似中軸線段，使肢體的判別產生錯誤。我們實驗的畫面解析度為 640×480，本系統取  $\delta$  為 5，若影像品質較好時，系統再取  $i$  為 1，如此使系統在偵測前景物的執行時間增快了 12 倍。系統擷取到的近似中軸線段後，再往近似中軸線段的兩端點連線方向，在畫面上擷取較正確的肢體末端位置，可以較快速地由粗而細擷取肢體末端位置。

本論文的實驗，如果在第  $i$  張畫面中偵測不到某些關節的位置，或是第  $i$  張畫面與第  $i-1$  張畫面的關節位置有不合理的差距變化時(一般情況下，關節瞬間移動量不大)，我們可加入時間因素考量。系統會依前幾張畫面的關節移動量變化，來預測第  $i$  張畫面的關節位置。由於肢體可以恣意的擺動，上述有時可能造成預測錯誤且較耗時。未來我們可以更進一步探討，如何有效的預測關節移動量。

表 6-1 影像來源、掃描線方向、肢體表示及辨識比較

	[15]	[16]	[17]	本論文
影像來源	灰階			彩色
掃描線方向	0°、45°、90°、145°			0°、90°
肢體表示	2D			3D
肢體辨識	無	無	有	有
部分遮蔽	不可	不可	不可	可
頭部位置	未限定	未限定	未限定	最高處



## 第七章 結論及未來工作

本篇論文中，我們將[15]、[16]、[17]研究中擷取近似中軸線段的條件放寬，使用水平( $0^\circ$ )及垂直( $90^\circ$ )的掃描線及膚色資訊，來建立適用於辨識人體姿態的前景物及膚色近似中軸線段，再藉由人體比例及基本運動學的概念，重建受測者的3D姿態人體模型。我們以腳部的外接矩形來偵測腳末端，並調整髌骨位置，使姿態較符合真實情況。系統會依人體比例來動態調整錯誤選取的頭部大小，並推算其它肢體的相對長度，此方法當使用者距離鏡頭甚遠時，亦有機會偵測使用者的3D姿態。由本論文提出的方法，在正身面對鏡頭時的正確率較高，因此適合應用在人機互動的操作上，而不需使用控制感應器即可操作簡單的指令。經實驗結果得知，在複雜的環境下或衣著與背景顏色相似的情況下，皆可偵測使用者的3D姿態。本系統使用畫面品質較好、影像擷取速度較快的攝影機作測試，速度可以達到20至24fps，其特定姿態(手部擺動、拳擊、抬腳、原地踏步、原地跑步等)的辨識率可達85%以上；畫面品質較差的攝影機，如筆電的網路攝影機，則會因影像的擷取速度較慢，降至8至15fps，其特定姿態的辨識率可達80%以上。

未來相關研究可探討人臉及身體在影像畫面中的方向，並進一步在不同環境光下對膚色區域做學習，以提高前景物上膚色區域的辨識率。另一方面，當受測者手舉高過頭的情況下，我們可依近似中軸線段的分布及平均截面的寬度，加以分析頭部及手部的確切位置，以建立更準確的3D人體模型。如果有兩個以上的受測者全身進入畫面中，在受測者不互相遮蔽的情況下，預期可以辨識多個受測者的3D姿態。相關研究未來可以融合具有感測景深的攝影機(ZCam)資訊，或在偵測環境中架設多台攝影機，來辨識更複雜的3D人體姿態，但所使用的軟、硬體設備較昂貴，該系統的效能及辨識率亦有待考量。

## 參考文獻

- [1] N. Thome, D. Merad and S. Miguet, “Human Body Part Labeling and Tracking using Graph Matching Theory”, *IEEE Int’l Conf. on Advanced Video and Signal based Surveillance*, pp 38-46, 2006.
- [2] D. Weinland, R. Rongard, E. Boyer, “Automatic Discovery of Action Taxonomies from Multiple Views,” *IEEE Conf.on Computer Vision and Pattern Recognition*, pp. 1639-1645, 2006.
- [3] D. Gavrilu and L. Davis, “3D Model-based Tracking of Humans in Action: A Multiview Approach,” *IEEE Computer Vision and Pattern Recognition*, pp. 73-80, 1996
- [4] C. Taylor. “Reconstruction of Articulated Objects from Point Correspondences in a Single Uncalibrated Image,” *IEEE Computer Vision and Pattern Recognition*, pp. 677-684, 2000
- [5] K. Onishi, T. Takiguchi and Y. Ariki, “3D Human Posture Estimation Using the HOG Features from Monocular Image,” *IEEE Pattern Recognition*, pp. 1-4, 2008.
- [6] A. Agarwal and B. Triggs, “3D Human Pose from Silhouettes by Relevance Vector Regression,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 882-888, 2004.
- [7] G. Mori and J. Malik, “Estimating Human Body Configurations Using Shape Context Matching,” *Proceeding of European Conference on Computer Vision*, vol. 3, pp. 666-680, 2002.
- [8] M. Lee and I. Cohen, “A Model-Based Approach for Estimating Human 3D Poses in Static Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 905-916, June 2006.
- [9] C. Yu, H. Cheng, J. Hwang, and K. Fan, “Human Body Modeling with Partial Self Occlusion from Monocular Camera” *IEEE Machine Learning for Signal Processing*, pp.193-198, 2008
- [10] L. Zhu, J. Hwang, C. Chen, M. Lin, and C. Yen, “Real-Time 3D Pose Reconstruction of Human Body from Monocular Video Sequences,” *IEEE Int’l Symposium on Circuits and Systems*, pp. 717-721, Taipei, Taiwan, May 2009
- [11] H. Horprasert, D. Harwood, and L.S. Davis, “A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection”, *IEEE International Conference on Computer Vision, Frame-Rate Workshop*, pp.1-19, 1999.
- [12] C. Garcia and G. Tziritas, “Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis,” *IEEE Transactions on Multimedia*, vol.

- 1, no. 3, pp. 264-277, 1999.
- [13] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving Shadow Suppression in Moving Object Detection with HSV Color Information," *Proceedings of IEEE Intelligent Transportation Systems*, pp. 334-339, Aug. 2001.
- [14] J. Kao, J. Chuang, and T. Wang, "Chromosome Classification Based on the Band Profile Similarity Based on the Band Profile Similarity along Approximate Medial Axis," *Pattern Recognition*, vol. 41, pp. 77-89, Jan. 2008.
- [15] J. Chuang, C. Lee, and K. Lo, "Human Activity Analysis Based on a Torso-less Representation," *19<sup>th</sup> International Conference on Pattern Recognition*, pp. 1-4, Dec. 2008.
- [16] N. Chu, "A Torso-less Representation of Human Activity with Multi-joint Trapezoids," Master's Thesis, Dept. Computer Science and Engineering, National Chiao Tung University, June 2008.
- [17] S. Hunag, "Human Posture Recognition System Based on Image Scanning," Master's Thesis, Dept. Computer Science and Engineering, National Chiao Tung University, June 2009.
- [18] Tilley, *The measure of man and woman*, New York: John Wiley and Sons, 2002.

