

視覺式體操動作辨識系統

Vision-based Gymnastics Motion Recognition System

Department of Computer Science and Information Engineering

National Taiwan Normal University

指導教授：方瓊瑤老師

組員姓名：顏羽君

(一) 系統簡介

本計畫設計一套以電腦視覺為基礎偵測人體全身肢體動作之系統，將連續之體操運動影像輸入電腦後做肢體動作判斷辨識其動作種類或發出錯誤動作訊息。為提升生活中資訊多元化的運用並與日常生活結合，利用攝影機拍攝動態人體行為後將影像輸入一套偵測人體各類動作之系統將會有多元化之用途。此計劃將以一台攝影機拍攝單一人物之體操運動辨識其動作及判斷是否正確，未來可運用在體育教學上將教師舞蹈動作錄製後，學生在自我練習運動動作時可即時與教師影像做比對，修正其動作之正確性；另外未來若發展為辨識手語，則利用攝影機拍攝手語，藉由判斷後翻譯手語內容，可成為身障人士與人溝通之橋樑，協助其在遇到緊急狀況時即刻與人求救並說明情況；而休閒娛樂方面可運用在跳舞機之全身肢體偵測、Wii 的影像偵測方法。

由於本計畫由基礎之國民健康操體操動作開始處理，將個人體操連續影像輸入此系統後，將建立與影像相同之模型，再利用由建立之肢體動作模型進行動作分類，最後判斷動作之正確性，適時要求學員進行重複練習或給予嘉賞。

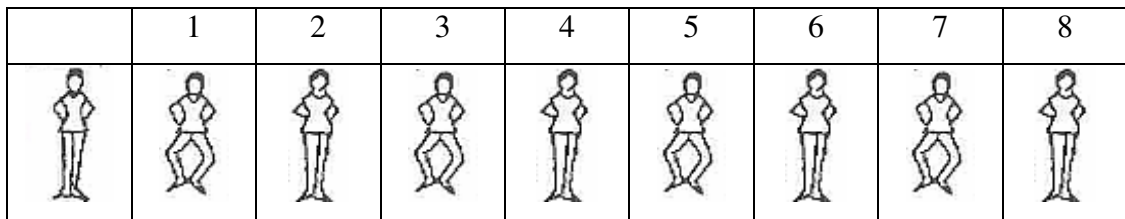
(二) 研究動機與研究問題

由於國立台灣師範大學早已設置體育系與競技系，平時觀察他們上課狀況後認為一套以辨識體操為出發點的肢體偵測系統將具備高度實用性，在體育教學方面，許多運動項目之動作都需要一定的標準程序才能成功，如網球手勢、體操...等等，若為此發展一套系統來偵測學生的肢體動作並給予適當建議，則可減少教師的負擔並幫助學生在家自學。最近蔚為流行之 Wii 電視遊樂器以具感知功能的控制桿作螢幕上人物動作之控制，但只限於有操作控制桿之肢體，因此容易受控

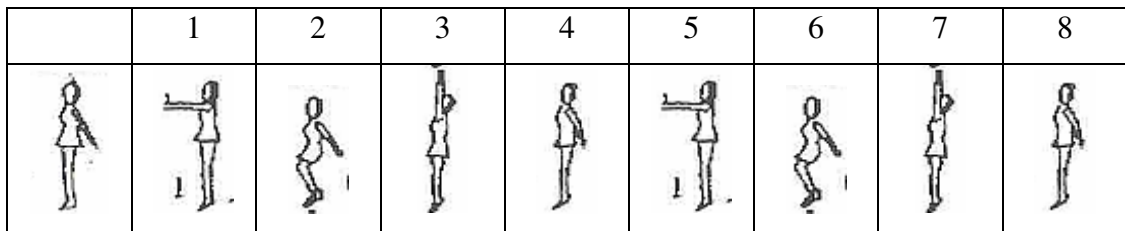
制桿數量的限制，若能結合電腦影像處理對全身肢體作辨識，更能夠增添其趣味性；坊間跳舞機的運用亦是相同道理，除了步伐正確性，根據全身肢體動作偵測與判定後，可給予全身舞蹈做適切評價。此計畫即為一可即時偵測體操運動正確性的系統，期許能在未來資訊與生活結合的世代能有更多元的日常生活應用。

此套系統將使用單一攝影機在室內達成辨識個人體操動作之功能，將連續彩色影像輸入此系統後可即時判斷出動作種類，雖然單一攝影機僅可做二維平面上之肢體動作，未來可藉由增加攝影機擴充進而辨識三維與跳躍之動作並可進行多人肢體動作辨識。此套系統擬辨識由六種國民健康操動作所組合之一套體操，其中動作共分六項，分別為下肢運動、上肢運動、胸部運動、左右彎體、四肢運動、轉體運動，而每項動作中又各細分為幾個姿勢，當一個動作從頭至尾所有姿勢之形狀與順序都無誤此動作才算正確。如圖一為本系統將辨識之六項國民健康體操分解動作[7]。

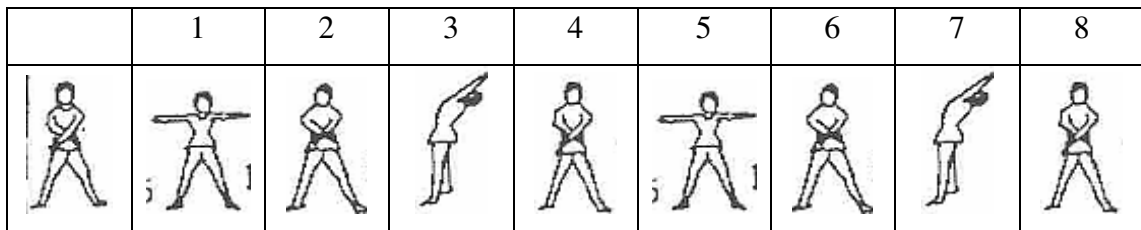
1. 下肢運動：



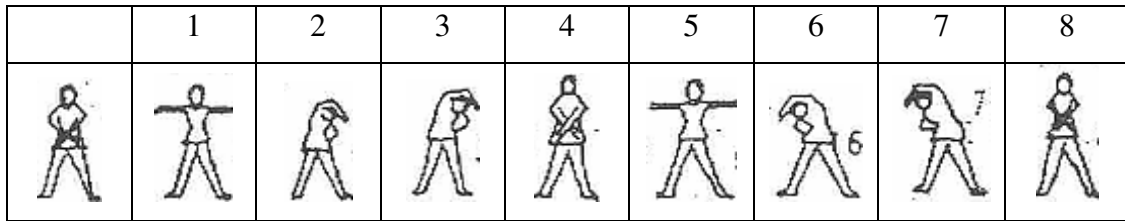
2. 上肢運動：



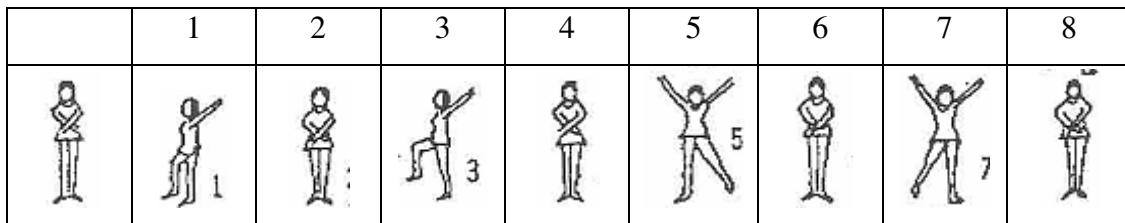
3. 胸部運動：



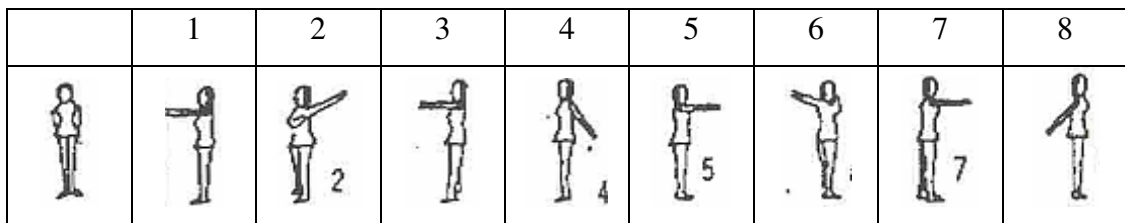
4. 左右彎體：



5. 四肢運動：



6. 轉體運動：



圖一、本系統欲辨識之六項國民健康體操動作分解示意圖，每個動作包含了一個起始姿勢，及八個拍子所各別代表的姿勢。

由於本計畫將會在室內進行攝影機架設，因此在影像處理與系統設計上，將會面臨許多需考慮的問題，以下為五項須面對之研究困難：

1. 光線因素：雖攝影地點在室內但室外光線投射入室內所造成之陰影變化、日光燈週期性閃動，以及人體體操過程中各部位造成之陰影都會造成研究上的困難。
2. 背景因素：當人物跳動影響地面引起得攝影機震動會產生背景的變化，背景的改變會造成影像校正困難。另外在體育場中其他人改變動作也會造成背景的改变。
3. 人物的衣著：人的姿態以及穿著的服裝會影響系統的辨識率，如果所穿著的服裝與背景顏色太過雷同，則會提高問題的難度。
4. 即時性：對於影像的分析處理，必須相當快速，否則將造成系統的實用性下降。
5. 由於人體所作之運動屬於形變物體(non-rigid object)的運動，有別於其餘固形物體(rigid objects)擁有不變的形體，要對形變物體進行動作偵測，亦為本計畫之研究困難。

(三)文獻回顧與探討

一般對影像的處理及分析，首先是利用「物體追蹤」(object tracking)的技術來追蹤動態物體，一旦我們追蹤到物體，就可以做進一步的肢體動作分析。而目前的物體追蹤是透過比對連續影像間物體的相似度來完成，於是物體特徵的擷取，相似程度的判別，目標物的搜尋，都涵蓋在此議題中。目前物體追蹤的方法可概分成四類[1]：區域式追蹤 (Region-Based Tracking)、主動式輪廓追蹤(Active Contour-Based Tracking)、特徵追蹤 (Feature-Based Tracking)、以及模型追蹤 (Model-Based Tracking)。

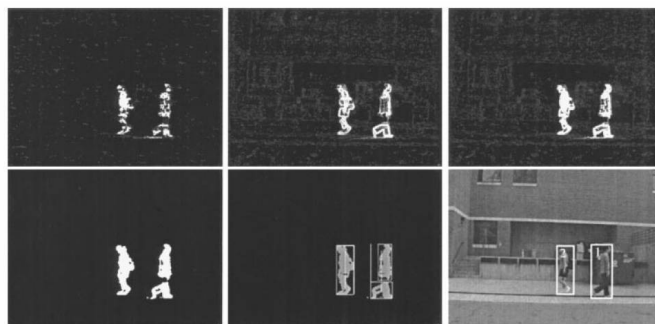
1.區域式追蹤 (Region-Based Tracking)

假設影像變動的區域即為目標物的位置，藉著偵測這些變動區域的位置來追蹤目標物。通常利用目前的影像與背景相減來偵測出變化的區域，然後再進一步的設立規則做篩選，合併，或劃分。McKenna 等人[2]設計了一種可調性背景相減法 (adaptive background subtraction)，針對人與人群做追蹤，其追蹤層級由小而大可分為三個層次：區塊 (region)、人 (people) 以及人群 (groups)，而每一個層次都可以再依照區塊色彩和表徵特性來進行合併或分解，可以達到追蹤單人或多人的效果。

此方法有其缺點，除了無法可靠地解決物體合併的問題，無法取得物體的 3D 動作外，對於雜亂有變動的背景亦不適用。圖一為兩人交錯所拍攝出支連續影像。圖二為使用背景消去法後產生目標物的方塊。



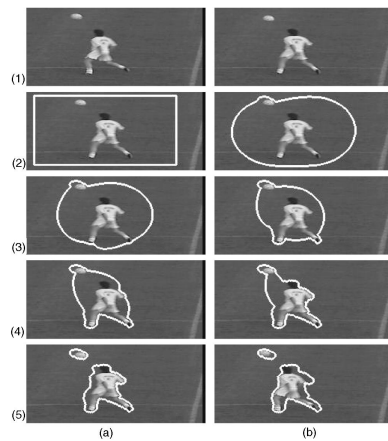
圖二、兩人交錯的連續影像[2]。



圖三、背景消去及產生目標物方塊[2]。

2.主動式輪廓追蹤(Active Contour-Based Tracking)

方法是用輪廓線 (contour) 來描述移動的物體，並利用輪廓線的改變進行追蹤，如[3]即使用這樣的方法。此方法的好處在於更有效率，且降低了複雜度，由於輪廓線是封閉曲線，即使物體交錯，也較容易解決。但其缺點仍然是缺乏 3D 的資訊，無法進行三度空間上的追蹤，且此演算法的追蹤效果對於初步輪廓的偵測或選取非常敏感，故較難用於全自動的偵測追蹤系統。圖四為足球場上主動式輪廓追蹤所呈現之結果。



圖四、足球場上的動作偵測。(1a)及(1b)為兩張連續的動作，輪廓線慢慢縮小的過程[3]。

3.特徵追蹤(Feature-Based Tracking)

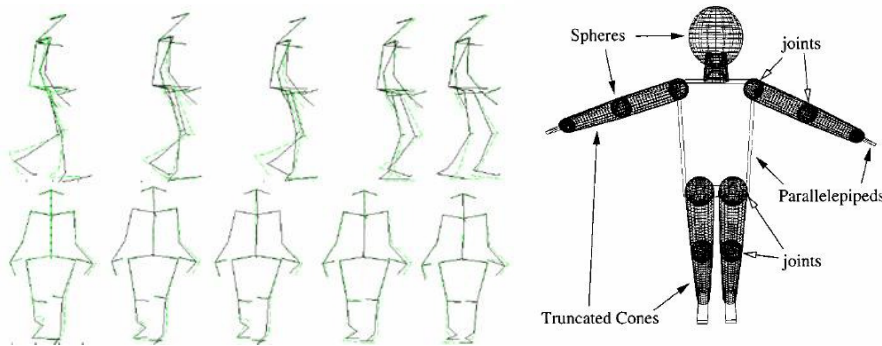
本方法是利用物體的特徵來進行追蹤，首先針對要追蹤的物體擷取特徵，這些特徵可分為三種：整體性特徵 (global feature-based)，如重心、色彩、面積等；局部性特徵 (local feature-based)，如線段、頂點等；相依圖形特徵 (dependence-graph-based)，如特徵間的結構變化等。當影像中目標物的低階特徵擷取出後，即可匯集成更高階的特徵資訊，利用比對連續影像間的高階特徵來追蹤該物體。

此方法可以用來即時追蹤多個物體(相依圖形特徵的方法除外)，利用運動特徵、局部特徵或相依的結構特徵來解決物體交錯的問題，但是使用運動特徵的方法穩定性不高，而使用局部特徵的方法則比較耗時費力。此演算法的另一缺點是，用 2D 影像對物體的辨識度並不高，且通常無法還原出物體的 3D 資訊。

4.模型追蹤(Model-Based Tracking)

模型追蹤的方法將會提供較精細與準確的判斷，因此需要良好的物體結構模型，可加入物體本身的運動特性，本質上，比較不受附近背景或其他物體干擾的影響，也較能對抗物體間交錯問題，但相對地運算量也比較大。通常模型追蹤的方式分成主要三步聚：(1)建立物體模型，(2)建立運動模型，(3)搜尋和比對。

(1)建立物體模型：以人為例，建立人形的姿勢結構模型，如棒狀圖形(stick figure)(圖五左)[4]所構成的人形。目前有愈來愈多研究偏向建立 3D 的 model，因為 2D 影像的主要缺點在於所得資訊受視角限制。Delamarre 等人[5]利用多台攝影機所得的剪影來建立 3D 模型(圖五右)。



圖五、左方為棒狀人形圖[4]，右方為 3D 人形圖[5]。

(2)建立運動模型。由於人體運動時四肢與關節的相對移動量不會很大，我們可以利用這些資訊來建立運動模型。建立 3D 的運動模型可以有效解決 2D 影像中物體間因為交錯而產生誤判的情況。

(3)搜尋和比對。目的為從連續影像中快速且正確找出物體位置，或是建立其他模型所需要蒐集的資訊。常用 Kalman filtering、Dynamical strategies、Taylor model strategies、Stochastic sampling 的技術。搜尋和比對的主要程序包括預測(predicting)物體的下一個狀態以縮小比對範圍，以及根據誤差更新(updated)系統的內部參數。針對比較複雜的情況下，使用 Levenberg-Marquardt algorithm[9]的技術來解決多參數的問題。另外，針對減少背景干擾的問題，Markov chain、Monte Carlo[10]、Genetic algorithms、condensation [11][12]等為常用的方法。

模型追蹤的方法與前三者方法相比，有幾項主要優點。由於此演算法從影像中所取得的即為 3D 資訊，故不需再多加處理。利用先前的 3D 資訊，可整合做為判斷下一個動作的根據，故即使物體互相交錯，也能準確地判斷。另外，此演算法在物體動作變化很大時仍可應用。

5. Using Temporal Templates

對於動態人體偵測部分，2001年Aaron F. Bobick等人另外提出了以temporal templates 做人類肢體動作辨識的方法，本系統將會採用 motion-energy images 以及 motion-history images做為比對的依據。

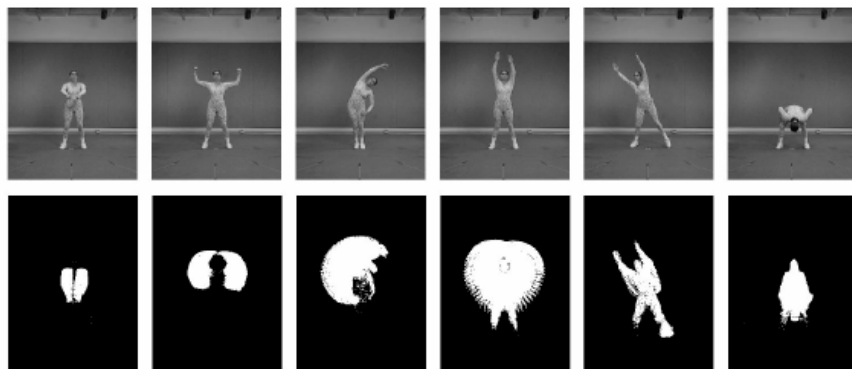
(1) Motion-Energy Images (MEI) :

以 $I(x,y,t)$ 表示連續影像序列， $D(x,y,t)$ 表示對應的 I 經過二值化後的連續影像序列，令 τ 為一段時間區間，下式即為時間 τ 之 MEI 擷取結果：

$$E_{\tau}(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t - i).$$

利用上述的式子可以將一段時間內，pixel value有變化的地方，即為動作發生的位置皆以白色表示，形成binary cumulative motion image，我們稱之為 motion-energy images。本系統將以MEI圖的比對做為肢體動作辨識的一部份。

下圖六即為六種不同體操動作所對應之MEI圖：



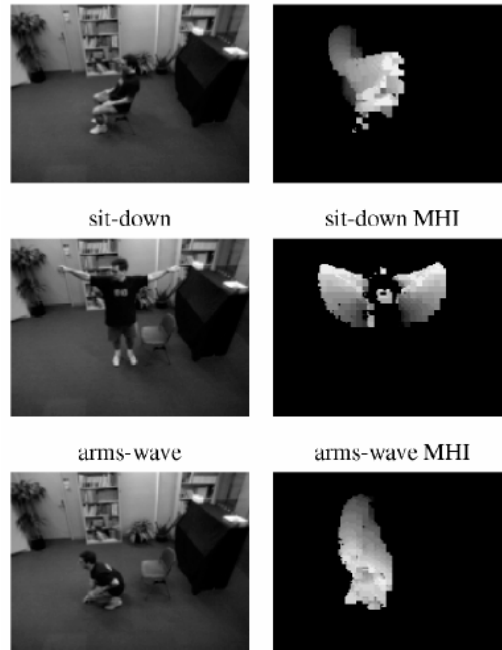
圖六、六種不同體操動作所對應之MEI圖[13]

(2) Motion-History Images (MHI) :

利用不同之 pixel value intensity 表示不同時間點動作變化的位置，其中 pixel value intensity 為 function $H(x, y, t)$ ，其值越高者表越靠近目前時間點動作的位置。 $H(x, y, t)$ 函式定義如下：

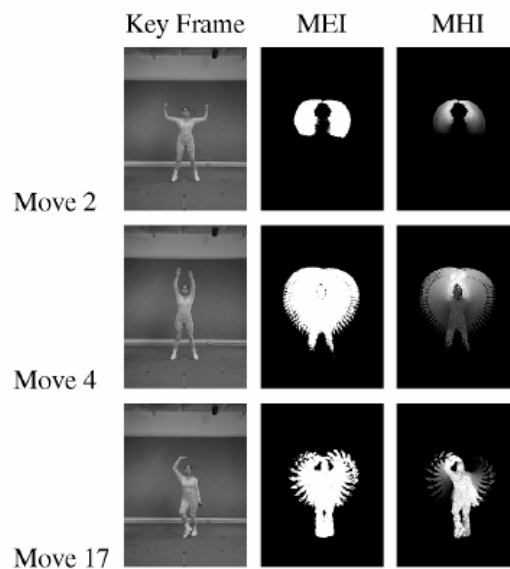
$$H_{\tau}(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_{\tau}(x, y, t - 1) - 1) & \text{otherwise.} \end{cases}$$

圖七為利用上式函式所計算出的三個不同動作所對應之 MHI 圖。[13]



圖七 三種不同動作所對應之 MHI 圖[13]

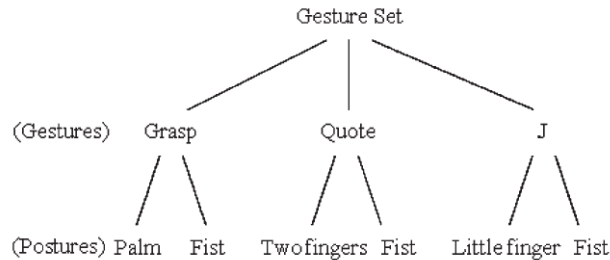
MEI 與 MHI 需要一起進行比對才可提升正確度，如圖八中，動作 2 與動作 4 之 MHI 相似度高，但觀察其 MEI 即可分辨兩者不同動作，而動作 4 與動作 17 之 MEI 相似度高，但經由 MHI 即可分辨兩不同動作。故取出兩圖同時進行比對是必要的。



圖八 不同動作所對應之 MEI 與 MHI 圖[13]

6. Stochastic context-free grammar(SCFG)

Stochastic context-free grammar 是 context-free grammar 的延伸，相較於 context-free grammar 來說更具彈性。經過一定數量的 training data 統計後，對每個 production rule 賦予一個機率的權重，使此方法更能有效精準的分析出正確的肢體動作。Qing Chen, Nicolas D. Georganas 等人在 Hand Gesture Recognition Using Haar-Like Features and a Stochastic Context-Free Grammar[14] 中運用 Stochastic Context-Free Grammar 的方法來辨識手部動作，將動作(gesture)細分成數個靜態的姿勢(posture)如下圖所示：



圖九、動作是由許多姿勢結合出的結果，此圖中包含三個動作及其各自包含的姿勢[13]

其SCFG 被定義為 $G_G = (V_{NG}, V_{TG}, P_G, S)$ ，其中 $V_{NG} = \{S\}$ ， $V_{TG} = \{t, p, l, f\}$ ，而 P_G 為

$$r_1 : S \xrightarrow{\frac{1}{4}} pt, \quad r_2 : S \xrightarrow{\frac{1}{2}} tf, \quad r_3 : S \xrightarrow{\frac{1}{4}} lf$$

其中的 t, p, f, l 分別代表各個姿勢的第一個英文字母， r_1, r_2, r_3 則為三條 production rules，機率 $P(r)$ 分別為 $1/4, 1/2, 1/4$ 。對於一個輸入串流 x ，其由任一條 production rule r 所推導出的機率，可以用下面兩公式算出

$$P(r \Rightarrow x) = D(z_r, x)P(r).$$

$$D(z_r, x) = \frac{\text{Count}(z_r \cap x)}{\text{Count}(z_r) + \text{Count}(x)}.$$

其中， Z_r 表示由 r 所推導出的字串，而 $D(z, x)$ 則表示輸入字串 x 與 Z_r 的相似程度，例如當收到輸入串流為 tt ，其 $D(z, tt)$ 的相似度及所推導出的機率分別如下：

$$D(z1, tt)=D(pt, tt)=1/(2+2)=1/4$$

$$P(r_1 \Rightarrow tt) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$$

$$D(z2, tt)=D(tf, tt)=1/(2+2)=1/4$$

$$P(r_2 \Rightarrow tt) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$$

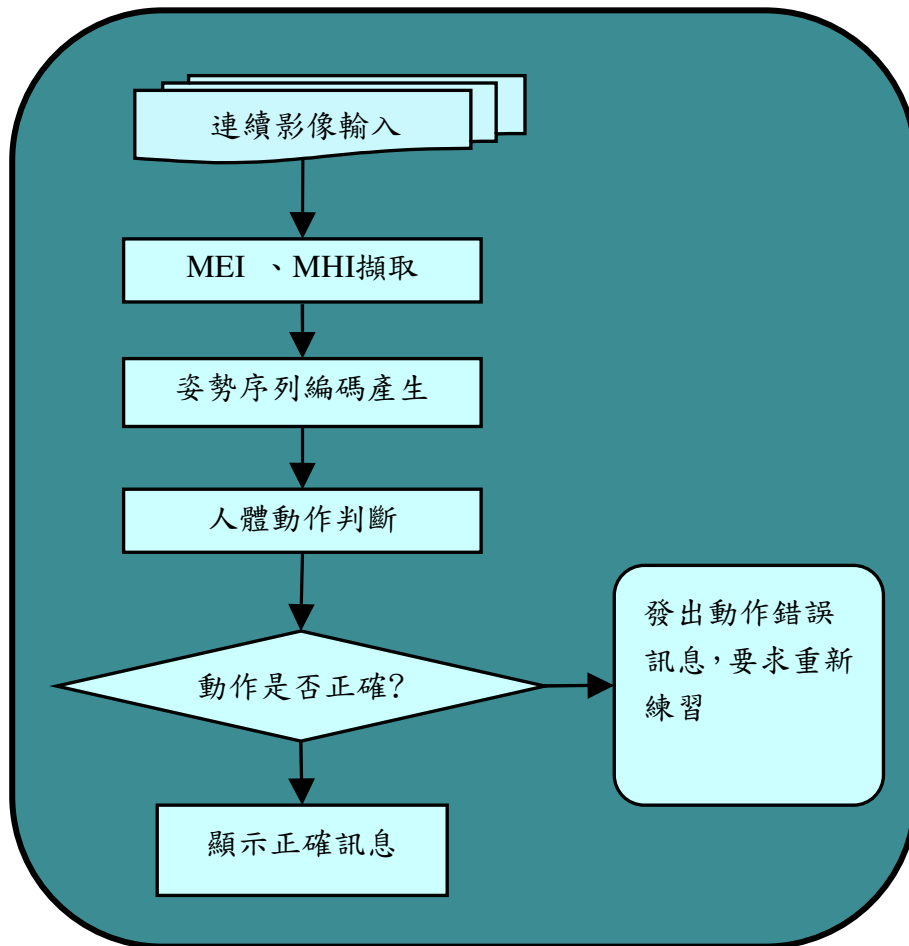
$$D(z3, tt)=D(lf, tt)=0/(2+2)=0$$

$$P(r_3 \Rightarrow tt) = \frac{0}{4} \times \frac{1}{4} = 0$$

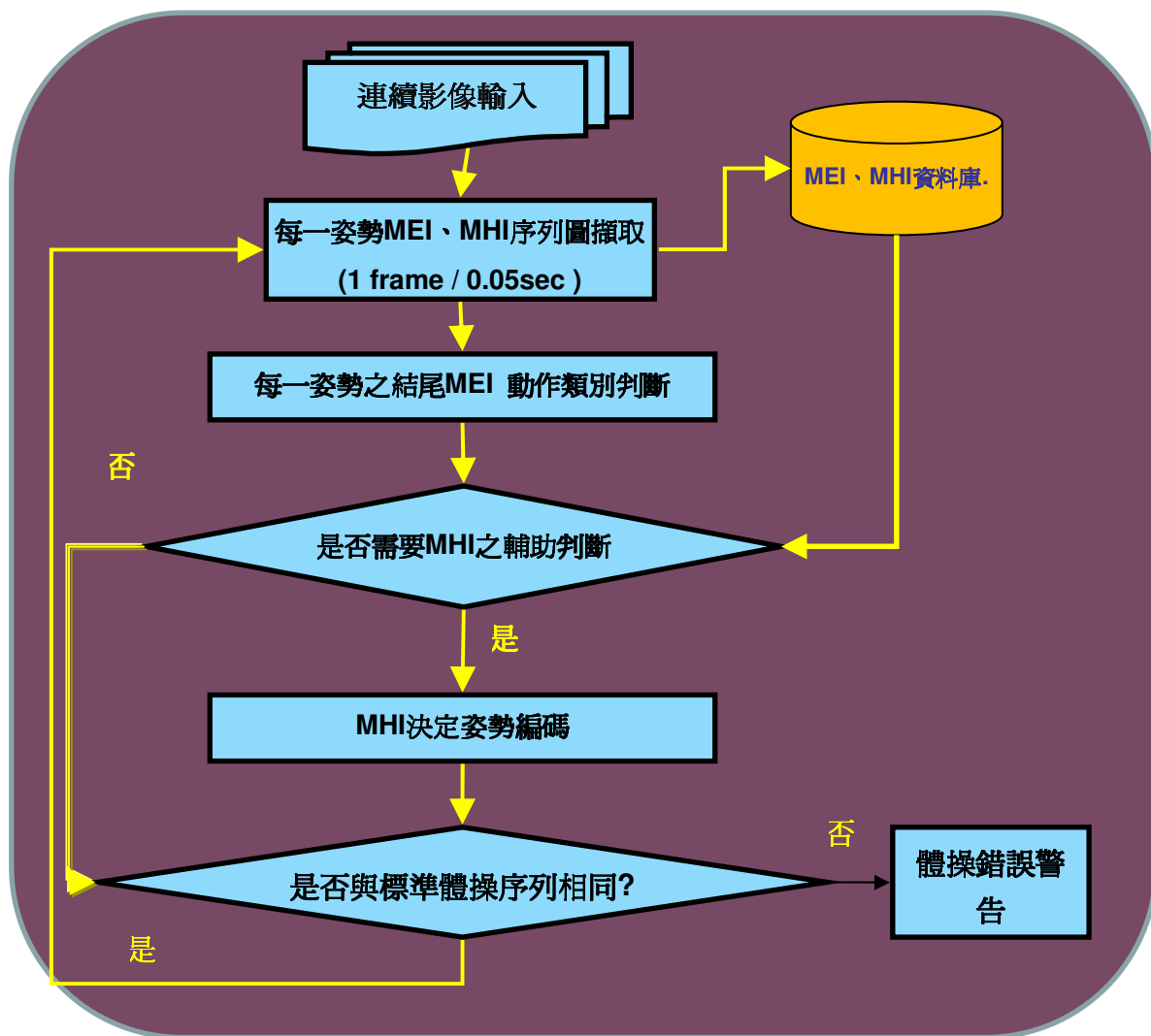
由上面所算出的最高機率決定，tt的動作由 r_2 所推導。

(四)研究進行方法

- 1.本系統將使用 Borland C++ Builder 作為開發軟體。
- 2.下圖六為本計劃之主要研究步驟：



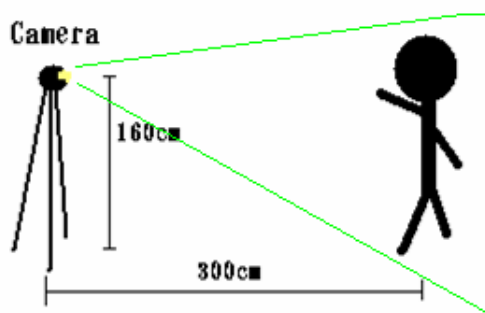
圖九、系統流程圖



圖十、人體動作辨識流程圖

(1) 連續影像輸入

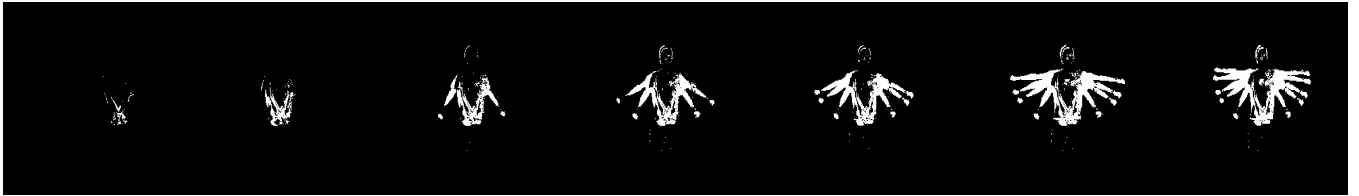
本系統將由架設在體育館內的網路攝影機取得連續的彩色影像，並針對連續影像序列進行分析。以下為攝影機架設圖，綠線表攝影機視框範圍。



圖十一、攝影機架設圖。

(3) MEI 的擷取

本系統可辨識之六套國民健康操體操動作共有 26 個姿勢，本系統對每個姿勢取其固定數量的 MEI 圖，再將每個姿勢的最後一張 MEI 送入資料庫中與所有資料苦中的 MEI 進行比對。當取得結束姿勢時，表一動作完成。如下圖

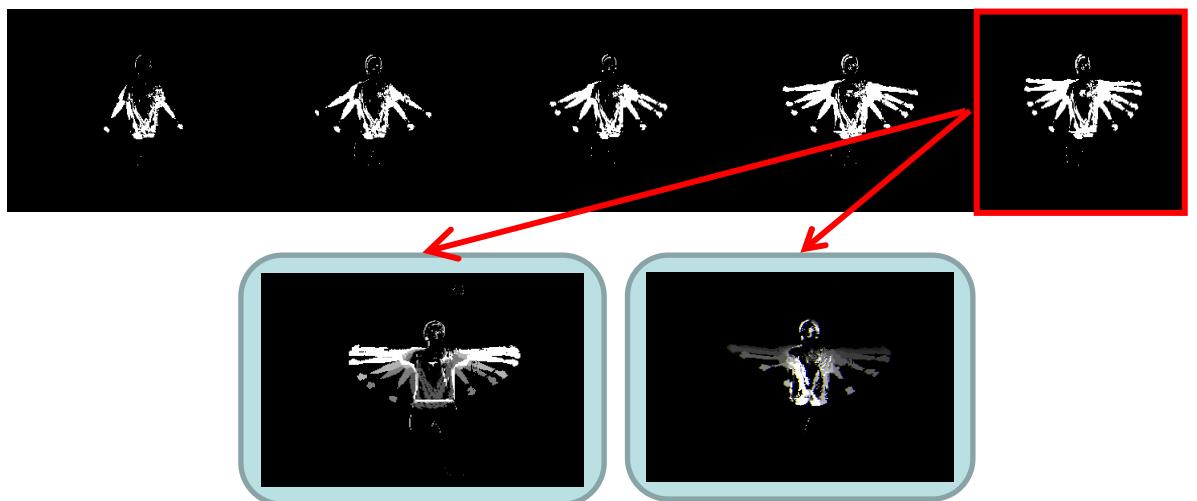


圖十二 (上圖) 完成同一姿勢之 MEI 圖序列

(下圖左->中) 動作序列 (下圖右)取得 MEI 圖

(2) MHI 的擷取

資料庫中的 MEI 圖間有些是很類似的，必須再將這個姿勢所形成的 MHI 序列取出加以辨識才能提高辨識效果。如以下的例子：

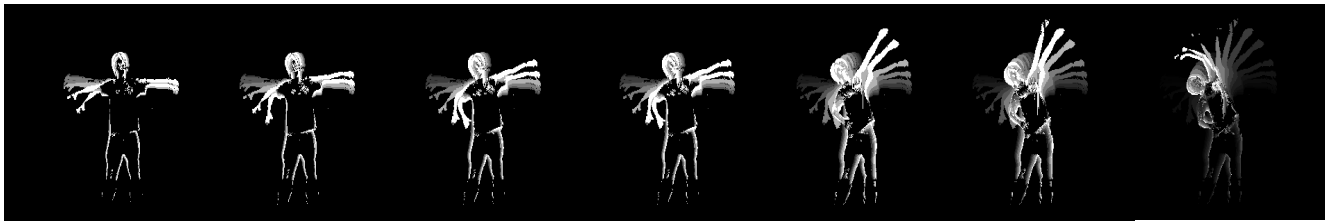


圖十三、MEI 所對應之不同姿勢

當取出此姿勢所形成的 MEI 後由於下方兩張 MHI 所對應的 MEI 都很

接近，因此將此段時間內所形成與下方之 MHI 分別比對，找到最相近的動作，在將其納入姿勢編碼。

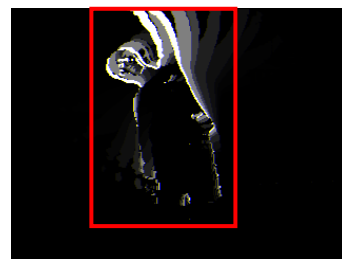
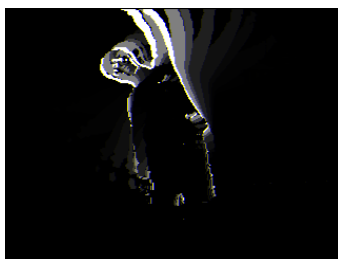
本系統可辨識之六套國民健康操體操動作共有 26 個姿勢，每個姿勢完成期間都會累計相同數量的 MHI 圖，此 MHI 圖將送入資料庫中進行比對，做為分辨擁有類似 MEI 圖形的姿勢，當比對完成後，即可將姿勢編號納入姿勢編碼中，當取得結束姿勢時，表一動作完成。如下圖為一姿勢的 MHI 序列，以下為某依姿勢所形成之 MEI 圖。



圖十四 (上圖) 完成同一姿勢之 MHI 圖序列

(下圖左->中) 動作序列 (下圖右)取得 MHI 圖

由於資料庫中的影像為固定大小，比對時兩張影像縮放為相同大小，將使結果的準確度較高，亦可簡化邊界問題。因此，取出 MHI 圖之後，將做正規化以利接下來資料庫比對的工作。



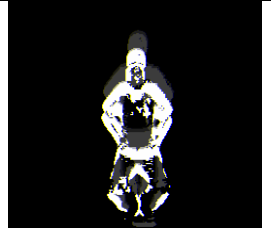


圖十五、左圖為擷取出之 MHI 圖，右圖紅框部分為抓取前景，將紅框內之影像做正規化後再與資料庫進行比對

(4) 資料庫比對

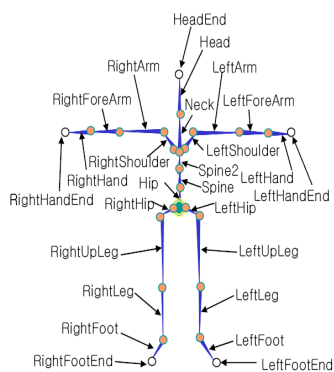
本系統之資料庫會預先進行分組，目的是為避免每張 current MHI 與所有資

料庫圖相比所造成的系統高額負擔，而分組依據是以 MEI 相似度高者被歸為同一類。本系統將會進行比對之 26 個姿勢，當每個 MHI、MEI 圖產生時即可送入資料庫中進行比對，當比對成功即記錄其編號，納入姿勢編碼中，當判斷其為結束姿勢時，將目前累積的姿勢編碼送入 Stochastic Context-Free Grammar。

編號	其所代表姿勢	類別	對應 MEI	對應 MHI
25		A		

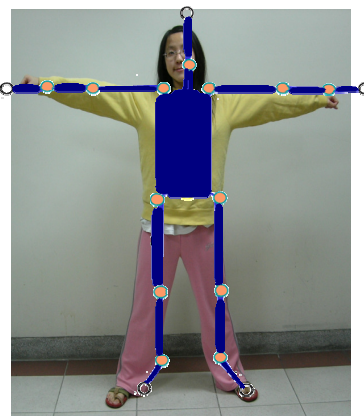
圖十六、資料庫內存取的資訊

(5) 判斷人體動作



圖十七、表示人體之 3D 結構模型[8]。

在判斷人體動作步驟中，會先對整套體操運動中的所有相異姿勢建構出一個 3D 模型並編號，所有模型會建成一個資料庫，在姿勢辨識時即以此資料庫作為比對之用。圖九為可表示人體之 3D 結構模型，圖中每個圈皆為一關節，每兩個



圖十八、模型調整後與原影像之比較的模型。

圈之間皆為人體的一小部位，資料庫中即儲存所有以此種 3D 模型為基礎的動作分類。

人體的各個部份，如脖子、關節、手彎曲角度等等稱為特徵參數，藉由影像的分析可擷取出人體各個部位的特徵參數資訊，此時將所有資訊送入一個 3D 模型，使模型做出與影像中人體相同的姿勢，稱為模型之調整，如圖十，即為調整模型動作後與原影像之比較。待模型調整完畢後即可將此模型送入資料庫中進行比對，資料庫事先會將所有模型姿勢進行分類，可採用 support vector machine(SVM)、類神經網路(neural network)等方式，分類完畢後即可用來判斷輸入姿勢是否為標準姿勢，若不是標準姿勢即會產生姿勢錯誤之警告，之後繼續處理下一張影像；若判定結果為標準姿勢後，再進行是否為動作結尾之判斷，若發現此姿勢並不為結尾姿勢，則繼續處理下張影像；若發現此姿勢為動作結尾，則將先前所完成之所有姿勢進行動作是否正確的判別。

判別動作可以用 HMMs 的相關技術進行，本系統將會設計一個 Stochastic Context-Free Grammar (SCFG)作為判別一連串姿勢是否可成功連結成正確動作的工具。當結尾動作被辨識出後，便將先前所有姿勢編碼成字串序列，而 SCFG 以字串序列作為輸入，當字串序列讀完時可被接受即判定動作正確並紀錄動作編號，若無法被接受即表示辨識動作失敗，系統會記錄錯誤訊息。此外，為方便於終止辨識功能，此體操系統將會要求受測人在整套體操動作完成後做出一有別於所有體操過程中的姿勢，當系統辨識到此終止姿勢即將所有動作與正確動作流程進行驗證，若流程正確即判定此次體操運動正確，若有錯誤則會顯示錯誤之動作。

(6) Stochastic Context-Free Grammar (SCFG)

當姿勢編碼產生時，即可作為 SCFG 的輸入串流。要得到一個姿勢編碼，必須將目前所得的 MEI 與 MHI 和資料庫中的影像做比對，並把相似度在 85%~100%之間的所有編碼與制定的 grammar 中各個 rules 的機率作結合以共同判斷。倘若目前姿勢編碼與資料庫中的 a_1 、 b_1 的相似度分別為 85%、95%，但包含 a_1 之 rule 機率為 0.8，而包含 b_1 之 rule 機率為 0.1，則結合兩者後， a_1 將為下一個輸入串流碼。以下為針對第一套體操動作所制定的三條 rules：



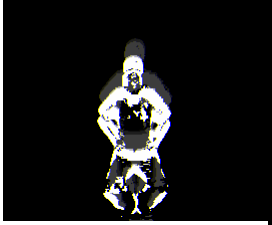


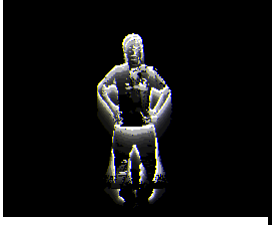



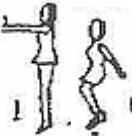






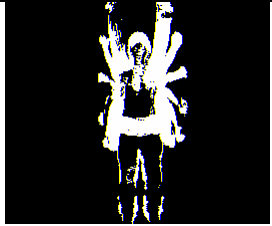
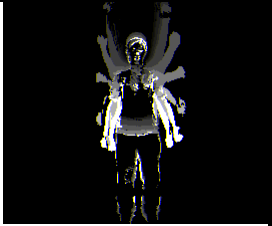
A \rightarrow a_1 B









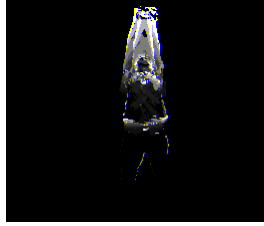


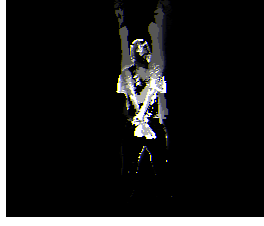


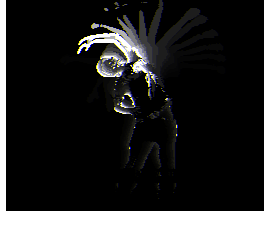






B \rightarrow a_2 A








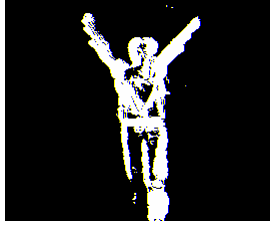
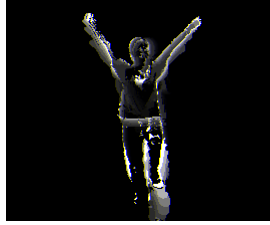












B \rightarrow a_2



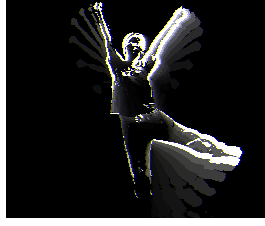



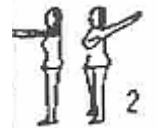


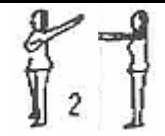
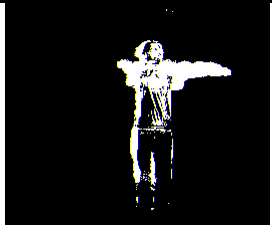
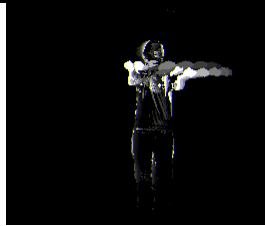
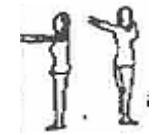
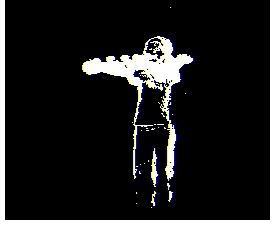
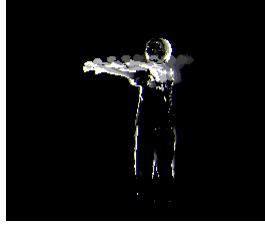

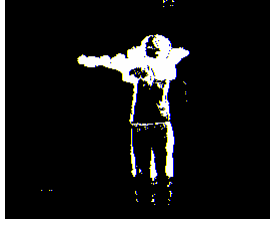

(五) 目前成果：

本系統已將二十五種姿勢對應之 MEI、MHI 圖接擷取完畢，以下整理此二十五組 MEI、MHI 圖所建立的完整資料庫。每一張圖皆由本系統所產生，未來將對此資料庫進行校正與去除雜訊，便可拿來作為標準姿勢庫：

編號	代表姿勢	類別	對應 MEI	對應 MHI
1		A		
2		A		
3		B		
4		B		
5		C		
6		C		

7		D		
8		D		
9		E		
10		E		
11		F		
12		F		
13		F		

14		F		
15		G		
16		G		
17		G		
18		G		
19		H		
20		H		

21		I		
22		I		
23		J		
24		J		
25		K		
26		K		

圖十九、由本系統所建立出之資料庫

分析六套動作後，可得到判斷動作分類的 Stochastic Context-Free Grammar (SCFG)，以下為針對各個動作建立出來的 rules：

Description	Production Rules	Description	Production Rules
下肢運動	$A \rightarrow a_1 B$ $B \rightarrow a_2 A$	四肢運動	$S \rightarrow WW$ $W \rightarrow LR$ $L \rightarrow e_1 M$ $\rightarrow e_5 N$
上肢運動	$C \rightarrow b_1 D$		

胸部運動	$D \rightarrow b_2 E$ $E \rightarrow b_3 F$ $F \rightarrow b_4 C$ $\rightarrow b_4$ $G \rightarrow c_1 H$ $H \rightarrow c_2 I$ $I \rightarrow c_3 J$ $J \rightarrow c_4 G$ $\rightarrow c_4$	轉體運動	$R \rightarrow e_3 O$ $\rightarrow e_7 P$ $M \rightarrow e_2$ $N \rightarrow e_6$ $O \rightarrow e_4$ $P \rightarrow e_8$ $D \rightarrow f_2 Q$ $Q \rightarrow f_3 T$ $T \rightarrow f_4 C$ $D \rightarrow f_5 U$ $H \rightarrow f_4$
左右彎體	$G \rightarrow c_1 H$ $H \rightarrow d_2 K$ $K \rightarrow d_2 H$ $I \rightarrow G$ $G \rightarrow d_4 H$ $H \rightarrow c_2$		

表一、為本系統欲辨識之體操動作所制定出 SCFG

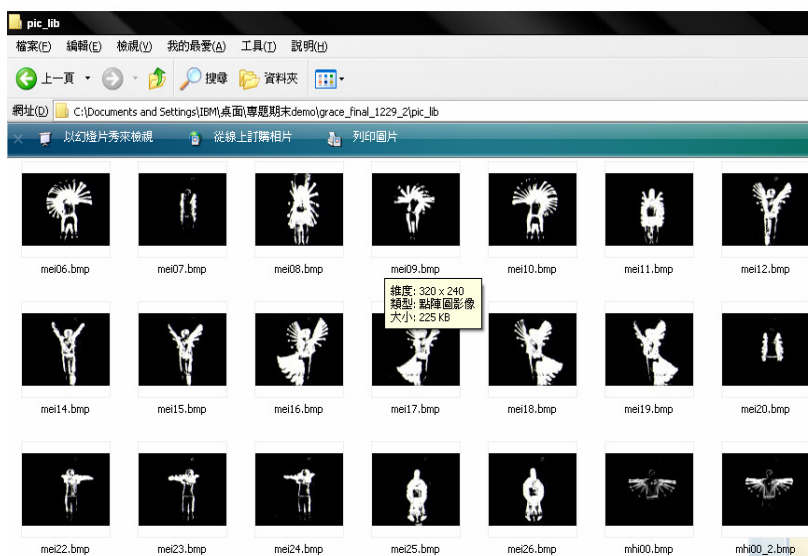
(六) 系統操作流程圖及相關說明

本系統實驗輸入之影像為為 320*240 pixel，產生 MEI、MHI 的速率為 1 秒 12 張。

以下為本系統操作步驟及各步驟結果：

1. 讀取資料庫：

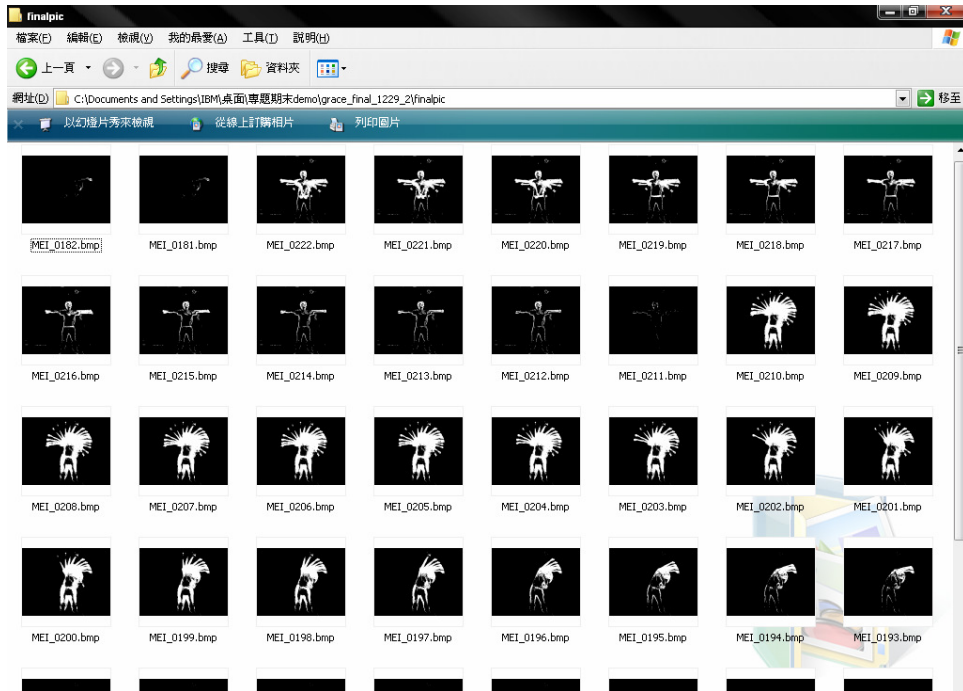
按下讀取資料庫後，本系統會將資料夾 pic_lib 中所存的資料庫都宰入系統中以便接下來比對的工作，下圖為本系統之資料庫(部分)：



圖二十、資料庫的儲存示意圖

2. 選取影片，並點選「開啟」開始 MEI、MHI 之擷取工作

開始擷取工作後之所有圖片都會存在資料夾 final_pic 中，由於體操動作之動作長度依據節拍進行，相當固定，因此本系統之姿勢間以時間作為切割根據，每個姿勢有 25 張 MEI、MHI 可進行比對。以下為開始擷取後資料夾 final_pic 中的儲存情形：



圖二十一、產生之 MEI 結果存放示意圖

3. 顯示比對結果



紅色框線中，上排表示八個姿勢完成後取得的最後一張 MEI 圖，下排則表示依

據所得的 MEI 所屬的資料庫類別進行 MHI 圖比對後所得到的姿勢編碼，此八張為資料庫所取出的結果而非由使用者所形成的。

(七)未來工作(國科會計畫繳件前)

1. 本計劃目前已可讀取包含八個姿勢的完整體操動作，並將各姿勢所形成的 MEI 及 MHI 圖，輸出圖檔並給予正規化處理，利用比對方式可判別出其姿勢編碼，未來可將蒐集的姿勢編碼做 String matching，分析使用者連續體操動作中錯誤及正確的子序列。
2. 找尋適切的 filter 對 MHI、MEI 圖片做 erosion、dilation 並計算辨識率。

(八)未來發展

本系統是判斷二維空間內的體操運動，未來可藉由增加攝影機進而辨識三維與跳躍之動作，並可朝向多人肢體動作辨識發展。此外，MHI 圖當中，可推斷某些關節移動的軌跡，甚至藉由深淺色塊多寡與區域大小，推斷此體操動作的速度，將來可根據這些資訊，讓系統除了動作比對外，還可判斷速度是否正確，而動作比對過程中，藉由紀錄測試者與資料庫的差值比例，可作動作精準度的計算，如此，將可讓系統更方便且實用。

(九)參考文獻

- [1] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man and Cybernetics, Part C*, Vol. 34, pp. 334–352, 2004.
- [2] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Computer Vision and Image Understanding*, Vol. 80, pp. 42–56, 2000.
- [3] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 22, pp. 266–280, 2000.
- [4] I. A. Karaulova, P. M. Hall, and A. D. Marshall, "A hierarchical model of dynamics for tracking people with a single video camera," *British Machine Vision Conf.*, pp. 262–352, 2000.
- [5] Q. Delamarre and O. Faugeras, "3D articulated models and multi-view tracking with physical forces," *Computer Vision and Image Understanding*, Vol.81, pp.328–357, 2001.
- [6] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups

- of people,” *Computer Vision and Image Understanding*, Vol.80, pp 42-56, 2000
- [7] <http://blog.freetimegears.com.tw/ryanblog/archives/000247.html> 國民健康操
- [8] B.W. Hwang, S. Kim, and S.W. Lee, “A Full-Body Gesture Database for Automatic Gesture Recognition,” *Automatic Face and Gesture Recognition*, pp 243- 248, 10-12 April 2006
- [9] D. G. Lowe, “Fitting parameterized 3-D models to images,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 441–450, 1991.
- [10] J. E. Bennett, A. Racine-Poon, and J. C. Wakefield, “MCMC for nonlinear hierarchical models,” *Markov Chain Monte Carlo in Practice*, pp. 339–357, 1996.
- [11] M. Isard and A. Blake, “CONDENSATION-Conditional density propagation for visual tracking,” *Int. J. Comput. Vis.*, Vol. 29, pp. 5–28, 1998.
- [12] M. Isard and A. Blake, “Condensation: unifying low-level and high-level tracking in a stochastic framework,” *Computer Vision in Proc. European Conf.*, Vol. 1, pp. 893–909. 1998.
- [13] James W. Davis, Aaron F. Bobick, "The Representation and Recognition of Human Movement Using Temporal Templates," *CPVR*, p. 928, 1997 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, 1997
- [14] Q. Chen, N. D. Georganas and E. M. Petriu, “Hand gesture recognition using Haar-like features and a stochastic context-free grammar”, *IEEE Transactions on Instrumentation and Measurement* , Vol. 57, no. 8, pp. 1562-1571, 2008

