

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※申請案號：96102113 606K 9/00 (2006.01)

※申請日期：96.1.19 ※IPC 分類：G06T 7/20 (2006.01)

一、發明名稱：(中文/英文) 606F 17/00 (2006.01)

結合時序姿態比對與模糊法則推論之人類動作辨識方法

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

國立交通大學

代表人：(中文/英文) 黃威

住居所或營業所地址：(中文/英文)

新竹市大學路 1001 號

國籍：(中文/英文) 中華民國 TW

三、發明人：(共 6 人)

姓名：(中文/英文)

1、張志永

2、呂志濤

3、卓建文

4、林進燈

5、蒲鶴章

6、劉得正

國籍：(中文/英文)

(均同) 中華民國 TW

四、聲明事項：

主張專利法第二十二條第二項 第一款或 第二款規定之事實，其事實發生日期為：95年7月20日。

申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

有主張專利法第二十七條第一項國際優先權：

無主張專利法第二十七條第一項國際優先權：

主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

主張專利法第三十條生物材料：

須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

五、中文發明摘要：

本發明是關於一種人類動作辨識方法，藉由結合時序姿態比對與模糊法則來完成人類動作的識別。首先，以減少照明對前景人物抽取的影響，每一張影像的前景人物利用一個基於前後影像比值而建立之統計背景模型抽取出來，並將抽取出來影像轉換成二值化的影像格式；再將二值化影像經由特徵空間及標準空間轉換，投影至標準空間；最後人類動作的識別在標準空間中完成，經由樣板比對的方法可將至少兩張影像序列，並轉換成轉變成一組時序姿態序列；利用模糊法則的推論方法，將這組時序姿態序列分類為某一個動作類別，使其不僅能夠結合時間序列上的資訊，並且可以容忍不同人做相同動作上的差異。因此，本發明提出的動作辨認方法較先前技術而言，具有較高的辨識正確率。

六、英文發明摘要：

The present invention relates to a human activity recognition method, particularly to a method which uses template posture matching and fuzzy rule reasoning to recognize an action. First of all, a foreground subject is extracted and converted to a binary image by a statistical background model based on frame ratio, which is robust to illumination changes. For better efficiency and separability, the binary image is then transformed to a new space by eigenspace and canonical space transformation, and recognition is done in canonical

space. At least two image frame sequence is converted to a posture sequence by template matching. The posture sequence is classified to an action by fuzzy rule inference. Fuzzy rule approach can not only combine temporal sequence information for recognition, but also be tolerant to variation of action done by different people. Therefore, the present invention is able to achieve higher recognition accuracy than the prior art.

七、指定代表圖：

(一)、本案代表圖為：第 1 圖

(二)、本案代表圖之元件代表符號簡單說明：

無

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

九、發明說明：

【發明所屬之技術領域】

本發明之主要目的係在提供一種人類動作辨識方法，特別是關於一種結合時序姿態比對與模糊法則推論來完成人類動作的識別。

【先前技術】

人類動作辨識在自動監視系統、人機界面、居家安全照護系統和智慧型居家環境等方面的應用中佔有主要的地位。許多人類動作辨識系統僅僅利用單一張影像的姿勢來辨別該動作。但是，在時間序列上，姿勢狀態轉換的關係是用來辨別人類動作的重要資訊。近來已經有不少人類動作辨識的方法被提出。大多數人類動作識別方法可以依據其方法所使用的特徵而分為兩大類。第一類是利用動量特徵(motion-based features)。在“IEEE Trans. Pattern Anal, vol. 23, no. 3, 2001”論文中，Bobick和Davis利用比較時序影像中樣板的動量能量(motion-energy)以及先前動量統計圖(motion-history)來辨視人類的動作。在“Proc. Conf. Comput. Vision Pattern Recog., vol. 4, pp. 38-45., 2003”論文中，R. Hamid等人採用了時空上的特徵，例如兩隻手之間相對距離和兩隻手之間的相對速度；此外還使用動態貝氏網路(dynamic Bayesian networks)來完成人類的動作識別工作。另一方面，則是使用二維跟三維的形狀特徵來識別人類動作。在“IEEE Comput. Soc. Workshop Models versus Exemplars in Comput. Vision, pp. 263-270, 2003”論文中使用肯尼邊緣偵測器(Canny edge detector)得到的資料來表示動作的輪廓形狀，並且對於每一個動作定義了關鍵畫面(key frame)。在“IEEE Int. Workshop on Anal. Modeling of

Faces and Gestures, pp. 74 - 81, 2003” 論文中的作者們提出了使用 SVM 以及不受視角影響的三維形狀特徵描述來達到人類動作的分類與辨識。

如果僅採用運動基礎和形狀基礎的特徵資訊來完成動作識別的工作，因為時序上資訊並未被考慮進去，所以還是有許多動作是無法清楚辨別的。因此，這促使我們設計一個穩定、強健的方法，利用人類動作中隱含固有的時序狀態資訊來達成較精準的動作識別。Hidden Markov Model (HMM) 能夠處理時序上的資料和能夠提供辨識上不受時間尺度改變影響的特性，Hidden Markov Model 也已經應用於手部姿勢識別和動作識別上。使用 HMM 的代價是效率上的問題以及必須收集大量的資料和花費許多時間在估計 HMM 相對應的參數。

本發明利用特徵空間轉換(Eigenspace transformation)及標準空間轉換(Canonical space transformation)提取圖像中的特徵，因而經由使用這些空間轉換後的向量可以將時態序列轉換成姿勢序列，這個姿勢序列可由樣板模板類別的索引值組合表示。因為如果在一個很短的時間間隔中所擷取的兩張影像之間的姿態變化差異會極少，這主要受到人類在做動作一個固有的自然頻率所限制，其頻率不會很高；因此我們採用減低取樣頻率影像而不使用所有畫面來做辨識。此外，本發明更提出了使用模糊法則推論方法，此方法不但可以結合人類動作上的姿勢序列資訊，而且模糊法則可涵蓋不同人在做相同動作時姿勢上些微差異，增加辨識效果的穩定性。

【發明內容】

本發明之主要目的係在提供一種人類動作辨識方法，特別是關於一種

結合時序姿態比對與模糊法則來完成人類動作的識別。

為達到上述目的，本發明之人類動作辨識方法包含以下步驟。首先設立一攝影裝置於一定點以捕捉該定點的原始畫面並建立學習一背景畫面，然後將該攝影裝置以固定頻率的時間擷取影像，並在降低抽樣頻率後化為一序列的影像輸入畫面，序列中的每一個輸入畫面並與該背景畫面做比對，以取出前景影像。再將每一張取出的前景影像實施空間轉換，然後將至少連續兩張已實施空間轉換後的前景影像轉變成一組時序姿態序列。最後將這組時序姿態序列透過模糊法則為基礎的人類動作識別來分類出成某一個人的動作類別。此方法不但同時可以結合人類動作上的姿勢序列資訊，而且模糊邏輯方法能涵蓋不同人在做相同動作時姿勢上些微差異。模糊法則的優點亦包含可以學習到姿勢轉換關係的隱藏模式。

底下藉由具體實施例配合所附的圖式詳加說明，當更容易瞭解本發明之目的、技術內容、特點及其所達成之功效。

【實施方式】

本發明之主要目的係在提供一種結合時序姿態比對與模糊法則推論來完成人類動作的識別的方法。

第 1 圖為本發明之系統架構圖，主要架構可以分成三個流程。第一個流程 S110 是前景影像人物的抽取。第二個流程 S120 是將影像經過空間轉換投影到一個維度較小、更容易辨識出姿態的空間上。第三個流程 S130 是單一張影像畫面的姿勢分類和使用姿勢序列來辨識未知動作所屬的類別。

人類動作識別系統的第一步是將前景人物部分抽取出來。在第一個流

程 S110 裡，首先第一步驟 S111 就是為前景人物抽取建造一個背景模型。本發明使用相連影像相除法來描述一個統計背景模型。透過計算統計最大、最小的灰階值和影像灰階值相除的最大比值，得以獲得一個背景模型。

本發明提出一種已經被證實對照明變化較不敏感的影像相除比值法，而不是使用影像差值法來建立背景模組。假設相機所捕捉到的畫面的影像強度可藉由方程式 1 表示。

$$I_i(x, y) = S_i(x, y) r_i(x, y) \quad (1)$$

其中 S_i 代表某一個像素位置的照度，而 r_i 代表某一個像素位置的反射率，而 i 代表影像序列中的索引值。如果只拍攝背景並且保持照像機的穩定，反射率的影響仍然會存在。不過如果使用影像相除的方法，則反射率的影響將可以被消除。兩張連續的影像相除可以將他們的像素強度比值寫成方程式 2。

$$\begin{aligned} \log\left(\frac{I_i(x, y)}{I_{i-1}(x, y)}\right) &= \log\left(\frac{S_i(x, y)r(x, y)}{S_{i-1}(x, y)r(x, y)}\right) = \log\left(\frac{S_i(x, y)}{S_{i-1}(x, y)}\right) \\ &= \log(S_i(x, y)) - \log(S_{i-1}(s, y)) \end{aligned} \quad (2)$$

所以本發明提出利用影像相除的比值來建造出背景模型。背景影像的每個像素都以三個統計數值來代表：最小灰階強度值 $n(x, y)$ 、最大灰階強度值 $m(x, y)$ 和最大相連影像灰階相除比值 $d(x, y)$ 。這三個數值的統計方式如方程式 3 所示。

$$\begin{bmatrix} m(x, y) \\ n(x, y) \\ d(x, y) \end{bmatrix} = \begin{cases} \begin{bmatrix} \max_i \{I_i(x, y)\} \\ \min_i \{I_i(x, y)\} \\ \max_i \{I_i(x, y)/I_{i-1}(x, y)\} \end{bmatrix} & \text{if } I_i(x, y)/I_{i-1}(x, y) \geq 1 \\ \begin{bmatrix} \max_i \{I_{i-1}(x, y)\} \\ \min_i \{I_i(x, y)\} \\ \max_i \{I_{i-1}(x, y)/I_i(x, y)\} \end{bmatrix} & \text{otherwise} \end{cases} \quad (3)$$

在建立出背景模型之後，步驟 S112 到 S113 是將前景人物部分可以從每個影像中分離出來。一張影像中的像素要被歸類於前景還是背景，主要依據方程式 4 所得到的結果。

$$B(x, y) = \begin{cases} 0, \text{ a background pixel if } \begin{cases} I_i(x, y)/m(x, y) < kd(x, y) \\ \text{or} \\ I_i(x, y)/n(x, y) < kd(x, y) \end{cases} \\ 1, \text{ a foreground pixel otherwise} \end{cases} \quad (4)$$

其中該方程式的 B 代表一張前景人物分離出來後的二值化影像， k 值為可調參數，一般來說， $k=1.4$ 。而前景人物的區塊則可以使用對 X 軸以及 Y 軸做投影統計後設定一個門檻值，依據此統計值的邊界位置就可以切割出來前景部分，切割出來的影像都會被調整到同樣的大小。

兩張影像如果是在相鄰間隔很短的時間內捕捉進來的話，這兩張影像中的姿勢差異性並不會太大。此外，人體約屬於剛體，因此有它的自然頻率；換言之，當做某一動作時，它有行動速度上的自然限制。在我們的方法中，我們每隔一段固定的間隔取一個影像，稱做為基本樣板影像。第 2 圖為挑選樣版影像的範例，該範例在大約固定時間間隔挑選出五組基本左

到右走路的樣版影像。這些基本樣版影像會經由特徵空間轉換以及標準空間轉換投影到一個新的空間上如第二個流程 S120 所示，而整個識別的流程是在標準空間中完成的。

在視訊以及影像處理上，圖像資料的維度通常會非常大。因為在這些圖像中有很多冗餘的部份，通常解決這些冗餘部分的方法是將圖像經由空間上的轉換投影到一個新的空間上。這些轉換方法大部分是透過較少的維度來逼近近似這個原始圖像。在第二個流程 S120 裡，本發明的轉換法結合了步驟 S121 的特徵空間轉換(Eigenspace transformation)及步驟 S122 的標準空間轉換(Canonical space transformation)。特徵空間轉換已經有效的使用在自動人臉辨識系統識別和走路姿態辨識等系統上。在特徵空間轉換後，再使用標準空間轉換來減少資料維度、最佳化於增強兩個類別間的可分離性和提升辨識系統的效能。識別是在標準空間中完成的。

假設總共有 c 個種類的學習資料。每一個種類都代表存在整個訓練影像資料中的某一特定動作的姿勢型態。 $x'_{i,j}$ 代表在第 i 個類別中的第 j 個影像，整個訓練集合的影像總數可以表示為 $N_T = N_1 + N_2 + \dots + N_c$ 。而整個訓練集合可以改寫為 $[x'_{1,1}, \dots, x'_{1,N_1}, \dots, x'_{2,1}, \dots, x'_{c,N_c}]$ ，其中每一個 $x'_{i,j}$ 都是含有 n 個像素的影像。

首先，每張影像的強度會先經過正規化的處理如方程式 5 所示。

$$\mathbf{x}_{i,j} = \frac{\mathbf{x}'_{i,j}}{\|\mathbf{x}'_{i,j}\|}. \quad (5)$$

透過方程式 5 可以得到每張影像的平均像素值並由方程式 6 來表示。

$$\mathbf{m}_x = \frac{1}{N_T} \sum_{i=1}^c \sum_{j=1}^{N_i} \mathbf{x}_{i,j}. \quad (6)$$

同時訓練影像集合可以改寫成一個維度是 $n \times N_T$ 的矩陣 \mathbf{X} ，也就是方程式 7。

$$\mathbf{X} = [\mathbf{x}_{1,1} - \mathbf{m}_x, \dots, \mathbf{x}_{1,N_1} - \mathbf{m}_x, \dots, \mathbf{x}_{c,N_c} - \mathbf{m}_x]. \quad (7)$$

假設矩陣 $\mathbf{X}\mathbf{X}^T$ 的秩是 K ，則矩陣 $\mathbf{X}\mathbf{X}^T$ 會有 K 的不為零的特徵值， $\lambda_1, \lambda_2, \dots, \lambda_K$ ，以及其相對應的特徵向量， $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_K$ ，並且符合方程式 8。

$$\lambda_i \mathbf{u}_i = \mathbf{R} \mathbf{u}_i, \quad i=1, 2, \dots, K \quad (8)$$

方程式 8 中的 $\mathbf{R} = \mathbf{X}\mathbf{X}^T$ ，而且這個 \mathbf{R} 是一個對稱的方陣。但是 $\mathbf{X}\mathbf{X}^T$ 的維度一般來說就是影像的尺寸大小，這樣維度會非常大，會增加矩陣運算很大的複雜度，基於奇異值分解理論(singular value decomposition theory)可以透過計算 $\tilde{\mathbf{R}}$ 這個矩陣來得到特徵值跟特徵向量，並透過方程式 9 來表示。

$$\tilde{\mathbf{R}} = \mathbf{X}^T \mathbf{X} \quad (9)$$

式子中 $\tilde{\mathbf{R}}$ 矩陣的維度是 $N_T \times N_T$ ，比原本 \mathbf{R} 矩陣的 $n \times n$ 小很多。這個 $\tilde{\mathbf{R}}$ 矩陣也會有 K 個不為零的特徵值， $\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_K$ ，和相對應的特徵向量， $\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \dots, \tilde{\mathbf{e}}_K$ ，接著可以利用方程式 10 來得到 \mathbf{R} 矩陣的特徵值跟特徵矩陣。

$$\begin{cases} \lambda_i = \tilde{\lambda}_i \\ \mathbf{e}_i = (\tilde{\lambda}_i)^{-\frac{1}{2}} \mathbf{X} \tilde{\mathbf{e}}_i \end{cases} \quad (10)$$

其中該式子中的 $i=1, 2, \dots, K$ ，則 K 則代表 K 個特徵向量為互相正交的向量。

基於主成分分析(principal component analysis)理論，每張影像可以利用 $k \leq K$ 最大的特徵值和相對應的特徵向量 $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k$ 來得到近似的結果。這些

k 個這徵象量可以延展出一個新的空間，並且可以利用方程式 11 將原本的影像投影到新的空間上。

$$\mathbf{y}_{i,j} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k]^T \mathbf{x}_{i,j} \quad (11)$$

式子中 $i=1, 2, \dots, c$ 而 $j=1, 2, \dots, N_c$ 。我們將 $[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k]^T$ 這個矩陣稱為特徵空間轉換向量。

基於標準空間分析法則，我們利用 $\{\phi_1, \phi_2, \dots, \phi_c\}$ 代表經過特徵空間轉換後的類別，而 $\mathbf{y}_{i,j}$ 則代表第 i 個類別的第 j 個向量。整個集合的平均向量則可使用方程式 12 計算出來。

$$\mathbf{m}_y = \frac{1}{N_T} \sum_i \sum_j \mathbf{y}_{i,j} \quad (12)$$

其中 $i=1, 2, \dots, c$ 而 $j=1, 2, \dots, N_i$ 。各類別的平均向量則可以使用方程式 13 計算。

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{\mathbf{y}_{i,j} \in \Phi_i} \mathbf{y}_{i,j}. \quad (13)$$

計算出這兩個平均向量後，我們可以定義下列三個變數： S_t 代表總散佈矩陣 (total scatter matrix)， S_w 代表群內散佈矩陣 (within-class matrix)， S_b 代表類別間散佈矩陣 (between-class matrix)。這三個變數可以透過下列方程式計算得到

$$\begin{aligned} S_t &= \frac{1}{N_T} \sum_{i=1}^c \sum_{j=1}^{N_i} (\mathbf{y}_{i,j} - \mathbf{m}_y)(\mathbf{y}_{i,j} - \mathbf{m}_y)^T \\ S_w &= \frac{1}{N_T} \sum_{i=1}^c \sum_{\mathbf{y}_{i,j} \in \Phi_i} (\mathbf{y}_{i,j} - \mathbf{m}_i)(\mathbf{y}_{i,j} - \mathbf{m}_i)^T \\ S_b &= \frac{1}{N_T} \sum_{i=1}^c N_i (\mathbf{m}_i - \mathbf{m}_y)(\mathbf{m}_i - \mathbf{m}_y)^T \end{aligned}$$

標準空間轉換的主要目的就是同時使群中散佈矩陣最小而類別間散佈矩陣最大。要得到這結果可以透過方程式 14 計算出來。

$$\mathbf{J}(\mathbf{W}) = \frac{\mathbf{W}^T \mathbf{S}_b \mathbf{W}}{\mathbf{W}^T \mathbf{S}_w \mathbf{W}}. \quad (14)$$

求解的方法就是將選擇一個極大解 \mathbf{W} 以符合方程式 15。

$$\frac{\partial \mathbf{J}}{\partial \mathbf{W}} = 0. \quad (15)$$

假設 \mathbf{W}^* 是最佳解，而 \mathbf{w}_i^* 向量是相對於第 i 個最大特徵值所得到的特徵向量。依照 “*Introduction to Statistical Pattern Recognition*, 2nd edition, 1990” 書中所提到的理論可以將方程式 15 改寫成方程式 16。

$$\mathbf{S}_b \mathbf{w}_i^* = \lambda_i \mathbf{S}_w \mathbf{w}_i^* \quad (16)$$

解完後可以得到 $c-1$ 個不為零的特徵值跟其相對應的特徵向量 $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{c-1}]$ 。利用這個基底，我們可以將一個在特徵空間上的一個點投影到標準空間上的另一個點如方程式 17 所示。

$$\mathbf{z}_{i,j} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{c-1}]^T \mathbf{y}_{i,j} \quad (17)$$

其中 $\mathbf{z}_{i,j}$ 代表投影到新空間的點，而正交基底 $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{c-1}]^T$ 則稱作標準空間轉換矩陣。

透過結合方程式 11 和 17，每張影像都可以透過方程式 18 投影到一個 $c-1$ 維的新空間上。

$$\mathbf{z}_{i,j} = \mathbf{H} \mathbf{x}_{i,j}. \quad (18)$$

第三個流程 S130 主要為辨識流程，在時間影像序列中，不同姿勢間的轉變關係是識別人類動作上一個非常重要的資訊。如果我們僅僅利用一張影像來做為動作辨識的依據，則分類結果很容易錯誤，因為兩個不同的動作中，可能會出現極類似的姿勢影像。人類動作的單一姿勢常常有模糊性，所以在步驟 S131 到 S133 裡，我們提出利用模糊法則來做人類動作識別，它不僅能夠結合時間序列上的資訊，而且可以容忍不同人做同一個動作的差異。相關使用模糊理論的論文如下。在“IEEE Trans. Syst., Man Cybern, vol. 22, no. 6”的論文中，Wang 和 Mendel 提出從例子中去學習產生模糊法則。在“IEEE Trans. Sys., Man Cybern, A, vol. 30, no. 2”的論文中，Su 提出了以模糊法則為基礎的方法來時序上的手部姿勢辨別。

假設 r_{ij} 是第 i 個姿勢類別中第 j 個人樣板影像的空間轉換向量，而 $\bar{\mathbf{a}}$ 是某一時刻影像的轉換向量。我們利用高斯型態的隸屬函數(membership function)來表示某一張影像相對於每一個姿勢類別的可能性，此隸屬函數可由方程式 19 所表示。

$$r_{ij} = \frac{1}{(2\pi)^{\frac{c-1}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\bar{\mathbf{a}} - \bar{\mu}_{ij})^T \Sigma^{-1} (\bar{\mathbf{a}} - \bar{\mu}_{ij}) \right] \quad (19)$$

其中 Σ 代表 $\bar{\mathbf{a}}$ 的協方差矩陣(covariance matrix)而 $\bar{\mu}_{ij}$ 代表的是平均向量。我們假設 $\bar{\mathbf{a}}$ 和 $\bar{\mu}_{ij}$ 的各維度都是互相獨立的，則方程式 19 可以改寫為方程式 20。

$$r_i = \arg \max_j \left\{ \prod_{m=1}^{c-1} \frac{1}{\sqrt{2\pi}\sigma_m} \exp \left[-\frac{1}{2} \frac{(\mathbf{a}^m - \mu_{i,j}^m)^2}{\sigma_m^2} \right] \right\} \quad (20)$$

其中 m 代表維度的索引， σ_m 是使用全部的平均向量去求得的第 m 維標準差，而 r_i 代表影像相對於第 i 個姿勢類別的隸屬函數值。最後可以求得這張影像所屬的姿勢類別並由方程式 21 所示。

$$i = \arg \max_i r_i. \quad (21)$$

每一張影像都會改用一個文字符號 P_i 來代表它，這邊的 i 指的是姿勢類別的索引值。

為了包含時序資訊，我們將三張相連的減低取樣頻率影像結合成為一組如 (I_1, I_2, I_3) 。如果這一組資料採用了太多張的影像，則一個快速完成動作週期的動作容易受到其相鄰動作的影響；如果採用太少張的影像，則時序上的資訊會不足。首先，這三張影像會投影到標準空間中，然後依照隸屬函數值將它轉換成姿勢序列。接著將姿勢序列與它相對應的動作類別作結合為 $(I_1, I_2, I_3; D)$ ， D 代為動作 (Action Category)。這樣的姿勢序列構成模糊法則系統在學習時的輸入輸出對。而系統經由這組輸入輸出對，會產生一組相對應的法則，它的形式如下

“IF antecedent conditions hold,

THEN consequent conditions hold.”

舉例來說，如第 2 圖所示，第一張影像的姿勢序列是屬於 P_{18} 、第二張影像屬於 P_{19} 、第三張影像屬於 P_{20} ，而其相對應的動作屬於從左往右走 (W_{LR})，則會由方程式 22 的輸入輸出對等，以第 2 圖為例，就會代表從左往右走。

$$(I_1, I_2, I_3; D) = (P_{18}, P_{19}, P_{20}; W_{LR}) \quad (22)$$

由此會產生下列的法則

“IF the activity's I_1 is P_{18} AND its I_2 is P_{19} AND its I_3 is P_{20} , THEN the activity is W_{LR} ”

第 3 圖顯示系統分類演算法的結構。首先，每間隔一段固定時間去讀取一張影像如 S_{300} 。這一張影像會經過空間的轉換到標準空間中並且計算出其相對應的隸屬函數值如 S_{305} 到 S_{310} 。接著如 S_{315} 到 S_{325} 步驟所顯示，每三張影像的隸屬函數值集成一個群組，去訓練所得到的模糊法則資料庫中尋找一組最相似的姿勢序列，這個影像序列會被分類到最類似的法則中所紀錄的動作類別。

我們的實驗環境是一間教室。光源是穩定的日光燈。採用僅有一張桌子的單純背景。攝影機固定在一個地點，並且不移動的拍攝同一場景。攝影機每秒鐘會拍攝 30 張 640×480 像素的影像。共有六個動作者，每位各自做相同的下面六個動作：「由左往右走」(W_{LR})、「由右往左走」(W_{RL})、「跳躍」($Jump$)、「蹲下」(C_{ROH})、「爬上」(C_{UP})和「爬下」(C_{DON})。其中五個人的影片為訓練用，剩下一個則當測試用，每個人的影片都會輪流拿來做測試。第 4a 圖 是一張由攝影機擷取的影像，第 4b 圖是前景人物分離出來並轉為二值化影像的結果。

我們各挑選了六類的基本姿勢樣板影像給「由左往右走」、「由右往左走」和「爬上」，五類給「爬下」，三類給「蹲下」，兩類給「跳躍」，總共是 28 類。在所有的測試影片都經過以上的訓練學習後，我們必須設定一個門檻值，這門檻值是用來摒除某些姿勢序列發生次數相對較少的法則，門

閾值的高低會影響法則數目的多寡。第 5 圖為不同門檻值與辨識結果的比較圖。圖中可以看到，當門檻值為二的時候辨識率最高，但在我們的實驗中，我們採用三為門檻值，這是因為如果採用的門檻值太低，則會產生太多的法則，而且有些法則的支持數會太少。此外，如果有一些互相矛盾的法則產生，我們則挑選在訓練時出現次數較多的法則。

表格一顯示我們的系統的認識率。我們的動作辨識系統目前是離線測試狀態，也就是我們的測試影片不是即時影片。因為有 5:1 降低抽樣影像的步驟，我們在測試的時候把由不同起始時序位置，即 1 至 5 的位置去讀取影像，做辨識的狀況也類似考慮進去，這與我們在訓練模糊法則時是一樣的，而且也比較相近於即時辨識時發生的狀況。舉例來說，我們影片中第一個、第二個、第三個、第四個或第五個影像開始去讀取做辨識。

在“IEEE CVPR, pp. 379 - 385, 1992”的論文中，Yamato 和 Ohya 提出使用 Hidden Markov Model 做人類動作的識別。Hidden Markov Model (HMM) 是一種狀態間轉換的機率模型，通常被拿來做時序上的資料分析。在實驗中我們將模糊法則方法與 HMM 方法的識別率做個比較。

表格一

所有動作者的動作辨識率

量測資料	辨識率 (%)					
	W _{LR}	W _{RL}	C _{ROH}	J _{UMP}	C _{UP}	C _{DOH}
第一人	100.0	92.3	71.0	78.4	78.1	94.6
第二人	100.0	82.5	97.1	61.8	100.0	94.3
第三人	100.0	100.0	74.4	94.1	100.0	45.3
第四人	100.0	93.7	100.0	91.3	93.6	76.7
第五人	100.0	100.0	100.0	100.0	90.7	100.0
第六人	100.0	100.0	97.6	100.0	100.0	100.0
平均值	91.78					

HMM 方法的第一部份是使用最近鄰居(nearest neighbor)分類法將影像序列變換成姿勢序列。每一張影像都被分類到距離最近的姿勢樣板類別，而姿勢樣板的挑選跟以上敘述模糊法則方法的樣版是一樣的。在學習的階段，各 HMM 必須經過訓練以產生最能代表某一類動作的姿勢轉換機率參數。在實驗中，我們使用 Baum-Welch 演算法來產生估計出 HMM 的相關參數，另外我們採用了前向鍵結方式(forward-chaining)，狀態的各數設定 28 個，最後觀測序列的長度設定為三。

在經過訓練資料的學過後，可以獲得六組 HMMs 分別代表一類的動作。在識別一個未知動作之觀測序列的動作時，將未知動作歸類於能從六個 HMMs 中產生最大機率值，也就是最相近的 HMM 類別中。我們採用前向演算法(forward algorithm)來計算出這個機率值。

使用 HMM 演算法和以模糊法則為基礎的演算法之間的辨識率比較顯示

在表格二。模糊法則演算法能夠獲得較高的正確率大約提高了 2.4%，這顯示了模糊法則演算法在人類動作的識別上有較好的辨識效果。

表格二

HMM 與模糊法則演算法辨識率比較表

	HMM	模糊法則演算法
第一人	81.18	84.61
第二人	88.33	91.03
第三人	86.25	87.15
第四人	90.00	93.33
第五人	93.80	96.71
第六人	96.90	97.85
平均值	89.41	91.78

實驗結果顯示，在沒有參考任何姿體所在位置、移動路徑和移動速度的狀況下，這六種動作的總辨識率可以達到 91.78%。模糊法則演算法與 HMM 方法相比較下提高了大約 2.4% 的正確率。

惟以上所述者，僅為本發明之較佳實施例而已，並非用來限定本發明實施之範圍。故即凡依本發明申請範圍所述之形狀、構造、特徵及精神所為之均等變化或修飾，均應包括於本發明之申請專利範圍內。

【圖式簡單說明】

第 1 圖為本發明之系統架構示意圖。

第 2 圖為挑選樣版影像的範例示意圖。

第 3 圖為系統分類演算法的結構示意圖。

第 4 圖為前景人物擷取範例示意圖。

第 5 圖為不同門檻值與辨識結果的比較示意圖。

【主要元件符號說明】

無

十、申請專利範圍：

1. 一種人類動作辨識方法，包含：

設立一攝影裝置於一定點以捕捉該定點的原始畫面並建立學習一背景畫面；

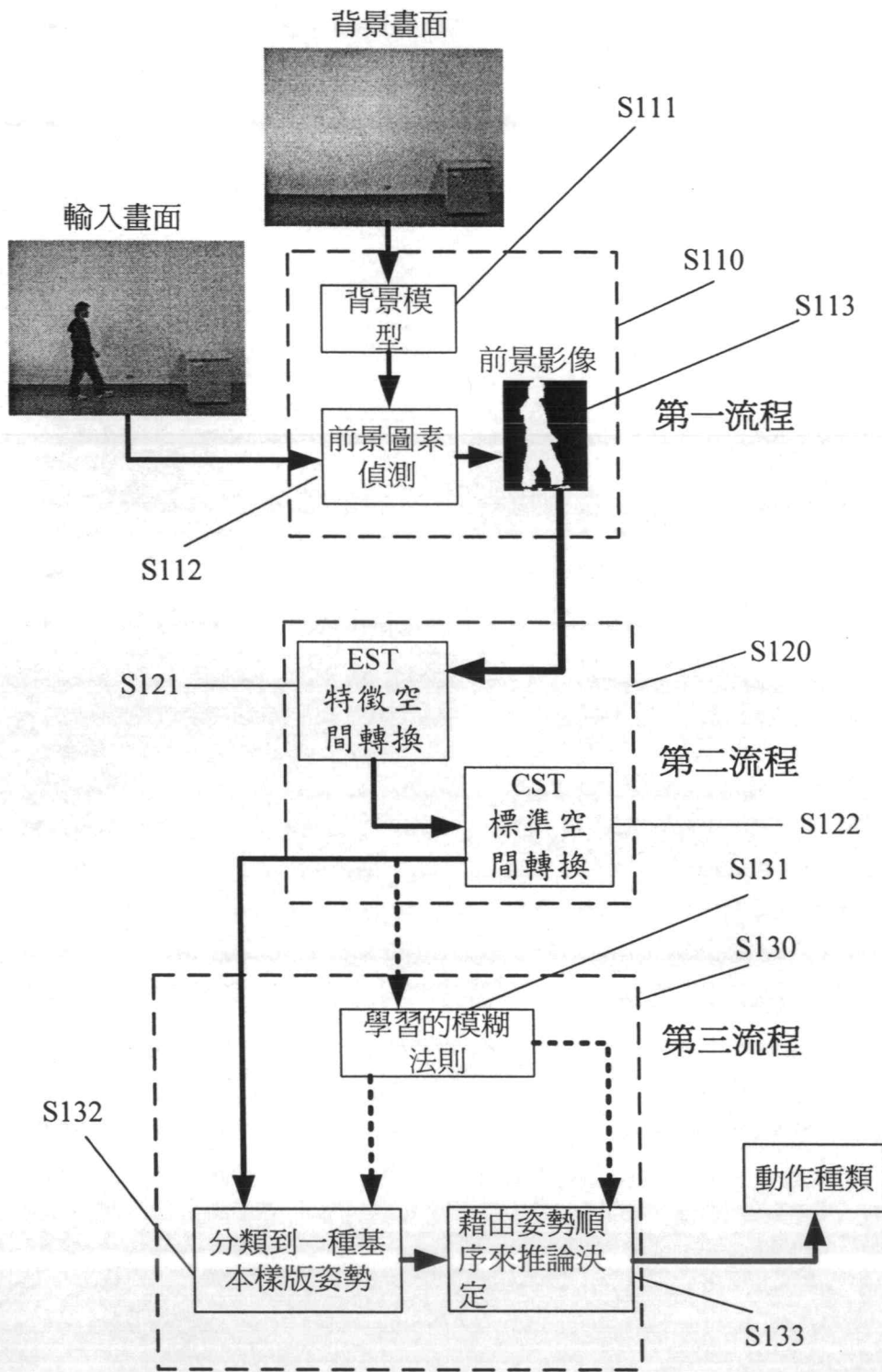
將該攝影裝置以固定頻率的時間間隔擷取影像，並在降低抽樣頻率後化為一序列的影像輸入畫面，序列中的每一個輸入畫面並與該背景畫面做比對以取出前景影像；

將每一張取出的前景影像實施空間轉換，然後將至少連續兩張實施空間轉換後的前景影像，轉變成一組時序姿態序列；以及

將這組時序姿態序列透過模糊法則推論為基礎的人類動作識別來分類出成某一個動作類別。

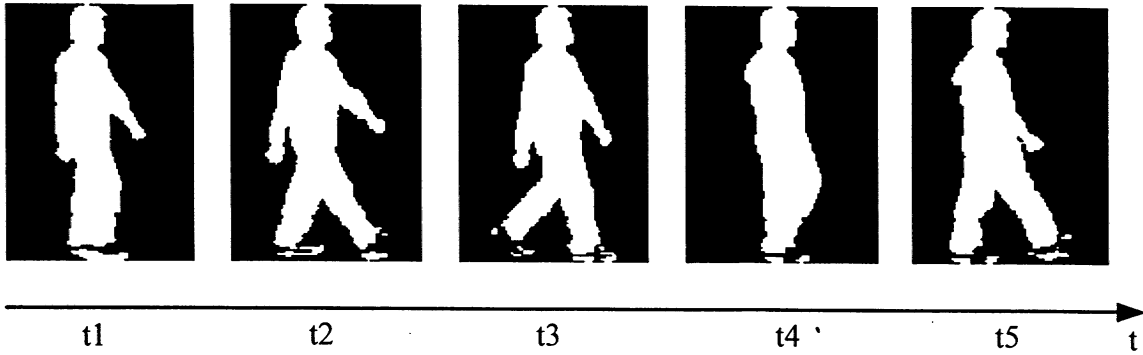
2. 如申請專利範圍第 1 項所述之人類動作辨識方法，其中該空間轉換包含特徵空間轉換與標準空間轉換。
3. 如申請專利範圍第 1 項所述之人類動作辨識方法，其中該取出的前景影像為二值化影像。
4. 如申請專利範圍第 1 項所述之人類動作辨識方法，其中該模糊法則透過將至少一張影像序列之較低抽樣頻率做為輸入，並結合至少一張前後基本人型姿勢所發展的人類動作辨識模糊法則。
5. 如申請專利範圍第 1 項所述之人類動作辨識方法，其中該模糊法則能夠結合時間序列上的資訊，並且可以容忍不同人做相同動作上的差異。
6. 如申請專利範圍第 1 項所述之人類動作辨識方法，其中該模糊法則包含可以學習到姿勢轉換關係的人類動作隱藏模式。

7. 如申請專利範圍第 1 項所述之人類動作辨識方法，其中該人類動作識別是在標準空間中完成的。
8. 如申請專利範圍第 1 項所述之人類動作辨識方法，其中該背景模型使用相連影像相除法來描述一個統計背景模型。
9. 如申請專利範圍第 8 項所述之人類動作辨識方法，其中該統計背景模型透過計算統計最大、最小的灰階值和相連影像灰階值相除的最大比值以獲得一個背景模型。
10. 如申請專利範圍第 8 項所述之人類動作辨識方法，其中該統計背景模型的背景影像像素都以三個統計數值來代表：最小灰階強度值、最大灰階強度值和最大相連影像灰階值相除比值。
11. 如申請專利範圍第 1 項所述之人類動作辨識方法，其中該人類動作識別利用高斯型態的隸屬函數，來表示每一張影像相對於每一個姿勢類別的可能性。

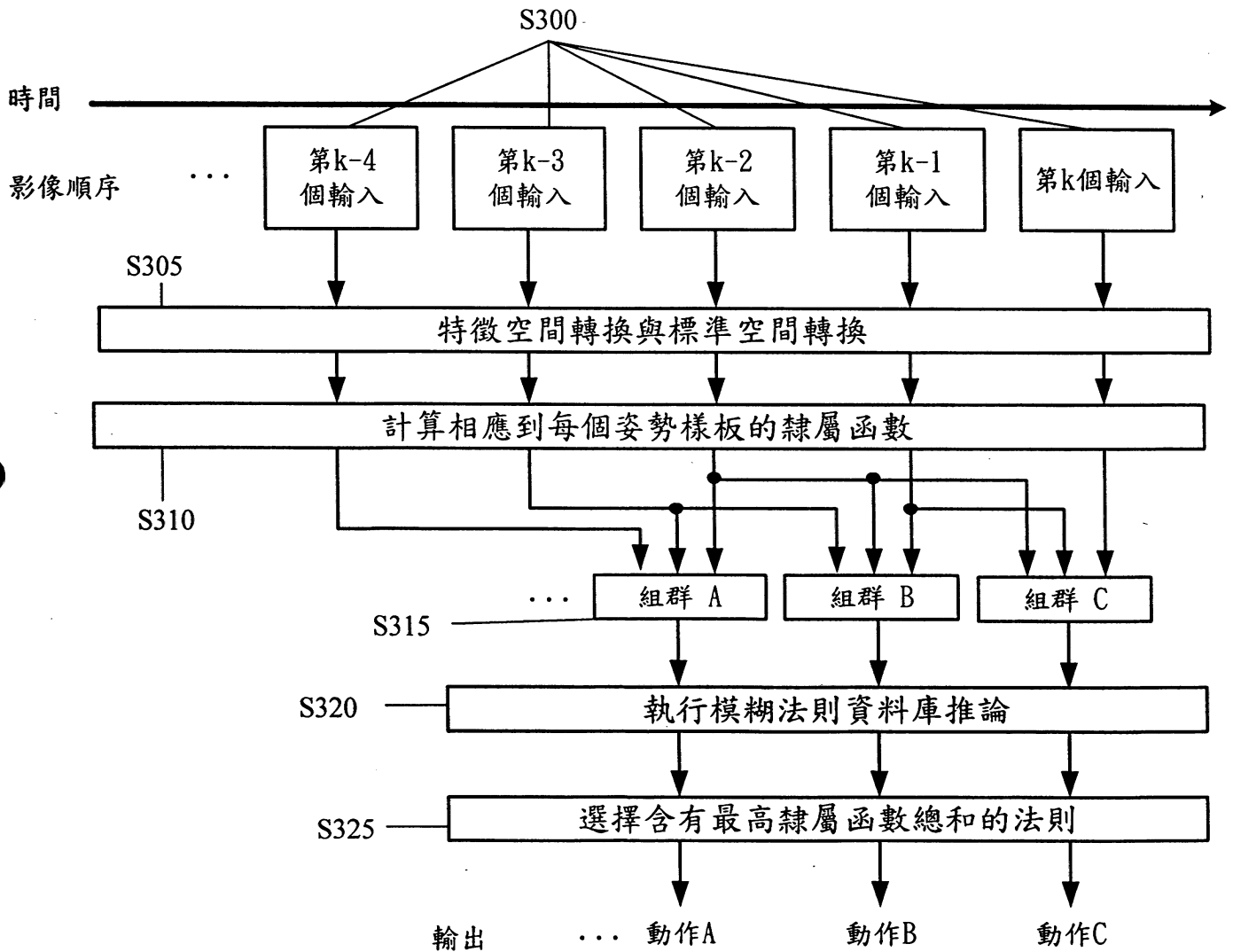


第1圖

樣版影像P₁₈ 樣版影像P₁₉ 樣版影像P₂₀ 樣版影像P₂₁ 樣版影像P₂₂



第2圖



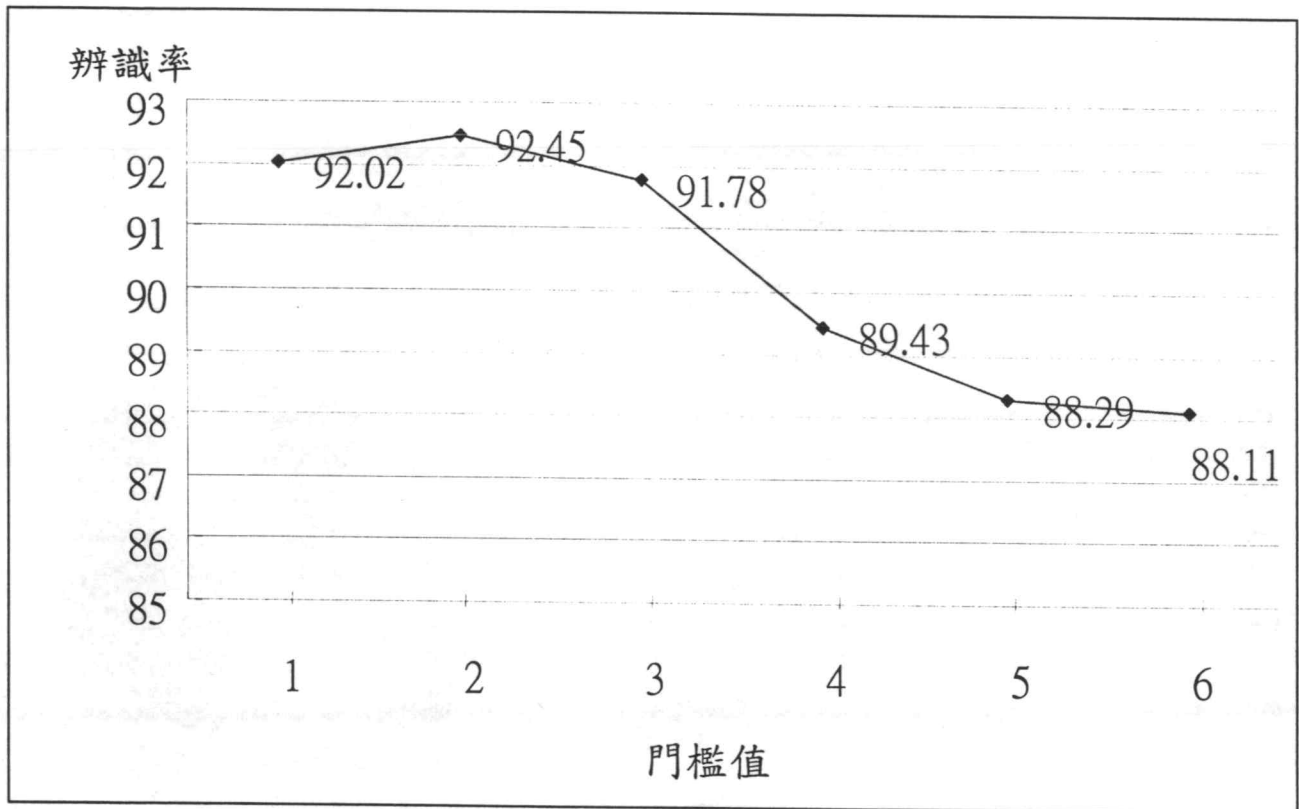
第3圖



(a)

(b)

第4圖



第5圖