コンピュータビジョンと
イメージメディア 103－14
（1997．1．24）

インタラクティブシステム構築のための実時間ジェスチャ認識の一手法

渡辺 孝弘 † 李 七雨 †* 谷内田 正彦 ‡

†(財) イメージ情報科学研究所
‡大阪大学 基礎工学部
* 全南大学 (韓国)
E-mail:twata@image-lab.or.jp

本論文では，インタラクティブシステム構築に適した動画像からの実時間ジェスチャ認識手法について述べる．本手法は，ジェスチャを行なう特定部位の画像をKL展開を用いて低次元のジェスチャ空間内で表現することによって，大量の動画像情報を効率的に実時間で処理するものである．まず，ジェスチャを行なう特定部位の画像がMaskable Tempalte Model(MTM) を用いて実時間でロバストに抽出される．MTMとは様々に変形する物体との柔軟なマッチングが行なえるように，従来のテンプレートマッチングの手法を我々が改良したものである．次に，その抽出された部位画像系列を低次元のジェスチャ空間にマッピングすることによりジェスチャ曲線として表現し，この曲線とモデルジェスチャ曲線を比較することによってジェスチャを認識する．ここでのジェスチャ空間は，モデル画像系列をKL展開することによって前もって構成されている．我々は，本手法を利用して実時間インタラクティブシステムー仮想指揮システムーを実現した．本システムは，ユーザが行なう指揮者のジェスチャを認識して，計算機が奏でる音楽を実時間で制御するもので，本手法の有効性を示すものである．

A Method of Real-Time Gesture Recognition for Interactive System

Takahiro Watanabe † Chil Woo Lee †* Masahiko Yachida ‡

† Laboratories of Image Information Science and Technology
1-4-2, Shinsenri-Higashimachi, Toyonaka, Osaka, Japan 565
‡ Faculty of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka, Osaka, Japan 560
* Department of Computer Engineering, Chonnam National University

This paper presents a novel gesture recognition method for a real time interactive system. This method recognizes a user's gesture by comparing a model gesture with an input one in a 'gesture space'. The space is a low-dimensional space constituted by the Karhunen-Loeve expansion of model image sequences which represents the human part to make a gesture. To extract an interesting region of input image robustly in real time, we developed the Maskable Template Model which is an improved template matching method for gesture recognition. Using our recognition method, we realize a real time interactive system, the Virtual Conductor System, which can control music played by a computer using gesture recognition results.

1 はじめに

人の意図や感情の表現手段の中でジェスチャの果たす役割は想像以上に大きい．そのため，コンピュータによるジェスチャの認識は，人を対象とする多くのアプリケーションの実現のために必要とされている．特にユーザにとってより自然で使い易いヒューマン・マシン・インタラクティブシステム構築のためには，連続画像から実時間でジェスチャを認識する実用的な手法が強く求められている．

これまでにも多くのジェスチャ認識手法が提案されているが，これらの手法は大きく，トップダウン的な処理によるモデルベース的手法とボトムアップ的な処理による特徴ベース的手法とに分けられる．モデルベース的手法としては，Rohr[1] や Kakadiaris[2] らのモデルを入力画像にフィッティングさせ関節角を求めて，その情報からジェスチャを認識するものなどがある．これらの手法は，対象の関節角が求められることから応用性は高いが，1) 処理が複雑なことが多く実時間処理が難しい，2) ノイズの多い実画像に対するフィッティングが安定しないなどの実用化への問題がある．

特徴ベース的な手法では以下のようなものが提案されている．大和 [3] らは，画像を細かく分割してそれぞれの領域から特徴ベクトルを求め，その時系列情報から HMM を用いてジェスチャを認識している．Darrell[4] らは，ジェスチャをビューモデルの集合と画像系列との相関パターンで表現し，入力画像系列の相関パターンとモデルジェスチャの相関パターンを Dynamic Time Warping(DTW) を用いてマッチングさせてジェスチャを認識する手法を提案している．また，この手法は特殊ハードウエアを用いて実時間処理が可能となっている．

これらの特徴ベース的手法は，実時間処理の可能性から最も実用的な手法である．しかしながら，ここで挙げた手法には認識すべき特定部分（例えば，手や腕など）のセグメンテーションについての効果的な手法は述べられていない．実用的なジェスチャ認識手法では，認識処理だけではなくセグメンテーションなども含めて一連の効果的な処理が必要である．また，ここで挙げた特徴ベース的手法は，ジェスチャの識別に主眼が置かれており，そのジェスチャがどのように行なわれたのかを認識するには適していない．しかしジェスチャ認識においては，ジェスチャの行なわれ方がユーザの感情や隠れた意図などを表すことがあるため，識別だけではなくジェスチャの行なわれ方も認識できるようなアプローチが望まれる．

本論文は，インタラクティブシステム構築に有効な実時間ジェスチャ認識手法について述べている．本手法は特定部位のセグメンテーションから認識処理までの一連の処理を実時間で行なう実用的な手法である．また，本手法はジェスチャの識別に加えてそのジェスチャがどのように行なわれたかを認識するのにも適している．以下，**2** では人の特定部位の抽出処理について述べる．本抽出処理は我々が開発した Maskable Tempalte Model(MTM) を用いて行なわれる．**3** ではその抽出された画像系列の認識処理について述べる．ここでの認識処理は，モデルジェスチャの画像系列を KL 展開することで低次元のジェスチャ空間を求めて，その低次元の空間内でモデルジェスチャと入力画像系列を比較することにより実時間ジェスチャ認識を可能にする．**4** では，本認識手法によるジェスチャ認識実験と本手法によって実現された実時間インタラクティブシステム－仮想指揮システム－について述べる．

2 Maskable Template Model による部位抽出

ジェスチャ認識における特定部位のセグメンテーションの問題を解決するために，我々は Maskable Template Model(MTM) を開発した [5]．MTM は従来の 2 値画像におけるテンプレートマッチングの手法を改良したもので，様々な形状に変化する腕などの特定部位を，少ないテンプレートでロバストに抽出する．

2.1 Maskable Template Model (MTM)

図 1は，左斜め前方から撮影した人（図3参照）の上半身を抽出するための MTM の例である．図 1に示すように，MTM は 3 値から構成される．入力 2 値画像が V_{p0}（黒領域）と V_{p1}（白領域）の 2 値をとるとすると，MTM は V_{p0} と V_{p1} の値をとるマッチング領域Ω_pと，V_{p0} と V_{p1} との中間値 V_m をとるマスク領域Ω_m（グレー領域）とから構成される．MTM を用いてテンプレートマッチングを行なうと，マッチング領域における相違度のみが求められ，マスク領域における相違度は無視されるようになる．しかも，その相違度を求める計算量は従来のテンプレートマッチング手法と等しく，特殊な処理を必要としない．

その理由を以下に示す．MTM：T_iと入力画像 I の座標 (x,y) における相違度 $d(T_i)$ は，マッチング領域における相違度 d_{Ω_p}とマスク領域における相違度

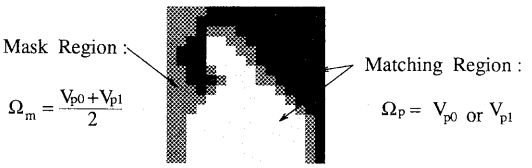


図 1: 上半身抽出用の Maskable Template Model の例

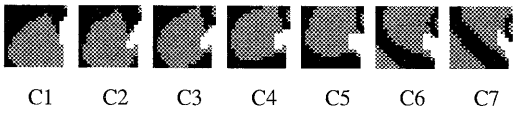


図 2: 右腕を抽出するための MTM の例.

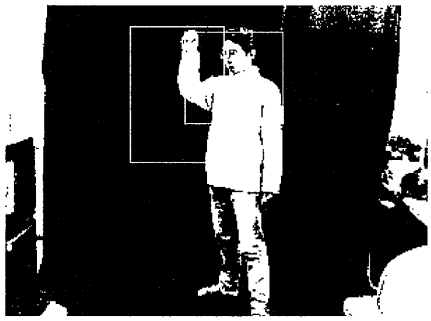


図 3: 切り出される上半身領域と腕領域の例.

d_{Ω_m} とを用いて以下のように表せる.

$$\begin{aligned} d(T_i) &= \sum_{m,n} |I(x+m,y+n) - T_i(m,n)| \\ &= \sum_{T_i(m,n)\in\Omega_m} |I(x+m,y+n) - T_i(m,n)| + \\ &\quad \sum_{T_i(m,n)\in\Omega_p} |I(x+m,y+n) - T_i(m,n)| \\ &= d_{\Omega_p}(T_i) + d_{\Omega_m}(T_i) \end{aligned} \quad (1)$$

ここで，マスク領域の値 V_mが，V_{p0}と V_{p1}の中間値であるため，d_{Ω_m}は以下のように入力画像の値とは無関係に，ある定数で表すことが出来る.

$$\begin{aligned} d_{\Omega_m}(T_i) &= \sum_{T_i(m,n)\in\Omega_m \wedge I(x+m,y+n)=V_{p0}} |V_{p0} - V_m| + \\ &\quad \sum_{T_i(m,n)\in\Omega_m \wedge I(x+m,y+n)=V_{p1}} |V_{p1} - V_m| \\ &= VS_{\Omega_{m(i)}} \\ &= K_i \end{aligned} \quad (2)$$

ただし,

$$V = \frac{|V_{p0} - V_{p1}|}{2} \quad (3)$$

で $S_{\Omega_{m(i)}}$は T_iのマスク領域の大きさを表す. よって, 相違度 $d(T_i)$ は以下の式のように書き換えられる.

$$d(T_i) = d_{\Omega_p}(T_i) + K_i \quad (4)$$

ここで，K_iの値は入力画像の値とは無関係であるため，MTM が複数ある場合は，各 MTM 間のマスク領域の大きさ $S_{\Omega_{m(i)}}$ を等しくなるように設定すれば式 (4) 中の K_iの値は $d(T_i)$ において無視できるようになる. よって最終的に式 (4) は以下のように書き換えられ，従来と同じ計算量でマッチング領域の相違度が求められる.

$$d'(T_i) = d_{\Omega_p}(T_i) \quad (5)$$

ここで，$d'(T_i)$ は単純化されたマッチング結果を表す.

2.2 Maskable Template Model による人体部位抽出

MTM におけるマスク領域の効果は，人体の部位抽出において以下のような効果をあげる.

図 1において，値 V_{p1}をとる白い領域は人の上半身に対応し，値 V_{p0}をとる黒い領域は背景に対応する. その上半身の領域の回りにマスク領域（グレー領域）をおくことにより，カメラの深さ方向の変化などに伴う見かけの大きさの変化への対応が可能になる. さらに，図 1の左側の右腕の可動範囲をカバーしているマスク領域によって，様々に右腕を動かしている人を抽出することができるようになる. つまり，図 1の 1 枚の MTM で，ある程度見かけの大きさが異なる場合でも，また右腕を適当に動かしている人でも抽出できるのである.

図 2は右腕を抽出するための MTM である. ただし，ユーザはここでも図 3に示すように左斜め前方から撮影されるものとする. これらの MTM において，白い部分は肩から肘までの腕の部分である上腕部を表し，それに接するグレーの扇型の部分は肘から先の腕の部分である前腕部の可動範囲を表している. (それ以外のマスク領域は各モデル間のマスク領域の面積を等しくするように，また，背景のノイズの影響を除くために設定されている.) つまりこれらの MTM は，上腕部の状態に合わせて変化し得る前腕部をマスク領域とすることによって，ユーザの腕が様々に変化しても，この 7 枚の MTM のみで右腕を抽出することができる.

図3はこれらのMTMを用いて右腕領域が切り出されるようすを表している．その手順は，まず図1の上半身MTMを用いて画像中から人の上半身を探索し，もしそれが検出されたら画像中に人が存在するとしてその位置から限定された範囲（右腕が存在すると思われる範囲）で図2の右腕MTMを用いて右腕が探索される．図中の右側の正方形が上半身のモデルが一致した位置を表しており，左側の長方形が切り出される右腕領域を表している．図2のMTMでは，各モデルの空間的な位置が合っていないので，各モデルの肩の位置によって空間的位置を合わせて切り出すために長方形になっている．この処理は我々の研究所で開発したReMOT-M[7]などのテンプレートマッチング専用ハードウエアを用いて実時間で行なうことができる．

3 ジェスチャ空間

MTMによってセグメンテーションされる部位画像は，あるジェスチャを行なう画像系列から多くの枚数が得られるため，そのままでは実時間内で処理するのは難しい．そこで我々は情報圧縮の手段としてKL展開を用いる．部位画像系列をKL展開することによって低次元のジェスチャ空間を構成し，その空間内でジェスチャを効果的に実時間で認識する．

3.1 KL展開によるジェスチャ空間の構成

ジェスチャ空間は，モデルとなるジェスチャ画像系列を以下に示すようにKL展開することによって構成される．

まず，切り出された一枚の部位画像を，ラスター走査してその画素値を要素とするベクトル$\mathbf{x}$を作る．

$$\mathbf{x} = [x_1, x_2, x_3, \cdots, x_m]^T \tag{6}$$

ただしここで，mは画素数（我々の実験では1536個）を表す．次にq枚の$\mathbf{x}$によって構成されるp番めのジェスチャ部位画像系列$\mathbf{X}_p$を以下のように表現する．

$$\mathbf{X}_p \equiv [\mathbf{x}_{p,1}, \mathbf{x}_{p,2}, \mathbf{x}_{p,3}, \cdots, \mathbf{x}_{p,q}] \tag{7}$$

すると，認識すべき全てのジェスチャ部位画像系列の集合は行列$\mathbf{X}$によって

$$\mathbf{X} \equiv [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \cdots, \mathbf{X}_P] \tag{8}$$

と表せる．そこで全てのジェスチャ部位画像系列の要素ベクトルから平均ベクトル$\mathbf{c}$を計算し，それぞれの部位画像系列のベクトルとその平均ベクトルとの差の集合を行列$\mathbf{Y}$として以下のようにして得る．

$$\mathbf{Y} \equiv [\mathbf{x}_{1,1} - \mathbf{c}, \cdots, \mathbf{x}_{2,1} - \mathbf{c}, \cdots, \mathbf{x}_{P,Q} - \mathbf{c}] \tag{9}$$

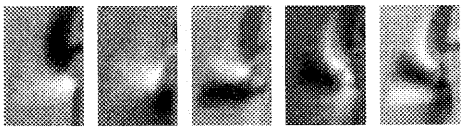


図4: 指揮者の2拍子叩き運動と3拍子平均運動による右腕部位画像の固有ベクトル

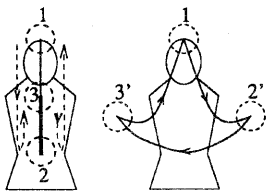


図5: (a) 2拍子叩き運動における腕の振りと (b) 3拍子平均運動における腕の振り

そして以下の式から$\mathbf{Y}$の共分散行列$\mathbf{U}$を求め，

$$\mathbf{U} \equiv \mathbf{Y}\mathbf{Y}^T \tag{10}$$

次の固有方程式

$$\lambda_i \mathbf{e}_i = \mathbf{U}\mathbf{e}_i \tag{11}$$

を解き，固有値λ_iと固有ベクトル$\mathbf{e}_i$を求めると，ジェスチャ空間（例えばk次元）は上位k個（$(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k)$）の固有値に対応する固有ベクトル$(\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k)$を基底ベクトルとすることにより得られる．

通常，各要素（ここでは一枚の画像）間の相関が高い場合は，上位の少ない次元でもとの集合（ここでは全ての部位画像系列）の情報をほとんど表すことができる．あるジェスチャを行なう場合の画像間でもかなり相関が高いので，上位の少ない次元の情報でもとのジェスチャを表すことができる．(4.1の実験で用いたデータでは，上位3次元で元のデータ1536次元の約65%，上位6次元で約81%を表していた．) そのため，ここでのジェスチャ認識に利用するのも上位の数次元のみである．(4.1の実験では上位3次元)

図4は，図3のように左斜め前方から撮影されたユーザが，指揮者の2拍子叩き運動と3拍子平均運動[6]を行なった時，図2のMTMによって切り出され

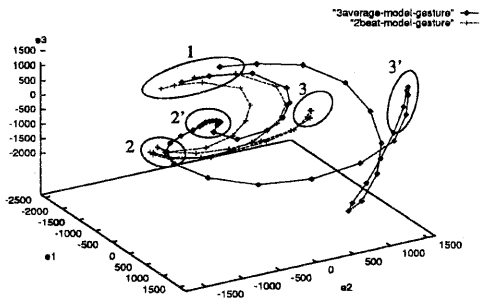


図 6: ジェスチャ空間上の 2 拍子叩き運動と 3 拍子平均運動のジェスチャ曲線

た右腕の部位画像系列から作成した，固有ベクトルを表したものである．2 拍子叩き運動とは図 5(a) に示すように腕を頭上中央から腹部前方へ（1 拍目），そして胸部前方まで上げてからもう一度腹部前方へ下ろし（2 拍目），そしてまた頭上中央へと比較的リズミカルに腕を振る運動である．3 拍子平均運動とは図 5(b) に示すように腕を頭上中央から左前方へ下ろしてさらに少し上げ（1 拍目），そして腹部前方を通り右前方へ（2 拍目），そして右前方で少し下げまた頭上中央へ（3 拍目）と，比較的静かに腕を振る運動である．

図 4から分かるように，ジェスチャによって腕が通る部分が上位の固有ベクトルでよく表現されている．

3.2 ジェスチャ曲線

ジェスチャ空間内では，あるジェスチャを表す部位画像系列は以下のようにして曲線として表現される．モデルジェスチャの画像系列中の 1 枚の画像を以下の式を用いて，ジェスチャ空間上に投影する．

$$\mathbf{g}_{p,q} = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k]^T (\mathbf{x}_{p,q} - \mathbf{c}) \qquad (12)$$

すると，1 枚の画像はジェスチャ空間上の 1 点として表せるので，あるジェスチャにおける部位画像系列をすべて投影すると，その空間内で離散的な点の集合が得られる．実際のジェスチャは連続であるため，本来はこれらの点も連続であると考えられることから以下の式によってそれぞれの点を時間的に平滑化して，

$$\mathbf{g}'_{p,q} = 1/n \sum_{i=0}^{n-1} \mathbf{g}_{p,q-i} \qquad (13)$$

それらの点を結ぶと，ジェスチャを表す部位画像系列は一つの曲線として表現される．ここで n は平均をとるためのフレーム数（実験においては n=5）を表す．図 6はモデルの 2 拍子叩き運動と 3 拍子平均運動を前節で求めたジェスチャ空間上（上位 3 次元）で表したものである．点線が 2 拍子叩き運動を表し，実線が 3 拍子平均運動を表している．それぞれの曲線において番号をつけた位置は図 5 の腕の位置番号と対応しており，ジェスチャの特徴がよく捉えられているのが分かる．

図 7において，実線は 3 拍子平均運動のモデルジェスチャ曲線を各固有ベクトル（上位 3 つ）に対する時間変化で表したものである．((a) が第 1 主成分，(b) が第 2 主成分，(c) が第 3 主成分の得点を表す．) 点線は 3 拍子平均運動を行なった入力ジェスチャ曲線を表している．

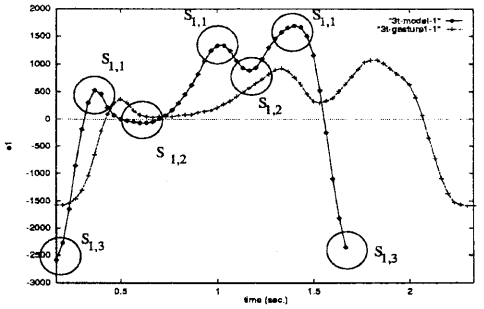
3.3 ジェスチャ空間におけるジェスチャ認識

ジェスチャ空間とモデルジェスチャ曲線を利用して，入力部位画像系列を認識する手法を以下に述べる．

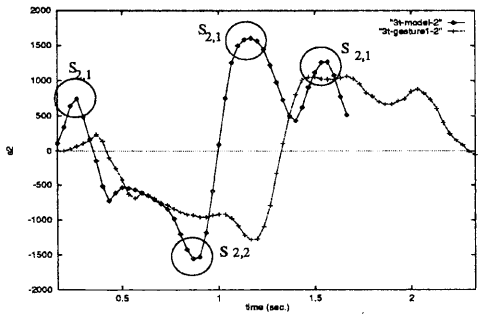
式 (12), (13) を用いて入力部位画像系列をジェスチャ空間に投影して，入力画像曲線を作成していく．そして認識処理を速く効果的に行なうために，それぞれの次元におけるジェスチャ曲線と入力画像曲線の特徴点について比較を行なう．ここでの特徴点とは，それぞれの次元のジェスチャ曲線においてある範囲内でとる極値を用いている．図 7において，円で囲まれた極値がその特徴点を表している．例えば図 7(b) において，特徴点 S_{21} は主成分得点が 0 から 2500 の間においてとる極大値を表し，特徴点 S_{22} は主成分得点が-1000 以下の極小値を表す．(その範囲外の極値は無視される．) これらの特徴点のとり方は，様々な人のジェスチャにおいても比較的安定した点を我々が経験的に選択したものである．このような特徴点によって，ジェスチャ曲線はジェスチャ特徴点系列 $S(p)$ として，例えば

$$S(p) \equiv \{S_{1,1}, S_{2,1}, \cdots\} \qquad (14)$$

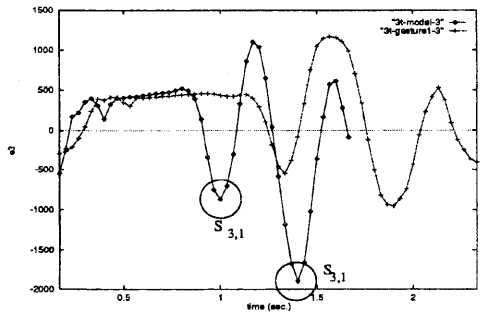
のように表せる．そこで，入力画像曲線によって作られるジェスチャ特徴点系列中にモデルジェスチャ曲線のジェスチャ徴特点系列が現れれば，そのモデルジェスチャが認識されるようにする．この手法では入力されたジェスチャとモデルジェスチャの速さが異なっていても認識可能であり，また，入力曲線が通過するある特徴点とある特徴点の間の時間を調べることによって，それら特徴点が表すポーズ間のジェスチャの速さを調べることも可能である．(4.2 のシステムではこの手法によって指揮者ジェスチャのテンポを認識している．)



(a)



(b)



(c)

図 7: モデルジェスチャ曲線（実線）とある入力ジェスチャ曲線（点線）をそれぞれの固有ベクトルにおいて時間軸上に表したグラフ

このように本手法は，まず大量の情報をもつモデルジェスチャ部位画像系列を KL 展開して低次元のジェスチャ空間構成し，その空間上でモデルジェスチャを表現しておく.（ここまでの処理の計算コストは非常に高いが，認識処理の前にあらかじめ行なうので実時間認識には影響しない.）そして，入力画像系列を専用画像処理装置などを用いて実時間でこの空間上に投影して，その空間上での特徴的な点において入力データを調べることによって効果的に認識処理を行ない，実時間ジェスチャ認識を実現している.

4 実験

4.1 ジェスチャ認識実験

これまでに述べてきた認識手法を用いて，ジェスチャ認識実験を行なった．認識するジェスチャとして，2 拍子叩き運動と 3 拍子平均運動を用いた.（図 5参照）ユーザは暗幕の前のほぼ決まった位置でジェスチャを行ない，位置，フォーカス，ゲインなどパラメータが固定された TV カメラによってジェスチャが撮影される．入力される画像は固定された閾値で 2 値化される．服装についての制限はしていない.（各人バラバラであった.）モデルジェスチャ曲線は一人のユーザがそれぞれのジェスチャを 3 回腕の振りの大きさを変えて行なって作成した．ジェスチャ特徴点 $s_{i,j}$ は固有ベクトルの上位 3 成分を用いて全部で 6 点（第 1 次元で 3 点，第 2 次元で 2 点，第 3 次元で 1 点）を選択した．3 拍子を認識する場合は，6 個の特徴点を用いて，長さ 13 のジェスチャ特徴系列 $S(1)$ を定義した．2 拍子の場合も 6 個の特徴点を用いて，長さ 14 のジェスチャ特徴系列 $S(2)$ を定義した.（ここでは，2 つのジェスチャ特徴点が等しいものであった.）また，切り出される右腕部位画像の大きさは原画像を縦横 1/2 に間引いて 32x48 の大きさとした.

実験はモデル作成者を含めて 6 人に対して，始めにジェスチャを説明して，それぞれのジェスチャを 5 回行なってもらった．モデル作成者のデータは，モデル作成時に用いたデータとは服装などにおいて異なるものを用いた．実験結果は，認識率が 80%，誤認率が 0%，どちらとも認識しなかったのが 20%であった．被験者のうち（モデル作成者も含めて）3 人では，90%以上の認識率であったが，1 人は認識率が 50%（2 拍子叩き運動は 100%認識されたが，3 拍子平均運動について全く認識されなかった.）であった.（結果の検討は 4.3.）

図 7において，点線は一人の被験者の 3 拍子平均運動のジェスチャ曲線を表す．かなりモデル曲線と

似た特徴になっているのが分かる．

4.2 アプリケーション

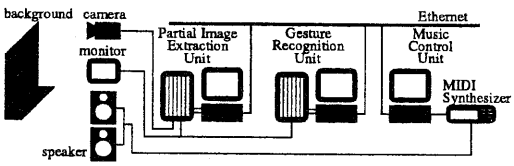


図 8: 仮想指揮システムの概略図

我々はこのジェスチャ認識手法を用いて，実時間インタラクティブシステム：仮想指揮システムを作成した．このシステムは，ユーザが行なう指揮者のジェスチャを認識して，さらにそのジェスチャからテンポ情報も抽出し，計算機が奏でる音楽を実時間で制御するものである．図 8は仮想指揮システムの概略図である．以下その構成ユニットについて説明する．

- **●部位画像抽出部：** 入力画像を閾値処理してシルエット画像を作成し，MTM を用いて部位画像を抽出する．この処理は我々の研究所で開発したテンプレートマッチング専用ハードウエア ReMOT-M[7] を用いて実時間で行なう．ここで抽出された部位画像が次のジェスチャ認識部に送られる．

- **●ジェスチャ認識部：** 前もって作成しておいた固有ベクトルを格納しておき，次々に入力される部位画像をジェスチャ空間にマッピングしていく．その処理は datacube 社の MaxVideo200 を用いて実時間で行なう．そして，本論文で述べたモデルジェスチャ曲線と入力ジェスチャ曲線との比較処理をホストコンピュータ (SS20) で行ない，ジェスチャ認識を行なう．その認識結果は Ethernet を通じて音楽制御部に送られる．

- **●音楽制御部：** 本ユニットは，MIDI シンセサイザーとホストコンピュータ（パーソナルコンピュータ PC9801FA）によって構成され，あらかじめ持っている音楽データを演奏する．ジェスチャ認識部から送られてくる認識結果によって，MIDI シンセサイザーにホストコンピュータから信号を送り，音楽を制御する．

本システムでは，ユーザに前もって演奏される音楽の拍子を知らせているので，それにあったジェスチャが正しく認識されれば，テンポ情報を認識して音楽が制御されるようになっている．また，テンポをとる腕（右腕）と反対の腕（左腕）について [5] によるポーズ推定手法を用いてポーズ推定を行ない，その情報によって音楽のボリュームも制御できる．

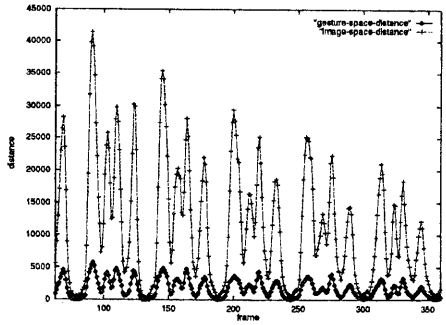


図 9: ジェスチャ空間上での距離変化と画像空間上の距離変化

4.3 考察

4.1 のジェスチャ識別実験において，認識率が 90% 以上のユーザがいる一方，ある一人は 50%の認識率であったのは，実験に用いたジェスチャが普段用いないものであり，説明からだけでは各人の受け取り方が微妙に異なったために，実際のジェスチャがかなり異なったためと思われる．しかしながら，認識率の低いユーザでも 4.2 の仮想指揮システムを使用していくと，初めのうちはうまく操作できなかったが，他のユーザのジェスチャを見たり，他のユーザからのアドバイスをもらったり，また何度か操作してジェスチャを習熟することにより，認識率は急速に向上した．実時間インタラクティブシステムにおいては，ユーザが試行錯誤しながら操作を繰り返すことで，そのシステムの特徴を素早く学習してそれに合わせた使い方ができるようになると思われる．このようなことから，本手法は現在の認識率で仮想指揮システムなどのインタラクティブシステムに十分活用できると思われる．

さらに，ジェスチャ空間は低次元でも基の画像空間の情報を十分に表しているため，ジェスチャ曲線を詳しく解析することで，ジェスチャについての様々な情報を得ることができる．例えばジェスチャの大きさ（腕の振りの大きさなど）の認識もその一つである．一般にジェスチャの大きさは画像空間におけるジェスチャ曲線の距離によって表現できる．KL 展開による線形変換では低次元で基の空間の情報を十分に表すことができるので，ジェスチャ空間での距離

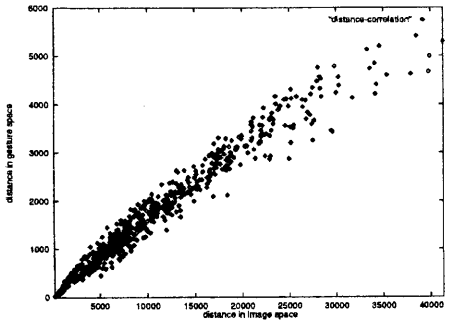


図 10: ジェスチャ空間と画像空間の距離の関係を表すグラフ

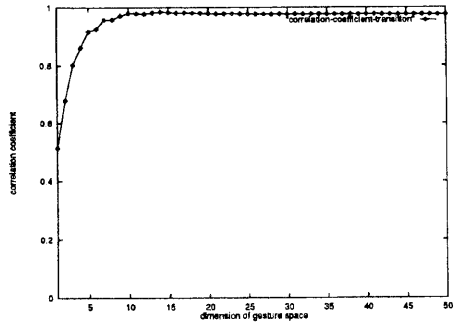


図 11: ジェスチャ空間の次元数と画像空間とジェスチャ空間との距離の相関係数を表すグラフ

を調べることでジェスチャの大きさを知ることができる．図9は3拍子のジェスチャの大きさを様々に変えて行なった場合に，画像空間（１５３６次元）上の距離の時間変化（点線，変化が大きい方）と上位１０次元のジェスチャ空間上での距離の時間変化（実線，変化が小さい方）を表したグラフである．このグラフから２つの空間上の距離の変化がほぼ一致しているのが分かる．図10はそれらの空間上での距離の関係を表したグラフである．横軸が画像空間上での距離，縦軸がジェスチャ空間上での距離を表している．このグラフから２つの空間上での距離の関係はが強い相関関係にあるのが分かる．(この場合の相関係数は0.98002である.）図11は距離を求めたジェスチャ空間の次元数と，そのジェスチャ空間での距離と画像空間（１５３６次元）での距離との相関係数をグラフに表したものである．この場合，上位１０次元のジェスチャ空間で画像空間距離とジェスチャ空間距離の相関係数がほぼ飽和しているので，上位１０次元でのジェスチャ曲線での距離を調べれば，ジェスチャの大きさを十分に認識できる．このようにして，ジェスチャ空間の上位数次元でのジェスチャ曲線を解析することで，その他の様々なジェスチャ情報も認識できるのではないかと思われる．

5 むすび

本論文はインタラクティブシステムのための実時間ジェスチャ認識手法について述べた．本手法は，ジェスチャを行なう特定部位の抽出，その抽出された特定部位画像の認識までの一連の処理を実時間で行なう．また，我々は本手法を利用した実時間インタラクティブシステム－仮想指揮システム－を実現し，本認識手法の有効性を示した．

今後，4.3 に述べたようなジェスチャ曲線の解析手法を検討し，様々なジェスチャ情報を抽出してより自然で使い易い実時間インタラクティブシステム構築に利用していくつもりである．

参考文献

[1] K. Rohr : "Towards Model-Based Recognition of Human Movements in Image Sequences", CVGIP: Image Understanding, vol.59, num.1, pp.94-115, (1994)

[2] I. A. Kakadiaris, D. Metaxas, Tuzena Bajcsy : "Active Part-Decomposition, Shape and Motion Estimation of Articulated Objects: A Physics-Based Approach",Proc. CVPR'96, pp.980-984 (1994).

[3] 大和, 倉掛, 伴野, 石井 : "カテゴリー別ＱＶを用いたＨＭＭによる動作認識" , 信学論 (D-II), **J77-D-II**, 7, pp.1311-1318 (1994).

[4] T. Darrell, A. Pentland : "Space-Time Gestures" ,Proc. CVPR'93,pp.335-340 (1993).

[5] T. Watanabe, C. W. Lee, A. Tsukamoto, M. Yachida : "A Method of Real Time Gesture Recogniton for Interactive System" ,Proc. ICPR'96, **III**, pp.473-477 (1996).

[6] 高階正光 : "指揮法入門１" , 音楽之友社 (1979).

[7] C. W. Lee, A. Tukamoto, K. Hirota, S. Tsuji : "A Visual Interaction System Using Real-Time Face Tracking" ,Proc. 28th Asilomar Conference on Signals, Systems & Computers, pp.1282-1286 (1994).