

## 動画像からの人物動作計測と認識

岩井 儀雄<sup>†</sup> 谷内田 正彦<sup>†</sup>

現代の計算機は音声や映像などの各種メディアを扱えるようになり格段に進歩した。しかし、人間の意図や感性にマッチした情報機器インタフェースはまだ実現していない。より良いインタフェースのためには人間情報を取り出し、理解する計算機が必要である。そこで、人に優しいマンマシンインタフェースの実現を目指して、身体ジェスチャを計測・認識する手法が研究、開発、実用化されている。本論文では、非接触な観測が可能であるカメラを利用した身体動作の計測・認識手法について筆者らの研究室で進めてきた研究について述べる。

### Measurement and Recognition of Human Motion from Image Sequences

YOSHIO IWAI<sup>†</sup> and MASAHIKO YACHIDA<sup>†</sup>

Computer can directly treat audio and visual information at present. However, human-computer interface which adapts to human intension or kansei is not realized. A system which understands human behavior is needed for better human-computer interface. Methods for sensing of human behavior are, therefore, researched, developed and applied for human-computer interface. In this paper, we describe methods for posture estimation and gesture recognition by using a vision system to actualize such interface.

#### 1. はじめに

日常社会において、人間は知らず知らずのうちにジェスチャをし、言語以上に意志を伝達することがよくある。人間は言語以外にも種々のコミュニケーション手段を持っており、それらが機能的に大きな役割を担っている証である。それらはノンバーバルコミュニケーションと呼ばれ、言語とは異なる性質を多く備えている<sup>1)</sup>。そして、人間はそれらのノンバーバルコミュニケーションを駆使して、情報交換、相互理解を行っている。

一方で、現代の計算機は音声や映像などの各種のメディアを扱えるようになり格段に進歩したが、機械とのインタフェースはマウスやキーボードといった従来の入力装置しか利用できない。人間と機械との間には、直接的で自然なインタフェースが確保され情報を円滑に受渡す必要がある。

自然なインタフェースとはいろいろ考えられるが、入力面においてだけ考察すれば、人間の行為が直接計算機に入力できることや、身体的拘束が少ないことなどがあげられる。人間の行為とは発話、ジェスチャ、

表情、思考などが考えられるが、本論文では身体ジェスチャに焦点を当てる。身体ジェスチャを利用したインタフェースとしては、対象の選択や指示、場所の指示、仮想対象の操作、身振りによる対話などが考えられる。これらは、機械との対話のためのコマンド、身振りインタフェースといえる。

身振りインタフェースを支える技術には装着型センサによるものと、非接触型センサとに大別できるが、接触型センサはユーザに身体的拘束を与え、人間本来の動作というものを制限してしまう。これでは、自然なインタフェースを実現したことはならない。一方、非接触型センサには動画像処理を用いて、身体情報を推定するものが多い<sup>2)~6)</sup>。非接触型装置の利点は、ユーザに身体的拘束をいっさい与えないことにあり、将来的には自然なインタフェースを実現できる装置であるといえる。

##### 1.1 画像処理による人物動作計測

人物を撮像しその動きを計測する研究は、コンピュータビジョンの分野では関節物体の運動として、1980年代以降にその研究成果が多く報告されている<sup>7)~18)</sup>。それらの研究のアプローチは物体中心モデル手法と観測者中心モデル手法に大別できる。物体中心モデル手法は人体に関する詳細なモデル化を行うことにより、部分的にデータが欠損してもモデルで補完して認識を行

<sup>†</sup> 大阪大学  
Osaka University

えるという利点があるが、複雑なシステムのモデル化が困難なこととその計算コストが大きいことが欠点としてあげられる。他方、観測者中心モデル手法は複雑なモデル化を避けて、センサ入力を直接的なモデルとして利用するアプローチであり、処理時間が節約できリアルタイムアプリケーションに向くが、計算時間の代わりにデータ量が増大してしまう欠点がある。

筆者らの研究室では、物体中心手法として形状モデルを用いて人体の姿勢を推定する手法について研究を行ってきた。人体モデルは自由度が多いので1枚の画像からでは一意な姿勢を定めることができない。そこで、ある定められた評価関数が最小になるパラメータを求めることで姿勢を推定する。また、動きを用いることで次フレームの姿勢を予測してパラメータ探索範囲を限定したり、区分線形的に近似して姿勢を推測したりして、最小になるパラメータを探索する手法を開発してきた。また、観測中心の手法としては、手形状モデルを利用してアスペクトを生成し手形状を認識する手法<sup>19)</sup>や、人物の動きを利用してジェスチャを認識する手法を開発してきた<sup>20)</sup>。筆者らの研究室で行われた観測者中心の手法は参考文献21)にまとめられているので、以降では、形状モデルを利用した物体中心の手法を主に紹介する。

## 2. 形状モデルによる姿勢復元

最初に物体の輪郭に着目し、形状モデルとの照合を行うことで関節物体の形状を復元する手法について紹介する。画像中には物体輪郭に相当するエッジが存在するので実際にはエッジ画像と形状モデルの輪郭との照合を行う。形状モデルを利用する際の欠点は、その計算コストと形状モデルの獲得であるが、本手法では画像上での特徴点を追跡し運動復元することで未知のモデルパラメータを減少させ、形状モデルの獲得については最初の画像から計測することで解決する<sup>22)</sup>。

### 2.1 手形状モデル

人体の姿勢を求めるためには人体の形状をモデル化する必要がある。人体形状のモデル化には様々な手法があるが、筆者らは姿勢を復元する目的であるので単純なモデルを用いる。手形状モデルを図1に示す。各指の中心軸に対し、指の幅部分を楕円近似し、指表面を3角形の平面で近似する。

### 2.2 画像特徴量

画像からは、手のシルエット輪郭、手の内部のエッジ強度画像、インタレストオペレータによるコーナ特徴を抽出する。コーナ特徴はフレーム間で追跡され、形状モデルの指先と指の付け根に対応付けられている。

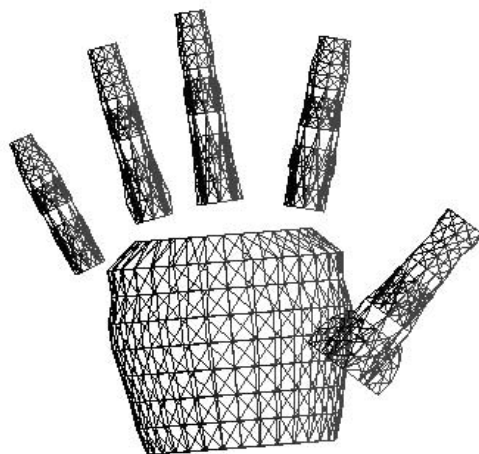


図1 手形状モデル

Fig.1 Shape model of hand.

手のシルエット輪郭とエッジ強度画像は形状モデルとの照合に利用される。

### 2.3 形状モデルと画像との照合

形状モデルと画像との最適な照合は評価関数  $C$  を最小化することで求める。最小化の手法は様々であるが、焼きなまし法を用いた。その理由は、本研究で用いた評価関数に平坦なところが多く、gradient 方向から最適解に収束することが難しいことと、焼きなまし法であれば、温度を上昇させることにより平坦部を脱出する可能性があるからである。欠点としては、ランダムな探索法であるので非常に計算時間がかかることである。

評価関数  $C$  は、特徴点(指先、指の付け根)の位置拘束と、形状モデルから生成される輪郭形状上に画像中にどの程度エッジが存在するかを表す輪郭線拘束と、各指の付け根が離れすぎないような拘束の3つの拘束の和からなる。評価関数が最小になるには、画像上の特徴点位置と、形状モデルの特徴点を画像に投影した位置とが一致し、形状モデルから生成した輪郭に完全にエッジが一致する必要がある。

手と指の位置および形状モデルから生成した画像を図2に示す。姿勢推定結果を入力画像に投影した画像を図3に示す。実際の手の場合、指の形状も指の屈曲に合わせて変形するので一般円筒では表現できない。しかし、デッサン用に用いられる手の木製模型を用いて本手法を適用した場合に得られる精度と比較しても、同程度の推定精度で復元できることが分かっている<sup>22)</sup>。

本実験では Indigo2 (SGI 社製, R4000 150 MHz) と画像入力装置 (SGI 社製, Galileo video), ビデオ

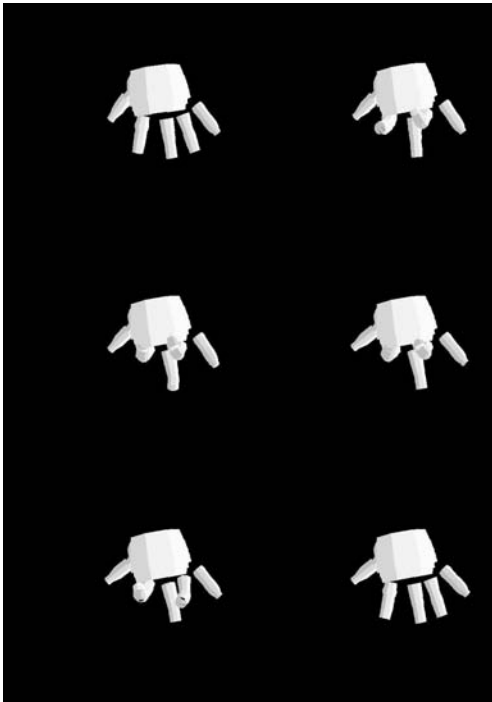


図 2 推定した形状と位置の生成画像

Fig. 2 CG images of estimated shape and positions.

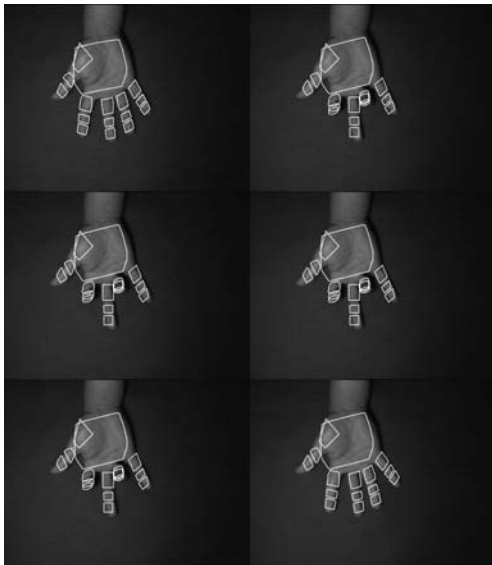


図 3 推定した姿勢の投影画像

Fig. 3 Input images on which estimated shape and positions are superimposed.

ディスク (SONY 製, LVR-3000AN) を用いた。処理時間は 1 フレームあたり画像入力から指の姿勢推定まで約 6 分であり, 評価関数を最小化する手法は計算時間がかかることが分かる。

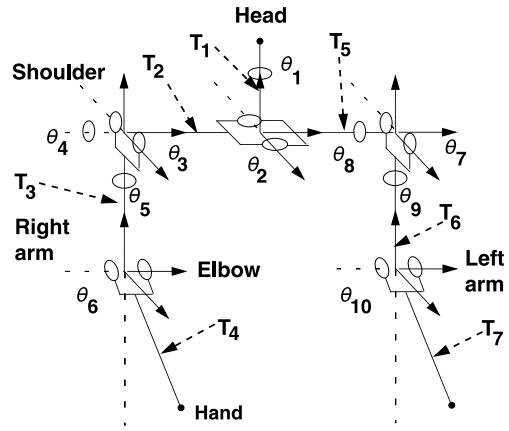


図 4 上半身形状モデル

Fig. 4 Structure model of upper body.

### 3. 動きモデルを利用した姿勢推定

評価関数の最小化処理は, 評価関数を繰り返し計算するので時間がかかる。最適化にかかる時間を少なくするためには, 繰返し数を減らしたり, 評価計算を並列化したりする必要がある。もし, 次フレームの姿勢を予測により最適姿勢の近傍に近付けることができれば, 繰返し演算が減少するので計算時間を減らすことができる。さらに, 探索においても評価関数のより良いパラメータ近傍に最適解が存在するとして, 優先的に探索すればより効率が上がると思われる。

一方, 人物の動きは関節物体モデルのモデルパラメータの時系列として表すことができ, 1 つの動きはモデルパラメータ空間内の 1 つの軌道で表現することができる。似たような運動を行えば, パラメータの時系列は空間内で似た軌道を描くと考えられる。したがって, この軌道を動きモデルとして複数の動きモデルを用意し, そのモデルから計算されたパラメータを対象の姿勢パラメータの探索初期値として利用すれば, 最適計算の効率を上げることができる。ここでは人体の上半身の動きのモデルを作成して, 動きモデルを選択しつつ次フレームの姿勢を予測する<sup>23),24)</sup>。

#### 3.1 上半身モデル

上半身の形状モデルとして図 4 のようなスティックモデルを構築する。各特徴点間, 関節間の長さが既知であるとき, モデルの姿勢はパラメータ  $\Theta = \{\theta_i | 1 \leq i \leq 10\}$  により決定される。 $\Theta$  の範囲は人の関節角がとりうる範囲を考慮して表 1 のように設定している。

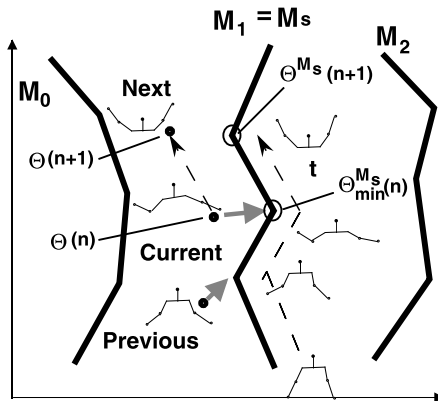
#### 3.2 動きモデルによる次フレーム予測

対象の動きに対応する動きモデル  $M_s$  をあらかじめ用意してある動きモデル  $M_i, (i = 1, \dots, n)$  のなか

表 1 上半身形状モデルパラメータの可動範囲

Table 1 Range of structure model parameters.

$\theta_i$ (deg.)	min	max
$\theta_1$	-180	180
$\theta_2$	-15	15
$\theta_3$	-90	90
$\theta_4$	-45	180
$\theta_5$	-90	90
$\theta_6$	0	180
$\theta_7$	-90	90
$\theta_8$	-45	180
$\theta_9$	-90	90
$\theta_{10}$	0	180



Model Parameter Space  $\theta$

図 5 モデル選択およびパラメータ予測

Fig. 5 Model selection and parameter prediction.

ら選択する．具体的には，モデルの生起確率  $P(M_i)$  を観測データをもとに計算し，その確率が最大となる  $M_s$  を選択する．

$$M_s = \arg \max_{M_i} (P(M_i)) \quad (1)$$

なお，モデルの生起確率  $P(M_i)$  の計算の詳細は，参考文献 23) に記述されているので簡単な説明にとどめる．観測データから各動きモデルまでの最短距離を計測し，それを重みとして逐次更新することでつねに最新の状態を維持する．

選択されたモデル  $M_s$ ，第  $n$  フレームにおいて推定されたパラメータ  $\Theta(n)$  より，第  $n+1$  フレーム目のパラメータ  $\Theta(n+1)$  を予測する．選択されたモデルの時系列上で次のフレームに対応する  $\Theta^{M_s}(n+1)$  を利用する (図 5)．

$$\Theta(n+1) = \Theta(n) + \Delta\Theta(t), \quad (2)$$

$$\Delta\Theta(t) = \Theta^{M_s}(n+1) - \Theta^{M_s}(n) \quad (3)$$

ここで， $\Theta^{M_s}(n)$  は選択されたモデル  $M_s$  上の点で，推定された姿勢パラメータ  $\Theta(n)$  の最近傍点である．

Priority Search Queue

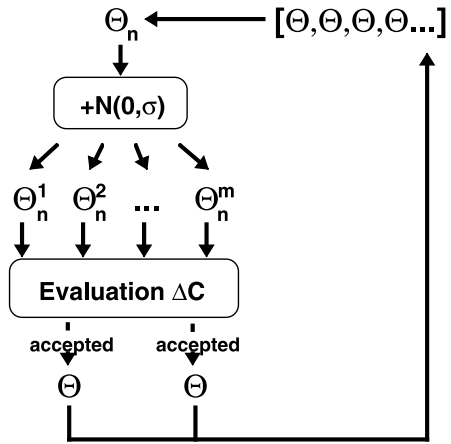


図 6 パラメータ最適化

Fig. 6 Parameter optimization.

### 3.3 パラメータ最適化

パラメータ予測により得られたパラメータ  $\Theta$  を初期値として  $\Theta$  を最適化する．画像特徴量としては，2章とは異なり，顔と手の3次元重心位置と人体のシルエット形状を用いる．シルエット領域と形状モデルの特徴点との照合および，上体から肘への距離を利用して最適なパラメータ探索を行う．動きモデルに基づき予測された  $\Theta$  を初期値とし，評価関数  $C$  を最小にする  $\Theta_{min}$  を探索する．探索には焼きなまし法を用いる．解候補部分は予測値 (初期値) の周辺にあるだろうという考えから，最大長  $Q_N$  の優先度付き待ち行列を焼きなまし法に加えてビーム探索を行う．また，評価関数  $C$  の計算は並列に処理できるので，可能な場合は同時に探索点を  $m$  点計算する．

なお，予測が解候補の近傍にあるためには， $P(M_i)$  の精度， $M_i$  の数，精度に依存し，近傍にない場合は収束するのに計算時間がかかってしまう．このようなときは，対象の動きに合わせた動きモデルを追加する必要がある．

図 6 に探索アルゴリズムの大まかな流れを示す．待ち行列から探索点  $\Theta_n$  を取り出し，乱数を生成して候補点  $\Theta_n^m$  を  $m$  点生成し，並列に評価計算を行い受理されたものをもう一度待ち行列に入れる．

### 3.4 特異点の回避

逆運動学で身体の動作を推定する場合，逆運動学の方程式のランクが不足し，解が得られないことがある<sup>25)</sup>．このようなランクが不足する特異点を回避する必要がある．特異点を回避する方法として  $\Theta$  の値

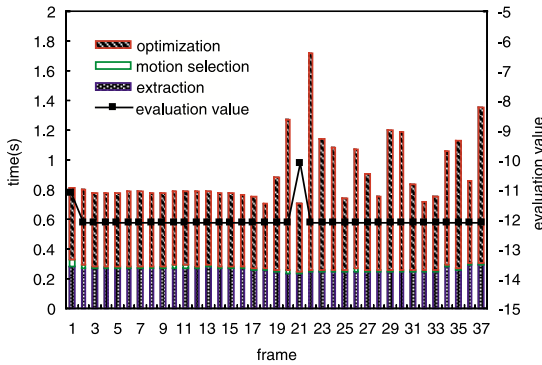


図 7 計算時間のサンプル

Fig. 7 The sample of computational time.

が特異点に落ちたとき、その範囲外に  $\Theta$  の値を誘導する。

図 7 に各処理の処理時間と最適化処理によって求めた最小の評価値を示す。推定されたパラメータが形状モデルのパラメータのとりうる範囲を超えた場合は、パラメータの最適化の探索回数に数えることなく再計算するため、通常より計算時間がかかることが分かる。また、特異点周辺では逆運動学により一意な姿勢が決定できないので、再計算が必要で時間がかかるのは不可避であるが、特異点を素早く避ける方法が構築できれば計算時間を縮小することは可能である。たとえば、形状モデルのパラメータがなす部分空間の境界部を平面パッチなどで近似し、入力データとの最近傍面に射影したりすることで素早く避けることが可能である。

320 × 243 画素の時系列ステレオ実画像に提案手法を適用した結果を図 8 に示す。パラメータ最適化において、探索しているパラメータは逆運動学で決定されない 4 パラメータであり、最大 500 回の繰返し探索を行った。パラメータ最適化に費やす計算時間は 0.6 秒で、そのときの姿勢推定誤差は 30 mm 程度であった<sup>23)</sup>。本手法は肌色分布をもとに 3 次元物体の対応付けを行っているために、頭と手の位置から上半身の姿勢だけを求めた。しかしながら、素足または靴など対応付けが容易な特徴がある場合には、下半身にも適用が可能である。そのような姿勢に対応するためには、人物モデルのパラメータを増加させる必要があるが、解の収束精度、速度に影響が生じる。本手法の応用分野としては、精密な姿勢を推定する必要はなく、ある程度姿勢誤差を吸収してくれる処理があるジェスチャ認識のようなアプリケーションが考えられる。



図 8 推定結果のサンプル

Fig. 8 Examples of estimation results.

#### 4. 区分的線形回帰モデルによる姿勢推定

手の位置や頭の位置などの観測値  $X$  と形状モデルの姿勢パラメータ  $\Theta$  の関係は順運動学問題を解くことにより写像  $A$  を用いて式 (4) で表すことができる。

$$X = A\Theta \quad (4)$$

入力される  $X$  空間上の点は式 (4) を逆に解くことにより  $\Theta$  空間上へ変換して直接姿勢を推定できるが、写像  $A$  は 1 対 1 写像の保証がないため解は一意に定まらない。

しかし、写像を区分的に近似して 1 対 1 写像を仮定することができれば姿勢を一意に推定することができる。そこで、あらかじめ人の動きモデルを大量に持っておきモデルとの類似度を用いて姿勢を推定することで効率的な推定を目指すことができる<sup>26)</sup>。

##### 4.1 姿勢推定のアイデア

姿勢推定システムを式 (6) に示す。

$$\tilde{\Theta}(t+1) = \begin{bmatrix} \mathbf{I} & \Delta \times \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tilde{\Theta}(t) \quad (5)$$

$$\hat{X}(t) = A\tilde{\Theta}(t) + u(t) \quad (6)$$

ただし、 $\tilde{\Theta}(t)$ 、 $\hat{X}(t)$  は時刻  $t$  における推定姿勢、観測値を表し、 $\mathbf{I}$  は単位行列、 $\Delta$  は 1/30[s]、 $A$  は順運動学行列、 $u$  はノイズを表す。

上のようなシステムの推定において広く用いられる方法としてはカルマンフィルタがある。しかしカルマンフィルタはシステムノイズが不偏な分布を持つことが仮定されており、今回のように  $A$  のランクが不足

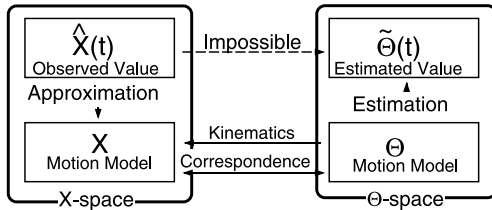


図 9 姿勢推定アイデア  
Fig. 9 Concept of posture estimation.

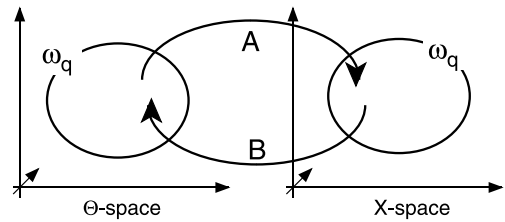


図 11 異次元空間の変換  
Fig. 11 Transformation between observation space and parameter space.

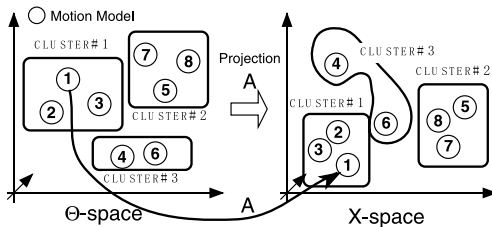


図 10 異次元特徴空間でのクラスタリング  
Fig. 10 Clustering of motion data and observation values.

し、 $u$  が不偏かどうか明らかでない場合は適していない。そこで図 9 に示すように、あらかじめ動きモデルからパラメータ空間  $\Theta$  と観測空間  $X$  との間に順運動学を用いて 1 対 1 に対応する代表点をいくつか求めておき、観測  $X$  に最も近い代表点から姿勢を推定する。

動きモデルのように大量データを整理する方法にはクラスタリングがある。そこで本論文では入力  $\hat{X}$  と動きモデル  $\Theta$  との類似度を計算するために、すべての  $\Theta$  を  $N$  個にクラスタリングする。具体的にはまず、すべての  $\Theta$  に関して式 (4) を適用することによって  $X$  を計算する。次にすべての  $\Theta$  について  $k$ -平均法により  $N$  個にクラスタリングする。評価距離にはユークリッド距離を用い、特徴量は位置と速度で正規化する (図 10 参照)。

クラスタリングの利点は、入力との類似パターン探索の高速化とジェスチャ生起確率のモデル化を同時に達成できる点である。まずクラスタ  $\omega_q$  に所属するジェスチャ  $M_i$  のクラス代表点として各ジェスチャのクラス中心を  $X$  空間で  $g_{q,M_i}$ 、 $\Theta$  空間で  $G_{q,M_i}$  とすると、 $\tilde{\Theta}$  は式 (7) で表せると仮定する。すなわち観測空間からモデル空間への写像行列  $B_{q,M_i}$  が求まれば姿勢を推定できることになる。

$$\Theta = B_{q,M_i} \begin{pmatrix} Y \\ 1 \end{pmatrix} + G_{q,M_i} + e \quad (7)$$

$$Y = X - g_{q,M_i} \quad (8)$$

ここで十分に近傍のモデルどうしならば  $\|e\|$  は十分に小さいと考えられ、 $X - g_{q,M_i}$  と  $\Theta - G_{q,M_i}$  に相関があると考えられる。そこで重回帰分析により求

められた  $B_{q,M_i}$  による写像を推定姿勢とする。図 11 に図示する。

本手法では生起確率最大のジェスチャだけでなく、その他のジェスチャの姿勢も推定する。生起確率が低くなるほど信頼できるデータ数が少なくなり、式 (7) の  $\|e\|$  は大きくなって写像の近似も悪くなると考えられるが、実際に選択されなかったジェスチャが起こった場合でも選択ジェスチャを切り替えることによって処理を続けることができる。姿勢推定を行う際、 $B_{q,M_i}$  が依存するのは所属クラスタ  $\omega_q$  と選択ジェスチャ  $M_i$  だけである。そこで毎フレーム重回帰分析を実行するのではなく、 $\omega_q$  と  $M_i$  に対応するすべての  $B_{q,M_i}$  をあらかじめオフラインで計算しておけばオンラインでは動きモデルを読み込んだり重回帰分析を行ったりすることなく姿勢を推定できる。これによって計算コストを抑え、効率良い処理を実現することができる。

#### 4.2 パラメータ最適化

3.3 節と同様に区分線形回帰により求めた値を探索初期値として最適化を行う。その際に、前フレームで最適化を行うことで得られた推定値も探索用の待ち行列に入れる。これは物体の静止を仮定した動きモデルである。動きが小さいときにはこちらの方が精度の向上が望める。なお、これらの初期探索点は優先度を決定するためにあらかじめ評価関数を計算しておく。

#### 4.3 姿勢予測

本手法による予測姿勢の誤差を右手について評価した結果を図 12 に示す。また最高生起確率の予測姿勢、静止を仮定した予測姿勢、線形予測による予測姿勢、本手法について真値と比較した結果を表 2 に示す。本手法では、静止モデルも動きモデルとして組み入れられているので、動きモデルだけの場合や、静止モデルだけの場合よりも性能が向上している。また、結果からは入力画像の人物の動きが小さかったため、静止モデルの方が優先的に選択されたが、動作速度が速い場合には動きモデルによる予測も効果を発揮した。

統計的な動きモデルを用いることによって各ジェス

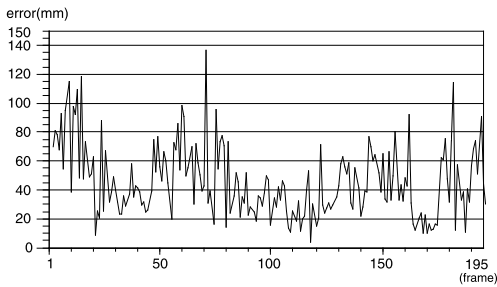


図 12 右手の推定値と真値との誤差

Fig. 12 Estimation errors of right hand position.

表 2 姿勢パラメータの推定誤差

Table 2 Average and covariance of estimation errors of model parameters.

-	動きモデル	静止	線形	区分線形回帰
平均 (mm)	116.5	47.9	116.9	45.9
標準偏差 (mm)	48.3	20.6	47.9	19.6

チャの生起確率を計算し、各ジェスチャに関して姿勢を推定することができる。そして、シルエット領域とのモデルマッチングによる最適化を行う際の探索初期点とすることができる。また実画像系列でも姿勢が推定できることを確認し、次元の異なる特徴空間の間に仮定した区分線形回帰モデルは十分に成り立つと考えられる。本手法はジェスチャを対象としたが、あらかじめ動きの分かっているような場合にはモデルを作ることによって様々な対象を扱うことができる。

## 5. 動き情報を用いたジェスチャ認識

ここでは、動き情報を利用してジェスチャを認識する手法を紹介する。動き情報は、対象のジェスチャによって発生している情報ならばどのようなものでもかまわない。今までの説明では、形状パラメータ  $\theta$  の動きを利用した物体中心の手法であったが、ここでは観測者中心の観測  $X$  を利用した手法を説明する<sup>20)</sup>。同様の議論は、形状パラメータの場合であっても成立する。

動き情報を利用するには人物を検出して動き情報の切り出しをする必要がある。ここでは、人物領域の切り出しが完了し、動き領域の切り出しができるとして話を進める。

### 5.1 動きベクトルの計算

動きベクトルの計算  $X$  には、勾配法、トークン法などがあるが、一般に計算時間がかかるのでリアルタイム処理には向いていなかった。しかし、近年のデジタルビデオ技術によりリアルタイムで動きベクトルを計算するハードウェアが普及してきた。ここでは、動



図 13 フロー抽出の例

Fig. 13 An example of flow estimation.

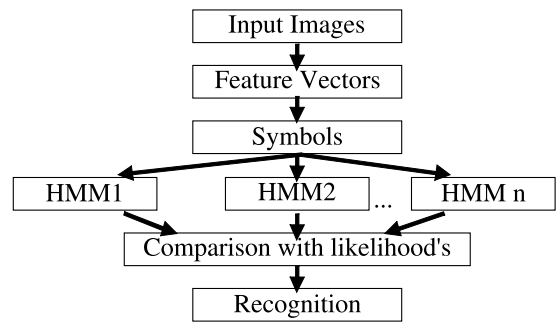


図 14 HMM を用いた認識のフレームワーク

Fig. 14 Framework for gesture recognition by using HMM.

きベクトルをリアルタイムで計算する富士通製トラッキングビジョン<sup>27)</sup>を用いる。図 13 にフローの抽出例を示す。

動きベクトルは大量のデータを発生するので、KL 展開を用いて入力データの圧縮を行う。ジェスチャは圧縮された部分空間上である軌道を描くので、その軌道と入力軌道との照合を行えばよい。

### 5.2 HMM によるジェスチャの認識

ジェスチャ認識は動きモデル  $M_s$  を選択することにより行うことができる。動きモデルの生起確率  $P(M_i)$  の構造は様々で、いろいろな実装が考えられるが、HMM を用いて学習により構成することもできる。HMM による認識の利点は、入力データの時間変動に対してロバストである点で、また、多人数に対しても学習により認識率を向上できる点である。ただし、この利点は欠点でもあり、認識率を高めるための学習データはかなりの量が必要である。

図 14 に HMM を用いた認識のフレームワークを示す<sup>20)</sup>。本手法は、カメラから得られる動画像において人物の動き情報に注目し、それをテンプレートマッ

チング法で動き領域を抽出することで、2次元のモーションベクトルを求める。それぞれのジェスチャにおいて各フレームごとにモーションベクトルを計算しKL展開することでジェスチャ空間を作成する。さらに、ジェスチャ空間に投影したモーションベクトルを量子化することで、シンボル列に変換する。変換されたシンボルの時系列をHMMに学習させ、各HMM内での尤度を計算し、最大となるHMMのモデルを選択することで認識を行う。

KL展開で得られた部分空間のクラスタリングには、VQ(ベクトル量子化)法で行う。代表点はベクトルの統計的分布を反映するように選ばなければならないが、この解析的手法は知られておらず、LBGクラスタリングアルゴリズム<sup>28)</sup>によって作成する。

認識用の各HMM  $M_i$  はあらかじめ学習画像を用いてそれぞれのジェスチャを学習しておく必要がある。観測されたシンボル系列  $O = O_1 O_2 \dots O_t$  から、各HMMが系列  $O$  を出力する確率  $P(O|M_i)$  を計算し、その確率を最大にするHMMを選択することにより、ジェスチャ  $M_s$  を選択する。

$$M_s = \arg \max_i (P_j(O|M_i)) \quad (9)$$

## 6. おわりに

近年、身体動作を計測する技術が多数開発されて、身体動作データを容易に利用できるようになった。現在の実用化されている技術は接触型インタフェースが主で、アクティブセンサや赤外光を発するものが主である。動画解析により身体動作を実時間で求めるには、まだ安定性や精度という点で劣っており改良を必要とする。

本論文では、筆者らの研究室でこれまで行ってきた動画による身体動作計測・認識の手法をまとめた。筆者らは物体中心のモデル化においては評価関数最小化フレームワークに基づき身体動作推定を行う手法を開発してきた。評価関数の良さと最適解への収束速度が推定精度の良さに反映されるのであるが、筆者らは動きのモデル化を行い、次フレームの予測を行って計算時間の短縮を図った。また、動き情報を利用してジェスチャの認識も可能であり、そのようなシステムの開発も行った。本論文で紹介した手法は、室内のような安定した人物抽出が可能な場合に有効である。屋外や複数人物の場合には、安定に人物領域を抽出する必要があるためにまだ安定性に欠ける。

今後は人とのインタラクションが可能なように屋外における人物の安定抽出や、複数人物の抽出、処理の

さらなる高速化が必要である。

謝辞 本研究の一部は、日本学術振興会未来開拓学術研究推進事業(JSPS-RFTF99P01404)、文部省科学研究費補助金重点領域研究・奨励研究・学術創成研究(課題番号13GS0018)の補助を受けた。

## 参考文献

- 1) 黒川隆夫：ノンバーバルインタフェース，オーム社，東京(1994)。
- 2) Bobick, A.F. and Wilson, A.D.: A State-based Technique for the Summarization and Recognition of Gesture, *Intl. Conf. on Computer Vision*, Massachusetts, pp.382-388, IEEE (1995).
- 3) Wu, H., Chen, Q. and Yachida, M.: A Fuzzy Theory Based Face Detector, *International Conference on Pattern Recognition*, Vol.3, pp.406-410 (1996).
- 4) 王 梅, 岩井儀雄, 谷内田正彦: 動きモデルによる顔特徴点追跡と表情認識, 電気学会論文誌D, Vol.119-D, No.5, pp.699-706 (1999).
- 5) Pramadihant, D., Iwai, Y. and Yachida, M.: Integrated Person Identification and Expression Recognition from Facial Images, *IE-ICE Trans. Inform. & Sys.*, Vol.E84-D, No.7, pp.850-866 (2001).
- 6) ダデットブラマディハント, 岩井儀雄, 谷内田正彦: フレキシブル特徴照合による人物同定と表情認識, システム制御情報, Vol.14, No.9, pp.421-429 (2001).
- 7) Horn, B.K.P. and Weldon, E.J.: Determining optical flow, *Artificial Intelligence*, Vol.17, pp.185-204 (1981).
- 8) Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections, *Nature*, Vol.293, pp.133-135 (1981).
- 9) Brooks, R.A.: Symbolic Reasoning Among 3-D Models and 2-D Images, *Artificial Intelligence*, Vol.17, pp.285-348 (1981).
- 10) Tsai, R.Y. and Huang, T.S.: Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces, *IEEE Trans. Pattern Recognition and Machine Intelligence*, Vol.PAMI-6, No.1, pp.13-27 (1984).
- 11) Asada, M., Yachida, M. and Tsuji, S.: Analysis of Three-Dimensional Motions in Blocks World, *Pattern Recognition*, Vol.17, No.1, pp.55-71 (1984).
- 12) 加藤克己, 杉原厚吉, 杉江 昇: 正射影速度場からの関節物体の構造復元アルゴリズム, 情報処理学会研究報告 CV42-3, 情報処理学会(1986).
- 13) 木本伊彦, 梶谷昭彦, 安田靖彦: スティックモデルに基づく単眼視動画画像からの人体歩行運動



- の解析, 電子情報通信学会論文誌, Vol.J74-D-II, No.3, pp.376-387 (1991).
- 14) Shakunaga, T.: Pose Estimation of Jointed Structures, *Intl. Conf. on Computer Vision and Pattern Recognition*, Lahaina, Maui, Hawaii, pp.566-572, IEEE (1991).
- 15) 亀田能成, 美濃導彦, 池田克夫: シルエットを利用した手指の三次元形状推定法, 画像の認識・理解シンポジウム, Vol.II, pp.II-239-II-246 (1992).
- 16) Heap, T. and Hogg, D.: Towards 3D Hand Tracking using a Deformable Model, *International Conference on Automatic Face and gesture Recognition*, Killington, pp.140-145 (1996).
- 17) 島田伸敬, 白井良明, 久野義徳, 三浦 純: 緩やかな制約知識を利用した単眼視動画像からの関節物体の形状と姿勢の同時推定, 電子情報通信学会論文誌, Vol.J81-DII, No.1, pp.45-53 (1998).
- 18) 島田伸敬, 白井良明, 久野義徳: 確率に基づく探索と照合を用いた画像からの手指の3次元姿勢推定, 電子情報通信学会論文誌, Vol.79D-II, No.7, pp.1210-1217 (1994).
- 19) 渡辺 賢, 岩井儀雄, 八木康史, 谷内田正彦: カラーグローブを用いた指文字の認識, 電子情報通信学会論文誌, Vol.J80-DII, No.10, pp.2713-2722 (1997).
- 20) 畠 直志, 岩井儀雄, 谷内田正彦: 動き情報と情報圧縮を用いた口バスタなジェスチャ認識手法, 電子情報通信学会論文誌, Vol.J81-DII, No.9, pp.1983-1992 (1998).
- 21) 谷内田正彦, 岩井儀雄: インタラクションのための人物動作解析と認識, PRMU 99-139, 電子情報通信学会, 那覇 (1999).
- 22) 岩井儀雄, 八木康史, 谷内田正彦: 単眼動画像からの手の3次元運動と位置の推定, 電子情報通信学会, Vol.J80-D-II, No.1, pp.44-55 (1997).
- 23) 大垣健一, 岩井儀雄, 谷内田正彦: 動きと形状モデルによる人物の姿勢推定, 電子情報通信学会, Vol.J82-DII, No.6, pp.330-338 (1999).
- 24) Iwai, Y., Ogaki, K. and Yachida, M.: Posture Estimation using Structure and Motion Models, *Intl. Conf. on Computer Vision*, Vol.I, Corfu, pp.214-219 (1999).
- 25) 川崎晴久: ロボット工学の基礎, 森北出版 (1991).
- 26) 萬上圭太, 岩井儀雄, 谷内田正彦: 統計的な動きモデルを用いた人物の姿勢推定, システム制御情報, Vol.14, No.6, pp.330-338 (2001).
- 27) 内山 隆, 森田俊彦, 沢崎直之, 浅田和徳, 佐藤雅彦, 井上博充, 稲葉雅幸: 動き追跡処理システム「トラッキングビジョン」, *JRSJ*, Vol.16, No.1, pp.52-53 (1998).
- 28) Rabiner, L.R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proc. IEEE*, Vol.77, No.2, pp.257-285 (1989).

(平成 13 年 12 月 25 日受付)

(平成 14 年 3 月 8 日採録)

(担当編集委員 八木 康史)



岩井 儀雄 (正会員)

平成 4 年大阪大学基礎工学部情報工学科卒業。平成 6 年同大学大学院基礎工学研究科修士課程修了。平成 9 年同大学院基礎工学研究科博士課程修了。現在, 同大学院基礎工学研究科助手。コンピュータビジョン, パターン認識に関する研究に従事。博士 (工学)。IEEE, 電子情報通信学会, 日本ロボット学会各会員。



谷内田正彦 (正会員)

昭和 46 年大阪大学大学院工学研究科修士課程修了。同年同大学基礎工学部制御工学学科助手。助教を経て同学部情報工学科教授, 平成 6 年同学部システム工学学科教授。昭和 42 年~43 年デンマーク原子力研究所留学。昭和 47 年~48 年米イリノイ大学にて Research Associate。昭和 55 年~56 年西独ハンブルグ大学 Research Fellow。昭和 57 年米ミネソタ大学 CDC Professor。ロボット学会, 人工知能学会等会員。著書「ロボットビジョン」(昭晃堂, 大川出版賞受賞), 「コンピュータビジョン」(丸善, 編著)等。コンピュータ・ビジョン, 画像処理, 人工知能, 移動ロボット等の研究を行っている。工学博士。