研 究 解 説

インタラクションのためのコンピュータビジョン

Computer Vision for Human Computer Interaction

岡　兼司*・佐藤洋一**

Kenji OKA and Yoichi SATO

1. はじめに

この20年程度でコンピュータ技術は大幅な進歩を遂げたが，HCI（Human Computer Interaction）の根幹はほとんど進化することがなかった．しかし，近年ではコンピュータの小型化，遍在化が急速に進み，現在主流のインターフェースであるGUI（Graphical User Interface）では不十分になりつつある．これは，GUIではユーザの意識がコンピュータに向けられていることを前提としていて，空間に遍在するコンピュータとの自然なインタラクションを想定していないためである．こういった状況の中で，ユーザがコンピュータとのインタラクションにとらわれず，実世界において行われる作業に集中できるように設計されたインターフェースが提唱されている．このような考えに基づくインターフェースは実世界指向インターフェース［48］やPUI（Perceptual User Interface）［36］と呼ばれ，近年盛んに研究が行われている．

実世界指向インターフェースやPUIを実現する上で重要となる技術は，ユーザの動作を理解する技術である．特に，手の動作を追跡，認識する技術は非常に重要となる．例えば，ユーザが実世界の中で物体をつかむなどの作業を行う場合，中心的な働きを担うのは手である．また，人間同士でコミュニケーションを行う場面では，言語によるコミュニケーションとともに手を用いたジェスチャが大きな役割を果たしている．

ユーザの手や指先の動作を計測する研究は比較的早くから行われてきた．例として，VPL ResearchのDataGlove［44］が挙げられる．この研究では，ユーザにグローブ状のデバイスを装着させ，指の関節角や手の位置などをグローブに付属したセンサにより計測している．このようにデバイスをユーザに装着させて動作を計測する手法は，これ以外にも幾つか報告されている．デバイスを使用した，いわゆる接触型の手法の場合，動作を高速かつ正確に追跡することが可能であるという利点がある．しかしながら，デバイスに付属したケーブルなどがユーザに拘束感を与えると考えられ，特にユーザの自然な使用感が要求されるHCIの分野での応用には適していない．

こういった理由から，コンピュータビジョンを利用することにより，ユーザに拘束感を与えないような動作計測手法が提案されてきた．これらの手法の流れを図1に示す．

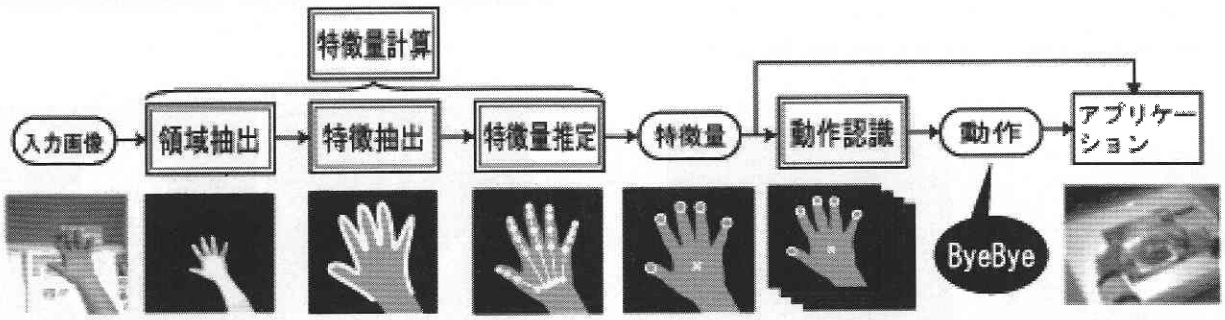

図1　画像処理の流れ

*東京大学生産技術研究所
**東京大学生産技術研究所　概念情報工学研究センター

まず，1台もしくは複数のカメラを用いてユーザを撮影する．次に，撮影された画像に対して数段階の処理を行うことにより動作の追跡，および様々な特徴量の計算を行う．さらに，得られた特徴量を利用して動作の認識を行う．以上の処理により得られた動作の特徴量や認識結果をHCIアプリケーションへの入力とする．これらの一連の処理ではユーザに特殊なデバイスを装着させるといったことがないために，ユーザの自然な使用感を損なわずに動作解析を行うことが可能となる．したがって，コンピュータビジョンを利用した非接触型の計測手法は，HCIの分野への応用に適した動作計測手法であると考えられている．

そこで本稿では，HCIへの応用に向けて必要となるコンピュータビジョンの各技術を図1の流れに沿って説明する．まず，第2章で手の動作の特徴量を計算する様々な技術を紹介し，次に第3章で，その結果を利用した手の動作認識技術について述べる．さらに第4章では，以上の技術をHCIに応用することを試みた研究例を紹介する．最後に，第5章でまとめを示す．

2. 動作の特徴量計算

手の動作の特徴量を計算する作業は幾つかの処理に分かれる．入力画像から追跡対象となるユーザ領域だけを抽出する処理，抽出されたユーザ領域画像から特徴を抽出する処理，そして内部モデルを利用して特徴量を推定する処理である．以下に，各処理で提案されている手法を紹介する．

2.1　ユーザ領域の抽出

ユーザ領域を抽出する手法としては非常に多くの手法が提案されている．以下でこれらの手法を紹介する．

・色情報の利用

色情報を利用する抽出手法としては，カラー画像（図2左）からユーザの肌の色に近い色領域を抽出することによりユーザ領域を特定する（図2右）手法，すなわち肌色抽出［30］［46］が代表的である．しかし，光源環境の変化によりユーザの肌色も変化するため，安定した領域抽出は困難である．この問題に対処するために，一定の照明を対象に向かって常に照射しておくなどの対策が取られる．

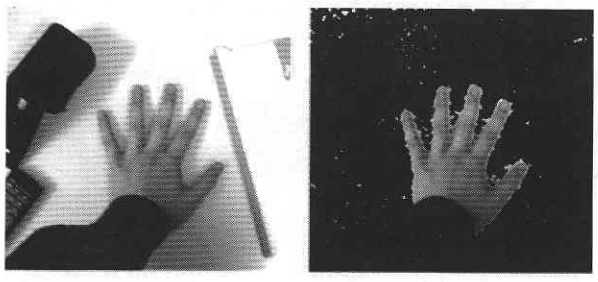

図2　肌色抽出

また，ユーザの手や指先にカラーマーカーを装着させ，指先などの抽出を安定かつ容易に行う手法［4］［19］も提案されている．しかしながら，ユーザにマーカーの装着を強制するという点でデバイスを利用する手法と同等であり，ユーザの自然な使用感を実現する上では問題がある．

一方で，色情報以外の要素も取り入れることにより，安定に領域抽出を行う手法が提案されている．例として，色と形状に関する統計的モデルを利用してユーザ領域を特定する手法［2］［22］［34］が挙げられる．

・画像間差分の利用

画像間の差分を利用する手法として，背景差分［31］［42］（図3）とフレーム間差分［40］［43］［50］（図4）が挙げられる．背景差分による抽出では，入力画像（図3左）と背景画像（図3中央）との差分を取ることにより前景画像だけを抽出する（図3右）．一方，フレーム間差分では，入力画像（図4左）において隣り合うフレーム同士の差分を取り，移動した領域の抽出を行う（図4右）．

これらの手法の場合，背景が複雑である場合や背景が一定でない場合，対象領域と背景が同様の色を持つ場合などにはユーザ領域だけを安定して抽出することが困難となる．そのため，背景を一定の状態や色に制限する必要があり，使用可能な環境が限定されるという問題が生じる．

・サーモグラフィの利用

サーモグラフィとは物体の温度差を輝度や彩度の変化で表した画像である．ユーザ領域を含む場面（図5左）を赤外線カメラにより撮影し，ユーザの体温付近の温度領域を抽出する（図5右）ことによりユーザ領域の抽出を行うことが可能である．この手法を利用した研究の1つにバーチャル歌舞伎［24］が挙げられる．バーチャル歌舞伎システ

図3　背景差分

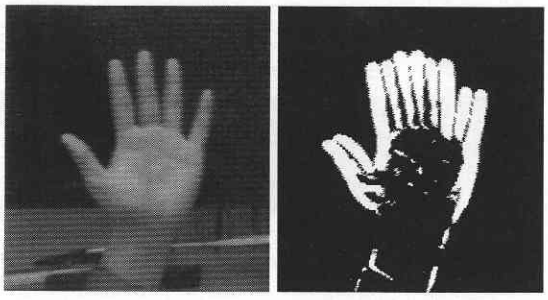

図4　フレーム間差分

ムでは，実時間でユーザの全身の姿勢推定を行っているが，ユーザ領域の抽出の際には赤外線カメラにより撮影された熱画像を利用している．また，Satoらはユーザの手領域を抽出するために同様の手法［29］［47］を用いている．

サーモグラフィを利用する場合，周囲の光源環境や背景の影響をほとんど受けないため，処理が安定しているという特徴がある．また，高速な処理も可能である．しかし，ユーザの体温と同等の温度を持つ領域も抽出されるので，面積フィルタの利用［29］［47］やカラー画像との併用によりさらに安定した領域抽出を行う必要がある．

・オプティカルフローの利用

オプティカルフロー［6］では，原画像（図6左）中の各画素の動きを表現した画像（図6右）を出力することが可能である．これにより対象の速度情報を取得できるので，抽出に成功した場合には同時に対象の追跡も可能となる．しかし，背景画像が動かないことや，追跡対象と背景の色が違うこと，追跡対象のテクスチャが鮮明であることなど制約条件も多く，また，厳密な領域抽出にもあまり適さない．

・深さ情報の利用

カメラからの距離（深さ）を利用してカメラに近い領域だけを抽出することにより，対象領域の抽出を行うことも可能である．深さ画像は抽出対象を含む場面（図7左）を2台のカメラで撮影し，ステレオマッチングを利用することによって生成される（図7右）．このようにして得られる深さ画像の1つとしてdense disparity map［14］が挙げられる．

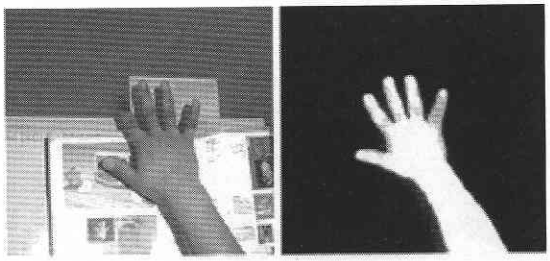

図5　サーモグラフィ

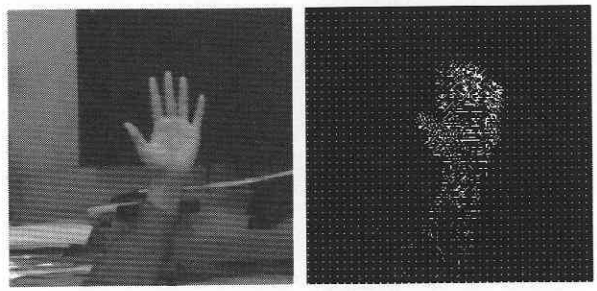

図6　オプティカルフロー

一方，非可視光領域を用いて深さを計測するカメラとして，モーションプロセッサ［3］が商品化されている．モーションプロセッサは搭載されているLEDから近赤外光を発射し，物体に反射した近赤外光をCCDカメラによって撮影する．このとき，遠方からの反射光は急速に弱まるため，深さの情報を得ることが可能となる．

深さ情報を利用する場合，周囲の光源環境などの影響が非常に少ないという利点があり，色情報による領域抽出の初期化にも用いられる［10］．一方で，ステレオマッチングによる深さ画像の場合，計算コストやステレオマッチング自体の精度などの問題点が挙げられる．

2.2　特徴抽出

前節で抽出したユーザ領域の画像からどういった特徴を抽出するかは最終的に必要となる特徴量や後節の特徴量推定の手法に応じて適切な選択が必要である．そこで本節では，抽出する特徴に関して分類し，それぞれの場合について紹介する．

・シルエット抽出

シルエットとは対象領域を2値化した画像である（図8）．ユーザ領域を抽出した画像からシルエットを抽出する作業は容易であるためしばしば用いられる［18］［32］［37］［39］［50］．しかし，カラー画像に2値化処理を行うことにより情報の欠落が起こり得るという問題がある．この問題に関して，手の3次元姿勢推定を行う際に，シルエット抽出により深さの情報が失われ，指の位置を正確に推定することが困難になるといった例が報告されている［18］．

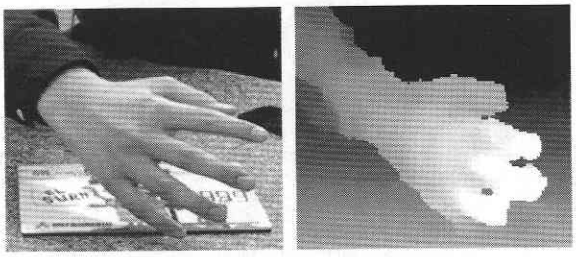

図7　深さ画像

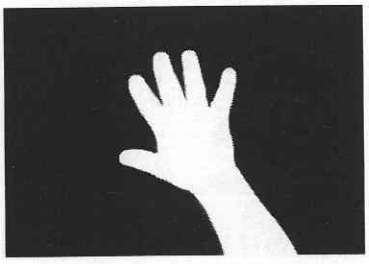

図8　シルエット抽出

・輪郭抽出

特徴として非常に多く用いられるものの1つに手や腕の輪郭（図9）が挙げられる．輪郭の抽出手法として，カラー画像や濃淡画像から直接的にエッジを抽出する手法［9］［49］が用いられる．また，エネルギー最小化の観点から物体の輪郭を抽出する手法としてはSnakes［16］が提案されていて，これを手の輪郭抽出に利用する研究［11］も行われている．抽出された輪郭は，後節の各モデルによる特徴量推定に広く利用される．

・指先抽出

指先も特徴として抽出されることが多い（図10）．指先抽出を行う際にはマーカー装着などを利用した装着型の手法がしばしば取られる［4］［19］．この手法の場合，比較的簡単な画像処理により高速かつ正確に指先を検出することが可能である．しかし，マーカーが自然なインタラクションの妨げとなることが予想される．

指先を抽出する非装着型の手法ではパターンマッチングを用いる手法が代表的である．テンプレートとしては，単純な円形テンプレート［29］［47］や実際の指先画像［5］などが用いられる．

別の指先検出手法としては，指先の形状的な特徴を利用した手法が提案されている．Maggioniらは，まず元の画像から手の輪郭を抽出し，その輪郭の曲率が一定以上である点を指先候補点として検出する手法［20］を用いている．

指先を特徴点として抽出する手法の最大の問題点はオクルージョンの問題（図11）である．これは，ある位置のカメラからは見えていた指先（図11左）が，別のカメラからは自分の手や指によって隠れてしまう現象（図11右）である．この問題の解決策として複数のカメラを選択的に利用する手法［46］が提案されている．これ以外にも，3次元構造モデルを利用してオクルージョンに対応することも可能である．

2.3　特徴量の推定

特徴量の推定とは前節で得られた特徴や手領域抽出画像，あるいは入力画像自体を用いて内部モデルとのマッチングを取り，必要となる特徴量を求める操作である．特徴量推定で使用するモデルには，3次元構造モデルと2次元形状モデル，画像モデルの3種類のモデルが存在する．各モデルを利用した特徴量推定手法について以下に説明する．

・3次元構造モデルの利用

3次元構造モデルとは，対象（手など）の3次元構造を表現したモデルである．特徴が抽出された画像と3次元構造モデルとのマッチングが取れるようにモデルを変形することにより特徴量推定を行う．一般的に，このマッチングは逆運動力学を解くことによって行われる．

3次元構造モデルは通常2種類に大別される．対象の3次元的な外見を表現したvolumetric model（図12）と，対象のリンク機構だけに着目したskeletal model（図13）である．

volumetric modelは，外見的にリアルな表現が必要なCGアニメーションの分野［21］や身体の大まかな構造をモデル化する場合［9］にしばしば用いられる．一方で，人間の手指のような関節の多い対象をモデル化する場合には特徴量空間が非常に大きくなり，膨大な計算コストがかかる

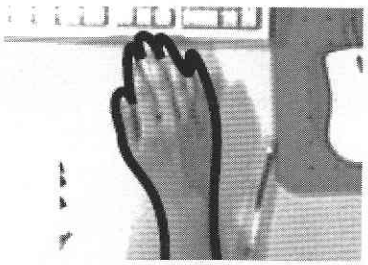

図9　輪郭抽出

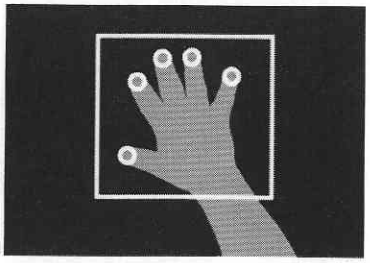

図10　指先抽出

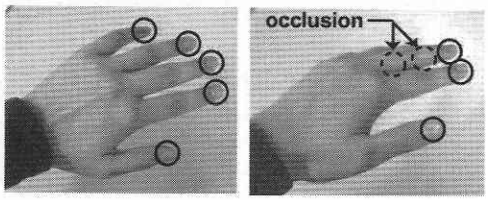

図11　オクルージョン

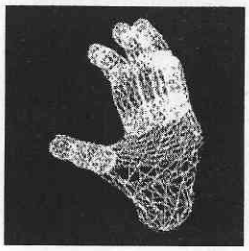

図12　Volumetric model

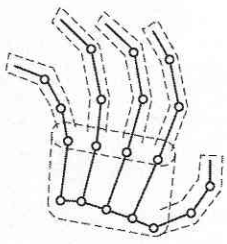

図13　Skeletal model

といった問題がある．

したがって，より高速に特徴量推定を行う場合には一般的にskeletal modelが利用される［28］［32］．skeletal modelでは，形態学や生物力学に基づき関節ごとに自由度や可動範囲を設定し，さらにリンクの長さを固定長にすることにより特徴量空間を小さくする工夫がなされている．これにより，volumetric modelを利用した場合と比較して，計算コストは大幅に低減される．

Shimadaらの研究［32］では，3次元構造モデルを利用して手の関節角や3次元位置を推定している．まず，手のシルエットと3次元モデルとのマッチングを取ることにより構造モデルの形状や姿勢を絞り込む．さらに，EKF（Extended Kalman Filter）［15］を適用して関節角と3次元位置を決定する．このとき，EKFには関節角の可動範囲を考慮した制限を加えている．

3次元構造モデルを利用する手法では，指先の3次元位置や手の姿勢などを比較的正確に推定することができ，オクルージョンの問題にも対処することができる．一方で，手のモデルのようにモデルの自由度が高くなると推定における蓄積誤差が大きくなることや，計算の複雑化による実時間処理の難しさが問題となる．今後，計算機の処理速度の向上に伴い，これらの問題が緩和されることが望まれる．

・2次元形状モデルの利用

2次元形状モデルとは，図14に示すように対象（手など）の輪郭や2次元形状などをモデル化したものである．このモデルでは，3次元構造モデルのようなリンク機構は考慮せずに，特徴が抽出された画像にモデルがフィットするようにモデルを変形，回転し，特徴量推定を行う．

2次元形状モデルとして輪郭を利用する手法では，各フレームで独立に輪郭を抽出するだけでなく，輪郭の動きをモデル化することにより輪郭モデルの安定化を図ることが多い．Heapらは Smart Snakesによって手の輪郭を検出すると同時に，PDM（Point Distribution Model）により輪郭の動きを学習し，輪郭の追跡を行っている［11］．また，Takahashiらは輪郭全体ではなく輪郭の特徴点だけをARモデルによりモデル化し，ARモデルのパラメータをKalman Filter［15］を使って推定することにより，体全体の姿勢を決定している［35］．同様に，CONDENSATIONアルゴリズムを利用して手の輪郭の追跡を行う研究［12］［13］も報告されている．

一方，Utsumiらは手の3次元姿勢推定に2次元形状モデルを利用している［37］．このモデルには楕円モデルを採用し，ユーザの手のひらを楕円形で近似する．そこで入力画像とモデルとのマッチングを行い，手の向きに対する回転角を推定する．

2次元形状モデルを利用する場合，3次元構造モデルを利用した場合と比較して，3次元姿勢や形状を精度良く推定する能力では劣っている．また，オクルージョンの問題が発生することも多い．しかし，高速な特徴量推定が可能であるために実時間処理には適している．また推定の精度に関しても，アプリケーションによっては実用に耐え得る精度を得ることが可能である．

・画像モデルの利用

画像モデルは画像そのものをモデル化したものである（図15）．3次元構造モデルや2次元形状モデルのように1つのモデルの変形を行うのではなく，複数のモデル画像とのマッチングを行い，動作認識に用いる特徴量を得る．

Darrellらのシステム［8］では，画像モデルとして様々な形状の手画像のセットを持っている．入力画像とモデルセットとの正規化相関を行うことにより，入力画像の特徴量として類似度のセットを計算し，動作認識に利用する．

画像モデルとして複数の画像を持つ場合，必要となる記憶容量が大きくなり，計算コストも非常に大きくなる．そのため，独自の特徴空間を用いてモデル画像を表現することによりモデルの圧縮を図る場合もある［39］［50］．

画像モデルを利用する場合，指先位置や関節角といった目に見える特徴量は得られない．そのため，手によるマニピュレーション動作のモデル化には適さない．実際，上で紹介した研究で扱っているのは，手の振り［8］や腕の振り［39］［50］などの大まかなジェスチャ認識であって，

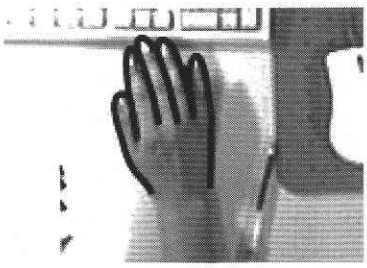

図14　2次元形状モデル

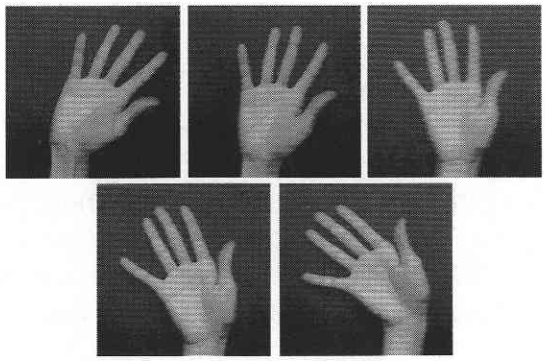

図15　画像モデル

手指による細かい動きは認識対象としていない．

3. ユーザ動作の認識

コンピュータビジョンによるユーザ認識は静止画像認識と動画像認識とに大別される．静止画像認識では，単一画像中の対象の形状が表現している内容を判別する．ここで一般的に行われる処理は次のようになる．まず，各クラスの1枚の標準パターンから抽出された特徴量と1枚の入力画像の特徴量との距離を計算する．そして，その距離に基づき入力画像の属するクラスの判別を行う．よって，一種のパターン識別［45］として考えられる．研究例として，マハラノビス距離による線形識別法を利用して数種類の手形状の判別を行う研究［10］やニューラルネットワークを利用して同様の判別を行う研究［30］などが存在する．

一方，動画像の認識の場合，時系列画像シーケンスを入力として持ち，画像中の対象の動きが表現している内容を認識する．動画像認識の例には，開発者の定義した簡単なジェスチャを認識する研究［1］［2］［8］の他に，ASLなどの手話を認識する研究［34］［38］［43］，手書き文字を認識する研究［23］，スポーツの動作を認識する研究［42］などが挙げられる．ここで使用される認識手法は非常に多岐に渡っているが，手や腕の動作の認識手法として用いられている手法のほとんどは次の2種類の手法に大別される．1つは1枚の画像に対するパターン識別を画像シーケンスに拡張した手法であり，もう1つは画像シーケンスを状態の集合とみなして統計的に処理する手法である．以下でこれら2種類の認識手法について説明する．

3.1 パターン識別に基づく時系列画像認識

1枚の画像に対するパターン識別とは，上で説明したように，1枚の標準パターンと1枚の入力画像との特徴量間距離に基づく識別である．これを時系列の連続画像に拡張する場合，時間軸方向の伸縮の取り扱いが問題となる．そこで，この問題に対処した様々な動作認識手法が提案されている．以下でこれらの手法について幾つか説明する．

ジェスチャ間の時間差を吸収するために提案された手法として連続DP（Continuous Dynamic Programming：CDP）［49］［50］が挙げられる．これは時間軸方向の非線型伸縮を行うことにより実現される．まず，入力画像シーケンスのt番目のフレームと標準パターンシーケンスのτ番目のフレームとの特徴量間距離$d(t, \tau)$を，図16上側の格子点(t, τ)に対応付ける．次に，入力シーケンスのあるフレームtが各標準シーケンスの終点$\tau = T$に対応していると仮定する．この仮定のもとに，それ以前の部分に対して時間的な非線型伸縮を行い，最小累積距離$S(t, T)$を求める．この操作を入力シーケンスの各時点tに対して行う．このとき，入力画像シーケンス中のジェスチャが終了する付近で入力シーケンスと正しい標準シーケンスが最適

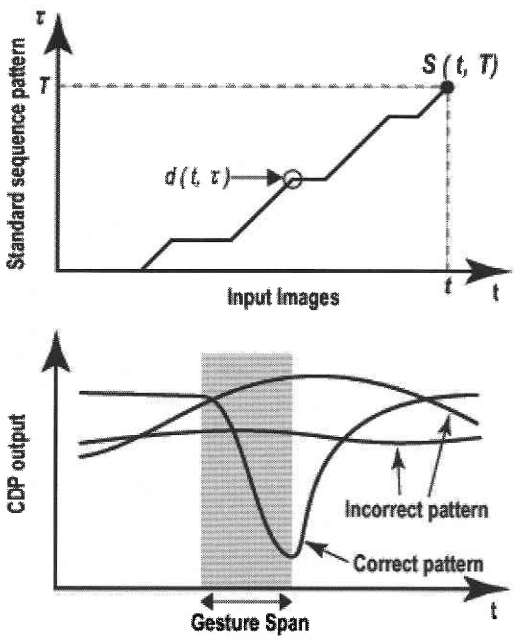

図16　連続DP

適合し，$S(t, \tau)$が局所的最小値を取る（図16下側）．この性質を利用して認識結果を定める．連続DPにはジェスチャ開始および終了についてのセグメンテーション（スポッティング）を認識作業と同時に行うことができるという利点がある．

連続DPと同様に，DTW（Dynamic Time Warping）［7］［8］も時間軸方向の対応を自由に取ることが可能である．この手法では，入力シーケンスと標準パターンシーケンスの開始点同士，終了点同士の両方あるいは片方を合わせて，時間軸方向に入力および標準パターンを非線形伸縮することにより特徴点間距離が最小になる経路（Time Warp）を求める．この作業を全ての標準パターンシーケンスに対して行い，距離が最小になるものを認識結果として得る．

ニューラルネットワークを時系列画像に拡張した手法としてはTDNN（Time-Delay Neural Network）［43］が挙げられる．ニューラルネットワークとは，入力層と出力層の間に複数の中間層を持つことにより，高い識別能力を発揮するパターン識別手法である．TDNNではこれらの層間の時間遅延を制御することにより，時系列画像に対応することが可能である．Yangらは動作の軌跡をTDNNに入力することによってASL（American Sign Language）の認識を行っている［43］．

また，固有空間上の動作の軌跡［39］やMHI（Motion History Image）［1］のように，時系列の特徴量を1つにまとめることにより，基本的なパターン識別のタスクとして扱うことが可能な手法も提案されている．

これらのパターン識別に基づく手法の最大の利点は，確率的な学習を行わないため学習データを大量に用意する必要がないということである．したがって，学習データが少ない場合には，大量の学習データを必要とする後節のHMMと比較して，これらの手法の認識率が高くなることも多い．また，比較的設計が容易であることも重要な特徴である．しかしながら，学習データ量や最適な設計など，整った環境下での認識率や特徴量空間内での変動吸収能力はHMMに劣る傾向がある．

3.2　統計的手法に基づく時系列画像認識

ここでは画像シーケンスの認識に統計的処理を取り入れた手法，その中でも代表的なHMM（Hidden Markov Model）について説明する．

音声認識手法として現在でも主流であるHMM［27］は，比較的最近になって連続画像認識に利用されるようになり，その有効性が証明されている［42］．先に紹介した連続DPなどと異なる大きな特徴は，1つのジェスチャを幾つかの「状態」に分割し，各状態間の遷移によってジェスチャを表現していることである．図17のように，各ジェスチャは状態の集合として表現され，各状態（q_i）は自分を含む各状態への遷移確率（$a_{i,j}$）と特徴量（v_k）の出現確率（$b_i(k)$）を保有している．認識の際には，入力画像シーケンスから抽出された特徴量シーケンスと各ジェスチャとのマッチングを行い，最大確率を得られたジェスチャを認識結果として出力する．さらに，入力が単一ジェスチャではなくジェスチャのシーケンスである場合には，ジェスチャ間の出現確率を表した文法モデルを用意し，その確率も利用するというのが一般的である．

近年のユーザ認識ではHMMを利用する手法［23］［34］［42］が非常に多く，通常のHMM以外にも様々な形のHMMを用いる手法が提案されている．手や腕の動作認識に利用している例として，ニューラルネットワークの識別能力を持ったIOHMM（Input-Output HMM）［22］，パラメトリックな性質を持つジェスチャに対応したPHMM（Parametric HMM）［41］，複数のジェスチャが同時に出現する状況に対応したCHMM（Coupled HMM）［2］やPaHMM（Parallel HMM）［38］などが挙げられる．また，動作の再構成にHMMの変形であるDBN（Dynamic Bayesian Network）を利用する手法［26］も提案されている．

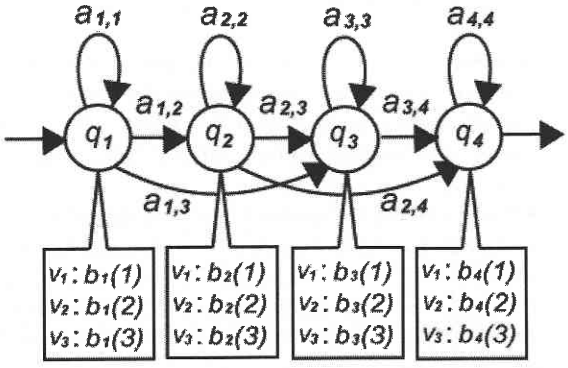

図17　HMM

HMMには時間軸方向の非線型伸縮性があるため，DTWなどの性質を包含していると言える．また，大量のデータを用いて学習したモデルを基に確率的な処理を行うために，特徴量空間における変動吸収能力も高い．さらに，文法モデルによるジェスチャ間の結合確率を導入することが容易であるため，より安定したジェスチャ認識が可能である．一方で，多くのパラメータに対して確率的な学習を行うために，学習データとして大量のデータを用意しなければならないという問題がある．

4. HCIへの応用

前章までにユーザの手を中心とする動作の追跡，認識手法について紹介してきたが，これらの手法を実際にHCIのアプリケーションとして利用する研究も多数存在する．以下では，それらの研究の中から幾つかのアプリケーションを紹介する．

Wellnerは机上での作業とコンピュータでの作業の融合を試みてDigitalDesk［40］の開発を行った．DigitalDeskは机上に投影されたコンピュータ画面を1本の指先などで操作するもので，ユーザは机上に投影されたイラストを切り取ってコピーしたり，机上に投影された電卓で計算を行うことが可能である．

同様の机型インターフェースとして，KoikeらはEnhancedDesk［17］を開発した．EnhancedDeskでは机上の紙書類に注目し，紙書類と電子情報の統合利用を試みている．ここではユーザの指の本数と各指先位置を検出し，その情報を基に机上に投影される画面や指先周辺を探索するカメラを操作している．実際のインタラクションとして，机上に投影されたコンピュータ画面を操作するために，直接手を用いたクリック操作などが実現されている．また，物体を囲む動作（図18）により机上物体との関連付けを行い，それを指差すだけでウェブページ等の関連情報を机

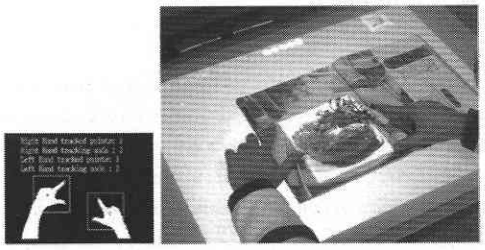

図18　EnhancedDesk

上に投影することが可能である．

3次元的な姿勢や位置推定を利用したインタラクション例としてはSeganらの研究［31］が挙げられる．この研究では，1方向からの照明をユーザの手に当てることにより影を生成し，手と影の関係を計測することによりユーザの手の3次元姿勢や位置を推定している．その結果を利用して，3次元的な位置制御が必要となるアプリケーションを操作する．具体的には，仮想空間内のロボットアームを3次元的に操作することによる仮想物体とのインタラクションや，手の形状と3次元位置を利用したビデオゲームの操作などのインタラクションが可能である．

同様に，Utsumiらは複数のカメラによる手の3次元位置や形状，位置推定を通して仮想空間内での描画操作や指差し操作を行い，また，両手による操作も行っている［37］．

Pentlandらのグループは，部屋単位での実世界指向インターフェースの実現を目指したSmartRooms［33］を開発している．SmartRoomsはカメラやマイクを始めとする様々なセンサを備え，ユーザの動作をあらゆる角度から認識することによりユーザの行動を補助している．ここには多数のアプリケーションが導入されているが，手の動作認識としてはASL認識が導入されている．これは，帽子に取りつけられた小型CCDカメラによりユーザの手を撮影し，HMMを利用したジェスチャ認識を行うというものである［34］．

5. おわりに

本稿ではユーザの手を中心とする動作の追跡，認識に関する様々な手法について説明した．さらに，それらの手法をHCIに応用した研究例を紹介した．

近年，ユーザ動作の追跡，認識に関して非常に多くの研究が行われているため，ここ数年でこれらの技術は大きな発展を遂げた．しかし，安定性や実時間性，使用条件などの制約は多数残っており，今後も研究の余地は十分にあると思われる．また，アプリケーションに関してもGUIの延長に過ぎないものや実際の生活に活用できるのか疑問の残るものが多く，GUIの枠を超えた有意義なアプリケーション開発が望まれる．

（2001年2月14日受理）

参考文献

1) A. Bobick, J. Davis, "Real-time recognition of activity using temporal templates," *Proc. IEEE International Conference on Automatic Face and Gesture Recognition '96*, 1996.
2) M, Brand, N. Oliver, A. Pentland, "Coupled hidden Markov models for complex action recognition," *Proc. IEEE Conference on Computer Vision and Pattern Recognition '97*, pp.994–999, 1997.
3) Cell-infortech Corporation: http://mptech.infortech.co.jp/
4) R. Cipolla, Y. Okamoto, Y. Kuno, "Robust structure from motion using parallax," *Proc. IEEE International Conference on Computer Vision '93*, pp.374–382, 1993.
5) J. Crowley, F. Berard, J. Coutaz, "Finger tracking as an input device for augmented reality," *Proc. IEEE International Workshop on Automatic Face and Gesture Recognition '95*, pp.195–200, 1995.
6) R. Cutler, M. Turk, "View-based interpretation of real-time optical flow for gesture recognition," *Proc. IEEE International Conference on Automatic Face and Gesture Recognition '98*, pp.416–421, 1998.
7) T. Darrell, A. Pentland, "Space-time gestures," *Proc. IEEE Conference on Computer Vision and Pattern Recognition '93*, pp.335–340, 1993.
8) T. Darrell, I. Essa, A. Pentland, "Task-specific gesture analysis in real-time using interpolated views," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 12. pp. 1236–1242, 1996.
9) D. Gavrila, L. Davis, "Towards 3D model-based tracking and recognition of human movement: a multi-view approach," *Proc. IEEE International Workshop on Automatic Face and Gesture Recognition '95*, pp.272–277, 1995.
10) R. Grzeszczuk, G. Bradski, M. Chu, J. Bouguet, "Stereo based gesture recognition invariant to 3D pose and lighting," *Proc. IEEE Conference on Computer Vision and Pattern Recognition '00*, pp.826–833, 2000.
11) T. Heap, F. Samaria, "Real-time hand tracking and gesture recognition using smart snakes," *Proc. Interface to Human and Virtual Worlds*, 1995.
12) T. Heap, D. Hogg, "Wormholes in shape space: tracking through discontinuous changes in shape," *Proc. IEEE International Conference on Computer Vision '98*, pp.344–349, 1998.
13) M. Isard, A. Blake, "Condensation- conditional density propagation for visual tracking," *International Journal Computer Vision*, Vol. 29, No. 1, pp.5-28, 1998.
14) N. Jojic, B. Brumitt, B. Meyers, S. Harris, T. Huang, "Detection and estimation of pointing gestures in dense disparity maps," *Proc. IEEE International Conference on Automatic Face and Gesture Recognition '00*, pp.468–475, 2000.
15) R. Kalman, "A new approach to linear filtering and prediction problem," *Journal of Basic Engineering*, Vol. 82, pp.35–45, 1960.
16) M. Kass, A. Witkin, D. Terzopoulos, "Snakes: active contour models," *IEEE International Conference on Computer Vision '87*, pp.259–268, 1987.
17) H. Koike, Y. Sato, Y. Kobayashi, H. Tobita, M. Kobayashi, "Interactive textbook and interactive Venn diagram: natural and intuitive interfaces on augmented desk system," *Proc. Human Factors in Computing Systems (SIGCHI2000)*, pp.121–128, 2000.
18) J. Kuch, T. Huang, "Vision-based hand modeling and tracking," *Proc. IEEE International Conference on Computer Vision '95*, 1995.
19) C. Maggioni, "A novel gestural input device for virtual reality," *Proc. IEEE Annual Virtual Reality International Symposium '93*, pp.118–124, 1993.

20) C. Maggioni, B. Kammerer, "Gesture computer- history, design and applications," *Computer Vision for Human-Machine Interaction* (*R. Cipolla and A. Pentland, eds.*), pp.23–51, Cambridge University Press, 1998.

21) N. Magnenat-Thalmann, D. Thalmann, *Computer Animation: Theory and Practice*. Springer-Verlag, 1990.

22) S. Marcel, O. Bernier, J. Viallet, D. Collobert, "Hand gesture recognition using input-output hidden Markov models," *Proc. IEEE International Conference on Automatic Face and Gesture Recognition '00*, pp.456–461, 2000.

23) J. Martin, J. Durand, "Automatic handwriting gestures recognition using hidden Markov models," *Proc. IEEE International Conference on Automatic Face and Gesture Recognition '00*, pp.403–409, 2000.

24) J. Ohya, "Virtual kabuki theater: towards the realization of human metamorphosis systems," *Proc. 5th IEEE International Workshop on Robot and Human Communication*, pp.416–421, 1996.

25) V. Pavlovic, R. Sharma, T. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.677–695, 1997.

26) V. Pavlovic, J. Rehg, T. Cham, K. Murphy, "A dynamic Bayesian network approach to figure tracking using learned dynamic models," *IEEE International Conference on Computer Vision '99*, pp.94–101, 1999.

27) L. Rabiner, B. Juang, *Fundamentals of Speech Recognition*. Prentice Hall, 1993.

28) J. Rehg, T. Kanade, "Model-based tracking of self-occluding articulated objects," *Proc. IEEE International Conference on Computer Vision '95*, pp.612–617, 1995.

29) Y. Sato, Y. Kobayashi, H. Koike, "Fast tracking of hands and fingertips in infrared images for augmented desk interface," *Proc. IEEE International Conference on Automatic Face and Gesture Recognition '00*, pp.462–467, 2000.

30) Y. Sato, M. Saito, H. Koike, "Real-time input of 3 D pose and gestures of a user's hand and its applications for HCI," *Proc. 2001 IEEE Virtual Reality Conference*, pp. 79–86, March 2001.

31) J. Segan, S. Kumar, "Shadow gestures: 3 D hand pose estimation using a single camera," *Proc. IEEE Conference on Computer Vision and Pattern Recognition '99*, pp.479–485, 1999.

32) N. Shimada, Y. Shirai, Y. Kuno, J. Miura, "Hand gesture estimation and model refinement using monocular camera-ambiguity limitation by inequality constraints," *Proc. IEEE International Conference on Automatic Face and Gesture Recognition '98*, pp.268–273, 1998.

33) Smart Rooms Project: http://vismod. www. media. mit. edu/vismod/demos/smartroom/

34) T. Starner, J. Weaver, A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 12, pp.1371–1375, 1998.

35) K. Takahashi, T. Sakaguchi, J. Ohya, "Real-time estimation of human body postures using Kalman filter," *Proc. International Workshop on Robot and Human Interaction*, 1999.

36) M. Turk, G. Robertson, "Perceptual user interfaces," *Communications of the ACM*, Vol. 43, No. 3, pp.33–34, 2000.

37) A. Utsumi, J. Ohya, "Multiple-hand-gesture tracking using multiple cameras," *Proc. IEEE Conference on Computer Vision and Pattern Recognition '99*, pp.473–478, 1999.

38) C. Vogler, D. Metaxas, "Parallel hidden Markov model for American sign language recognition," *Proc. IEEE International Conference on Computer Vision '99*, pp.116–122, 1999.

39) T. Watanabe, M. Yachida, "Real time gesture recognition using eigenspace from multi input image sequences," *Proc. IEEE International Conference on Automatic Face and Gesture Recognition '98*, pp.428–433, 1998.

40) P. Wellner, "Interacting with paper on the Digital Desk," *Communications of The ACM*, Vol. 36, No. 7, pp.87–96, 1993.

41) A. Wilson, A. Bobick, "Parametric hidden Markov models for gesture recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 9, pp.884–900, 1999.

42) J. Yamato, J. Ohya, K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," *Proc. IEEE Conference on Computer Vision and Pattern Recognition '92*, pp.379–385, 1992.

43) M. Yang, N. Ahuja, "Recognizing hand gesture using motion trajectories," *Proc. IEEE Conference on Computer Vision and Pattern Recognition '99*, pp.466–472, 1999.

44) T. Zimmermann, J. Lanier, C. Blanchard, S. Bryson, Y. Harvill, "A hand gesture interface device," *Proc. ACM Conf. Human Factors in Computing Systems and Graphics Interface*, pp.189–192, 1987.

45) 石井健一郎，上田修功，前田英作，村瀬洋，パターン認識．オーム社，1998.

46) 内海章，宮里勉，岸野文郎，大谷淳，中津良平，"距離変換処理を用いた多視点画像による手姿勢推定法，" 映像メディア学会誌，Vol. 51, No. 12, pp.2116–2125, 1997.

47) 岡兼司，小林貴訓，佐藤洋一，小池英樹，"複数指先軌跡の実時間計測とHCIへの応用，" 情報処理学会コンピュータビジョンとイメージメディア研究会報告，2000-CVIM-123-7, pp. 51–58, 2000.

48) 小池英樹，Bit別冊ビジュアルインタフェース—ポストGUIを目指して—．共立出版，2.1章，pp.24-44, 1996.

49) 高橋勝彦，関進，小島浩，岡隆一，"ジェスチャー動画像のスポッティング認識，" 信学論，Vol. J 77-D-II, No. 8, pp.1552–1561, 1994.

50) 西村拓一，向井理朗，野崎俊輔，岡隆一，"白黒動画像からの形状特徴を用いたジェスチャのスポッティング認識システム，" 信学論，Vol. J 81-D-II, No. 8, pp.1812–1821, 1998.