情報処理学会論文誌　コンピュータビジョンとイメージメディア　Vol. 1　No. 1　53–63 (June 2008)

手指ジェスチャの画像計測手法とその応用

島田伸敬[†1]　白井良明[†1]

人間の体は多関節構造物として非常に複雑な物体であり，その形状や姿勢を画像列から非接触に計測することは，長年コンピュータビジョンの分野におけるチャレンジングな課題となっている．本論文では，著者らのグループでこれまでに研究してきた画像列から手指形状を計測する手法について述べる．三次元形状モデルを画像に照合する手法と，画像の見えを記録しておいて照合する方法，さらに複雑背景下で領域抽出と照合を行う手法を紹介する．またこれらの手法を応用した手話認識技術もあわせて述べる．

Image-based Measurement of Hand Gesture and Its Applications

Nobutaka Shimada[†1] and Yoshiaki Shirai[†1]

Since the human body has the highly articulated structure, it has been a challenging problem in the computer vision field to estimate its posture from an image sequence in a touchless way. This paper introduces the 3-D hand posture estimation methods using image sequence which the authors have been developing. The methods are diveided into two major categories: one is the appearance-based method and the other is the 3-D-shape-model-based method, which can be suitably selected according to applications. The paper also introduces their applications to sign language recogntion.

1. まえがき

計算機を援用したヒューマンインタフェースの 1 つとして，ジェスチャ認識が注目されて久しい．バーチャルリアリティにおいては，特別な手袋などの装着デバイスを用いずに手指の状態を入力できるように，TV カメラからの画像から人体の三次元形状と姿勢を非接触に認識することが望まれている．また，コンピュータの操作や手話による対話などにも直観的なジェスチャインタフェースが有効な場合がある．

画像に基づくジェスチャ認識手法は 1990 年代に入ってから研究がさかんになり，CV の問題としては関節物体としての人体のモデルフィッテング問題として広く取り扱われている[38]．その結果，背景や照明の制御およびマークや距離情報の利用が可能な場合には，画像に基づくいわゆるモーションキャプチャのような形状姿勢推定が可能となり，また簡単なジェスチャに限れば実時間で認識できるようになった．

しかし，道具や物体を手で操作する様子を識別したり，手話のような複雑な手の形と動きを推定したりすることは，背景と手指領域の峻別や，手指とカメラの位置関係による見え方の変化，手形状自体の個人差，ジェスチャ自体に揺らぎがあることなどから，単純な見えの照合では困難である．この問題に対処するために，これまで大きく分けると 3-D-model-based と 2-D-appearance-based の 2 種類のアプローチが研究されてきた．

前者の手法は，画像から突起状領域の短点や骨格などの三次元形状に由来する特徴を抽出して，その特徴に対して三次元形状モデルをあてはめている[2),15)]．この方法は，最小二乗基準に基づいた高精度な姿勢の推定を試みているが，セルフオクルージョンなどによって見えが多様に変化する姿勢に対しては一意的な特徴の対応付けが困難になりロバスト性に欠ける．そこで多視点の画像から対応付けに最も適切な画像を選ぶ手法[23)]がある．また，画像の局所的な特徴ではなく多視点画像から作成されたボクセルに対する三次元モデルのあてはめにより姿勢を推定する方法が提案されている[12),22)]．この方法では，ICP（Iterative Closest Point）アルゴリズムに基づいて追跡を行うが，ステレオカメラや多視点カメラの配置により視野が限定されることや，セルフオクルージョンが起こるとボクセルが正しく作成されないという問題がある．そこで近年では単眼もしくはステレオの入力画像と三次元モデルから生成した見えとの照合尤度を計算し，particle filter などの探索手法を応用して，照合のロバスト性を確保しつつ広いパラメータ空間を効率的に探索する手法が提案されている[3)–5)]．

一方後者の方法では，記録されている対象物体の様々な二次元の見えの中から入力の手指の見えに最も照合するものが選び出される[6),8)]．これは三次元の形状特徴どうしを対応付けるのではなく，入力画像中の見えと記録されているモデルの見えを比較照合するものであり，セルフオクルージョンに対して比較的ロバストであるとされる．見えの画像を主成分分析（PCA）で圧縮すれば，照合のロバスト性を向上させつつ計算時間の短縮が可能であ

†1 立命館大学情報理工学部知能情報学科
Department of Human and Computer Intelligence, Ritsumeikan University

り[7]，見えの画像特徴を用いた boosting によって手形検出器を構成することも行われている[9]．しかし，これらの方法は入力を限られた種類の二次元パターンのクラスへの分類にとどまっており，三次元情報の抽出は行っていない．Black ら[10] はこのアプローチを二次元の位置と方向の推定に拡張したが，三次元までは至っていない．見えと三次元姿勢パラメータとの間の写像関係を学習しておけば高速に姿勢を推定することができるが[14),21]，学習されるべき写像の非線形性の強さから，きわめて多くの学習用画像サンプルを必要とするという問題点がある．

最近年においては，比較的複雑な背景から人体領域を検出する問題も扱われるようになった[34)–36]．関節物体としての人体姿勢の自由度の高さゆえ，複雑背景中の人体領域にいきなり三次元形状モデルをあてはめると，画像特徴とモデルの対応付けの曖昧さのために検証すべき姿勢候補が爆発的な数にのぼってしまう．そこで二次元モデルなど非常にシンプルな記述のあてはめ結果に基づいて，次第により自由度の高い記述力のあるモデルをあてはめていくことによって解決を試みる手法も提案されており[34]，今後の発展が期待されている．検出率の問題や非常に高い計算コストを必要とするなど，顔検出における Viola-Jones 法[33] のような決定版的手法はいまだ確立されていない．

著者らの研究グループでは，これまで主に画像追跡の手法に基づいて，三次元形状モデルと入力画像のシルエットマッチングによる関節物体の姿勢を計測する手法について研究してきた[16),17),19),20),25),26]．3-D-model-based と 2-D-appearance-based の方法との橋渡しをする考え方として，起こりうる見えを三次元形状モデルから生成し，その見えと入力の見えとの照合を行う “Estimation by Synthesis” [11),13] の考え方に基づき，三次元モデルから生成した見えを効率良くデータベースに登録し，三次元的な見えの変化を考慮した照合を行うことによって，従来の単純なジェスチャだけではなく，多様に変化する手指の形状を三次元的に推定することができる．

また上記の手法は関節物体のシルエットがきれいに抽出できることを前提にしているため，実際のジェスチャ推定に適用するには人体領域抽出問題を解決することが必要となるが，複雑な背景下では人体の形の推定結果を手がかりにしないと雑音や解釈のあいまい性のために領域抽出や特徴の対応付けの誤りが頻発する．そこで，誤った照合が起こる可能性を照合度の評価に取り入れた確率的評価法を導入することによって，切り出しを行いながら形状推定する手法を提案している[27]．

ジェスチャ認識やインタフェースへの応用面を考えると，必ずしも三次元的な形状を必要としない代わりに，複雑な背景の下での比較的高速な動作が求められる場合もある．そこで著者らのグループでは，手話単語の認識を対象に，ジェスチャ中の手指の二次元的な見えの変化を遷移ネットワーク上に登録し，領域抽出と手形状識別を並行的に行いながら手形状変化の系列として手話単語を認識する研究を行っている[1),28),32]．

本論文では人間の手指形状を画像から推定する問題に関して，上記の著者らの研究内容についてまとめて紹介する．なお紙面の都合により，各手法の技術的詳細については引用する個々の論文を参照されたい．

2. 手領域輪郭に基づく実時間手指形状計測システム

三次元構造モデルを利用して，あらかじめありうる手指の見え方とその時間変化を学習させておき，入力画像と照合することで手指形状を識別しつつ画像上の領域を追跡することが可能となる．

あらかじめモデルから生成した手形状シルエット輪郭をデータベースに登録する．姿勢推定時には，入力画像のシルエットから輪郭を抽出し，輪郭形状がマッチする候補をデータベースから検索し，見つかった候補を生成する際に用いた形状モデルの関節パラメータが入力画像の推定姿勢パラメータであるとする（**図 1**）．次の時刻における画像フレームの推定では，姿勢パラメータが大きく変動しないという仮定に基づいて近傍を探すことにより効率的な識別を行う．また，一時的な識別誤りに対応するため，ビーム探索による並列探索を今回の実装では用いたが，近年よく用いられるサンプリング法である particle filter を利用す

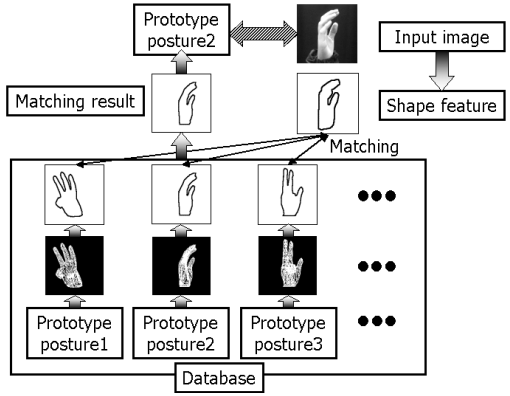


図 1　輪郭形状に基づく手指姿勢推定方法
Fig. 1　Overview of contour-based hand posture estimation.

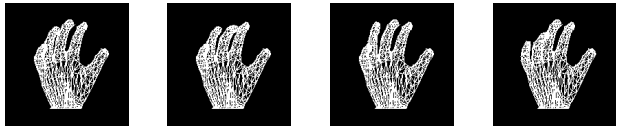
ることもできる．

ここで，手指の関節の総自由度は二十数自由度と高く，様々な仮定によって指の動きを限定したとしても 5〜8 自由度程度はあるため，すべての見え方モデルをデータベースに登録することは現実的でない．したがって，各自由度ごとにある程度粗くサンプリングしたスパースな見え方データベースとなる．すると，入力画像の照合時には必ずしも見え方モデルとよく一致する見え方が得られるとは限らず，登録したモデルの中間的な姿勢が入力されることが多くなる．これを従来法のまま照合すると照合の誤りが非常に多くなる．

そこで本手法では，データベースに登録された見え方モデルの近傍で三次元姿勢パラメータ（関節角度）を変動させ（**図 2**），見え方のシルエットがどのように変動するかをあらかじめ学習し（**図 3**），許容される変動とそうでない変動に分けてから照合を行う方法をとる．

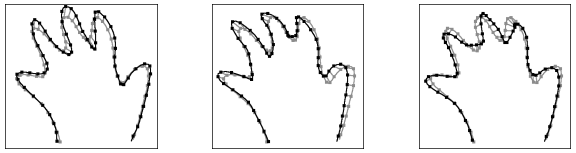
具体的には，見え方モデルごとにパラメータを変動させて生成したシルエット輪郭上に等間隔に一定数の点をサンプルし，シルエットの重心を基準としたそれらの座標を並べた特徴ベクトルをつくる．特徴ベクトルを PCA に基づく次元圧縮を行うことで，許容できる輪郭の変動成分が固有値と固有ベクトルの形で得られる．個々の見え方モデルごとに PCA を行うので，パラメータ空間中の非線形な空間を小さい線形空間のパッチでつないで近似的に表現することができる．照合時には，入力シルエットからモデルと同様に生成した特徴ベクトルを，前時刻で推定されたモデルの近傍に存在するモデルの線形空間とのマハラノビス距離が十分小さい候補を選択する．

これにより三次元構造モデルから許される見え方変動であれば多少大きくても許容して照合し，逆に許されない見え方変動は小さくても棄却することが可能となり，照合の精度が向上した（**図 4**）．



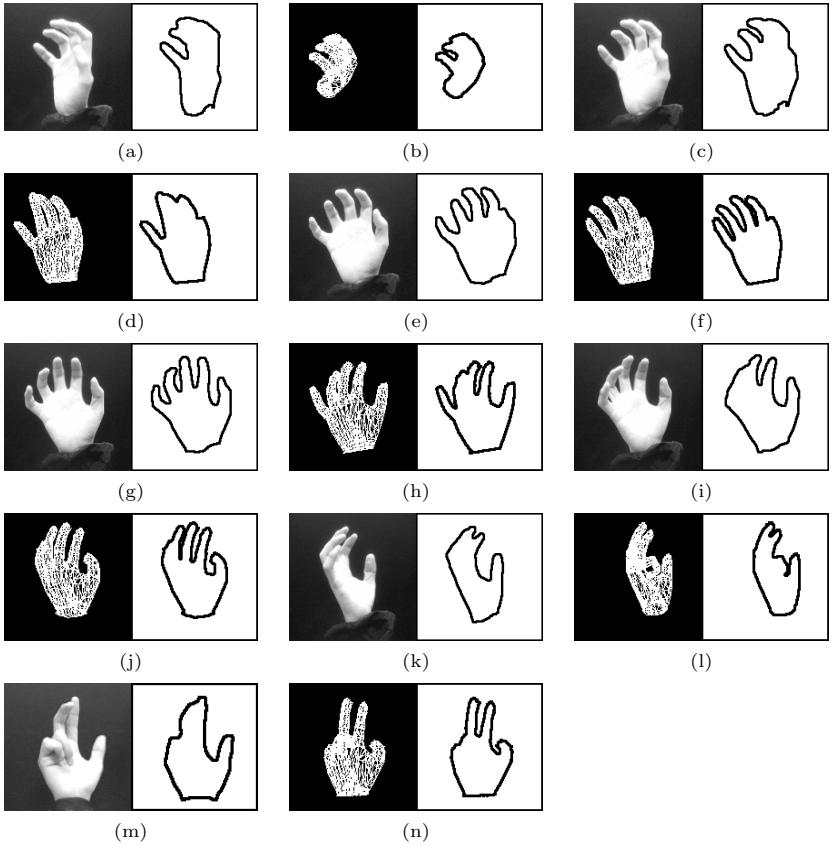
(a) Prototype posture　(b) Variation1　(c) Variation2　(d) Variation3

図 2　典型姿勢とバリエーション
Fig. 2　A prototype posture and its variations.



(a) 1st eigenvector　(b) 3rd eigenvector　(c) 5th eigenvector

図 3　学習された許容変動の例（固有ベクトル）
Fig. 3　Extracted possible variations (eigenvectors).



(a)　(b)　(c)
(d)　(e)　(f)
(g)　(h)　(i)
(j)　(k)　(l)
(m)　(n)

図 4　許容変動を考慮した照合結果
Fig. 4　Matching results by considering possible variations.

隠蔽が発生する候補では，微小な関節角度変化が大きな見え方変化をもたらし，見え方変動が単純な正規分布モデルでは記述できないことが分かった．そこで，輪郭変形の学習時にクラスタリングを行って複数の正規分布の混合モデルで記述する方法を採用し，また輪郭上に等間隔にとったサンプル点が手形状の個人差によってずれることを考慮して学習する方法と組み合わせてより頑健な照合が可能となった．

以上のアルゴリズムを，ハイエンド PC に大容量メインメモリを搭載し，相互に Gbit 高速 LAN で接続した PC クラスタシステム上に実装した．これにより約 16 万通りの手指姿勢を 10 fps 程度で推定するシステムを構築することができた[25),26)]．

3. 誤り照合尤度に基づく一般背景における手指形状推定

一般背景の下では，手指領域の問題と形状の推定問題を分離することが困難なため，両者を同時に解決する必要がある．すなわち列挙された手指の形状候補と画像中の特徴との対応付けをまず仮定し，候補の画像に対する照合度を評価することになる．対応付け手法として，画像上の距離が近い特徴点を探して対応付ける Chamfer matching がよく用いられる[24)]．しかしこの単純な評価法はテクスチャが多く，顔のような手指と混同しやすい背景の下でしばしば誤推定の原因となる．

図 5 の例では，入力の手指形状と比較して指がより曲がったモデル形状が照合した．このように手の皺や背景のエッジのために正しい対応を求めることは一般に困難なため，Chamfer matching による照合度評価が高い候補が必ずしも正しいとは限らない．

この問題は，入力画像とあるモデル候補の見えとの単純な照合度だけを考慮して推定を行うことに原因がある．そこで，候補を評価する際，複数の参照画像に対する照合度を考慮することが考えられる．たとえば図 5 の例では，候補モデル (a) と (c) 両方の特徴が入力画像に照合してしまう（(b)，(d) を参照）．しかし，正解が (a) である手指画像に誤ってモデル (c) をあてはめたときの照合度，ならびにその逆の場合の照合度に差異があれば，両モデルを入力画像にあてはめたときの照合度をともに考慮することで両者が識別できる可能性がある．

この原理に沿った評価法の 1 つである Embedding アプローチ[24)] をベイズ推定，あるいは最尤推定の枠組みで一般化することで，上記の識別原理を確率論的に定式化する[27)]．

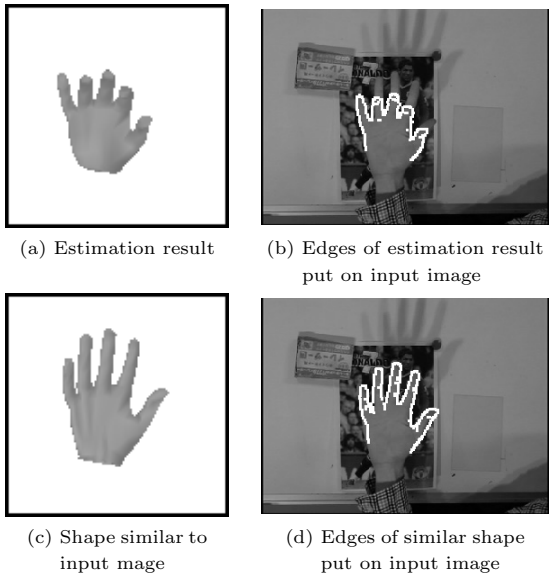
入力画像特徴（実験ではエッジ画像と肌色領域画像を用いた）を I とする．最尤推定では，候補モデル Θ_j に対する入力画像 I の尤度 $P(I|\Theta_j)$ を最大化する候補モデル $\hat{\Theta} = \arg\max_{\Theta_j} P(I|\Theta_j)$ を選択する．

ここで，本来 Θ_j に対応すべき形状が存在する画像 I の中に，背景や手のテクスチャの影響でたまたま，Θ_k の理想的な画像特徴 A_{Θ_k} が存在してしまう場合を想定しよう．Θ_j のもとで $A_{\Theta_1}, A_{\Theta_2}, \ldots$ は互いに排反と仮定すると，尤度 $p(I|\Theta_j)$ は $A_{\Theta_1}, A_{\Theta_2}, \ldots$ が存在する場合に場合分けすることによって

$$\begin{aligned} & p(I|\Theta_j) \\ &= \textstyle\sum_k p(I, A_{\Theta_k}|\Theta_j) \\ &= \textstyle\sum_k p(I|A_{\Theta_k}) p(A_{\Theta_k}|\Theta_j) \end{aligned} \tag{1}$$

と展開できる．入力の画像特徴 I と候補モデルの理想的な画像特徴 A_{Θ_k} どうしの照合尤度モデルは，入力画像と候補モデルから描画される見え画像特徴の差異に基づき定義され，モデル Θ_j のもとで A_{Θ_k} が出現する尤度は CG による照合シミュレーションによってあらかじめ見積もることができる．

一般の最尤推定は，式 (1) において $k = j$ の場合のみを考慮することに相当し，本手法では $k \neq j$ のときの尤度，すなわち「誤り照合尤度」を新たに考慮していることが従来法と



(a) Estimation result　(b) Edges of estimation result put on input image

(c) Shape similar to input mage　(d) Edges of similar shape put on input image

図 5　Chamfer maching による誤推定
Fig. 5　Wrong estimation due to Chamfer matching.

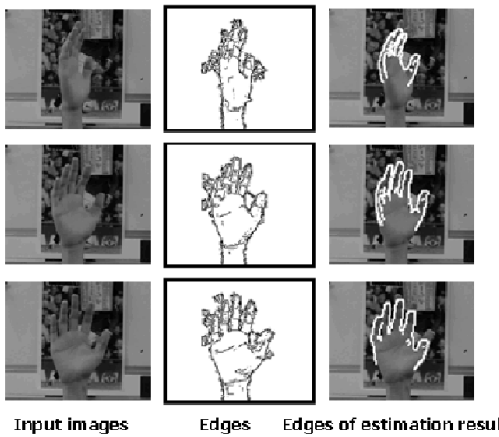


図 6　誤り照合尤度を考慮した形状推定結果 1
Fig. 6　Experimental results considering Mistakenly Matching Likelihood 1.

異なる．図 5 の例では，入力画像中の手指形状は Θ_c であるが，A_{Θ_a} があてはまる．Θ_c の手指画像には誤って A_{Θ_a} があてはまるということを考慮すれば，入力画像に対する Θ_c の尤度が大きくなる．

式 (1) の評価により選ばれた候補（前章の手法と同様に複数の候補を考慮する）について，得られたエッジ特徴の対応結果に基づいてモデルパラメータ（関節角度および形状メッシュの頂点座標）をフィッティングによって変更し，より詳細な形状パラメータ推定を行う．

250 枚の入力画像に対して推定実験を行い，一般の最尤推定と本手法による比較を行ったところ，正答率（正解モデルは目視により主観的に決定した）はそれぞれ 70.4%，82.0%となった．時系列画像について推定実験を行った結果を**図 6**，**図 7** に示す．セルフオクルージョンのため指の一部が隠され，また形状の変形によって新たな背景エッジが現れるが，正しく推定できていることが分かる．

4. 形状モデルのオンラインモデル詳細化

本手法で用いる手指の三次元形状モデルは実際の人体をレーザレンジファインダで取り込んだものをもとにしているが，各パーツの形状や関節の位置が固定のジェネリックなものであり，実際のユーザの手とは完全に一致しない．そのため，データベース中の手形状輪郭も実際の入力画像とはそもそも完全に一致しないので，姿勢（関節角度）の推定精度は一定以

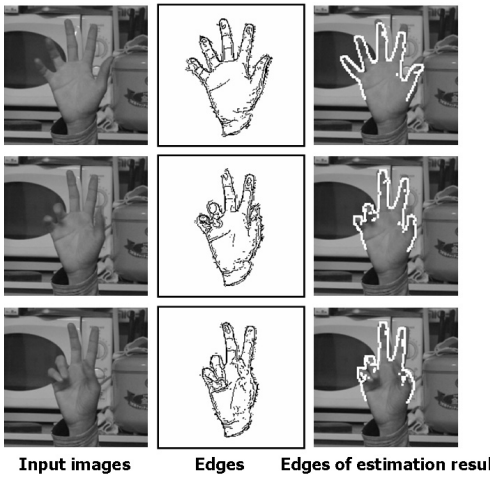


図 7　誤り照合尤度を考慮した形状推定結果 2
Fig. 7　Experimental results considering Mistakenly Matching Likelihood 2.

上には向上しない．

しかし，推定時には次々と画像が入力されていくので，推定中に各部の寸法が変化しないという仮定を用いれば，姿勢を推定しながらオンラインでモデル形状を詳細化してユーザにあわせていくことが可能となる．三次元形状を測定するには一般に 2 つ以上の方向から見た画像が必要とされるが，本研究ではこれを 1 つのカメラで行うことを検討した．物体の形状と姿勢がまったく変化しなければ 1 つのカメラでも物体を回転して見る方向を変えながら画像を入力すれば形状測定が可能であるが，関節物体は関節が動くため，毎時刻変化した姿勢の画像が入力される．本質的にこの種の推定問題は不良設定であり，奥行きのあいまい性のため解を求めることができない．

しかし，この場合でも，

(1)　関節の動きの速さに上限がある（動きの連続性），
(2)　各部の寸法（長さ，太さなど）が変化しない（部分剛体性），
(3)　関節の可動範囲，ならびに各関節どうしの許容角度差が既知，
(4)　各部の寸法の上限と下限，ならびに各部の寸法の許容差が既知，

などの条件が仮定できる場合には，単眼視動画像からでも関節物体の形状を詳細化できることが分かった．上の 4 つの条件は，いずれも形状と姿勢のパラメータ空間における不等

式制約として記述されるので，パラメータ空間中の閉領域を構成する．もしフィルタリングなどの手法で推定されたパラメータがこの閉領域の境界付近にある場合，境界によってパラメータの存在範囲が一時的に狭められることが期待できる．1 度時間不変なパラメータの範囲が限定されると，それより増加することはない．通常画像から得られる観測によって，時間変動する関節パラメータと，時間不変である形状パラメータ（各部の寸法）の推定値の間には相関が生じるため，時間変動パラメータの推定精度もまた向上することになる．

したがって，不等式制約と画像から得られる観測値（誤差を含む）の両方を満足するパラメータの範囲をいかに幾何的に記述し，更新するかが課題となる．この計算は多次元空間ではきわめて大きな計算量を必要とするため，カルマンフィルタにおける平均値と共分散楕円の組[17]，あるいは多次元楕円体と平行超平面の組[18] によって可能なパラメータ群を近似的に記述する．前時刻でのパラメータ群と時系列観測，上述の不等式制約条件を分布切取法（**図 8**）によって情報統合することによって，パラメータのとりうる範囲をしだいに限定する．これによりオフラインではあるが現実的な計算量で，単眼動画像から手指の形状（長さや太さ）と姿勢（関節角度）を推定することができた．

図 9 は合成画像による正解 (a)，通常のカルマンフィルタを用いた推定結果 (b)，本手法による不等式制約を用いた推定結果 (c) である．(b) では関節の角度の曲がる方向が奥行きのあいまい性のために間違っているが，(c) では正しく推定できている．

手指実動画像に対する実験例を**図 10** に示す．この例では，各関節の動きが 30 度/フレー

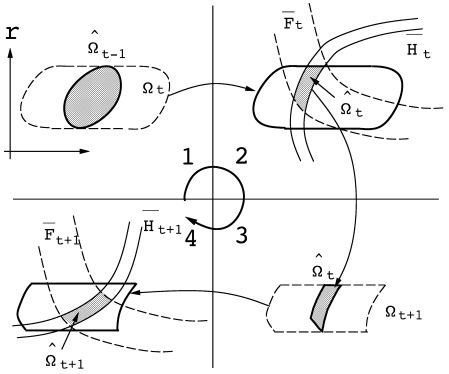


図 8　パラメータ領域の逐次的な更新
Fig. 8　Incremental update of the parameter region.

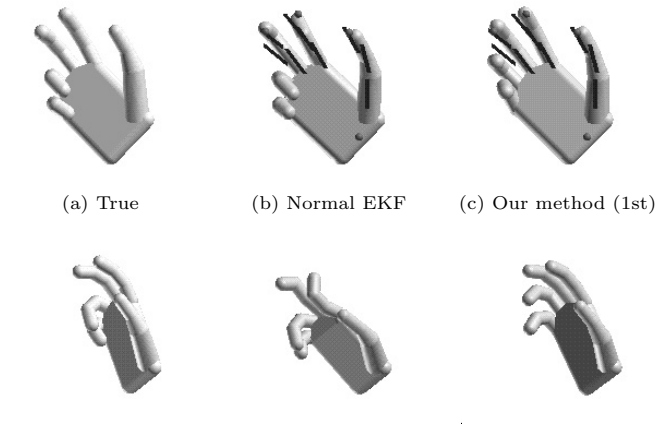


図 9　三次元の関節物体の推定
Fig. 9　Estimaion of 3-D articulated object.

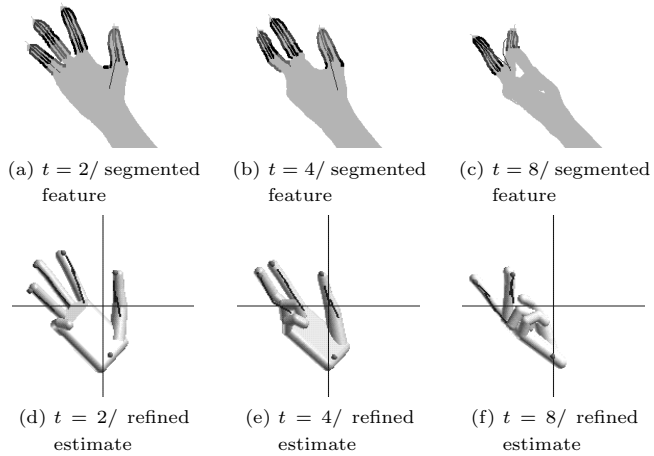
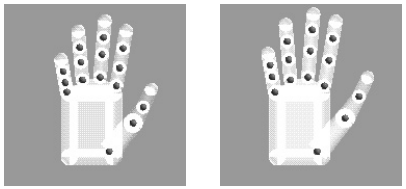


図 10　実画像からの推定例
Fig. 10　Estimation results for real hand images.



(a) 初期形状　　(b) 形状推定結果

図 11　形状推定結果
Fig. 11　Shape estimation result.

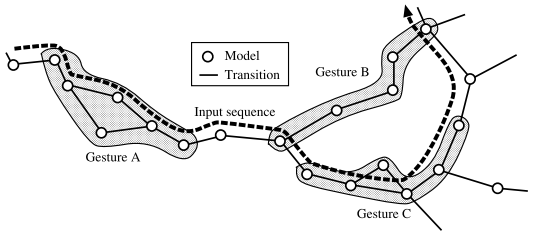


図 12　遷移ネットワークに基づくジェスチャ認識
Fig. 12　Gesture recognition based on a transition network.

ム程度の速い動きの入力画像を用いた．図 10 (a)〜(c) は入力シルエットとそれから抽出された指状の突起特徴である．はじめに大まかな形状モデルをあたえ，突起特徴を利用したシルエットマッチング[16] によりオクルージョンを解決し，大まかな姿勢推定を得る．それをもとに各指節の対応する中心軸を突起特徴から切り出し，大まかな推定結果を初期値として，形状と姿勢を同時に推定する．推定結果が (d)〜(f) である．**図 11** に，最初に与えた大まかな形状モデルが本手法によって修正された結果を示す．

5. 複雑背景下での手指追跡と手話単語認識

手指形状推定技術の応用として，手話の認識と翻訳があげられる[37]．手話認識はヒューマンインタフェースとして実時間的に動作することが最終的に求められるが，前章までに述べた手法では主として手指形状や姿勢のモデルパラメータを取得することを主眼としていたため，そのための計算リソースの消費が大きかった．

よりシンプルに手話認識に適合した手法を開発するために，著者らはまず，限定的なシチュエーションにおける手話単語認識では発生しうるジェスチャの数が比較的限られている，という仮定をおき，必要な単語を HMM モデルによってあらかじめ学習して，推定時の探索範囲を限定することでより高速に形状推定，ならびに単語の認識を行った[31]．

しかし背景を単純にしたり，肌色をキーに領域分割をしてから追跡を行ったりする方法では，机の上や本だなの前などのきわめて複雑な背景下では切り出しに失敗して認識できない．そのような場合に対応するために，あらかじめ登録されたジェスチャモデルに基づいて手の形状を背景から切り出しつつ形状姿勢推定を行う手法を検討した．

手の形状はダイナミクスに従って連続的に変化するが，ダイナミクスパラメータはジェスチャごとに離散的に変化する，との仮定から，同時に複数のダイナミクスによる予測を考慮しつつフィルタリングによる予測と推定を行う Switching Linear Model を導入した[29),30)]．このモデルでは確率的に複数のジェスチャの可能性を考慮しつつ，連続的に変化する動的輪郭を追跡でき，複雑な背景下での手指ジェスチャを追跡，識別するのにこのモデルが有効であることを確認した．しかし数個のジェスチャを登録するのに多量の学習計算を必要とし，認識時にも実時間化が困難な計算量が必要であった．

そこで複数の手話単語学習シーケンスから見えの遷移ネットワークを構成し，遷移ネットワークを可能な遷移をたどることによって効率的に照合を行う手法を提案した[32]．

実際に演じられる手話の画像は以下のような特徴がある．

[高速性]　高速な手指の形状変化と移動をともなうため，画像が動きでぼける．

[複雑背景]　画像上で手指と似た色の背景と重なる．

つまり，従来のモデルフィッティングに基づく追跡手法や，色，動きベクトルによる領域抽出法の適用が困難な状況といえる．

本手法では，あらかじめ手話単語のサンプルシーケンスを用意しておき，ありうる二次元的見え方形状とそのありうる変化を遷移ネットワークとして学習させ，これを用いて照合と追跡を行う．

手指の高速な移動と変形をともなうジェスチャ動画像に対して，速度に応じて適宜選択された特徴を持つ 2 種類のモデルから構成される遷移ネットワーク（**図 12**）をあらかじめ生成しておき，そこから見えのマッチング候補を選択する．手指形状が重要な意味を持つ瞬間では手指の移動速度は遅い．一方，手指の移動速度が速い瞬間では手指輪郭は不明瞭となるが，手指形状は重要でない．そこで，手指の移動速度が遅いときは形状と位置と速度の 3 つを特徴とするモデルを登録する．また，手指の移動速度が速いときは明瞭な輪郭形状が得られないので，位置と速度のみを特徴とするモデルを登録する．登録されたモデルノードは

適当な粒度でクラスタリングを施すことによって，複数のサンプルジェスチャ間でノードを共有する．またサンプルジェスチャ動画中で連続するフレームに対応するノード間にはリンクを生成するので，手指形状や動きが必ずしも類似でないモデルどうしでもサンプルジェスチャ中にその遷移が存在すればネットワーク上でその遷移が反映され，比較的高速な手指移動や変形をともなうジェスチャの手指形状追跡を行うことが可能となる．

さらに，この遷移ネットワークを以下のように拡張して，ジェスチャ区間の事前抽出を必要としないジェスチャ認識を行うことができる．遷移ネットワークは遷移区間も含むすべての学習用サンプル画像列から 1 つの大規模なネットワークが構築される．切り出し済みの単一のサンプルジェスチャ（手話単語）の画像列を遷移ネットワークに通して遷移ノード列を求めると，そのノード列はこのネットワーク上で部分経路を構成する．そこで，遷移ネットワーク上に各サンプルジェスチャに対応する部分経路を登録しておき，認識対象画像列の形状追跡経路がジェスチャのどの部分経路を通過したかを検出することで，ジェスチャを認識する（図 12）．遷移ネットワークに基づく形状推定によりジェスチャ間の遷移区間も含めてつねに形状変化の追跡が行われ，その推定結果からジェスチャの開始，追跡，終了を検出するので，あらかじめジェスチャ区間を切り出す必要がない．

遷移ネットワーク中のモデルと画像の照合では，手指輪郭および背景からエッジ点が観測される確率に基づく評価基準によって複雑背景下で正しい照合を実現した．とくに，従来の評価では，あるモデルの評価にはそのモデルの照合度のみを考慮し，他のマッチング候補が間違いであるかどうかは評価していないが，本手法では真の手指輪郭上，および背景中のエッジ点の存在確率に基づき，すべてのマッチング候補のモデル輪郭点を評価する評価基準を尤度分布として定義し，ベイズ推定の枠組みにより各モデルを評価する．

手話単語 20 単語（各単語 3 シーケンス）の計 2,390 フレームの時系列画像から，形状を持つモデル数が 187，形状を持たないモデル数が 101 の合計 288 モデルからなる遷移ネットワークを自動生成し，速い動作や顔と手の重なりを含む手指形状の追跡を正しく行うことができた．**図 13** に学習画像に含まれる手指形状の例を示す．**図 14** に入力画像に対して推定されたモデルを示す．**図 15** に時系列画像の推定結果例を示す．結果画像はビーム探索で探索された経路のうち，最終フレームで推定されたモデルに到達可能な経路のモデルを描いている．

テスト画像列 60 個に対して単語の認識実験を行った結果，60 個のテスト画像列のうち，44 個について正解ジェスチャのみが検出された．また，複数のジェスチャが候補として検出された場合では，時系列の連続性を考慮したジェスチャ照合の評価基準に基づいて最適ジェスチャを選択することで，13 個について正解ジェスチャが検出された．残りの 3 個は，三次元的な見えの変化のために，正しく照合するモデルが存在せず追跡に失敗したものである．

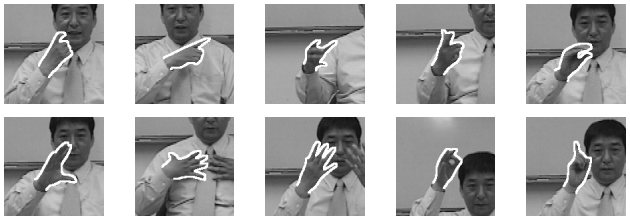
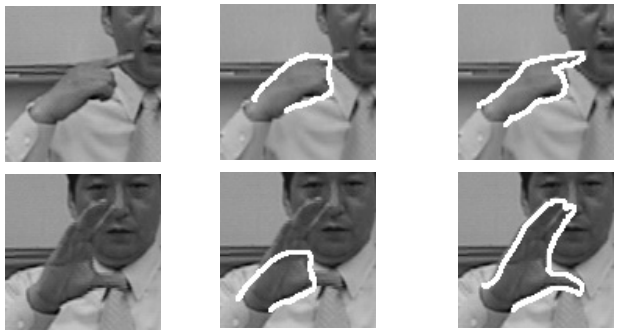


図 13　学習する形状の例
Fig. 13　Shape samples for learning.



(a) Input image　(b) With ave. contrast　(c) With our criterion

図 14　提案手法による照合結果
Fig. 14　Matching results.

6. ま と め

本論文では，複雑な関節物体としての手指の形状と姿勢を画像列から非接触に推定する手法について，

(1)　三次元手指構造モデルを用いた形状変動の学習に基づく手指形状識別
(2)　誤り照合尤度に基づく一般背景における手指領域切り出しと形状推定
(3)　形状モデルの不確実さを補償するオンライン式モデル詳細化手法

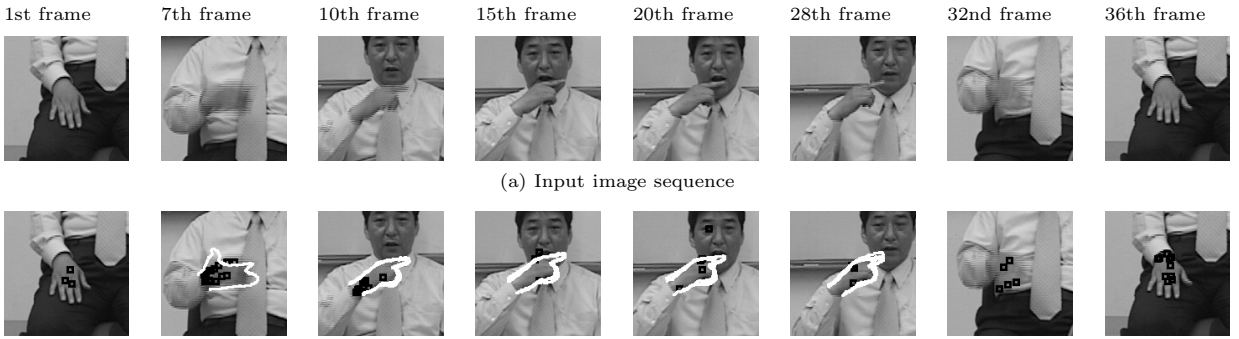


図 15　手話動画像に対する領域追跡および形状推定結果
Fig. 15　Segmentation and tracking results for real sign language word.

(4)　手話知識を利用した遷移ネットワークに基づく手指領域の追跡と単語認識
の 4 つのトピックを取り上げた．

三次元手指構造モデルを用いた形状変動の学習に基づく手指形状識別では現在約 16 万通りの姿勢パターンが登録されているが，手指関節の自由度を 8 自由度に制限しているので，自由度を増やして実際の手指のあらゆる形を識別しようとするとパターンの数が指数的に増大するという問題点がある．これを解決するには，手指輪郭全体をパターンマッチするのではなく，モデル中の独立に稼働する部分ごとに画像輪郭との対応付けを行い，部分ごとのマッチングを行う必要がある．

誤り照合尤度に基づく手指領域切り出しと形状推定については，モデルマッチングとして従来から一般に行われている，単純に候補モデルと画像を照合させてベストな候補を選ぶのではなく，個々の候補が他の候補と取り違えて照合してしまう可能性を考慮した評価基準を最尤推定の枠組みで定式化しその有効性を示した．現時点ではエッジと色特徴のみを利用しているので，差分や動きといった特徴に本手法を拡張することや，いったん背景であると分かった部分の画像情報を有効に利用して追跡を行うなどの課題があげられる．

形状モデルのオンラインモデル詳細化手法については，従来不可能とされてきた単眼視時系列画像から関節物体の腕の長さの推定について，関節の可動範囲や各部の長さの相関を利用することにより推定できる場合があることを原理的に示すことができた．しかし，不等式制約がそもそも不正確にしか与えられない場合には得られる推定パラメータ範囲も不正確となるため，テクスチャの対応などより多くの制約を考慮する必要があると考えられる．

遷移ネットワークを利用した複雑背景下における手指領域の抽出と追跡については，手指の三次元姿勢変化が微少であっても，学習に利用した手話サンプルが 2-D の輪郭であるため，カメラと手話演者との位置関係によって輪郭形状がしばしば大きく変化してしまうことが原因で照合・追跡に失敗するケースがある．それらに対応するには，部分的に上述の三次元手指モデルを用いた形状変動の学習方法を適用することが必要と考えるが，実際のシステムを実装するうえでは，計算リソースや実時間性の問題を考慮しつつ，性能と必要資源のバランスをとることが必要である．

さらに，今後取り組んでいかなければならない方向として，道具や対象物体をハンドリングする様子を切り出し，形状推定し，動作認識する，といった課題が存在する．物体認識において動作と物体自身の関わりを積極的に利用する試みがいくつか行われており[39)–42)]，そのような機能が家庭用ロボットのような人間の作業を支援するシステムの開発には欠かせないものになってきている．それらに活用できるような本格的な人体形状推定システムを構築するには，本論文で述べたような個々の要素技術を現実的な計算資源の中で統合することが必要であるが，現時点では個々の要素技術が個別にできあがった段階であり，これらを実用的なシステムとするには，ジェスチャ認識と特徴抽出や追跡の処理が相互に補完しあう構造に組み直す方策も今後検討する必要があるだろう．

参 考 文 献

1) 浜田康志，島田伸敬，白井良明：手話認識のための複雑背景における手指形状推定，電子情報通信学会パターン認識とイメージメディア研究会技術報告，PRMU2003-152, pp.7–12 (Nov. 2003).
2) 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，電子情報通信学会論文誌 D-II，Vol.J80-D-II, No.1, pp.44–55 (1997).
3) Stenger, B., Thayananthan, A., Torr, P. and Cipolla, R.: Filtering using a tree-based estimator, *International Conference on Computer Vision*, Vol.2, pp.1063–1070 (2003).
4) Wu, Y., Lin, J. and Huang, T.S.: Analyzing and Capturing Articulated Hand Motion in Image Sequences, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.25, No.12, pp.1910–1922 (2005).
5) Okada, R., Stenger, B., Ike, T. and Kondoh, N.: Virtual Fashion Show Using Real-Time Markerless Motion Capture, *Proc. Asian Conference on Computer Vision*, Vol.2, pp.801–810 (2006).
6) Liu, X. and Fujimura, K.: Hand Gesture Recognition using Depth Data, *Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition*, pp.529–534 (2004).
7) Imagawa, K., Taniguchi, R., Arita, D., Matsuo, H., Lu, S. and Igi, S.: Appearance-based Recognition of Hand Shapes for Sign Language in Low Resolution Image, *ACCV2000*, pp.943–948 (2000).
8) Cui, Y. and Weng, J.: Hand Segmentation Using Learning Based Prediction and Verification for Hand Sign Recognition, *Proc. IEEE 2nd Int. Conf. on Automatic Face and Gesture Recognition*, pp.88–93 (1996).
9) Kolsch, M. and Turk, M.: Robust hand detection, *Proc. IEEE Int. Conf. on 6th Automatic Face and Gesture Recognition*, pp.614–619 (2004).
10) Black, M.J. and Jepson, A.D.: EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation, *Int.J.of Computer Vision*, Vol.26, No.1, pp.63–84 (1998).
11) 亀田能成，美濃導彦，池田克夫：シルエット画像からの間接物体の姿勢推定法，電子情報通信学会論文誌 D-II，Vol.J79-D-II, No.1, pp.26–35 (1996).
12) Delamarre, Q. and Faugeras, O.: Finding pose of hand in video images: A stereo-based approach, *Proc. IEEE 3rd Int. Conf. on Automatic Face and Gesture Recognition*, pp.585–590 (1998).
13) Cui, J. and Sun, Z.: Visual Hand Motion Capture for Guiding a Dexterous Hand, *Proc. IEEE 6th Int. Conf. on Automatic Face and Gesture Recognition*, pp.729–734 (2004).
14) Rosales, R. and Sclaroff, S.: Algorithms for Inference in Specialized Maps for Recovering 3D Hand Pose, *Proc. IEEE 5th Int. Conf. on Automatic Face and Gesture Recognition*, pp.143–148 (2002).
15) Rehg, J.M. and Kanade, T.: Visual Tracking of High DOF Articulated Structures: An Application to Human Hand Tracking, *Proc. European Conf. on Computer Vision '94*, pp.35–46 (1994).
16) 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三次元姿勢推定，電子情報通信学会論文誌 D-II，Vol.J79-D-II, No.7, pp.1210–1217 (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 D-II，Vol.J81-D-II, No.1, pp.45–53 (1998).
18) Shimada, N., Shirai, Y. and Kuno, Y.: Model Adaptation and Posture Estimation of Moving Articulated Objects Using Monocular Camera, *Proc. Int. Workshop on Articulated Motion and Deformable Objects*, LNCS 1899, pp.159–172 (2000).
19) Shimada, N., Kimura, K., Shirai, Y. and Kuno, Y.: Hand Posture Estimation by Combining 2-D Appearance-based and 3-D Model-based Approaches, *Proc. IAPR Int. Conf. on Pattern Recognition '00*, pp.709–712 (2000).
20) Shimada, N., Kimura, K. and Shirai, Y.: Real-time 3-D Hand Posture Estimation based on 2-D Appearance Retrieval Using Monocular Camera, *Proc. Int. Workshop on RATFG-RTS*, pp.23–30 (2001).
21) Athitsos, V. and Sclaroff, S.: An Appearance-based Framework for 3D Hand Shape Classification and Camera Viewpoint Estimation, *Proc. IEEE 5th Int. Conf. on Automatic Face and Gesture Recognition*, pp.40–45 (2002).
22) Ueda, E., Matsumoto, Y., Imai, M. and Ogasawara, T.: A Hand-pose Estimation for Vision-based Human Interfaces, *IEEE Trans. Industrial Electronics*, Vol.50, Issue 4, pp.676–684 (2003).
23) 内海　章，大谷　淳，中津良平：多数カメラを用いた手形状認識法とその仮想空間インタフェースへの応用，情報処理学会論文誌，Vol.40, No.2, pp.585–593 (1999).
24) Athitsos, V. and Sclaroff, S.: Estimating 3D Hand Pose from a Cluttered Image, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, II, pp.432–439 (2003).
25) 島田伸敬，今井章博，白井良明：単眼画像入力による非接触ビデオレート手指形状推定システム，第 8 回画像センシングシンポジウム，pp.313–318 (2002).
26) 今井章博，島田伸敬，白井良明：輪郭の変形の学習による 3-D 手指姿勢の認識，電子情報通信学会論文誌 D-II，Vol.J88-D-II, No.8, pp.1643–1651 (2004).
27) 今井章博，島田伸敬，白井良明：複雑背景下におけるモデルの照合誤りを考慮した手指形状推定，電子情報通信学会論文誌 D，Vol.J91-D, No.3, pp.784–792 (2008).
28) Hamada, Y., Shimada, N. and Shirai, Y.: Hand Shape Estimation under Com-

plex Backgrounds for Sign Language Recognition, *Proc. IEEE 6th Int. Conf. on Automatic Face and Gesture Recognition*, pp.589–594 (2004).
29) Jeong, M.H., Kuno, Y., Shimada, N. and Shirai, Y.: Recognition of Shape-Changing Hand Gestures, *IEICE Trans. Division D*, Vol.E85-D, No.10, pp.1678–1687 (2002).
30) Jeong, M.H., Kuno, Y., Shimada, N. and Shirai, Y.: Recognition of Two-Hand Gestures Using Coupled Switching Linear Model, *IEICE Trans. Division D*, Vol.E86-D, No.8, pp.1416–1425 (2003).
31) 谷端伸彦，島田伸敬，白井良明：複雑背景下における手指特徴抽出と手話認識，画像の認識・理解シンポジウム MIRU2002，Vol.II, pp.105–110 (2002).
32) 浜田康志，島田伸敬，白井良明：遷移ネットワークに基づく複雑背景下での手指ジェスチャの認識，情報処理学会 CVIM 研究会研究報告 2005-150-2，pp.9–16 (2005).
33) Viola, P. and Jones, M.: Rapid Object Detection Using A Boosted Cascade of Simple Features, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (*CVPR2001*), Vol.1, pp.511–518 (2001).
34) Zhang, J., Collins, R. and Liu, Y.: Baysian Body Localization Using Mixture of Nonlinear Shape Models, *Proc. Int. Conf. on Computer Vision* (*ICCV2005*), pp.725–732 (2005).
35) Wu, B. and Nevatia, R.: Detection and Tracking of Multiple, Partially Occluded Humans by Baysian Combination of Edgelet based Part Detectors, *Int. J. of Computer Vision*, Vol.75, Issue 2, pp.247–266 (2007).
36) Sabzmeydani, P. and Mori, G.: Detecting Pedestrians by Learning Shapelet Features, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (*CVPR2007*), CDROM (2007).
37) Ong, S.C.W. and Ranganath, S.: Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning, *IEEE Trans. Pattern Analysis and Mechine Intelligence*, Vol.27, No.6, pp.873–891 (2005).
38) 島田伸敬，有田大作，玉木　徹：関節物体のモデルフィッティング（サーベイ（2）），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研究報告 CVIM, Vol.2006, No.51(20060518) 2006-CVIM-154-(40), pp.375–392 (2006).
39) Higuchi, M., Aoki, S., Kojima, A. and Fukunaga, K.: Scene Recognition based on Relationship between Human Actions and Objects, *Proc. 17th Int. Conf. on Pattern Recognition*, Vol.3, pp.73–78 (2004).
40) 尾関基行，宮田康志，青山秀紀，中村裕一：作業支援システムのための人工エージェントとのインタラクションを援用した物体認識，第 10 回画像の認識・理解シンポジウム（MIRU2007）論文集，OS-B1-02, CDROM (2007).
41) 大田博義，木村朝子，島田伸敬，田中弘美：Analysis by Reality-Based Simulation に基づく関節物体の力学的機能推定，電子情報通信学会論文誌 D，Vol.J90-D, No.7, pp.1799–1811 (2007).
42) 片山憲昭，牧　和宏，島田伸敬，白井良明：画像に基づく部屋内シーン変遷の自動検知と対話的イベント検索システム，第 10 回画像の認識・理解シンポジウム（MIRU2007）論文集，IS-1-20, CDROM (2007).

(平成 19 年 9 月 14 日受付)
(平成 20 年 1 月 24 日採録)

(担当編集委員　中澤 篤志)

島田　伸敬（正会員）

平成 4 年大阪大学工学部電子制御機械工学科卒業．平成 7 年同大学大学院工学研究科電子制御機械工学専攻博士後期課程修了．博士（工学）．同年同専攻助手．平成 13 年同研究科研究連携推進室情報ネットワーク部門講師，同助教授を経て，平成 16 年より立命館大学情報理工学部知能情報学科助教授（現，准教授）．平成 19 年より 1 年間米カーネギーメロン大学ロボティクス研究所客員准教授．コンピュータビジョン，ジェスチャ認識，ヒューマンインタフェースの研究に従事．電子情報通信学会，人工知能学会，IEEE 各会員．

白井　良明（正会員）

昭和 39 年名古屋大学工学部機械工学科卒業．昭和 44 年東京大学大学院工学系機械工学専攻博士課程修了．同年電子技術総合研究所研究官．昭和 54 年同視覚情報研究室長．昭和 60 年同制御部部長．昭和 46 年より 1 年間米 MIT，AI lab. 客員研究員．昭和 63 年大阪大学工学部（後に大学院工学研究科）教授．平成 8〜11 年東京大学大学院工学研究科教授併任．平成 14〜18 年情報学研究所客員教授．平成 17 年より立命館大学情報理工学部教授．知能ロボット，画像処理，ヒューマンインタフェースの研究に従事．電子情報通信学会，人工知能学会，日本ロボット学会，計測自動制御学会，ヒューマンインターフェイス学会，IEEE 各会員．