

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

今井 章博[†] 島田 伸敬^{††} 白井 良明^{††}

[†] 大阪大学大学院工学研究科電子制御機械工学専攻 〒 565-0871 大阪府吹田市山田丘 2-1

^{††} 立命館大学情報理工学部知能情報学科 〒 525-8577 滋賀県草津市野路東 1-1-1

E-mail: [†]imai@cv.mech.eng.osaka-u.ac.jp, ^{††}{shimada,shirai}@ci.ritsume.ac.jp

あらまし 本論文では、一般背景下で撮影された画像から三次元手指形状を推定する方法について述べる。従来法では、入力とモデルとの間の画像特徴の誤対応によって、手指形状が異なるにもかかわらず入力とモデルとが照合してしまうことがある。そのため本研究では、各形状の見えに対して別の形状の見えとの照合のしやすさを求める。入力の見えに対して、入力画像中の手指とは形状が異なるモデルの見えが照合しても、誤った照合の起こりやすさから入力画像中の手指の形状を予測することができる。入力に対する各形状の照合結果と誤った照合の起こりやすさから最適な形状を選ぶ。画像特徴としてエッジと肌色領域の形状を用いた実験によって本手法の有効性を示す。

キーワード 手指形状推定, 三次元モデル, 複雑背景, エッジ画像, ベイズ推定

Hand Posture Estimation in Complex Backgrounds by Considering Mis-match of Model

Akihiro IMAI[†], Nobutaka SHIMADA^{††}, and Yoshiaki SHIRAI^{††}

[†] Dept.of Computer-Controlled Mechanical Systems, Osaka University Yamadaoka, Suita, osaka 565-0871, Japan

^{††} Dept.of Human and Computer Intelligence, Ritumeikan University 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

E-mail: [†]imai@cv.mech.eng.osaka-u.ac.jp, ^{††}{shimada,shirai}@ci.ritsume.ac.jp

Abstract This paper proposes a novel method of estimating 3-D hand posture from images observed in complex backgrounds. Conventional methods often cause mistakes by mis-matches of local image features. Our methods compute possibility of the mistakes between each model appearance and the other model appearances. Appropriate model is selected from the mistake possibility and matching results between input image and model appearance, even if image feature mis-matches occur. Finally, by using edges and skin region as image feature, experimental results show effectiveness of the method.

Key words hand posture estimation, 3-D model, complex background, edge image, Bayesian estimation

1. ま え が き

仮想世界や実世界での物体のジェスチャによるマニピュレーション, 手話自動翻訳などを実現するために, 微細な手指の動きを画像情報から非接触で計測する研究が行われている [1] [2] [3] [4] [5] [6].

一般に手指画像の背景は複雑であり, 手指領域の切り出し問題と形状の推定問題を分離することが困難なため, 2つの問題を同時に解決する必要がある. 現実的な計算資源による解法を得るため, 1) 手指形状の候補をどのように限定するか, 2) 形状候補と画像特徴の間の照合度をどのように評価するか, の

観点から, さまざまな手法が提案されてきた.

探索すべき見えの形状を限定するために, Active Shape Models [7] ではあらかじめ学習しておいた輪郭変形モデルのパラメータを, 時間的な変化が微小であると仮定して探索する. また, 三次元形状の変化による輪郭形状の不連続な遷移をあらかじめ学習しておく方法 (switching linear model [8]) もある. 見えではなく三次元形状の近接関係から直接的に候補を限定する方法もある [5]. いずれの手法も, 一時的な推定誤りに対する耐性を得るため, 前時刻で可能性の高い複数の候補からの遷移を並列探索することが一般的であり, ビームサーチや particle filter が用いられる [5] [9] [10] [11] [13] [14] [15].

上述のような探索空間削減の取り組みは多々行われているのに対し、画像とモデルの照合度の評価方法は簡易な方法が利用されている。まずモデルの探索候補の見えを描画し、モデル輪郭の法線方向に入力画像の特徴点を探す、あるいは chamfer matching [12] [16] を用いて最も画像上の距離が近い特徴点を探して、特徴の対応付けを行う。次に対応する特徴点間の画像上距離の総和に基づいて候補モデルの画像への照合度を評価し、最大評価を与える候補を選択する。しかしこの単純な評価法は、テクスチャが多く、顔のように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。

chamfer matching による誤推定の例を図 1 に示す。この例では、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。入力と推定結果のモデルとのエッジ特徴点対応の模式図を図 2 に示す（灰色が入力、黒色がモデル）。モデルの指先端のエッジ点近くに入力画像中の指の側面のエッジが存在するために chamfer matching による距離は小さくなり、照合度は高く評価される。手の皺のエッジ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推定結果ではない。

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。例えば図 1 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) を当てはめたときの照合度、その逆の場合の照合度のいずれかに差異があれば、両モデルを入力画像に当てはめた時の照合度と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一つである Embedding アプローチ [12] では、複数の参照画像に対する照合度の列が入力画像のものと同じモデルを選択するが、照合度の差の単純な二乗和で評価するため、必ずしもその推定の最適性が保障されず、正しい推定が得られないことが多かった。

文字認識において、特定の文字同士間でのみ誤認識の確率が高くなることをコンフュージョンマトリックスで表し、それを用いて誤認識文字を訂正することができる誤り訂正符号を形成する方法がある [20]。

本論文では、ある手指形状に誤った候補モデルが照合する場合を想定した、誤り照合尤度を新たに導入し、上述の識別原理をベイズ推定の枠組みで定式化しなおして、確率論的に最適な照合度評価基準を導出する。

2. 照合の候補

2.1 三次元手指モデル

本研究で用いている三次元手指モデルは、元々手の表面を三角形パッチの集合で構成したワイヤフレームモデルである。CG 画像からエッジを抽出するために、このモデルをシェーディングを施したモデルへ改良した。伸縮の動作に対する指の各関節の角度を $\theta_{b,t,1}, \theta_{b,t,2}, \theta_{b,t,3}, \theta_{b,i,1}, \theta_{b,i,2}, \theta_{b,i,3}, \dots$ で定義する



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



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

図 1 従来の方法の誤推定

Fig. 1 Mistake of conventional method

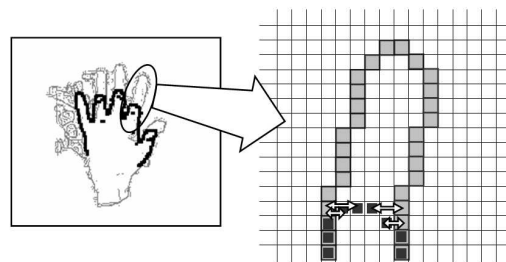


図 2 エッジの対応

Fig. 2 Corresponding of edges

(図 3 に示す)。根元の関節だけは z 軸回りの回転の動作も行う ($\theta_{o,t}, \theta_{o,i}, \dots$)。手指モデル全体が平行移動 (t_x, t_y, t_z)、回転 ($\theta_{r,x}, \theta_{r,y}, \theta_{r,z}$) することにより姿勢が変化する。手指モデルは、これらの合計した 26 個の変数から成り立っている。26 個の変数をまとめて

$$\vec{\theta} = (\theta_{b,t,1}, \dots, \theta_{r,z}) \quad (1)$$

で定義する。

典型的な自由度の変数から作成した手指モデルの CG 画像の例を図 4 に示す。指の 3 個の関節に関して伸縮の関節角度は連動している [5]。一部の例外を除いて、人指し指、中指、薬指、小指については隣り合う指の伸縮の角度にも拘束がある。図 4(e) のように伸縮と垂直な方向の曲げについても各指の関節角度は連動するとしている。CG 画像からエッジを抽出すると図 5 のようになる。

2.2 候補モデル

初期フレームでの手指形状、位置、姿勢は既知であると仮定している。様々な指の長さ、太さの手指モデルの見えを画像に当てはめて、最も画像に合うものを選択する。直前のフレームの推定手指形状 $\vec{\theta}$ から $\Delta\theta^c, \Delta t^c$ 単位で 3 段階に変化させた候補を生成する。変化の量子化幅を表 1 示す (表中では、 θ_b の量子化幅を $\Delta\theta_{bend}^c$ 、また中指以外の θ_o のそれを $\Delta\theta_{open}^c$ と表し

表 1 探索の中心からの量子化幅

Table 1 Quantization width from search center

$\Delta\theta_{bend}^c$ [°]	$-\zeta$	0	ζ
$\Delta\theta_{open}^c$ [°]	-6	0	6
$\Delta\theta_{o,m}^c$ [°]	-	0	-
$\Delta\theta_{r,x}^c$ [°]	-30	0	30
$\Delta\theta_{r,y}^c$ [°]	-15	0	15
$\Delta\theta_{r,z}^c$ [°]	-6	0	6
$\Delta t_x^c, \Delta t_y^c$ [mm]	-8	0	8
Δt_z^c [mm]	-40	0	40

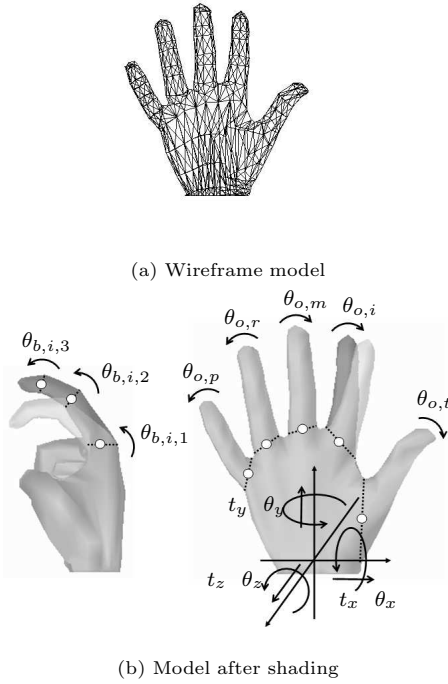


図 3 手指モデル

Fig. 3 Hand model

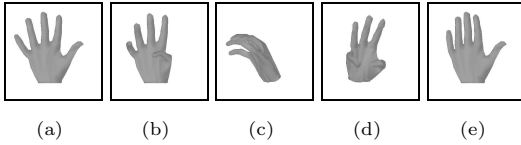


図 4 典型的な手指モデルの CG 画像

Fig. 4 CG of typical hand model

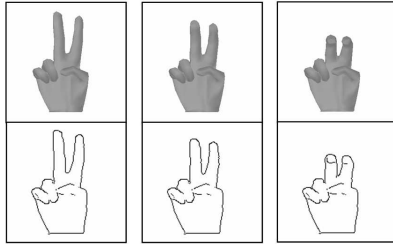


図 5 モデルのエッジ画像

Fig. 5 Edge image of hand model

た). 手指の場合, 指が伸びているときと比べて, 縮んでいるときの方が θ の単位変化当りの見えの変化が小さい. そこで見えを均一に変形させるために, 指が伸びているほど大きくなるように $9[^\circ]$ から $15[^\circ]$ の間で ζ の値を設定した.

3. 画像特徴に対する各候補の尤度を考慮した最適な候補の推定

入力画像をエッジ特徴と肌色領域の特徴で表わし, それを I とする. 最尤推定法では, 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

$$\hat{\Theta} = \arg \max_{\Theta_j} P(I|\Theta_j) \quad (2)$$

を選択する. 尤度のモデルは, 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る. 本節ではまず本研究で使用した画像特徴の相違度を示す. 次に本手法による識別原理を説明したあと, ベイズ推定の枠組に基づいて具体的なモデル選択基準を導出する.

3.1 画像特徴の相違度

本手法では, 手指領域のシルエットとエッジ特徴に基づいて入力画像と候補モデルの見えの相違度を評価する.

シルエットに注目した相違度 $f_{skin}(A_{\bar{\theta}}^{(s)}; I^{(s)})$ は, 入力画像の肌色領域 $I^{(s)}$ のうち, 候補モデル $\bar{\theta}$ から CG で生成したシルエット $A_{\bar{\theta}}^{(s)}$ に重ならない部分の面積で評価する.

またエッジ特徴に注目した相違度を評価するため, chamfer matching [1] [12] [16] を改良し, 勾配方向が近いエッジ同士で chamfer matching [14] [18] によってエッジ特徴の対応づけを行う. また各エッジ点の信頼性をエッジ強度で評価して, 信頼性の高いエッジの寄与率を高くする. 本論文では, エッジの強度と勾配方向を考慮した相違度 f_{dist} として以下の式を用いた. 候補モデル $\bar{\theta}$ から生成したエッジ $A_{\bar{\theta}}^{(e)}$ の j 番目のエッジ点を $\bar{x}_{\bar{\theta},j}$ とし, 入力の k 番目のエッジ点を $\bar{x}_{I^{(e)},k}$ として

$$f_{dist}(A_{\bar{\theta}}^{(e)}; I^{(e)}) = \sum_j w_{\bar{\theta},j} \min_k (||\bar{x}_{\bar{\theta},j} - \bar{x}_{I^{(e)},k}|| + f_{I^{(e)},k} + g(j,k)). \quad (3)$$

ここで, $\min(\cdot)$ の中の第一項 $||\bar{x}_{\bar{\theta},j} - \bar{x}_{I^{(e)},k}||$ は 2 次元のユークリッドノルムである. $w_{\bar{\theta},j}$ はモデルの j 番目のエッジ点の強度に基づいた寄与率で, モデルのエッジ点のエッジ強度を $d_{\bar{\theta},j}$ とすると

$$w_{\bar{\theta},j} = \frac{d_{\bar{\theta},j}}{\sum_l d_{\bar{\theta},l}} \quad (4)$$

である. 第二項 $f_{I^{(e)},k}$ は強度が低い入力のエッジに対するペナルティ項であり, 入力エッジの強度 $d_{I^{(e)},k}$ を用いて次式で定義

する。

$$f_{I^{(e)},k} = -w_d d_{I^{(e)},k}. \quad (5)$$

ただし, w_d は重みを表わす定数で経験的に決める. 第三項 $g(j,k)$ はモデルのエッジと入力のエッジとの対応を評価するための勾配方向の相違の度合いである. j 番目のモデルの勾配方向を $\phi_{\bar{\theta}_j}$, k 番目の入力の勾配方向を $\phi_{I^{(e)},k}$ とし,

$$g(j,k) = w_\phi \|\phi_{\bar{\theta}_j} - \phi_{I^{(e)},k}\|. \quad (6)$$

ただし, w_ϕ は重みを表わす定数である. この改良した chamfer matching は距離変換を用いて高速に計算することができる.

3.2 識別原理

1. 節で述べた誤推定の例を用いて, 本手法における手指形状の識別原理を説明する.

図 1(a),(c) に相当するモデルをそれぞれ θ_a, θ_c とする. θ を CG レンダリングして生成した理想的な画像の見えを A_θ とする. 画像中の手指の形状が θ_c の場合に A_{θ_c} が出現する確率を考える. モデルと見えが同一手指形状なので, A_{θ_c} に対する θ_c の尤度 $p(A_{\theta_c}|\theta_c)$ は十分大きな値をとる.

つぎに, A_{θ_a} が出現する確率を考える. 図 2 に示すように, 形状が異なるにもかかわらず大部分が θ_c の画像に含まれるので, その尤度 $p(A_{\theta_a}|\theta_c)$ も大きくなってしまふ.

そのため, 本来 θ_c が正解となるべき実際の入力画像 I が得られたとき, I にモデルの見えを当てはめる場合にも, $A_{\theta_a}, A_{\theta_c}$ がともに I に対して照合することが予想される. この時, 撮像時のノイズや形状モデルの不正確さ, モデル変数の量子化誤差などのために, たまたま A_{θ_a} の照合の評価の方が大きかった場合には, 誤って θ_c が選択される.

このような従来法の問題を解決するために, ここで, それぞれのモデルの見えに対する θ_a の尤度 $p(A_{\theta_a}|\theta_a)$, $p(A_{\theta_c}|\theta_a)$ に着目する. この場合, $p(A_{\theta_a}|\theta_a)$ は $p(A_{\theta_c}|\theta_c)$ と同様に大きく, θ_a の CG 画像中に見え A_{θ_c} は出現しないため $p(A_{\theta_c}|\theta_a)$ は小さくなることが予想される.

θ_a については $p(A_{\theta_a}|\theta_a)$ は大きく $p(A_{\theta_c}|\theta_a)$ が小さくなり, θ_c については $p(A_{\theta_a}|\theta_c)$, $p(A_{\theta_c}|\theta_c)$ が共に大きくなることを考慮すると, 見え A_{θ_a} と A_{θ_c} が共に画像中に観測されたのなら, その画像はモデル θ_c と推定され, A_{θ_a} のみが観測されたのなら, その画像はモデル θ_a と推定されるべきである (図 6 参照).

したがって, あらかじめ可能性のある見え A_{θ_k} に対する各候補モデル θ_j の尤度関数 $p(A_{\theta_k}|\theta_j)$ を見積もっておき, 画像 I に対する各見え A_{θ_k} の尤度 $p(I|A_{\theta_k})$ を確率的に統合評価することで, 正しい候補モデルを選択することができる.

$k \neq j$ の場合の尤度 $p(A_{\theta_k}^{(e)}|\theta_j)$ は, 本来モデル θ_j に推定されるべき画像に誤って θ_k の見えを当てはめたときの尤度 (以下誤り照合尤度と呼ぶ) である.

3.3 誤り照合尤度の評価に基づくモデル選択

前節で述べた識別原理をバイズ推定の枠組みによって導出する.

$I^{(e)}$ に対する θ_j の尤度を $p(I^{(e)}|\theta_j)$ で表わす. $A_{\theta_1}^{(e)}, A_{\theta_2}^{(e)}, \dots$

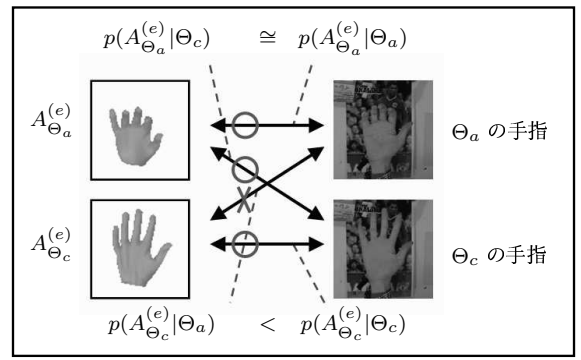


図 6 エッジ画像に対する尤度

Fig. 6 Likelihood for edge images

は理想的なエッジを表わす. θ_j の元で $A_{\theta_1}^{(e)}, A_{\theta_2}^{(e)}, \dots$ は互いに排反と仮定できるので

$$\begin{aligned} p(I^{(e)}|\theta_j) &= \sum_k p(I^{(e)}, A_{\theta_k}^{(e)}|\theta_j) \\ &= \sum_k p(I^{(e)}|A_{\theta_k}^{(e)}, \theta_j) p(A_{\theta_k}^{(e)}|\theta_j) \end{aligned} \quad (7)$$

エッジ $A_{\theta_k}^{(e)}$ がエッジ画像 $I^{(e)}$ を生成する情報を全て含んでいるので条件 θ_j を省いても良く, $p(I^{(e)}|A_{\theta_k}^{(e)}, \theta_j) = p(I^{(e)}|A_{\theta_k}^{(e)})$ と仮定して

$$p(I^{(e)}|\theta_j) = \sum_k p(I^{(e)}|A_{\theta_k}^{(e)}) p(A_{\theta_k}^{(e)}|\theta_j) \quad (8)$$

従来の最尤推定法は, $k = j$ の場合のみを考慮することに相当し, 本手法では $k \neq j$ のときの誤り照合尤度を新たに考慮していることが従来法と異なる点である. 図 1 の例では, 入力画像中の手指形状は θ_c であるが, $A_{\theta_a}^{(e)}$ が当てはまる. θ_c の手指画像には誤って $A_{\theta_a}^{(e)}$ が当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.

同様に, 入力画像から求めた肌色領域を表わす画像を $I^{(s)}$ とし, θ の手のシルエットを $A_\theta^{(s)}$ とする. $I^{(s)}$ に対する θ_j の尤度は

$$p(I^{(s)}|\theta_j) = \sum_k p(I^{(s)}|A_{\theta_k}^{(s)}) p(A_{\theta_k}^{(s)}|\theta_j) \quad (9)$$

である.

θ_j の元では $I^{(e)}$ と $I^{(s)}$ は独立としてよく,

$$p(I|\theta_j) = p(I^{(e)}|\theta_j) p(I^{(s)}|\theta_j) \quad (10)$$

を候補の選択の基準とする.

$p(A_{\theta_k}^{(e)}|\theta_j)$ の求め方を 3.4 節で説明して, $p(I^{(e)}|A_{\theta_k}^{(e)})$ の求め方を 3.5 節で説明する. シルエットについても $p(A_{\theta_k}^{(s)}|\theta_j)$, $p(I^{(s)}|A_{\theta_k}^{(s)})$ はエッジと同様の求め方を用いた.

3.4 見えに対する手指形状候補の尤度モデル

離散化された変数の候補 θ_k のモデルのエッジ $A_{\theta_k}^{(e)}$ に対する θ_j の尤度は,

$$\begin{aligned} p(A_{\theta_k}^{(e)}|\theta_j) &= \int_{\theta_j} p(A_{\theta_k}^{(e)}, \vec{\theta}_j|\theta_j) d\vec{\theta}_j \\ &= \int_{\theta_j} p(A_{\theta_k}^{(e)}|\vec{\theta}_j, \theta_j) p(\vec{\theta}_j|\theta_j) d\vec{\theta}_j \\ &= \int_{\theta_j} p(A_{\theta_k}^{(e)}|\vec{\theta}_j) p(\vec{\theta}_j|\theta_j) d\vec{\theta}_j. \end{aligned} \quad (11)$$

ただし, $\vec{\theta}_j$ は θ_j の中心で, 離散化された変数の候補 θ_j のも

とでは $p(\theta_j|\Theta_j)$ が一様分布として差し支えない。また量子化幅が十分小さければ、 $p(A_{\Theta_k}^{(e)}|\theta_j^*)$ は $\theta_j \in \Theta_j$ に対して一定であると仮定して $p(A_{\Theta_k}^{(e)}|\Theta_j) = p(A_{\Theta_k}^{(e)}|\theta_j^*)$ とできる。

ここでは、画像特徴の相違度 f_{dist} の分布を正規分布と仮定して、画像特徴の照合度 $p(A_{\Theta_k}^{(e)}|\theta_j^*)$ を評価する [13] [14] [17]。 $\bar{\theta}$ の CG 画像から求めたエッジ画像を表わすレンダリングモデル $I_r^{(e)}(\bar{\theta})$ を導入して、

$$p(A_{\Theta_k}^{(e)}|\theta_j^*) = \alpha_{\theta_j^*}^{(e)} \exp(-(d_M^{(e)}(k, j))^2) \quad (12)$$

ここで、

$$d_M^{(e)}(k, j) = \frac{f_{dist}(A_{\Theta_k}^{(e)}; I_r^{(e)}(\theta_j^*))}{\sigma_M^{(e)}} \quad (13)$$

で、 $f_{dist}()$ 関数は、 $A_{\Theta_k}^{(e)}$ を基準にした $I_r^{(e)}(\theta_j^*)$ の距離である。その距離の分散を $\sigma_M^{(e)2}$ としている。 $\sigma_M^{(e)}$ の値は経験的に決定した。 $\alpha_{\theta_j^*}^{(e)}$ は正規化定数で、

$$\alpha_{\theta_j^*}^{(e)} = \left(\sum_k \exp(-(d_M^{(e)}(k, j))^2) \right)^{-1}. \quad (14)$$

である。

シルエットについては、 $p(A_{\Theta_k}^{(s)}|\Theta_j)$ を求める。 $f_{dist}(A_{\Theta_k}^{(e)}; I_r^{(e)}(\theta_j^*))$ の値の代わりに、CG 画像から求めたシルエット画像 $I_r^{(s)}(\theta_j^*)$ にシルエット $A_{\Theta_k}^{(s)}$ を重ね合わせて求めた $f_{skin}(A_{\Theta_k}^{(s)}; I_r^{(s)}(\theta_j^*))$ を用いて、 $p(A_{\Theta_k}^{(s)}|\Theta_j)$ を求める。

3.5 画像特徴に対する見えの尤度

エッジ画像 $I^{(e)}$ に対する見え $A_{\Theta_k}^{(e)}$ の尤度は

$$p(I^{(e)}|A_{\Theta_k}^{(e)}) = \beta_{\Theta_k}^{(e)} \exp\left(-\frac{(f_{dist}(A_{\Theta_k}^{(e)}; I^{(e)}))^2}{\sigma_I^{(e)2}}\right). \quad (15)$$

$\sigma_I^{(e)2}$ は chamfer matching の距離の分散で、経験的に決定する。 $\beta_{\Theta_k}^{(e)}$ は正規化定数で、 Θ_k によって値が異なる。 $p(i^{(e)}|A_{\Theta_k}^{(e)})$ を、全てのエッジ画像 $i^{(e)}$ について積分する

$$\int p(i^{(e)}|A_{\Theta_k}^{(e)}) di^{(e)} = 1. \quad (16)$$

ほとんどの $i^{(e)}$ について $p(i^{(e)}|A_{\Theta_k}^{(e)}) = 0$ で、可能性のある候補についてだけ大きな値をとると仮定してよい。ゆえに以下のように見積もることができる。

$$\begin{aligned} \int p(i^{(e)}|A_{\Theta_k}^{(e)}) di^{(e)} &\approx \int p(I_r^{(e)}(\theta_l)|A_{\Theta_k}^{(e)}) d\theta_l \\ &\approx \sum_l p(I_r^{(e)}(\theta_l^*)|A_{\Theta_k}^{(e)}) \cdot \delta \\ &= \beta_{\Theta_k}^{(e)} \sum_l \exp(-(d_M^{(e)}(k, l))^2) \cdot \delta \\ &\equiv 1 \cdot \delta \end{aligned} \quad (17)$$

ただし、 δ は Θ の量子化幅を表わす。したがって

$$\beta_{\Theta_k}^{(e)} = \left(\sum_l \exp(-(d_M^{(e)}(k, l))^2) \cdot \delta \right)^{-1} \quad (18)$$

となる。 $A_{\Theta_k}^{(e)}$ が他の形状の手指画像 $I_r^{(e)}(\theta_l)$ の多くに当てはまりやすい（照合誤りが起きやすい）ときには、 $\beta_{\Theta_k}^{(e)}$ が小さくなり、逆に $A_{\Theta_k}^{(e)}$ が他の形状の画像にあまり当てはまらないときは（照合誤りが起きにくい）、 $\beta_{\Theta_k}^{(e)}$ が大きくなる。図 6 においては、モデルのエッジを基準とした入力のエッジ画像の chamfer matching の結果が同じくらいでも、 $p(I^{(e)}|A_{\Theta_k}^{(e)})$ よりも $p(I^{(e)}|A_{\Theta_k}^{(e)})$ の方が大きくなる。

4. 精確な関節角度の推定

照合結果をもとに、更に詳細に指の関節角度や手全体の回転角度の変化を推定する方法について以下に述べる。候補モデルの変数の量子化誤差により、照合の最良候補の変数と入力画像中の手の変数とは少しの誤差がある。そのため更に精確に関節角度を推定する必要がある。

見えが入力画像に合うようにワイヤフレームの手指 CG モデルを変形させることにより、精確な手指の三次元形状を復元する方法がある [19]。ただし、この方法は手の表面の曲面形状を復元しているだけで関節角度の推定は行っていない。我々は、まずこの方法によって見えが入力画像に合うように CG モデルを変形させる。その CG モデルのワイヤフレームの頂点の座標値から関節角度を求めることにより精確に関節角度を推定する。精確な関節角度推定の処理の手順を以下に示す。

1. 画像に重ね合わせた最良候補の CG モデルのエッジと入力画像中の手のエッジとを対応付ける。

2. モデルのエッジの座標を入力画像中の手のエッジの座標に近づけるために、エッジの対応付けから見えの変形を求める。

3. 手の自由度は大きく探索空間は広いので、あらかじめ学習した見えの変形を考慮して、画像から求めた見えの変形から CG モデルを画像平面内で変形する。

4. 1. に戻り、最良候補の CG モデルの代わりに、3. で変形させた CG モデルを入力画像に重ね合わせて、再び見えの変形を求め CG モデルを変形させる。この操作を繰り返す。

5. 入力画像に合わせたモデルと手指モデルの変数を変化させて作成した手指モデルとの間で、ワイヤフレームの各頂点において二次元座標のユークリッド距離を求める。その距離の総和を求めて、最急降下法を用いて頂点の距離の総和が最小となる関節角度を求める。その関節角度が推定結果である。

5. 実 験

250 枚の入力画像に対して推定実験を行った結果、 $p(I|\Theta_j) = p(I|A_{\Theta_j}^{(e)})p(I|A_{\Theta_j}^{(s)})$ を基準として照合を行った場合と提案手法の推定の成功率はそれぞれ 70.4%、82.0% となった。成功率からも提案手法の有効性が確認できる。

照合誤りを考慮することにより、従来法の照合の間違いが改善された例を図 7 に示す。従来法では指が曲がった形状が誤って認識されているが、本手法それが改善されている。

時系列画像について推定実験を行った結果を図 8 に示す。手が横に傾いているときはセルフオクルージョンのため指の一部が見えなくなる。また、手以外の背景エッジが現れている。このような例でも、本手法では正しく推定できている。別の時系列画像に対する推定結果を図 9、図 10 に示す。こちらでも正しい推定が行われている。

6. おわりに

本研究では、画像特徴の照合による複雑背景下での手指形状の推定方法について述べた。誤った候補モデルが照合する場合を想定して、見えに対する手指形状候補の尤度を求めた。その

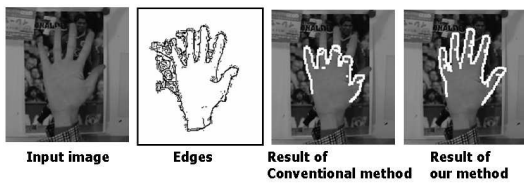


図 7 実験結果
Fig. 7 Experimental result

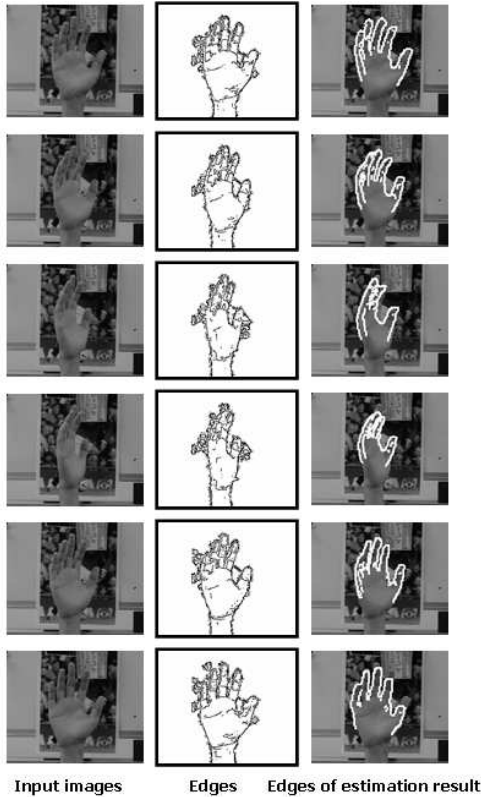


図 8 実験結果
Fig. 8 Experimental result

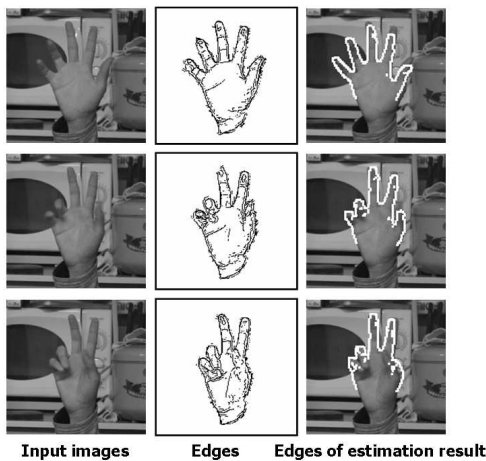


図 9 実験結果
Fig. 9 Experimental result

尤度を考慮して，入力画像に対する見えの照合結果から最適な候補を選択することができた．実験によって本手法の有効性を示した．

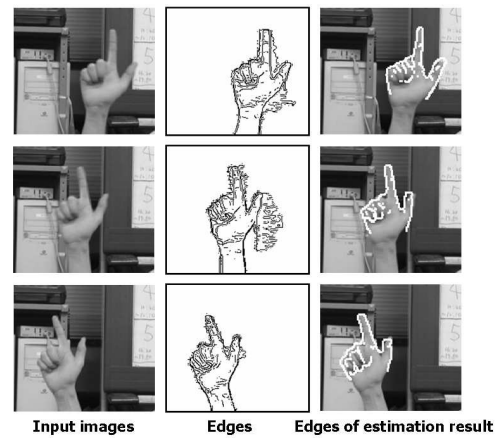


図 10 実験結果
Fig. 10 Experimental result

今後の課題として，エッジやシルエット以外の画像特徴の利用や時系列での探索法の改良を考えている．また，握むや握るといった動作の認識を考えている．

文 献

- [1] X. Liu, K. Fujimura. "Hand Gesture Recognition using Depth Data". In *Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition*, pp. 529-534, 2004.
- [2] 岩井儀雄, 八木康史, 谷内田正彦. "単眼動画画像からの手の3次元運動と位置の推定". *信学論 (D-II)*, D-II Vol.J80-D-II, No.1, pp.44-55, 1997.
- [3] Sung Uk Lee, Issac Cohen. "3D Hand Reconstruction from a Monocular View". In *Proc. 17th Int. Conf. on Pattern Recognition*, Volume 3, pp.310-313, 2004.
- [4] 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". *信学論 (D-II)*, D-II Vol.J79-D-II, No.1, pp.26-35, 1996.
- [5] N. Shimada, K. Kimura, and Y. Shirai. "Real-time 3-D Hand Posture Estimation based on 2-D Appearance Retrieval Using Monocular Camera". In *Proc. Int. Workshop on RATFG-RTS*, pp. 23-30, 2001.
- [6] 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". *信学論 (D-II)*, D-II Vol.J88-D-II No.8, pp.1643-1651, 2005.
- [7] T. F. COOTES, C. J. TAYLOR, D. H. COOPER, AND J. GRAHAM. "Active Shape Models-Their Training and Application". In *COMPUTER VISION AND IMAGE UNDERSTANDING*, Vol.61, No.1, pp. 38-59, 1995.
- [8] M. Jeong, Y. Kuno, N. Shimada, and Y. Shirai. "Recognition of shape-changing hand gestures". In *IEICE Trans. inf. & Syst.*, Vol. E85-D, no.10, pp.1678-1687, 2002.
- [9] M. Isard and A. Blake. "Visual tracking by stochastic propagation of conditional density". In *Proc. European Conf. Computer Vision*, pp.343-356, 1996.
- [10] M. Isard and A. Blake. "ICONDENSATION:Unifying low-level and high-level tracking in a stochastic framework". In *Proc. European Conf. Computer Vision*, pp.767-781, 1996.
- [11] T. Heap and D. Hogg. "Wormholes in Shape Space:Tracking through Discontinuous Changes in Shape". In *6th Int. Conf. on Computer Vision*, pp. 344-349, 1998.

- [12] V. Athitsos, and S. Sclaroff. "Estimating 3D Hand Pose from a Cluttered Image". In *IEEE Conf. on Computer Vision and Pattern Recognition*, II, pp. 432-439, 2003.
- [13] H. Zhou and T. S. Huand. "Tracking Articulated Hand Motion with Eigen Dynamics Analysis". In *9th Int. Conf. on Computer Vision*, Vol.2, pp. 1102-1109, 2003.
- [14] B. Stenger, A. Thayananthan, P.H.S. Torr, and R. Cipolla. "Model-Based Hand Tracking Using a Hierarchical Bayesian Filter". In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.28, No.9, pp. 1372-1384, 2006.
- [15] Y. Wu, J. Lin, and T.S. Huang. "Analyzing and Capturing Articulated Hand Motion in Image Sequences". In *IEEE TRANS. ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, Vol.27, No. 12, pp. 1910-1922, 2005.
- [16] H. G. Barrow, J. M. Tenenbaum, R. C. Bolles, and H. C. Wolf. "Parametric correspondence and chamfer matching: Two new techniques for image matching". In *Proc. 5th Int. Joint Conf. Artificial Intelligence*, pp. 659-663, 1977.
- [17] A.Blake and M.Isard. "Active Contours". In *London:Springer-Verlag*, 1998.
- [18] R. Navaratnam , A. Thayananthan , P. H. S. Torr , R. Cipolla. "Hierarchical Part-Based Human Body Pose Estimation". In *Proc. British machine Vision Conference*, 2005.
- [19] T. Heap and D. Hogg "Towards 3D Hand Tracking using a Deformable Model". In *2nd Int. Conf. on Automatic Face and Gesture Recognition*, pp.140-145, 1996.
- [20] 稲葉宏幸, 寺島史雄, 若杉耕一郎, 笠原正雄" 誤り訂正符号を適用した文字認識システムに関する一考察". *信学論 (D-II)* , D-II Vol.J77-D-II, No.2, pp.353-361, 1994.