

照合の誤り予測と見え変化の学習に基づく複雑背景下での手指形状推定

今井 章博† 島田 伸敬‡ 白井 良明‡

†大阪大学大学院工学研究科機械工学専攻

‡立命館大学情報理工学部知能情報学科

†〒565-0871 大阪府吹田市山田丘 2-1 imai@cv.mech.eng.osaka-u.ac.jp

‡〒525-8577 滋賀県草津市野路東 1-1-1 {shimada,shirai}@ci.ritsumeimei.ac.jp

本論文では、一般的な背景の下で撮影された人の手指の時系列画像から、その手指形状を認識する方法について述べる。従来の active contour による輪郭追跡では、セルフオクルージョンによる急激な輪郭の変形を追跡するのは難しい。また、入力画像の特徴と手指モデルの画像特徴とを照合する方法では、モデルの手指形状が入力と異なっているにもかかわらず、マッチしてしまうことがある。複雑背景やセルフオクルージョンによってエッジが抽出されなかったり、皺などのテクスチャでエッジが抽出されるため、本来のオクルーディングエッジでないエッジを誤って照合してしまうことが原因である。

入力画像からの手指エッジの検索の補助とするために、輪郭やオクルーディングエッジを用いて入力画像と候補モデルとの照合を行って、最もマッチする候補モデルを選ぶ必要がある。入力画像とマッチする候補モデルを見つけるために、画像特徴に対する候補モデルの尤度モデルを構築して照合を行う。照合結果の手指モデルを手がかりとして入力画像中のエッジとモデルのエッジとを対応付けする。手指モデルの変数を変化させたときの見えの変化を学習しておくことにより、前時刻の推定結果をもとにモデルのエッジが入力画像中のエッジに合うように、手指形状を変形させる。得られた手指の三次元形状から関節角度を決定する。実験結果によって本手法の有効性を示す。

Hand Posture Estimation in Complex Background by Mis-match Prediction and Appearance Variation Learning

Akihiro IMAI† Nobutaka SHIMADA‡ Yoshiaki SHIRAI‡

†Dept.of Mechanical Engineering, Osaka University

‡Dept.of Human and Computer Intelligence, Ritumeikan University

†2-1 Yamadaoka, Suita, Osaka 565-0871 imai@cv.mech.eng.osaka-u.ac.jp

‡1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577 {shimada,shirai}@ci.ritsumeimei.ac.jp

This paper proposes a novel method of estimating 3-D hand posture from image sequences captured in complex background. Conventional active contour methods are essentially weak in appearance changes due to self-occlusion. Model-based methods also cause mistakes. Often edges are not observed due to complex background and self-occlusion. Edges are observed from texture such as crease. Because model edges match with something different from occluding edges, an invalid model matches to an input image.

Contour and occluding edges are used for finding the best-matching model candidate to an input image. The best model is selected considering likelihood of a model for image features. Hand edges are selected from an input image guided by the model. Hand model is deformed in order to fit model edges with input edges considering learned possible appearance variation. Model DOF variables are determined from the deformed hand model.

1 はじめに

最近ではマウスやキーボードなどのこれまでの入力デバイスに代わるものとして、ジェスチャによるヒューマンインターフェースが注目を集めている。仮想世界や実世界での物体のマニピュレーション、手話認識などのアプリケーションを実現させるためには、データグローブで計測できるような微細な手指の動きを画像情報から計

測する必要がある。

背景を単純と仮定してジェスチャ認識を試みた研究には、手指のアピアランスをモデルとして利用した方法と三次元形状をモデルとして利用した方法の二種類のアプローチがある。前者の方法では、記録されている対象物体の様々な姿勢の二次元の特徴から入力の手指の特徴と最も照合するものが選ばれられる [1]。後者の方法では、画像から局所的な特徴を抽出して、その特徴に対して三

次元形状モデルを当てはめている [2]。また、典型的な三次元形状の手指モデルから輪郭を生成し、入力の手指と照合を行う方法もある [3][4][5]。ただし、これらの方法をそのまま複雑背景で撮影された画像に適用することはできない。

複雑背景の画像から特徴を抽出した例を図 1 に示す。(a) が入力画像、(b) が肌色領域を抽出した画像、(c) がエッジ画像、(d) が肌色領域と隣接しているエッジを二値化した画像である。背景に手以外の肌色領域があると、色情報から手領域を限定することはできない。また、手が背景の肌色領域と重なると、その部分のエッジがうまく抽出されないことがある。複雑背景で撮影された画像から手の特徴だけを抽出するのは非常に難しい。

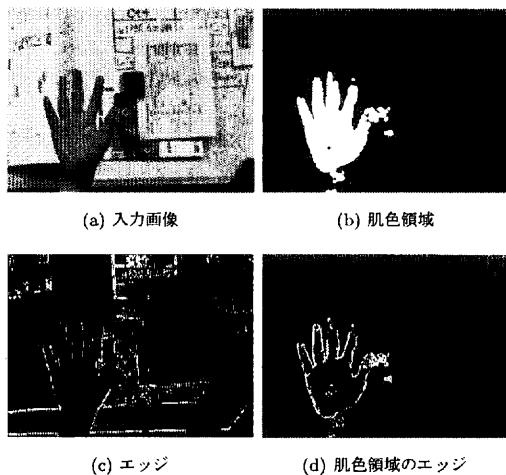


図 1: 複雑背景下での特徴抽出

複雑背景下での推定方法も、モデルとしてアピアランスを利用した方法と三次元形状を利用した方法とに分類できる。前者の方法の一つに active contour を用いて手の輪郭を抽出する方法がある [6]。この方法では、輪郭線は時系列で滑らかに変化すると仮定して、少し前の時刻で発見された手指の輪郭線の近傍領域を探索し、画像に合うように輪郭線を変形させていく。これを時系列で繰り返すことによって、手指の輪郭線を追跡している。この方法は輪郭線が滑らかに変化している場合には有効であるが、隠蔽が生じると手指の輪郭線は急激に変化する。例を図 2 に示す。(a) では小指と薬指の間に輪郭が現れているが、(b) では小指により薬指が隠蔽されているので輪郭が現れていない。薬指と中指の間と中指と人指し指の間も同様に輪郭の一部が見えなくなっている。active contour による方法は、このような場合には検出を誤ることがある。変形による形状遷移の可能性を実際の手指動画像から学習し、得られた形状同士の隣接関係を元に探索を行う方法が提案されている [7]。しかし、遷移可能な形状の組み合わせの数は膨大なものとなるので、手動で連結性を定義するのは困難である。

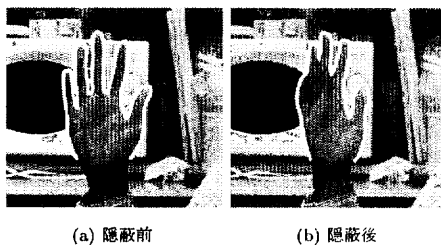


図 2: 隠蔽による輪郭の変形

Ong らの方法 [8] では、登録してある手指画像の中から入力画像と最もマッチするものを選んで、手指形状を認識している。これは、入力画像に対して局所的な特徴を検出する複数のオペレータを施し、それらの結果を統合して入力画像と最もマッチする画像を選んでいる。しかし、アピアランスをモデルとして利用する方法では、手動で実画像を収集する必要があるため代表的な手指形状のモデルしか記録しておくことができない。細かい関節角度を知るためには、三次元モデルを用いて自動で特徴の作成を行う必要がある。

後者の方法では、典型的な手指形状の三次元モデルから画像を生成し、その画像からエッジ画像や直線セグメントを計算する。入力との間でそれらを照合し、最適なものを選ぶことにより手指形状を推定している [9]。また、手指の形状（関節角度）を状態とし、パーティクルフィルタ [10][11] と同様に時系列での追跡を組み込んだ方法も提案されている [12][13][14]。

この方法には照合の評価の方法に問題があり、誤推定を起こすことがある。誤推定の結果を図 3 に示す。(a) が入力画像、(b) が推定結果の CG モデル、(c) が入力画像に推定結果の形状のモデルから抽出したエッジを重ねた画像、(d) が入力と同じ形状の CG モデルである。この例では、入力の手指形状と比較して指が曲がった手指形状の特徴が入力画像とマッチした。照合結果は、指の形状が入力のものと異なっているが、この指の大部分が前の指によって隠蔽されている。指の形状が異なってもその部分が隠蔽されていれば、入力画像と比較しても見えの相違は生じないのでマッチしてしまう。したがって、モデルの見えが画像とマッチした形状を選んで、それが正しい推定結果であるとは限らない。

Jeong らの方法 [7] では、輪郭の変形を学習するために一枚一枚画像を収集する。収集した画像の中で連結のある画像の組み合わせに対して手動で形状遷移の可能性を定義する。変形が滑らかでなくても連結性のある輪郭形状も入力画像中で探索することによりセルフオクルージョンに対応している。

本研究では、画像特徴として輪郭、オクルーディングエッジを用いる。あらかじめ特徴に対する候補の形状の尤度を計算して、誤りが起こりうることを予測しておく、そのことを考慮して最適な候補の形状を選ぶ。セルフオクルージョンに対応するために三次元モデルを用いて自動で見えの変化を学習しておく。学習した見えの変化を考慮して画像平面内のエッジの移動ベクトルから三次元形状の変形を推定し、関節角度を計算する。

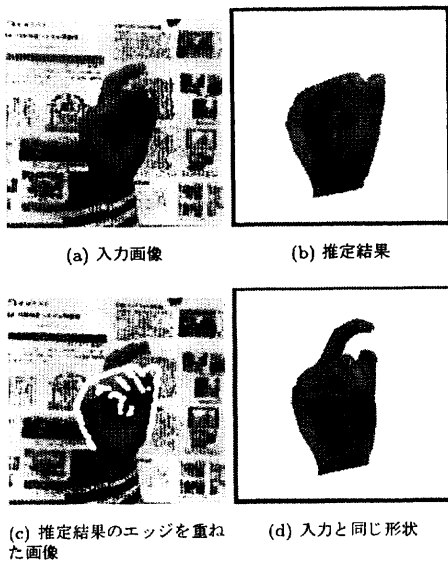


図 3: 従来の方法の誤推定

2 本手法の概要

初期フレームでの手指形状、位置、姿勢、サイズは既知であると仮定している。その手指形状の特徴を三次元モデルから作成し、初期フレームの入力画像に重ね合わせて、手指領域内の画素の色を取得する。肌色の分布を正規分布と仮定して、取得した色から肌色モデルを作成しておく。また、様々な指の長さ、太さの手指モデルの見える画像に当てはめて、最も画像に合うものを選択する。これにより個人によって異なる指の長さ、太さを決めておく。

推定方法の手順を図 4 に示す。手順は大まかに二つに分けられる。まず入力画像と候補モデルとの照合により初期形状を決定して、入力画像中から特徴の検索を行う(3 節で述べる)。検索された特徴から手指の三次元形状を復元して特徴の対応付けを修正する。これを繰り返す。最終的に得られた三次元形状から関節角度を決定する(4 節で述べる)。

特徴の検索では、まず入力画像からエッジ、肌色領域を抽出する。あらかじめ照合の誤り予測尤度分布を CG モデルシミュレーションによって学習する。入力と候補モデルの CG 画像との間でエッジ、領域とを照合し、誤り予測尤度分布を考慮して最良の候補を選ぶ。その候補のエッジの近傍領域で入力の手指のエッジを検出する。

推定のために、検索したエッジから移動ベクトルが必要になる。画像平面内の最良候補の位置に、初期形状として直前のフレームでの推定結果の手指形状を合わせる。初期形状のエッジから検索された入力のエッジまでの移動ベクトルを求める。移動ベクトルから三次元空間での手指形状の変形を計算する。計算した変形に基づき直前のフレームでの推定結果の手指形状を変形させる。初期形状から移動ベクトルを求めたのと同様に、変形し

た手指形状から移動ベクトルを求め、同様の手指形状の変形を行い、序々に画像のエッジに合うように手指形状を変形させていく。最終的に画像のエッジに合うように復元された手指の三次元形状から関節角度、姿勢、サイズの変数を決定する。

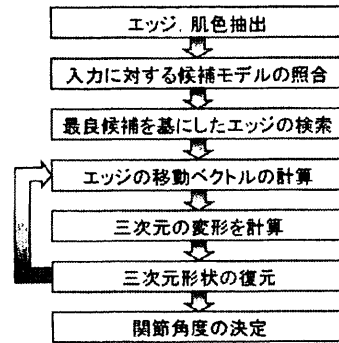


図 4: 推定の手順

3 モデルの照合に基づく手指特徴の検索

3.1 照合の候補モデル

本研究で用いている三次元手指モデルを図 5 に示す。このモデルは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。このモデルに対してシェーディングを行う(図 6)。指は各関節が曲がることにより伸縮の動作を行う。根元の関節だけは伸縮と垂直方向にも曲がる。手指モデル全体が回転することにより姿勢が変化する。また、画像画面内の平行移動と手全体のサイズについて自由度を持っている。関節角度、回転角度、画像平面での座標、サイズを合計して 26 個の変数から成り立っている。

各変数間には、拘束がある。例えば、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。拘束を考慮し

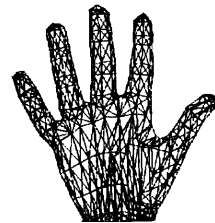


図 5: ワイヤフレームモデル

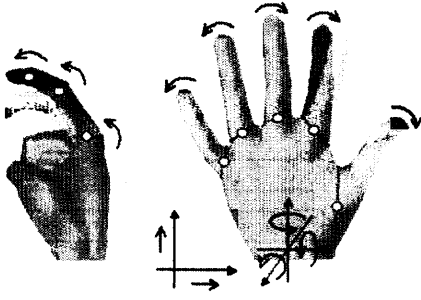


図 6: シェーディング後のモデル

て典型的な変数の組み合わせだけを選びだして、その三次元形状を候補のモデルとする。候補のモデルのCG画像の例を図7に示す。また、CG画像からエッジを抽出すると図8のようになる。

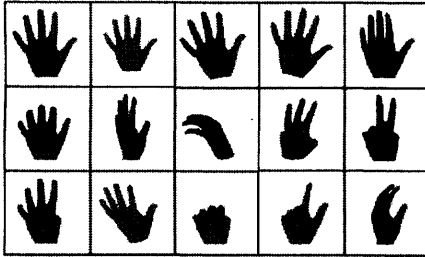


図 7: 候補のモデルのCG画像

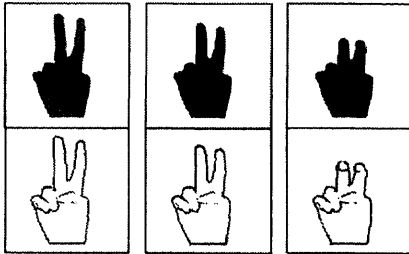


図 8: モデルのエッジ画像

3.2 特徴の照合

入力画像からの手指エッジの検索の補助とするために、入力画像と候補モデルとの照合を行って、最もマッチする候補モデルを選ぶ必要がある。この節では、入力画像とマッチする候補モデルを見つけるための、入力画

像と候補モデルとの照合方法について説明する。エッジと領域を表すテンプレートとを画像特徴として照合に用いる。

入力画像から肌色領域とそれに隣接するエッジを抽出する。同様に候補モデルのCG画像から領域を表すテンプレートとエッジを抽出する。入力のエッジ点とモデルのエッジ点とで勾配方向が近いものを対対応付けるために、勾配方向によってエッジ点をグループ分けする。 $\frac{1}{6}\pi$ ごとになるようにグループ分けしたモデルのエッジ点の座標を $\{\bar{x}_{ij}\} (j = 1, \dots, J_i), (i = 1, \dots, 12)$ とする。勾配方向がグループの境界付近となっているエッジが正しく対応付けされるように、勾配方向が $(i-1.5)\frac{1}{6}\pi$ 以上 $(i+0.5)\frac{1}{6}\pi$ 未満の入力のエッジ点をグループ i のエッジ点とする。その座標を $\{\bar{y}_{ij}\} (k = 1, \dots, K_i)$ とする。モデル、入力のそれぞれのグループ分けの仕方を図で示すと図9のようになる。

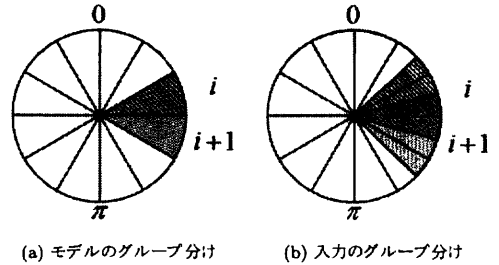


図 9: 勾配方向によるエッジ点のグループ分け

入力と候補のモデルとのエッジの相違度を各勾配方向ごとの chamfer matching[9]の総和と定義する。それを式で表すと以下のようになる。

$$d_e = \sum_i \frac{1}{J_i} \sum_j \min_k \|\bar{x}_{ij} - \bar{y}_{ik}\| \quad (1)$$

また、候補モデルのCG画像の領域で、入力画像の肌色領域からはみ出している部分の面積を U_s とする。

モデルから抽出したエッジを A_e 、領域を表すテンプレートと A_s とする。入力画像を I で表す。エッジ A_e を入力画像 I に重ね合わせて、エッジが入力画像にマッチしているときの d_e の分布を正規分布と仮定する。平均、分散を μ_{eT}, σ_{eT} とおき、正規分布を表す関数を

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (2)$$

とすると、確率密度は $f(d_e, \mu_{eT}, \sigma_{eT})$ となる。エッジが入力画像にマッチしていないときの平均、分散を μ_{eF}, σ_{eF} で表す。 d_e の分布を正規分布と仮定すると、確率密度は $f(d_e, \mu_{eF}, \sigma_{eF})$ である。同様に U_s の分布も正規分布と仮定する。画像にマッチしているとき、していないときの平均、分散をそれぞれ $\mu_{sT}, \sigma_{sT}, \mu_{sF}, \sigma_{sF}$ とすると、確率密度はそれぞれ $f(U_s, \mu_{sT}, \sigma_{sT}), f(U_s, \mu_{sF}, \sigma_{sF})$ である。

変数が θ の候補モデルから抽出したエッジを $A_{e\theta}$ 、領域を表すテンプレートを $A_{s\theta}$ とする。入力画像中にエッ

ジが存在している確率は以下ようになる。

$$p(A_{e\bar{\theta}}|I) = \frac{p(I|A_{e\bar{\theta}})p(A_{e\bar{\theta}})}{p(I|A_{e\bar{\theta}})p(A_{e\bar{\theta}}) + p(I|\bar{A}_{e\bar{\theta}})p(\bar{A}_{e\bar{\theta}})} \quad (3)$$

$p(A_{e\bar{\theta}}) \approx \beta p(\bar{A}_{e\bar{\theta}})$, ($\beta < 1$) と仮定して

$$\approx \frac{\beta p(I|A_{e\bar{\theta}})}{\beta p(I|A_{e\bar{\theta}}) + p(I|\bar{A}_{e\bar{\theta}})} \quad (4)$$

この候補モデルを入力画像に重ね合わせたときのエッジの相違度を $d_{e\bar{\theta}}$ として, $p(I|A_{e\bar{\theta}}) = f(d_{e\bar{\theta}}, \mu_{eT}, \sigma_{eT})$, $p(I|\bar{A}_{e\bar{\theta}}) = f(d_{e\bar{\theta}}, \mu_{eF}, \sigma_{eF})$ より

$$p(A_{e\bar{\theta}}|I) = \frac{\beta f(d_{e\bar{\theta}}, \mu_{eT}, \sigma_{eT})}{\beta f(d_{e\bar{\theta}}, \mu_{eT}, \sigma_{eT}) + f(d_{e\bar{\theta}}, \mu_{eF}, \sigma_{eF})} \quad (5)$$

となる。領域を表すテンプレートについても同様に、領域が重なっていない部分の面積を $U_{s\bar{\theta}}$ とすると、入力画像中に領域を表すテンプレートが存在している確率は次式で表すことができる ($\gamma < 1$)。

$$p(A_{s\bar{\theta}}|I) = \frac{\gamma f(U_{s\bar{\theta}}, \mu_{sT}, \sigma_{sT})}{\gamma f(U_{s\bar{\theta}}, \mu_{sT}, \sigma_{sT}) + f(U_{s\bar{\theta}}, \mu_{sF}, \sigma_{sF})} \quad (6)$$

3.3 誤り予測尤度を用いた最良な候補の推定

照合度が最大の候補でも図3のように入力とは形状が異なることがある。図3(b),(c)の候補モデルの変数をそれぞれ $\bar{\theta}_b$, $\bar{\theta}_c$ とする。以下の議論は A_e を A_s に置き換えても成立する。 $A_{e\bar{\theta}_b}$ に対する $\bar{\theta}_b$ の尤度 ($p(A_{e\bar{\theta}_b}|\bar{\theta}_b)$) は当然大きい。 $A_{e\bar{\theta}_c}$ に対する $\bar{\theta}_c$ の尤度 ($p(A_{e\bar{\theta}_c}|\bar{\theta}_c)$) も大きいので、 $A_{e\bar{\theta}_b}$ の照合結果から $\bar{\theta}_b, \bar{\theta}_c$ の優劣を決めることはできず照合を誤ることがある。ここで、 $p(A_{e\bar{\theta}_b}|\bar{\theta}_b), p(A_{e\bar{\theta}_c}|\bar{\theta}_c)$ に着目する。 $p(A_{e\bar{\theta}_b}|\bar{\theta}_b)$ は小さく、 $p(A_{e\bar{\theta}_c}|\bar{\theta}_c)$ は大きい。そのことと $A_{e\bar{\theta}_c}$ が画像にマッチしていることから、 $\bar{\theta}_c$ を推定結果として選択することができる。したがって、あらかじめ特徴に対する候補モデルの尤度モデルを作成しておき、その尤度モデルと入力画像に対する特徴の照合結果を総合して入力の手指形状を推定する。この尤度モデルのことを本論文では「誤り予測尤度」と呼ぶ。

誤り予測尤度の計算方法について説明する。探索範囲内のパラメータ候補の集合を Θ とする。候補 $\bar{\theta}_k \in \Theta$ から、その三次元モデルのCG画像 $I_{\bar{\theta}_k}$ を作成する。候補 $\bar{\theta}_j \in \Theta$ の特徴を $I_{\bar{\theta}_k}$ に重ね合わせて $p(A_{e\bar{\theta}_j}|I_{\bar{\theta}_k})$, $p(A_{s\bar{\theta}_j}|I_{\bar{\theta}_k})$ を求める。 $p(A_{e\bar{\theta}_j}|I_{\bar{\theta}_k})$, $p(A_{s\bar{\theta}_j}|I_{\bar{\theta}_k})$ を用いて誤り予測尤度を以下で定義する。

$$p(A_{e\bar{\theta}_k}|\bar{\theta}_j) = p(A_{e\bar{\theta}_k}|I_{\bar{\theta}_j}) \quad (7)$$

$$p(A_{s\bar{\theta}_k}|\bar{\theta}_j) = p(A_{s\bar{\theta}_k}|I_{\bar{\theta}_j}) \quad (8)$$

探索範囲内の各 $j, k \in \Theta$ に対して $p(A_{e\bar{\theta}_k}|\bar{\theta}_j)$, $p(A_{s\bar{\theta}_k}|\bar{\theta}_j)$ を計算しておく。

入力画像中の手の変数が候補 $\bar{\theta}_j$ である確率は、誤り予測尤度と入力画像に対する見えの照合結果を用いて次

式のように表される(付録参照)。ここで α は定数である。

$$p(\bar{\theta}_j|I) = \alpha \prod_{k=1}^N \{p(A_{e\bar{\theta}_k}|\bar{\theta}_j)p(A_{e\bar{\theta}_k}|I) + p(\bar{A}_{e\bar{\theta}_k}|\bar{\theta}_j)p(\bar{A}_{e\bar{\theta}_k}|I)\} \\ \times \{p(A_{s\bar{\theta}_k}|\bar{\theta}_j)p(A_{s\bar{\theta}_k}|I) + p(\bar{A}_{s\bar{\theta}_k}|\bar{\theta}_j)p(\bar{A}_{s\bar{\theta}_k}|I)\} \quad (9)$$

$p(A_{e\bar{\theta}_k}|I), p(A_{s\bar{\theta}_k}|I)$ は式(5),(6)から計算し、 $p(\bar{A}_{e\bar{\theta}_k}|I) = 1 - p(A_{e\bar{\theta}_k}|I), p(\bar{A}_{s\bar{\theta}_k}|I) = 1 - p(A_{s\bar{\theta}_k}|I)$ としている。各 j に対して $p(\bar{\theta}_j|I)$ を計算し、 $n = \arg \max_j p(\bar{\theta}_j|I)$ を選び出す。

この節では、候補モデルと画像とを照合する方法について述べた。照合結果をもとに、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次の節で述べる。

4 エッジの移動ベクトルからの関節角度の推定

4.1 エッジの移動ベクトルの計算

関節角度を推定するためにエッジの移動ベクトルを用いている。エッジの移動ベクトルを求めるためには照合で得られた最良候補が手がかりとなる。エッジの移動ベクトルを求める方法を図10に示す。最初に、最良候補から計算したエッジと最良候補から計算した三角形パッチの各頂点(図5)とを、入力のエッジを二値化した画像の距離変換画像に投影する。その頂点の近傍で最良候補のエッジ点を見つける。距離変換画像から、見つかった最良候補のエッジ点の座標における、入力のエッジに向かう移動ベクトルがわかる。近傍領域内のエッジ点での移動ベクトルの平均を求めて、それを頂点の移動ベクトルとする。最良候補と直前のフレームで推定された手指形状の間の自由度の変数の関係から対応する頂点間の変位ベクトルが求まる。それらから入力とモデルとのエッジを合わせる方向に向かう、直前のフレームでの形状の各頂点の移動ベクトルを計算しておく。直前のフレームの形状からの i 番目の頂点の移動ベクトルを $(dx_i, dy_i) (i = 1, \dots, P (= 600))$ とする。

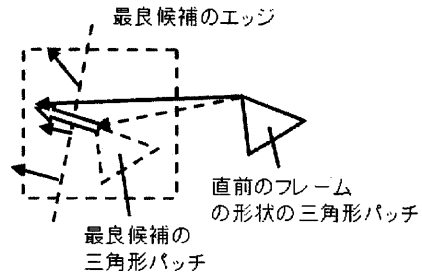


図10: エッジの移動ベクトルの計算

4.2 見えの学習

各候補モデルに対して、自由度の変数の値を少し変化させて三次元モデル（バリエーション）を作成する。図11に候補モデルとそのバリエーションの例を示す。候補モデルに対して、バリエーション1は手全体が傾いている。バリエーション2は指が内側に曲がっている。バリエーション3は指が内側に曲がり、さらにバリエーション1とは別の方向に手全体が傾いている。候補モデルでは中指と薬指の間の輪郭や薬指と小指の間の輪郭が見えているが、バリエーションではこれらの輪郭の一部もしくは全部が見えなくなっている。

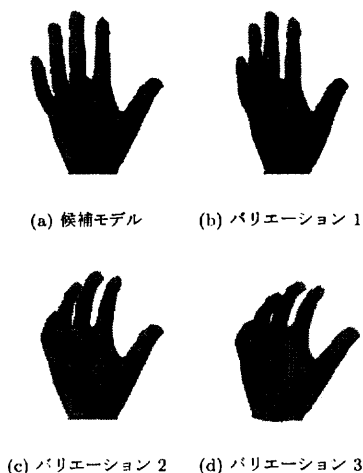


図 11: 候補モデルとバリエーション

三次元モデルの三角形パッチの頂点の座標を (x_i, y_i, z_i) とする。手指の三次元形状を $\vec{s} = (x_1, \dots, x_P, y_1, \dots, y_P, z_1, \dots, z_P)$ で表す。

候補モデルごとにバリエーションの \vec{s} の平均と共分散行列を求め、 \vec{s} の Q 次元の固有空間を作成する。 λ_{ij} , \vec{e}_{ij} をそれぞれ i 番目の候補モデルについて計算したバリエーションの \vec{s} の固有空間の j 番目に大きい固有値、 j 番目の固有ベクトルとし、 $\vec{e}_{ij} = (e_{ij1}, \dots, e_{ij3P})$ とする。

4.3 三次元形状の復元

各頂点の移動ベクトルから画像平面内の手指形状の変形を表すベクトル

$$\vec{ds}_{in} = (dx_1, \dots, dx_P, dy_1, \dots, dy_P, 0, \dots, 0) \quad (10)$$

$$= (ds_{in1}, \dots, ds_{in3P}) \quad (11)$$

を求める。学習した見えの変形を考慮して \vec{ds}_{in} から手指の三次元形状を復元する方法について説明する。

\vec{ds}_{in} は理想的な手指の変形のデータではなく、入力のエッジ点と最良候補のエッジ点との誤対応などの手指の

変形以外の影響を受けている。これらの影響を軽減する必要がある。Active Shape Models[6]では、あらかじめ輪郭の変形を学習した固有空間を作成する。入力画像から観測された輪郭の変形をその固有空間に投影して高次の成分の値を小さくすることによって、観測されたデータに混在している輪郭の変形以外の成分を軽減している。 \vec{ds}_{in} から手指の変形以外の影響を軽減するために同様の方法を用いる。

\vec{ds}_{in} を最良候補の固有空間に投影したベクトルを $\vec{b} = (b_0, \dots, b_Q)$ とする。 $b_j = \vec{ds}_{in,j} \cdot \vec{e}_{nj}$ である。最良候補に対して得られた固有値を対角成分に並べた行列を W とする。 $W_{ii} = \lambda_{ni}$, $W_{ij} = 0 (i \neq j)$ である。 $d_m = \sqrt{\sum_{i=1}^Q (b_i^2 / \lambda_{ni}^2)}$ とする。手指の変形以外の影響を軽減するために

$$\vec{b}' = \frac{\beta}{d_m} W \vec{b} \quad (12)$$

を計算する。ここで β は定数である。

\vec{b}' から三次元形状の変形を求めることができる。元々奥行きに対応する成分が0であるベクトルを投影して \vec{b}' が得られたので、元の空間に戻すときに $\frac{1}{\sum_{k=1}^{2P} e_{n,jk}^2}$ 倍する必要がある。

$$ds_i = \sum_{j=1}^Q \frac{1}{\sqrt{\sum_{k=1}^{2P} e_{n,jk}^2}} b'_j e_{n,ji} \quad (13)$$

を計算して、 $\vec{ds} = (ds_1, \dots, ds_{3P})$ が画像平面内での頂点の移動ベクトルから求められた三次元形状の変形である。

求められた \vec{ds} を初期形状に加えて三次元形状を変形させる。変形後の三次元形状は初期形状と比べて入力のエッジに合うように変形している。初期形状から移動ベクトルを求めたのと同様に、変形した三次元形状から移動ベクトルを求め、同様の手指形状の変形を行い、序々に画像のエッジに合うように三次元形状を変形させていく。これを $M (= 30)$ 回繰り返して最終的に得られた三次元形状が復元された手指形状である。

4.4 関節角度の決定

復元された三次元形状から関節角度を決定することができる。手指モデルの変数を変化させて作成した三次元形状と復元された三次元形状との間で各頂点の三次元座標のユークリッド距離の総和を計算する。復元した三次元形状との距離の総和が最小となる時の変数の関節角度が推定結果である。

画像平面内の平行移動は照合結果の値をそのまま採用する。全探索で最小となるものを選ぶと時間がかかる。そのためまず手指モデルの手の平の部分だけを用いて、姿勢の回転角度とサイズの変数を変化させて距離の総和が最小となるものを選び、姿勢の回転角度とサイズの変数を決定する。次に各指について、指の根元の部分を用いて曲げの関節角度と根元の伸縮の関節角度を決定する。各指の中間の関節の伸縮の角度を決定し、最後に指の先端の伸縮の角度を決定する。

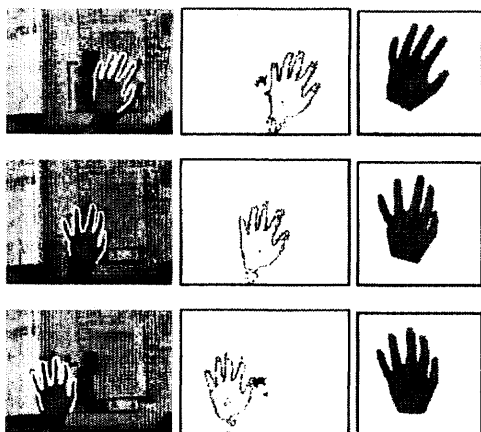


図 12: 実験結果

5 実験

一般的な背景の下で撮影した手指の時系列画像を入力として、手指形状の推定実験を行った結果を図 12 に示す。左は入力画像に推定結果の自由度の変数から作成した手指モデルのエッジを重ね合わせた画像、中央は肌色領域と隣接するエッジ、右は推定結果の自由度の変数から作成した手指モデルである。一番下の段の画像は図 1 と同じ画像で背景に肌色領域が存在するが、正しく手指形状を推定できている。

図 13 は別の時系列画像を用いて実験を行った結果である。この例は観測されるエッジが隠蔽によって大きく変化するが、正しく推定できている。

6 まとめ

本論文では、輪郭やオクルーディングエッジを用いて複雑背景下で手指形状を推定する方法について述べた。誤り予測尤度を考慮することにより、画像特徴の照合結果から入力画像中の手指形状とマッチする候補モデルを選びだすことができた。また、手指モデルを用いて自動で見えの変形を学習しておき、入力画像から観測された見えの変形からセルフオクルージョンが生じている手指の形状を推定した。

今後の課題としては、テクスチャの利用、個人差に影響されない推定、握むや握るといった動作の認識を考えている。

謝辞

本研究で用いている CG モデルを提供してくださった大阪大学大学院情報科学研究科の岸野 文郎教授、北村 喜文助教授に深く感謝致します。

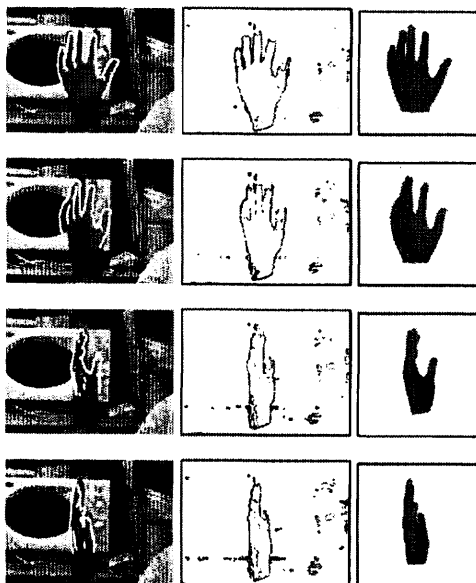


図 13: 実験結果

付録

式 (9) の導出について説明する。以下の式展開では $E_k \in \{A_{e\theta_k}, \bar{A}_{e\theta_k}\}$, $S_k \in \{A_{s\theta_k}, \bar{A}_{s\theta_k}\}$ とする。

$$\begin{aligned}
 p(\theta_j|I) &= \sum_{E_1, S_1, \dots, E_N, S_N} \{p(\theta_j|E_1, S_1, \dots, E_N, S_N, I) \\
 &\quad \times p(E_1, S_1, \dots, E_N, S_N|I)\}
 \end{aligned}$$

ここで

$$p(\theta_j|E_1, S_1, \dots, E_N, S_N, I) = p(\theta_j|E_1, S_1, \dots, E_N, S_N)$$

と仮定して、

$$\begin{aligned}
 &= \sum_{E_1, S_1, \dots, E_N, S_N} \{p(\theta_j|E_1, S_1, \dots, E_N, S_N) \\
 &\quad \times p(E_1, S_1, \dots, E_N, S_N|I)\}
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{E_1, S_1, \dots, E_N, S_N} \left\{ \frac{p(E_1, S_1, \dots, E_N, S_N|\theta_j)p(\theta_j)}{p(E_1, S_1, \dots, E_N, S_N)} \right. \\
 &\quad \left. \times p(E_1, S_1, \dots, E_N, S_N|I) \right\}
 \end{aligned}$$

この式に対して、 $p(\theta_j)$ は j に依らず一定として

$$\frac{p(\theta_j)}{p(E_1, S_1, \dots, E_N, S_N)} = \alpha (\text{定数}),$$

また

$$p(E_1, S_1, \dots, E_N, S_N | \theta_j) \\ = p(E_1 | \theta_j) p(S_1 | \theta_j) \dots p(E_N | \theta_j) p(S_N | \theta_j),$$

$$p(E_1, S_1, \dots, E_N, S_N | I) \\ = p(E_1 | I) p(S_1 | I) \dots p(E_N | I) p(S_N | I)$$

と仮定して,

$$= \alpha \sum_{E_1, S_1, \dots, E_N, S_N} \{ p(E_1 | \theta_j) p(S_1 | \theta_j) \dots \\ \times p(E_N | \theta_j) p(S_N | \theta_j) \\ \times p(E_1 | I) p(S_1 | I) \dots \\ \times p(S_N | I) p(S_N | I) \} \\ = \alpha \sum_{E_1} \{ p(E_1 | \theta_j) p(E_1 | I) \} \times \sum_{S_1} \{ p(S_1 | \theta_j) p(S_1 | I) \} \dots \\ \times \sum_{E_N} \{ p(E_N | \theta_j) p(E_N | I) \} \times \sum_{S_N} \{ p(S_N | \theta_j) p(S_N | I) \} \\ = \alpha \prod_{k=1}^N \{ p(A_{e\theta_k} | \theta_j) p(A_{e\theta_k} | I) + p(\bar{A}_{e\theta_k} | \theta_j) p(\bar{A}_{e\theta_k} | I) \} \\ \times \{ p(A_{s\theta_k} | \theta_j) p(A_{s\theta_k} | I) + p(\bar{A}_{s\theta_k} | \theta_j) p(\bar{A}_{s\theta_k} | I) \}$$

参考文献

- [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] 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.
- [4] 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.
- [5] 今井章博, 島田伸敏, 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.
- [6] 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.
- [7] 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.
- [8] Eng-Jon Ong and Richard Bowden. "A Boosted Classifier Tree for Hand Shape Detection". In *Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition*, pp.889-894, 2004.
- [9] 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.
- [10] M. Isard and A. Blake. "Visual tracking by stochastic propagation of conditional density". In *Proc. European Conf. Computer Vision*, pp.343-356, 1996.
- [11] 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.
- [12] 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.
- [13] B. Stenger, A. Thayananthan, P.H.S. Torr, and R. Cipolla. "Filtering Using a Tree-Based Estimator". In *9th Int. Conf. on Computer Vision*, Vol.2, pp. 1063-1070, 2003.
- [14] 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.