

## 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

浜田康志 島田伸敬 白井良明

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:{hamada,shimada,shirai}@i.ci.ritsume.ac.jp

一般に手指ジェスチャは衣服や顔などを含む複雑背景下で形状が提示される。我々は学習用画像列中の典型的な手指形状と可能な形状遷移から遷移ネットワークを作成し、形状遷移を追跡することで手指形状を推定する手法を提案した。本稿ではこの遷移ネットワークを拡張し、ジェスチャを認識する手法を提案する。遷移ネットワークの生成では全てのジェスチャの学習用画像列を用いて学習することで一つの大規模なネットワークが生成され、各ジェスチャは遷移ネットワークの部分経路として学習される。そこで入力画像列の各フレームで推定される手指形状の系列から遷移ネットワーク上の経路を求め、ジェスチャの部分経路を通じたことを検出し、ジェスチャを認識する。

## Hand Gesture Recognition under Complex Backgrounds Based on Transition Network

Yasushi Hamada, Nobutaka Shimada, and Yoshiaki Shirai

Dept. of Human and Computer Intelligence, Ritsumeikan University

1-1-1 Nojihigashi, Kusatsu, 525-8577, Japan

E-mail:{hamada,shimada,shirai}@i.ci.ritsume.ac.jp

In hand gesture recognition, hand shapes are generally shown under complex backgrounds including cloths and a face. We proposed a method to estimate hand shapes based on a transition network. The transition network is generated from training sequences in a learning phase, and a hand shape is tracked by using the network in an estimation phase. In this paper, we propose a method to recognize hand gestures from shape estimation results by using the enhanced transition network. Each gesture is registered as a partial path on the transition network by gesture learning. Also, Shape estimation results construct paths on the network. Since a gesture is recognized by detecting the gesture path which is passed by the estimation results.

## 1 まえがき

近年、新たなインターフェイスの一つとしてジェスチャ認識の研究が行われており、様々な応用が期待されている。ジェスチャ認識においては、ジェスチャ提示者にわずらわしさを感じさせない、画像を用いた非接触手法が適している。本論文では、高速な手指の形状変化と移動を伴い、画像上で手指と似た色の背景と重なるジェスチャを認識対象とする。

ジェスチャ認識の処理は特徴抽出段階と認識段階の二つの段階に分けられる。まず特徴抽出段階にお

いて、手指の形や面積、位置、動きなどの認識に有効な特徴を各フレームの画像から抽出する。次に認識段階において、入力画像系列とジェスチャの特徴系列を照合し、ジェスチャを認識する。

特徴抽出段階では、認識に有効な手指形状の特徴を抽出する。本論文で対象とするジェスチャは手指形状が複雑背景下で提示されるため、特徴抽出が容易でない。このようなジェスチャにおける特徴抽出では、時系列画像の入力に対する形状変化の追跡による予測と、予測される輪郭形状と入力画像エッジ

との照合による形状推定が有効である。しかし、高速な形状変化と運動を伴うジェスチャでは手指形状の変化量が大きく、手指画像のぼけがしばしば生じるため追跡が困難である。また、手指が肌色の物体と重なると境界のエッジが弱くなるため、推定をしばしば誤る。これらの問題に対し、著者らは遷移ネットワークに基づく形状推定手法を提案した [2]。遷移ネットワークは学習用サンプル画像列から典型形状と可能な遷移を学習し、また手指の速度が速い瞬間では位置と速度のみを特徴として持つモデルを登録することで、手指の素早い移動と変形の追跡を可能とした。また、真の手指輪郭と背景のエッジ点存在確率を考慮することで、肌色物体と重なる手指形状を正しく推定した。

ジェスチャ認識段階では、入力画像列とジェスチャの特徴系列を照合し、最適ジェスチャを選択することで認識を行う。入力画像系列中のジェスチャ区間の開始終了位置が与えられる場合、各ジェスチャの特徴系列モデルと入力画像系列の照合は容易である。しかし、手話文章のように入力系列中のジェスチャ区間が与えられない場合、照合は容易ではない。この問題に対し、スポッティング認識による手法が提案されている [3]。スポッティング認識は入力系列の各時刻をジェスチャの終了時点と仮定して照合を行い、最適対応を求める手法である。しかし、この手法ではすべての時刻でジェスチャとの照合を行うため、計算量が大きい。谷端ら [4] は、動き情報に基づいてジェスチャの区切り候補を検出し、区切りの組み合わせによって分割される区間を照合する手法を提案している。しかし、この手法においても組み合わせの数が多く、計算コストが大きい。連続音声認識では単語 HMM を連結した文章 HMM を入力系列全体と照合する手法が提案されている。しかし、手指ジェスチャではジェスチャ間に遷移区間が存在するため、遷移区間もモデル化しなければならない [5]。

本論文では、遷移ネットワークを拡張し、遷移ネットワークに基づく形状推定結果からジェスチャの認識を行う。

複雑背景下で提示される手指ジェスチャに対し、遷移ネットワークに基づく形状推定を行う。従来法 [2] では、各フレームにおいて最適モデルを選択するた

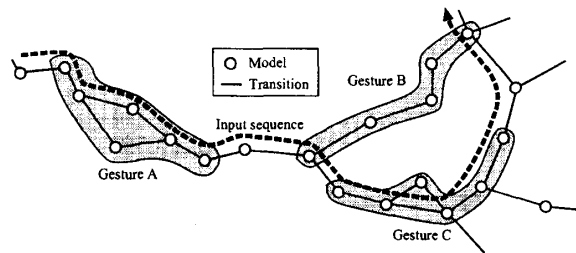


図 1: 遷移ネットワークに基づくジェスチャ認識

めにエッジ点存在確率に基づいて真の手指輪郭と重なる部分の長さを推定している。しかしジェスチャ認識では評価値の時系列での統合が必要である。そこで本論文では、新たにエッジ点存在確率に基づく尤度を評価基準として定義し、時系列での評価値を求める。

ジェスチャ区間切り出しの問題に対し、遷移ネットワークを拡張することで、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。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一つの大規模なネットワークが構築される。入力画像の形状推定結果の系列はこのネットワーク上で経路を構成し、また各ジェスチャは遷移ネットワーク上の部分経路に対応する。そこで、形状推定結果の経路がジェスチャの部分経路を通過したことを検出することで、ジェスチャを認識する (図 1)。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常に形状変化の追跡が為され、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。

2 章ではモデルとジェスチャの学習について、3 章では形状推定の手順と最適モデル選択のための評価基準について述べる。4 章でジェスチャ認識手法について述べ、5 章でジェスチャ認識実験の結果を示す。

## 2 モデルの学習

### 2.1 手指状態の表現

本論文では右手のみを扱う。学習段階は背景や衣服に制約を与えることが可能なので、学習用時系列画像から手指の輪郭や顔領域を自動で抽出する。も

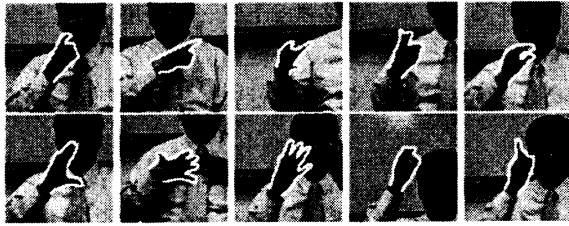


図 2: 学習する形状の例

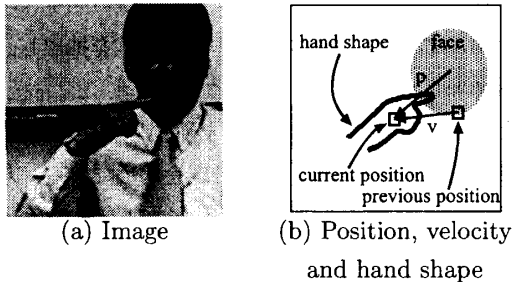


図 3: 形状モデルの要素

し必要ならば手で修正を加える。図 2 に学習画像に含まれる手指形状の例を示す。

各学習画像から顔領域と手指の輪郭を抽出し、手指の位置、速度、形状を求める(図 3)。手指位置は顔領域の中心位置からの相対位置  $p$  とする。手指の速度は学習用時系列画像における前フレームから現フレームまでの手指領域の重心の移動ベクトル  $v$  とする。手指形状は手指領域の重心を原点とする座標系での手指輪郭各点の二次元座標列  $x = (x_1, y_1, \dots, x_l, y_l)$  ( $l$  は輪郭長さ) とする。

手話のような手指の動きと形状を組み合わせたジェスチャでは、手指形状が意味を持つ瞬間と、手指の動きが意味を持つ瞬間からなる。動きが意味を持つ瞬間では、手指の動きが速く、図 4 のように手指画像にぼけが生じ、その輪郭が不明瞭となるため、モデル輪郭を画像と照合することができない。そこで、

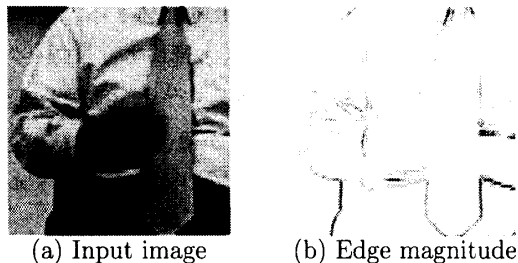


図 4: 手指画像のぼけ

手指の移動速度  $v$  が閾値以下の画像からは三つの特徴全てを持つモデル、 $v$  が閾値以上の画像からは位置と速度だけを持つモデルを登録する。

## 2.2 形状遷移ネットワークの生成

学習用時系列画像に現れる代表的な手指形状と、それらの間の可能な遷移を学習し、遷移ネットワークを作成する [2]。遷移ネットワークは学習画像中の典型的な形状をモデルとして登録する。そして、各学習画像とモデルの対応を取り、連続する二枚の学習画像に対応の取られたモデル間にリンクを生成する。

典型的な形状のモデルを登録するために、学習段階では学習用画像から抽出される特徴と似たモデルが既に存在すればそのモデルと対応を取り、存在しなければ新たにモデルとして追加する。学習時は輪郭形状が与えられるので、輪郭形状の相違度を求め、相違度が小さいほど形状が類似とする。また、位置と速度の相違度は、位置ベクトルと速度ベクトルそれぞれの差の大きさをを用いる。速度が遅い瞬間の画像では、登録されている形状を持つモデルと形状、位置、速度の相違度を求め、全ての相違度がしきい値以下となるモデルが存在すれば対応を取り、存在しなければ新たなモデルとして登録する。速度が速い瞬間の画像では、登録されている形状を持たないモデルと位置、速度の相違度を求め、全てがしきい値以下となるモデルが存在すれば対応を取り、存在しなければ新たなモデルとして登録する。

モデルの登録とリンクの生成を全ての学習用時系列画像に対して行うことにより、二種類のモデルから構成される遷移ネットワークを生成する。

## 2.3 ジェスチャの学習

ジェスチャを認識するためには、入力画像列の形状推定結果から各時刻でどのジェスチャが提示されているかを検出しなければならない。また、ジェスチャが開始、途中、終了状態を正しく経過したことを追跡しなければならない。そこで各形状モデルが属するジェスチャの種類と各ジェスチャ系列中で生じるモデル間遷移を遷移ネットワークに登録する。

遷移ネットワークの生成により、各ジェスチャのサンプル画像列の各フレームは形状モデルと対応が取られる。そこで、サンプルのジェスチャの種類を対応が取られたモデルに登録する。また、ジェスチャ系列中の可能な遷移を登録する。ジェスチャは学習用ジェスチャサンプルと入力系列の比較により認識されるが、ジェスチャの速度は学習時と認識時とで変化する。そこで、学習時の二倍までの形状変化速度に対応するために、ジェスチャ系列中で二つ以内のリンクで遷移可能なモデルからの遷移をジェスチャ内遷移として登録する。学習時より認識時の速度が遅いときは同じモデルが繰り返されることで追跡される。ある一つの形状が特定の意味を持つとき、その形状は一定時間、提示される。このような形状は、そのジェスチャを認識する場合に必ず提示される必要がある。そこで、学習時に繰り返し提示された形状については、その形状モデルを飛び越える遷移を除外する。

図 5 に、F,C,G と遷移するジェスチャ a と A,B,B,C,D,E と遷移するジェスチャ b を遷移ネットワークに登録する例を示す。例ではジェスチャ a の二番目とジェスチャ b の三番目は同じモデル C に対応するので、形状モデル C に a と b の二つのジェスチャの種類が登録されている。また、図中の矢印はジェスチャ b のジェスチャ内遷移を示す。ジェスチャ b のサンプルではモデル B が複数回繰り返されているので、モデル B を飛び越える遷移が除外される(図中の破線矢印)。

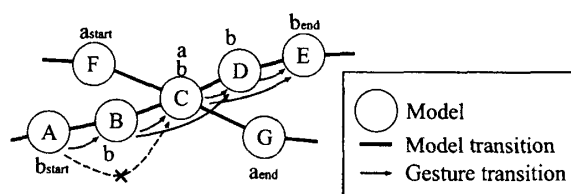


図 5: ジェスチャの学習

### 3 手指形状推定

モデル輪郭を入力画像のエッジと照合し、最適モデルを入力画像中の手指形状と推定する。

最適モデルを効率的に探索するために、まず色と手指の動きを用いて手指が存在する画像中の領域を

限定する。また、遷移ネットワークを用いて、マッチング候補を決定する。各候補について、エッジ点存在確率に基づいて照合を行い、最適モデルを決定する。

#### 3.1 色情報と動きの連続性による位置の限定

手指は肌色なので、画像中の肌色領域を抽出し、手指候補領域とする。

あらかじめ、手動で抽出した肌と服の色サンプルから RGB 空間での平均と共分散を求めておく。人物領域内の各画素について、肌色および衣服の色とのマハラノビス距離をそれぞれ求め、衣服色より肌色に近い画素を肌色領域として抽出する。

時系列画像の連続するフレーム間では手指の動きの連続性が仮定できる。そこで、手指の動きの連続性による手指位置の予測を用いて手指位置の探索範囲を限定する。ここでは等速直線運動の仮定により直前の二つのフレームから現在的手指位置を予測する。予測される手指位置は加速度による誤差を含むので、学習用時系列画像中の加速度の最大値を半径とする円の領域内を予測される手指位置の領域とする。

肌色領域内で、かつ動きの連続性から予測される円領域内を手指位置候補領域とする。

#### 3.2 遷移ネットワークによるマッチング候補の決定

遷移ネットワークを用いてマッチング候補を決定する。学習時と推定時では形状変化速度に差が生じることを考慮し、前フレームで推定されたモデルとそのモデルから二つ以下のリンクで到達可能なモデルを選択する。ここでは形状変化を追跡するので、2.3 節のジェスチャ内遷移ではなく、遷移ネットワークのすべてのリンクからモデル選択を行う。例えば、図 6 では、前フレームでモデル A が適合したので、現フレームのマッチング候補はモデル A 自身とモデル B に限定される。

限定されたモデルのうち、モデルの位置から手指位置候補領域までの最短距離がしきい値以下のもの

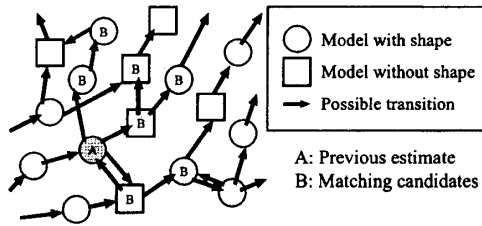


図 6: 遷移ネットワークに基づくマッチング候補

をマッチング候補として決定する。

実画像では、エッジ強度にばらつきがあり、また個人差による輪郭形状の違いも生じるため、正解モデルの評価値が最大とならないことがある。そこで、入力時系列画像の各フレームにおいて推定された上位  $N$  位以内のモデル  $A_1, \dots, A_N$  を追跡候補とし、ビーム探索を行う。

### 3.3 手指位置と向きへの決定

モデルは手指位置を特徴として持つが、入力画像上の位置と完全には一致しない。そこで各マッチング候補ごとに位置と向きを補正しながら照合する。

形を持つモデルでは、モデル輪郭と画像エッジを照合し、位置と向きを決定する。まず、肌色領域内で入力画像の明度からソーベルオペレータによりエッジ強度を求める。次に、各マッチング候補について位置と向きを変化させながら入力画像に当てはめ、モデル輪郭上のエッジ強度の平均  $c_a$  が最大となる位置と向きを探索結果とする。

位置の探索結果を用いてマッチング候補をさらに限定する。各マッチング候補について、前フレームからの手指位置の変化と、そのモデルの速度との差を求め、しきい値以上となるモデルをマッチング候補からあらかじめ除く。残ったマッチング候補について、エッジ点存在確率に基づくより詳細な照合を行う。

### 3.4 エッジ点存在確率に基づく最適モデル選択のための評価基準

前節で位置と向きが決定されたマッチング候補の中から、形状の類似度を求めて最適なモデルを選択する。

従来手法では、モデル輪郭上のエッジ強度の平均  $c_{ave}$  が評価基準としてしばしば用いられる [1]。しかし、図 9(a) の入力画像に対して  $c_{ave}$  が最大のモデルを推定結果とすると図 9(b) のモデルが選択される。これは、顔と重なる輪郭ではエッジ強度が弱く、エッジ強度の強い一部の輪郭を通る長さの短いモデルほど選ばれやすいためである。従来の評価では、図 9(b) のようなモデルを評価する場合、このモデルのみに注目しており、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。そこで、本論文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により形状推定する。

本論文では、エッジ強度がしきい値以上の点（エッジ点）に注目する。マッチング候補のモデル輪郭と重なるエッジ点の観測  $o$  が与えられたときにマッチング候補の中のあるモデル  $S_i$  が正しい確率  $P(S_i|o)$  をベイズの定理により変形すると以下の式になる。

$$P(S_i|o) = \alpha P(o|S_i)P(S_i) \quad (1)$$

事前確率  $P(S_i)$  は全てのマッチング候補に対して一様に与えられるとする。このときの尤度  $P(o|S_i)$  を考える。

真の手指輪郭上、および輪郭近傍の背景にはそれぞれ一様にエッジ点が分布すると仮定する。モデル  $S_i$  が正しいモデルであるとき、モデル  $S_i$  の輪郭上では真の手指輪郭上でのエッジ点存在確率  $P_T$  でエッジ点が観測される。一方、マッチング候補中のモデル  $S_i$  以外のモデル輪郭は背景と重なるので、背景のエッジ点存在確率  $P_F$  でエッジ点が観測される。そこで尤度  $P(o|S_i)$  を以下の式により定義する。

$$\begin{aligned} P(o|S_i) &= P(o_1, o_2, \dots, o_N|S_i) \\ &= \prod_{j \in Q_i} P(o_j|S_i) \prod_{j \in \bar{Q}_i \cap Q} P(o_j|S_i) \\ &= \prod_{j \in Q_i} P_T(o_j) \prod_{j \in \bar{Q}_i \cap Q} P_F(o_j) \quad (2) \end{aligned}$$

ここで、 $o_j$  は点  $j$  でエッジ点が観測される事象とされない事象の確率変数、 $Q_i$  はモデル  $S_i$  の輪郭点の集合、 $Q$  はマッチング候補全ての  $Q_i$  の和集合  $Q =$

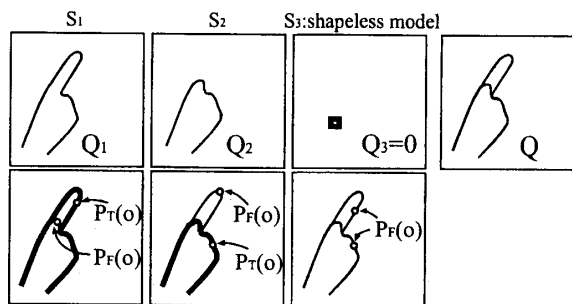


図 7: エッジ点存在確率に基づく評価基準

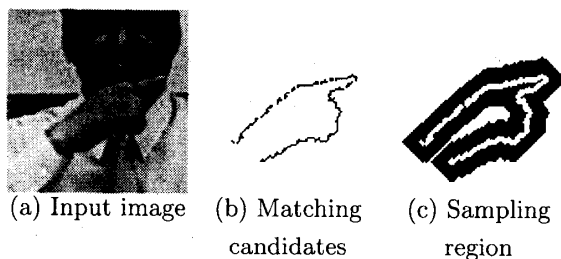


図 8: 輪郭の周辺領域

$Q_1 \cup \dots \cup Q_M$  である。もし  $S_i$  が形を持たないモデルならば  $Q_i = \emptyset$  である。図 7 に形状を持つモデル  $S_1, S_2$  と形状を持たないモデル  $S_3$  を評価するときのそれぞれの点でエッジ点が観測される確率を示す。もし正解が  $S_1$  で指が顔と重なる場合、指のエッジが弱くなるため従来法では  $S_2$  の評価値が高くなる。一方、本手法では  $S_2$  の評価値が大きくなるためには指の輪郭のエッジ点の観測が背景のエッジ点存在確率に従う必要がある。もし指輪郭のエッジが弱くても、エッジ点の観測が背景より真の手指輪郭のエッジ点存在確率に従っていれば、 $S_1$  を正しく推定できる。

式 (1) の対数を取り、エッジ点存在確率に基づく評価基準  $c_{prob}(S_i)$  と定義する。

$$c_{prob}(S_i) = \log P(S_i|o) \quad (3)$$

$c_{prob}(S_i)$  が大きいほどより入力画像に適合するモデルとする。

$c_{prob}(S_i)$  を求めるには確率  $P_T(o_j), P_F(o_j)$  を与える必要があるが、手指は画像中を移動し、その周囲の背景が変化するので、事前に与えることができない。そこで各フレームについて、マッチング候補  $m$  が正解モデルであると仮定したときの  $P_T^{(m)}(o_j), P_F^{(m)}(o_j)$  を入力画像から推定し、それらから  $P_T(o_j), P_F(o_j)$  を推定する。

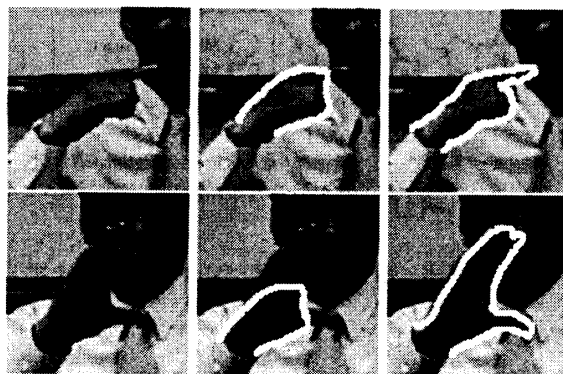


図 9: エッジ強度平均による推定結果と提案手法による推定結果

図 9: エッジ強度平均による推定結果と提案手法による推定結果

マッチング候補  $m$  の輪郭長さに対するモデル輪郭上で観測されるエッジ点数の比を  $P_T^{(m)}(o_j)$  の推定値とする。また、モデル輪郭を膨張させて得られる領域から輪郭を除いた輪郭周辺領域を求め、この領域の面積に対する領域内のエッジ点数の比を  $P_F^{(m)}(o_j)$  の推定値とする。図 8(a) の入力画像に対して選ばれたある候補モデルの輪郭と輪郭周辺領域を図 8(b),(c) に示す。

候補モデルの存在する位置は手指候補領域内に既に限定されている。位置が近ければ背景の明度分布はほぼ等しいので  $P_T(o_j), P_F(o_j)$  が等しいと仮定する。このとき、誤ったモデルでは輪郭が背景を通ることで  $P_T^{(m)}(o_j)$  の推定値が真の値よりも小さくなる。そこで  $P_T^{(m)}(o_j)$  が最大のモデル  $m_{max}$  を選び、 $P_T^{(m_{max})}(o_j), P_F^{(m_{max})}(o_j)$  を全てのモデルの  $P_T(o_j), P_F(o_j)$  の推定値とする。

各マッチング候補について、推定された  $P_T(o_j), P_F(o_j)$  を用いて評価基準  $c_{prob}(S_i)$  を求め、最適モデルを選択する。図 9(c) に図 9(a) の入力画像に対して評価基準  $c_{prob}$  を用いて推定されたモデルを示す。

#### 4 ジェスチャ認識

前節の形状推定により、入力画像系列の各フレームに対してビーム探索により追跡された形状モデル

とそれぞれの評価値が得られる。この形状モデル系列は遷移ネットワーク上の経路を構成する。そこで、学習段階で遷移ネットワークに登録されたジェスチャの種類とジェスチャ内遷移に基づいて、入力画像系列のネットワーク上の経路がジェスチャの部分経路を通過したことを検出し、ジェスチャを認識する。各フレームにおいて以下の処理を行う。

**開始検出** 推定結果モデルの中に開始状態に登録されたジェスチャが存在する場合、そのジェスチャが開始されたとし、追跡中ジェスチャに追加する。

**ジェスチャ追跡** 推定結果モデルの中に追跡中ジェスチャに属するモデルが存在する場合、ジェスチャ内遷移に基づいて、そのモデルに遷移可能なモデルが直前フレームで推定されていたかを調べ、もし存在すれば追跡を継続する。存在しなければ追跡中ジェスチャから除く

**終了検出** 追跡中ジェスチャがジェスチャ終了状態に到達した場合、そのジェスチャが検出されたとする。

形状推定はビーム探索により複数候補を追跡するため、複数のジェスチャが検出されることがある。そこで、複数のジェスチャが検出されたときは以下の評価基準により評価する。

$$c_{ges} = \frac{1}{N} \sum_{S^{(t)} \in R} c_{prob}(S^{(t)}) \quad (4)$$

ここで  $R$  は、形状推定結果系列のうち、検出されたジェスチャに対応する各時刻での形状モデル  $S^{(t)}$  の集合、 $N$  は  $R$  内の形状モデル数である。 $c_{ges}$  が最大のジェスチャを認識結果とする。

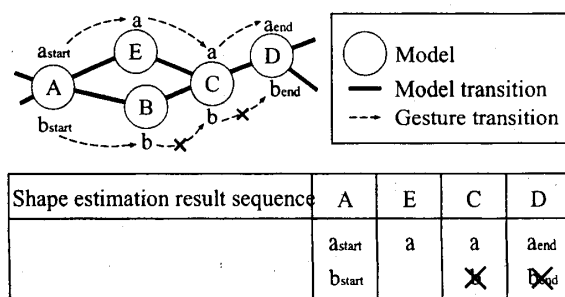


図 10: ジェスチャ検出

図 10 のようにジェスチャ a とジェスチャ b が登録されており、入力画像の推定結果系列が AECD であった場合、ジェスチャ a は最終状態まで追跡され、検出結果として得られる。一方、第三フレームのモデル C は a, b どちらのジェスチャにも属するが、ジェスチャ b のジェスチャ内遷移の遷移元が前フレームに存在しないため追跡が中断される。

## 5 ジェスチャ認識実験

手指の高速な移動と変形を伴うジェスチャとして、手話動画をを用いてジェスチャ認識実験を行った。手話単語 20 単語、各単語について毎秒 15 フレームで撮影した同一の手話者による時系列画像を 3 つずつ、計 2390 枚の画像を用意した。各単語は手指を膝に置いた姿勢（初期姿勢）から始まり、単語を提示した後、初期姿勢へと戻る。

学習段階では各単語あたり一つの時系列画像を用いて、形状モデルを登録し、遷移ネットワークを生成とジェスチャの学習を行った。全学習用画像から登録されたモデル数は形状を持つモデル数が 187、形状を持たないモデル数が 101 であった。また、この 288 モデルから成る遷移ネットワークを自動生成した。

### 5.1 動画画像の形状推定結果

用意した全ての時系列画像を用いて遷移ネットワークによる手指形状変化の追跡実験を行った。初期フレームでは入力画像の肌色領域内に存在するモデル全てをマッチング候補とし、以降のフレームでは遷移ネットワークでマッチング候補を限定した。途中でマッチング候補がなくなるか、あるいは全てのマッチング候補の評価値がしきい値以下になったときは、初期フレームと同様の処理により初期化を行った。また、追跡は上位 10 位以内のモデルを追跡候補としてビーム探索を行った。

図 11 に評価基準  $c_{prob}$  による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。画像中の白い輪郭形状は形状を持つモデルのうち、各フレームで評価

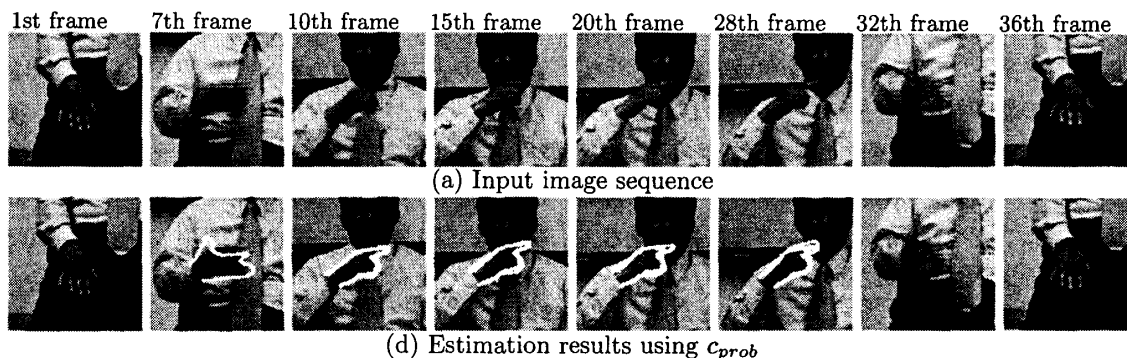


図 11: 動画像の形状推定結果

表 1: ジェスチャ認識結果

形状推定で 正解を追跡	正解のみ検出	44	
	複数検出	評価値で正解を 検出	13
		評価値で正解を 検出できず	0
形状推定で 追跡に失敗		3	

値が大きい順に上位四つの輪郭を描いたもの、黒い四角は形状を持たないモデルの位置である。

## 5.2 ジェスチャ認識結果

テスト画像列 60 個に対して認識実験を行った。認識結果を表 1 に示す。

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検出された場合は、ジェスチャ照合の評価基準  $c_{ges}$  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 個は形状推定段階において追跡に失敗した。この結果より、形状推定に成功したジェスチャは全て、認識に成功した。

## 6 むすび

複雑背景下で提示されるジェスチャの認識のための遷移ネットワークに基づく手指形状推定とジェスチャ認識を行った。複雑背景下での形状推定とジェ

スチャ系列の照合のために、エッジ点存在確率に基づく尤度を評価基準とした。また、手指形状の推定結果に対応する遷移ネットワーク上の経路がジェスチャの部分経路を通過したことの検出による、ジェスチャの事前切り出しを必要としないジェスチャ認識手法を提案した。

今後の課題として、サンプル数やジェスチャの種類を増やすこと、手話文章の認識を行うことが挙げられる。また、両手や顔の表情の情報を統合したジェスチャの認識も課題である。

## 参考文献

- [1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int. J. Computer Vision*, vol. 1, no. 4, pp. 321-331, 1988.
- [2] 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," *信学技報, PRMU2003-152*, pp. 7-12, Nov. 2003.
- [3] 高橋勝彦, 関進, 小島浩, 岡隆一, "ジェスチャ動画像のスポッティング認識," *信学論 (D-II)*, vol. J77-D-II, no. 8, pp. 1552-1561, Aug. 1994.
- [4] 谷端伸彦, 島田伸敬, 白井良明, "手話認識のための手指抽出と単語認識," *信学技報, WIT2001-22*, pp.37-42, 2001.
- [5] W. Gao, G. Fang, D. Zhao, and Y. Chen, "Transition Movement Models for Large Vocabulary Continuous Sign Language Recognition," *Proc. of Int. Conf. on Automatic Face and Gesture Recognition*, pp. 553-558, 2004.