手の動きと形を用いた動作分割による手話認識

あらまし　本論文では手の動きと形を用いた動作分割による手話認識について述べる．手話認識には HMM を用いる．よく似た特徴量をもつフレームが同じ状態に属するように分割して HMM の入力とする．状態分割のために手の重心の動きを分割する手法が提案されているが，手を振るような重心がほとんど動かず，手領域の形状のみが変わるような動きは分割できない．そこで本論文では，手領域の形状の変化にも注目して分割を行う．この形状の変化をとらえるために見かけの動きの大きさを反映した特徴を定義する．さらに，認識に用いる形状の特徴にこの定義した特徴を加えて認識を行う．結果，提案法の有効性が示された．
キーワード　手話認識，動画像処理，時系列解析, HMM，動き抽出，状態分割

Sign Language Recognition by Motion Segmentation using Hand Movement and Hand Shape

Abstract　We propose a method to automatically segment a motion for recognizing a word in sing language from a sequence of images. The motion segmentation uses not only a hand movement but also a shape in order to represent a hand shape feature. To segment a motion by the hand shape, we define the feature in which a length of a appearance of a movement was reflected. In addition, we recognize a word by using this feature. With experiments, we show the effectiveness of the proposed method.
Key words　sign recognition, video processing, time series analysis, HMM, motion extraction, state splitting

1. はじめに

画像系列からの手話認識処理は特徴抽出処理と抽出された特徴の識別処理から成る．特徴の識別手法としては手の位置や速度，形状等を特徴とし，各手話単語の特徴を学習した隠れマルコフモデルを用いて識別を行う手法が広く用いられてきた [1], [2]．HMM は複数の状態の遷移関係と各状態での特徴量の分布の組で表され，手話認識においては各々の状態が「手を上げる」「手を振る」などの 1 つの意味のある動きに対応している (図 1)．学習

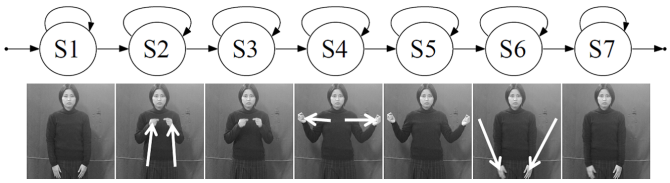


図 1　手話認識における HMM

時には各手話単語に対応したモデルを作成し，認識時は入力特徴量を出力する尤度が最も高いモデルに対応する単語を認識結果とする．

手話認識の問題の１つとして，単語によって動作の複雑さが大きく異なることが挙げられる．川東ら [3] や松尾ら [4] の手法では，各モデルの状態数を単語動作から自動的に推定している．そのため単純な動作に対しては状態数の少ないモデルを，複雑な動作に対しては状態数の多いモデルを用いて学習を行える．この推定は画像系列を手の重心の動きを用いて複数の状態に分割することで必要な状態数を推定している．重心は変化せず手の形状のみが変化するような手話動作に対しては，必要な状態数より少ない状態数のモデルを用いて学習を行うため，手の形状の変化を反映した学習が難しい．

本論文では，手の形状の変化にも注目して状態を分割することで，単語動作の複雑さをより反映した学習を行う．形状の変化をとらえるための特徴を定義し，この特徴と手の重心の動きを用いて状態を分割し，認識に用いる特徴にこの特徴を加えて認識を行う．手話は手の動きや身体全体の動きで表現するため，動きや形の目立ちやすさが重要となる．そこで形状の変化をとらえるための特徴は，見かけの大きさを反映させる．また，手話には手の速度が大きいときは手の動きが重要であり，速度が小さいときは手の形状が重要であるという傾向がある．そのため状態の分割では，この傾向を考慮して分割を行う．

2. 手領域の抽出

本論文で用いる手領域の抽出は [3] の方法で抽出する．背景差分によって人物領域の抽出を行う．抽出した人物領域から肌色を用いて両手と顔の肌色領域を抽出する．しかし画像中の肌の色は撮影状態や個人によって異なる

ため，固定的な閾値で様々な話者の肌の色を判定することは難しい．そこで画像系列ごとに撮影初期のフレームから話者の肌の色を取得することによって，様々な話者に対応する．

3. 認識に用いる特徴

前節の画像処理によって，両手と顔の各領域の位置や形状を得ることができる．これを HMM によって学習・認識するためには，領域の位置や形状を数値として表す必要がある．

3.1 手の位置・速度の特徴

本論文では以下の川東ら [5] の特徴を用いる (図 2)．

（1） 顔からの距離の対数
（2） 顔からの距離の対数の変化
（3） 顔からの方向
（4） 顔からの方向の変化
（5） 左手からの右手の相対距離

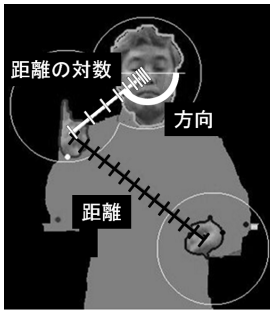


図 2 手の位置・速度の特徴

一般に手話においては，手の細かい動きに重要な意味がある場合，手は顔に近い場所で動かされる傾向がある．そこで，手領域の重心位置を表す特徴量として顔領域の重心からの相対座標を基準に対数的に変換したものを用いる．これにより同じ大きさの動きでも顔に近いほど，大きな動きとしてとらえることができる．

3.2 手の形の特徴

用いる特徴を以下に示す (図 3)．

（1） 面積
（2） 円形度 c
（3） 回転量 (慣性主軸の回転角速度)$\Delta\theta_p$
（4） 見かけの回転量 L

手の形状の特徴には面積・円形度・慣性主軸方向・突起数・周囲長などがある．本論文では手領域を画像系列から抽出しているため，手の向きや形状の少しの変化で指同士の領域がくっついたり離れたりする．そのため，これに大きな影響を受ける突起数と周囲長は用いない．

川東ら [5] は慣性主軸の方向ベクトルを用いていたが，主軸の方向は $\pm\pi(rad)$ の自由度があるため，学習・認識が難しい (図 4)．そこで本論文では方向ベクトルの代わりに回転量を用いる．

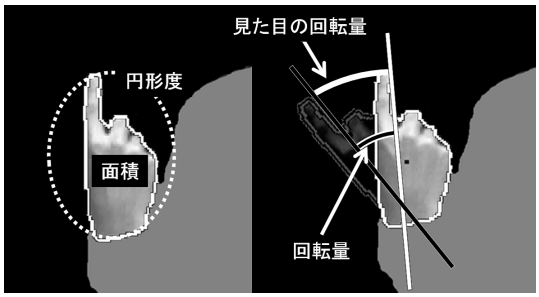


図 3 手の形の特徴

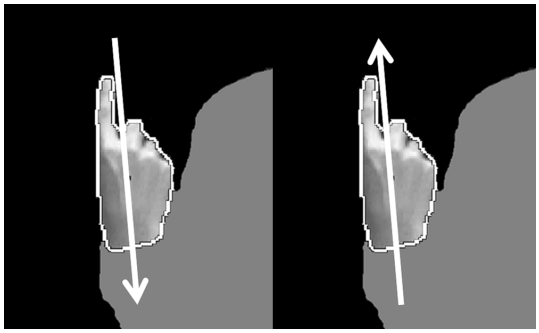


図 4 慣性主軸の方向ベクトルの自由度

手話は手の動きや身体全体の動きで表現するため，動きや形の目立ちやすさが重要となる．川東ら [5] も手の速度に関連した特徴を，同じ大きさの動きでも顔に近いほど大きな動きとしてとらえている．そこで回転量も同様に目立ちやすさを考慮した，見かけの動きの大きさを反映した回転量を用いる．

3.2.1 見かけの動きの大きさを反映した回転量

手領域の回転を考えたとき，同じ角度だけ回転しても，真円に近い形より細長い形のほうが大きく回転しているように見える (図 5)．そこで，見かけの動きの大きさを

(a) 真円に近い形　　(b) 細長い形

図 5 形と見かけの動きの大きさ (20 °回転)

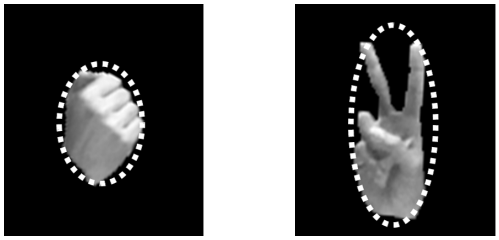
反映できるように，回転により描かれる弧の長さにあたる回転量 L_0 を以下のように定義する．

$$L_0 = \frac{\Delta\theta_p}{\sqrt{c}} \tag{1}$$

θ_p は慣性主軸方向を表し，ここでは円形度 c は

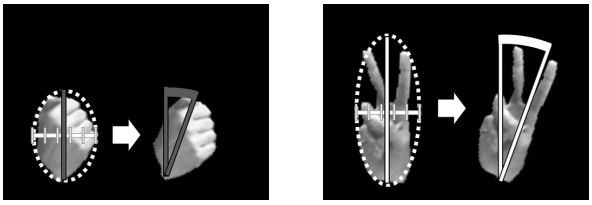
$$c = \frac{\min_\theta I(\theta)}{\max_\theta I(\theta)} \tag{2}$$

とし，$I(\theta)$ は θ 方向の軸まわりの慣性モーメントを表す (図 6)．$1/\sqrt{c}$ は長軸 / 短軸に相当し，長軸 / 短軸を用いることで L_0 は短軸の長さを揃えたときの長軸の描く弧の長さに相当する (図 7)．



(a) $c = 0.49$ $(1/\sqrt{c} = 1.42)$　(b) $c = 0.18$ $(1/\sqrt{c} = 2.36)$

図 6　円形度の例



(a) 真円に近い形 $(L_0 = 0.24)$　(b) 細長い形 $(L_0 = 0.40)$

図 7　長軸の描く弧の長さ (20 °回転)

形状が真円に近いときは少しの形状の変化で主軸の取り方が変わり，回転していなくても主軸方向が大きく変化する (図 8)．定義した回転量 L_0 は弧の長さにあたる特徴量であるため，真円に近いときは回転していないにも関わらず回転量は大きくなることがある．そもそも真円の場合に主軸はない．そのため真円の場合は回転量を 0 とすべきである．よって，さきほどの回転量 L_0 を以下のように再定義する．

$$L = \left(\frac{1}{\sqrt{c}} - 1\right)\Delta\theta_p \tag{3}$$

これにより真円に近いほど回転量 L を 0 に近づけることができ，真円に近い場合の不安定さを抑えた見かけの大きさを反映した回転量 L を定義できた (図 9)．以降は，この L を見かけの回転量 L と呼ぶこととする．

4. 学習用サンプルの区間分割

学習用サンプルから「手を上げる」「手を振る」などの 1 つの意味のある動きに対応したフレームに分割する．松尾ら [4] の方法は，停止区間，移動区間，振動区間に分割する (図 10)．振動区間とは移動区間ではあるものの短時間内に向きが頻繁に変わる区間を表す．これに形

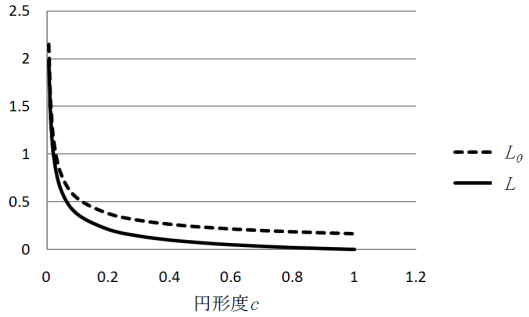


図 9　L_0 と $L(\Delta\theta_p = 0.17)$

状変化区間を加えることで単語動作の複雑さをより反映した学習を行う．手話には手の速度が大きいときは手の動きが重要であり，速度が小さいときは手の形状が重要であるという傾向がある [3]．そこで静止区間中に手の形状が変化したときに，形状変化区間に再分割する．また形状変化区間が移動区間と隣接している場合は，その形状変化区間は手の移動の始まりととらえることができるので，2 つの区間を統合して 1 つの移動区間とする．本論文では，分割に用いる手の形状の特徴に前節で定義した見かけの回転量 L を用いる．静止区間中で見かけの回転量 $|L|$ が十分大きいときに，回転区間として再分割する (図 11)．

区間分割の流れを以下に示す．

（1） 手領域の重心位置の移動速度によって各フレームを静止フレーム，あるいは移動フレームに類別する．

（2） 連続する静止フレームは静止区間としてまとめる．連続する移動フレームは，短時間の内に向きが頻繁に変わるものは振動区間としてまとめ，それ以外を移動区間としてまとめる．

（3） 静止区間内で手の形状が変化しているフレームを形状変化フレームに再類別する．

（4） 連続する形状変化フレームを形状変化区間としてまとめる．

（5） 形状変化区間が移動区間と隣接している場合は，統合して 1 つの移動区間とする．

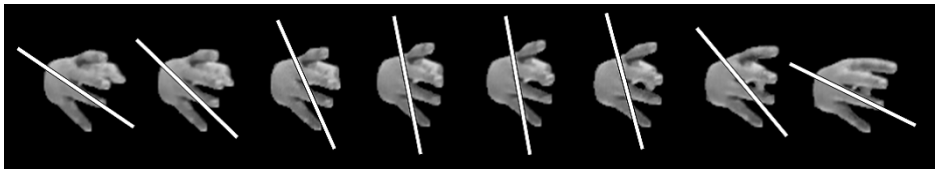


図 8　真円に近い場合の慣性主軸方向

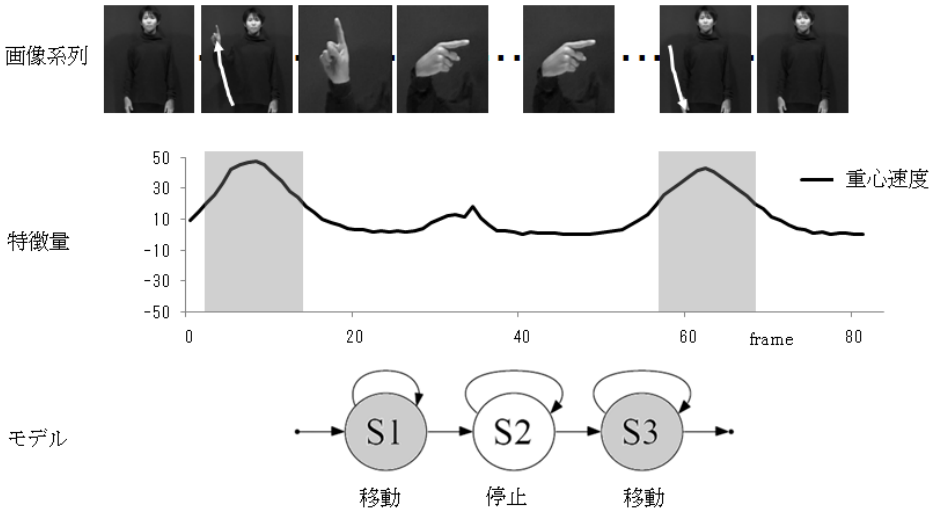


図 10　従来法の区間分割

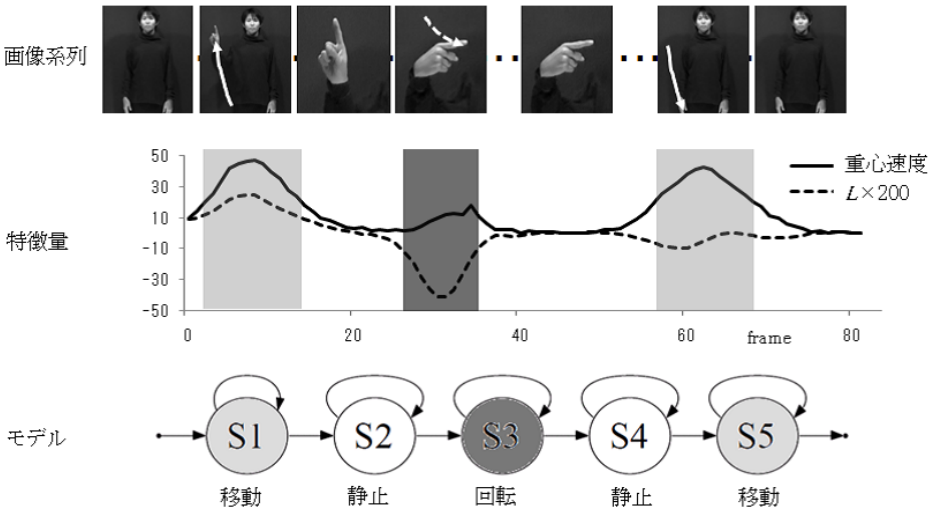


図 11　提案法の区間分割

5. 状態遷移構造の生成

本論文では松尾ら [4] の方法で生成する．HMM による手話認識における重要な問題として，同じ意味の手話単語動作であっても，そこから得られる特徴量系列が発話によって大きく異なるという点がある．そこで，単語動作の複雑さに応じた枝分かれを含む状態遷移構造を持つ HMM を用いる (図 12)．まず手話単語動作を撮影した画像系列を，１つの意味のある動きの区間列に分割し，各区間を状態の候補として扱う．複数の発話に渡って状態候補を収集し，そこから同じ動きと見なせるものを見つけて統合することで状態遷移構造を生成する．

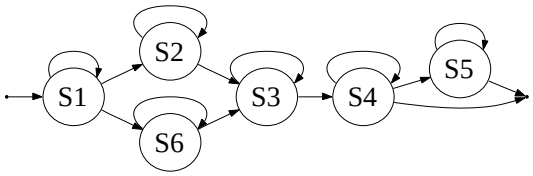


図 12　動作の複雑さに応じた HMM

6. 認識実験結果

この実験では，3 人の話者に各単語につき 3 回ずつ動作を行ってもらい．35 単語について単語動作を集めた．単語は手話技能検定 5・6・7 級から手の重心がほとんど動かない単語は 20 語，重心が動いている単語は 15 語を選んだ．各単語につき 1 つを発話の認識対象，残り 8 つうち 6 つを学習用サンプルとして認識対象のサンプルを変えながら，9 通りの認識実験を行った．また，比較のため川東らの方法でも同様の実験を行った．

表 1　手話単語動作の認識結果

	話者 1	話者 2	話者 3
従来法	86(81.9%)	98(93.3%)	76(72.4%)
提案法	98(93.3%)	104(99.0%)	91(86.7%)

7. おわりに

単語動作の複雑さをより反映した学習を行うための手の形状の変化を用いた状態分割を提案した．形状の変化をとらえるために定義した特徴は見かけの大きさを反映した特徴となっており，状態分割は手話の傾向を考慮した分割となっている．提案法を用いれば，各々の単語の持つ手の形状の変化を反映したモデルを自動的に生成することができる．

文　献

[1] K. Grobel and M. Assan: “Isolated sign language recognition using hidden markov models”, Systems, Man, and Cybernetics, 1997. ’Computational Cybernetics and Simulation’., 1997 IEEE International Conference on, **1**, pp. 162–167 vol.1 (12-15 Oct 1997).

[2] T. Starner, J. Weaver and A. Pentland: “Real-time american sign language recognition using desk and wearable computer based video”, IEEE Transactions on Pattern Analysis and Machine Intelligence, **20**, 12, pp. 1371–1375 (1998).

[3] 川東, 白井, 島田, 三浦：“手話の HMM 作成のための状態分割”, 信学技報, 第 105 巻 of WIT2005-21, pp. 55–60 (2005).

[4] 松尾, 白井：“手話認識のための動作の多様性に応じた HMM 構造生成”, 信学技報, 第 109 巻 of PRMU2009-179, pp. 161–166 (2010).

[5] K. Kawahigashi, Y. Shirai, J. Miura and N. Shimada: “Automatic synthesis of training data for sign language recognition using hmm”, Proc. 10th Int’l Conf. on Computers Helping People with Special Needs (ICCHP), pp. 623–626 (2006).