社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
PRMU2004-16,MI2004-16,WIT2004-16(2004-05)

HMMを用いた手話単語の認識

金山　和功† 　白井良明† 　島田伸敬†

† 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
E-mail: †kanayama@cv.mech.eng.osaka-u.ac.jp, ††{shirai,shimada}@mech.eng.osaka-u.ac.jp

あらまし　本論文では時系列画像から HMM を用いた手話単語の認識について述べる．画像から，肘を追跡し，隠蔽を考慮した顔や手の領域を決定し，手話の特徴をよく表す手指の特徴量を抽出する．抽出した特徴量を用いて HMM で手話単語を認識する．手話に応じた HMM の状態数の決定方，認識時の手の動きを用いた候補の限定の方法および実験結果を示す．

キーワード　隠れマルコフモデル，手話，画像処理

Recognition of Sign Language using HMM

Yasutaka KANAYAMA†, Yoshiaki SHIRAI†, and Nobutaka SHIMADA†

† Department of Computer-Controlled Mechanical Systems Graduate School of Engineering Osaka University
2-1 Yamadaoka,Suita,Osaka 565-0871,JAPAN
E-mail: †kanayama@cv.mech.eng.osaka-u.ac.jp, ††{shirai,shimada}@mech.eng.osaka-u.ac.jp

Abstract　This thesis states recognition of sign language using HMM. We track elbow,decide hand and face region,and extract features of finger. These features are used by HMM recognition. We state decision rule of number of state,and result of experiment.

Key words　HMM,Sign Language,Image Processing

1. はじめに

聾唖者のコミュニケーション手法として，手話がある．しかし，公共の場所では手話を使える人はほとんどおらず，手話を日本語，また逆に日本語を手話に翻訳するシステムの必要性は高いといえる．

手話の翻訳の流れは，手話のデータ取得，手話の認識，認識結果出力と分けられる．この中で，前二つが困難にしている部分である．前者について，手の形状データをデータグローブで得るものがあり，試験的に実用化されている [1]．しかしデータグローブを用いる方法は被験者に装着等の負担が生じる．それに比べ，本論文で用いるカメラ画像からデータを得る方法は，被験者はデータグローブの取り外しの煩わしさから解放される．

一方，画像から手の形状や動きのデータを得る研究は，ジェスチャや手話認識のために広く研究されている．手のシルエットと 3 次元 CG の手とをマッチングすることにより，手の形状を求める方法がある．[2] しかし，単純な背景下で行っているため，いろいろな状況に対応するのは難しい．手話認識の分野 [3], [4] では，肌色の範囲より手の領域を抽出しているものがあるが，肌色の範囲は被験者や周りの状況に影響されるので，あらかじめ決められた肌色の範囲では，手の領域を抽出するのは困難である．

また，画像ベースでの手の特徴を抽出する方法は，肌色に似た背景の一部により，エラーを含んでいるため，認識が困難である．エラーを含んだ特徴列をロバストに認識する方法として，Hidden Markov Model(以下 HMM) [5] がある．HMM は，音声認識，表情認識 [6]，ジェスチャー認識 [8]，行動の認識 [9] の分野でよく用いられている．多くの学習サンプルからモデルを構築でき，時間軸の伸縮が可能であるので，スピードの違う手話や，動きが完全に一致しないものでもロバストに認識可能であると考えられる．

本研究では，画像からの手話認識のために，カメラ画像からの手指の特徴量抽出と手話単語認識方法について述べる．

2. 特徴量抽出のための画像処理

2.1 人物領域の抽出

複雑背景下で撮影された手話画像から人物の領域のみを抽出するには，背景差分を用いる．人物の領域を正確に抽出するためには，人物の影の領域は抽出しないような背景差分を考えなくてはならない．ここでは単純に HSV 色空間で差分をとるの

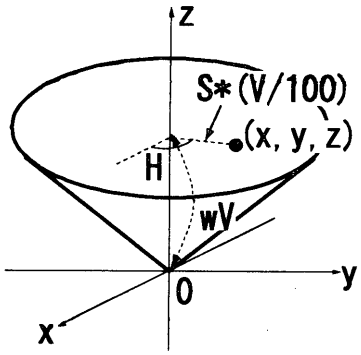
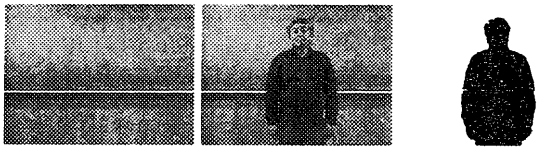


図 1　*HSV* 色空間の円錐 *x-y-z* への変換



(a) 背景画像　(b) 入力画像　(c) 差分画像

図 2　人物領域の抽出

ではなく，HSV 色空間を明度の変化をあまり考慮に入れないような色空間に変換してから差分を取る方法を用いている．変換式は，以下である（図 1 参照）

$$\begin{aligned} x &= S*(V/100)*cos(H) \\ y &= S*(V/100)*sin(H) \\ z &= w*V \\ 0 &\leqq H \leqq 2\pi, 0 \leqq S \leqq 100, 0 \leqq V \leqq 100 \\ 0 &\leqq w \leqq 1 \end{aligned} \quad (1)$$

である．ここで w はV に対する重みである．本論文では実験的に w を 0.5 としている．2 画素 (x_1, y_1, z_1) と (x_2, y_2, z_2) の差分 d_c は以下のように定義される．

$$d_c = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 \quad (2)$$

d_c が閾値以上（実験的に 35）であれば，人物領域として抽出する．明るさが小さいと色相や彩度の変化を重視しなくなるため，差分が出にくくなり影の影響を緩和することができる．

このように手話画像と背景画像との差分を取り抽出された領域を収縮膨張してノイズを少し減らした結果を図 2 に示す．

2.2　肌色抽出

人物領域から顔と手の領域を人物領域から肌色部分として抽出する．しかし，様々な被験者の肌色を判定することは難しい．そこで，撮影初期のフレームから被験者の肌色の情報を取得することによって，様々な被験者に対応する方法を提案する．

まず被験者に最初に両手を太腿に置いてもらう．このようにしてもらうことで手や顔の位置は想定できているので，この部分から色情報をサンプルすることにより，その被験者の肌色の情報を得ることができる．

肌色の分布は，HS 色空間において正規分布であると仮定して，90% の等確率楕円内に入っており，肌色として取りうる明るさを固定的に与える．その域にある色を肌色と決定する．前節で説明した人物領域内であり，かつ手の位置に等加速度運動を仮定してその予測位置近傍のみを探索範囲として肌色を抽出する．なお，本研究では顔と手の肌色は微妙な違いがあることを考慮して，それぞれ別の肌色として情報を保存し，そのどちらかで肌色と判定される色を肌色と判定する．肌色情報を初期フレームから取得した結果と，このようにして得た肌色領域の抽出結果を図??に示す．

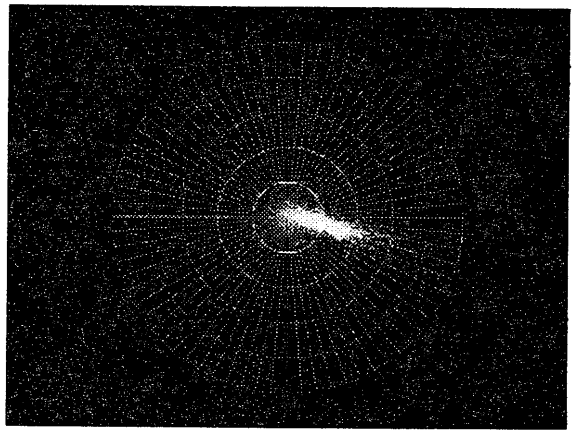


図 3　肌色部分のプロット（原点を中心とした半径が彩度，角度が色相）

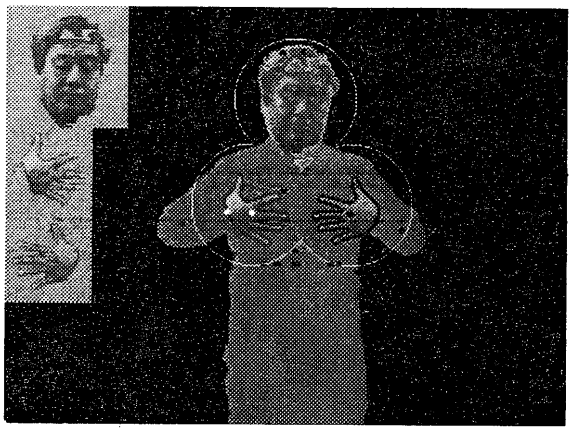
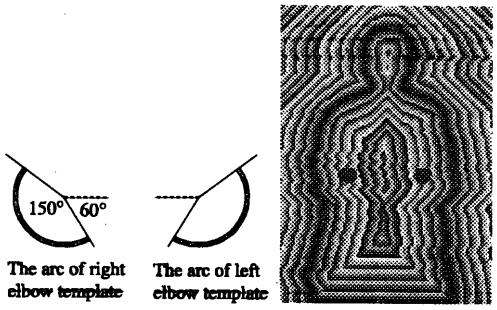


図 4　肌色領域の抽出結果（左部に抽出された顔，手を表示）

2.3　肘・手首の検出

手話の特徴量である手の突起数や，手の方向を計算する場合に手首の位置が重要である．手首は手領域で肘にもっとも近い点であるとし，肘を発見することで手首を発見する．

肘は円弧状の形状をしていると仮定し，円弧のテンプレート (図 5) と人物領域の輪郭線とをマッチングすることによって抽出する．円弧の大きさや向きは動きによって変化するので複数のテンプレートを用いて，最もマッチするものを選択する．ただ，輪郭線との単純なマッチングでは，精度良く正確な肘の場所を発見することが難しいので，マッチングには輪郭線画像を距離変換した画像を用いて，円弧テンプレート上の距離の総和が最も小さくなるような場所の円弧の中心を肘の位置とする．そのようにして発見した肘の場所から，手領域（領域の決定方



(a) 肘のテンプレート　(b) 輪郭線の距離変換画像（点が肘の位置）

図 5　肘の抽出結果

図 6　隠蔽がおこっている状態

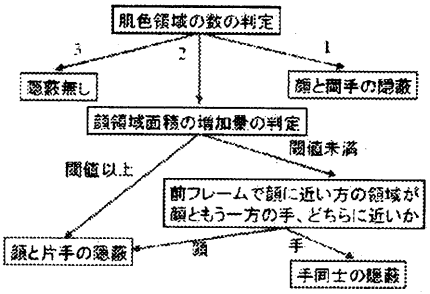


図 7　隠蔽の判定方法

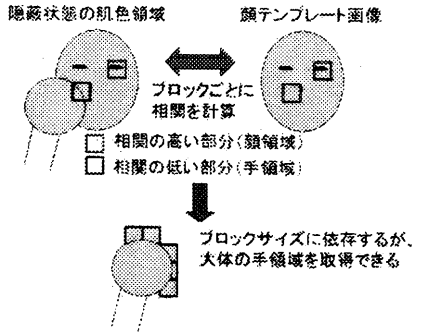


図 8　顔と手の隠蔽時の処理

については後述）の中で肘から最も近い点を手首として検出する．

肘のテンプレートおよび輪郭線の距離変換画像の例とマッチングの結果を図 5 に示す．距離変換画像にポイントされた点が肘の位置である．

2.4　テンプレートの保存

顔の前方で手話をしたり，手同士が重なったりするときには各領域を正確に抽出することは難しい．そこで，隠蔽が起っていない直前の状態の顔や手の形状をテンプレートとして保存しておく．隠蔽に関係のないと判断された領域の重心を中心に 80×80 の大きさ (実験用画像 640×480 に対して) の矩形内でその領域の画像を保存しておく．例えば隠蔽が全くおこっていないときには顔と両手すべて，右手と顔が重なっている時には左手のテンプレートのみが更新することができる．隠蔽が起ったことの判断等については次節で述べる．

2.5　領域のラベル付けと隠蔽時の判定とその処理

前節までの方法によって，肌色領域を抽出することができるが，それ以外の余計な領域が抽出されてしまうこともある．また，取得した領域が左手であるのか右手であるのか，また顔であるのかといった判定をしなくてはならない．さらに，手と手が重なったときや，顔と手が重なった時 (図 6) の判定が必要である．

2.5.1　各領域のラベル付けと隠蔽の判定

まず，手領域や顔領域は一定以上の大きさを持っているはずであるので，面積の小さい領域を除去する．

初期フレームで手の初期位置を指定しているので，最初は両手の位置および顔の位置はわかっている．このことから最も前フレームの位置から近い領域をその領域としてその領域とする．

人物領域内の肌色領域の数と，顔領域の位置はあまり変化しないという仮定のもとでその面積や，前フレームの各領域の位置情報を用いて図 7 のようにして隠蔽の判定を行う．

2.5.2　顔と片手の隠蔽

顔の前方で手話を行うことは多いので，顔と手の隠蔽はよく起こる．そこで，

- 肌色領域が 2 つ
- 顔領域のラベルとどちらかの手のラベル番号が一致
- 顔領域の面積が 10% 以上増加

か，

- 肌色領域が 2 つ
- 前フレームでの手の位置がもう一方の手領域よりも顔領域に近い

という条件を満たしたときに，顔と手の隠蔽が起ったと判定する．

このとき，顔と手の領域が一つとして抽出されるので，手領域と顔領域を分離しなければならない．これを前節で説明したテンプレートを用いて解決する．顔と手はそれぞれ直前の隠蔽が起っていないフレームのテンプレートを保持している．特に顔テクスチャはほとんど変化しないのでそのテンプレートを現在の顔と仮定する．このことから顔のテンプレートを肌色抽出した画像とマッチングすることによって，現在の顔の位置とすることができる．

顔のテンプレートとブロックごとに明度相関算 cor の高いブロック (実験的に 0.05 以上) は顔領域であるとするようにすることで，ブロックサイズの大きさ程度の精度（本研究ではブロックサイズは 5×5）で手領域を抽出できる．相関 cor はある

図 9　隠蔽時の相関値と分離結果

ブロックについて，顔テンプレート画像の明るさ $ftemp(i,j)$，分離したい画像の明るさ $org(i,j)$ とすると以下の式によって計算している．

$$cor = \sum_{i=0}^{5}\sum_{j=0}^{5}(ftemp(i,j) - org(i,j))^2 \qquad (3)$$

このようにして得た顔領域を除去した肌色画像内と保存していた手のテンプレートとマッチングを行う．マッチングはまず顔領域を除いた肌色領域の肘から最も近い点を手首候補点としてそこを中心として，テンプレートを回転させてマッチングを行い，最もマッチした位置を手の位置とする．

2.5.3　手同士の隠蔽

両手を近づけて行う手話では手同士が隠蔽することがままある．これを処理する方法について述べる．

人物領域の探索範囲内で肌色領域が二つであり，さらに顔との隠蔽でないと判断された時を手同士の隠蔽であると判定する．

手同士の隠蔽の時には保存していた両手のテンプレートを用いて両手の分離を行う．テンプレートマッチングは手の候補となる領域中で，左右の肘から最も近い輪郭上の点を左右の手首位置として，テンプレートの手首位置をその位置に合わせ回転のみのテンプレートマッチングを行い，最もマッチした位置を手の位置とする．

3.　手話の特徴をよく表すような特徴量

前章の画像処理によって，両手と顔の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，領域の位置や形状の情報を数値データとして表す．この章では手話の特徴をよく表す特徴量について述べる．

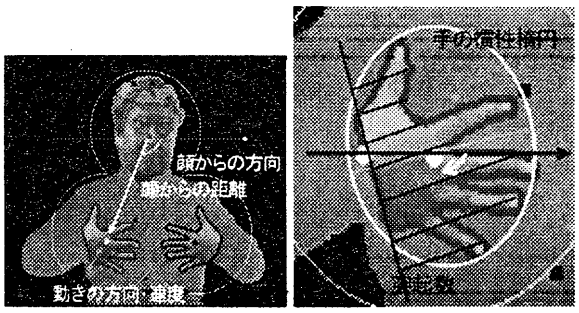
手話の特徴として，「手が動いているときは手の位置の変化が重要であり，手が止っている時には手の形状が重要である．」という傾向がある．このような傾向をうまく表現できるような特徴量を定義する．

3.1　手の位置，動きに関する特徴量

手話単語は手の動きによってどの単語かをおおまかに分類できることが多い．ここで，手の動きを表す特徴量として，両手の顔からの相対位置，前フレームからの動きの方向などを特徴量として用いることにする．(図 10(a)) また手が顔から離れるときには，手の位置はあまり重要ではないことを考慮して，手の位置を極座標系で表す．顔からの距離が一定以上遠いときには変化を少なくするように変更する．顔から手の距離を r とすると特徴量としての顔からの距離 R は，(式)

$$R = \begin{cases} r & (r <= R_a) \\ \sqrt{R_a}\sqrt{r} & (r > R_a) \end{cases} \qquad (4)$$

R_a は顔から離れているとする距離で，本研究では実験的に 150 ピクセルとしている．



(a) 位置，動きに関する特徴量　　(b) 形状に関する特徴量

図 10　手話特徴量

3.2　手の形状に関する特徴量

手の形状を表す特徴量として，領域の面積，領域を楕円近似したときの円形度（長軸/短軸），慣性主軸の方向，手領域の突起数を用いる．(10(b)) 突起数は手首線（手首と重心を結んだ線の垂線）からの輪郭線の距離を測定し，山と谷の距離が一定以上であり，山の両隣の谷の距離が一定以下の極値の数とする．このとき，極値とその両隣の谷までの距離を見て，長いものは握りこぶしであるとして突起数 0 とする．

3.3　特徴量の正規化

HMM を用いてモデルを構築する際，様々な被験者の手話を用いて学習した方が認識性能がよくなる．このとき前節までで述べた特徴量をそのまま使うと，撮影の状況が異なったり，腕の長さなどに個人差があるため不都合が起る．そこで，特徴量を正規化することを考える．

正規化しなければならない特徴量は顔からの距離と手領域の面積である．これらを正規化するために，顔からの距離は初期状態の顔と手の距離で除算し，手領域の面積は，初期状態の顔領域の面積で除算したものを特徴量として正規化を行う．

図 11,12 に正規化を行った結果を示す．上で説明した方法を使ってある単語について被験者 A のデータと被験者 B のデータがどれほど一致しているかを示したグラフである．横軸がフレーム，縦軸が各特徴量（顔からの距離，手の面積）である．実線が A の元のデータであり，細かい点線が正規化を行ったものである．正規化した後，別の被験者データによくマッチしていることがわかる．(状態ごとの定常値が近くなっている) このようにすることで複数の被験者によって学習のデータベースを構築することと，別の被験者のデータを認識に用いることができるようになる．

4.　HMM の状態数の決定

HMM は学習対象ごとに適切な状態数を設定しないとうまく

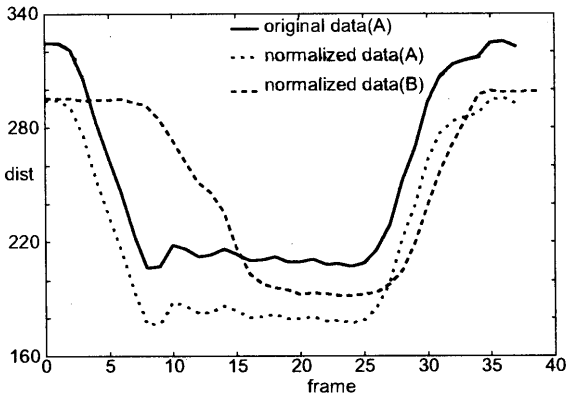


図 11　顔からの距離の正規化

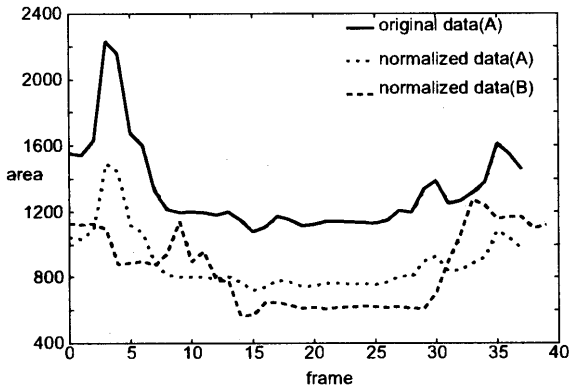


図 12　手の面積の正規化

機能しない．そこで各単語を学習する時の状態数の決め方を述べる．HMM の出力ベクトルは各状態正規分布に基づく出力確率があるので，明らかに異なる特徴量が出力されるフレームが一つの状態に属さないように分離したい．しかし状態数が多すぎると各状態で学習に用いることのできるデータが少なくなりすぎる．

手話は移動中は手の形はさほど重要ではなく，静止中の形状変化は重要である．このことから，次のような時をそれぞれ 1 つの状態と考える．

- 初期状態
- 一定以上の速度で同じ方向に動いているとき
- 一定以下の速度中に手の形状が大きく変化したとき
- 終了状態

このようにすることで，移動中にも 1 つの状態があるため，移動前と移動後の場所が二つの単語でいっしょであった場合でも経路が異なった場合異なる学習結果を得ることができる．

ここで実際の単語例に対する状態数の決定について述べる．図 13 は「かばん」という単語であるが，まず初期状態（状態 1 ）があり，右手をわき腹のあたりまで移動する移動中の状態（状態 2 ）があり，わき腹のあたりで手を握って静止している状態（状態 3 ）があり，最終状態（もとの位置）まで移動する移動中の状態（状態 4 ）があり，最終状態（状態 5 ）があるため，「かばん」という単語の状態数を 5 と設定する．また両手で

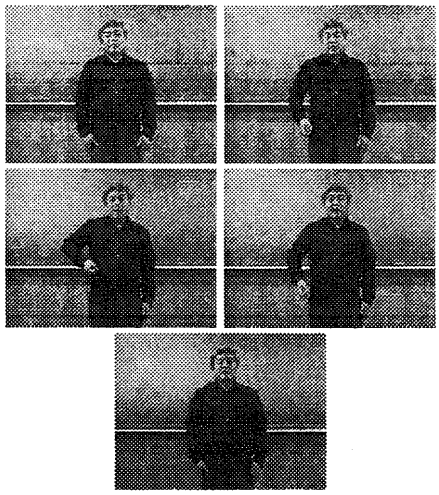


図 13　かばんの状態遷移 (左上，右上，左中央，右中央，下の順)

行う手話についてはそれぞれの状態遷移を確認し，同時に起っている状態遷移は一つと数え，ほかは同様に決定する．

5. 認識時の候補の限定

5.1 HMM を用いた手話単語の認識（マッチング）

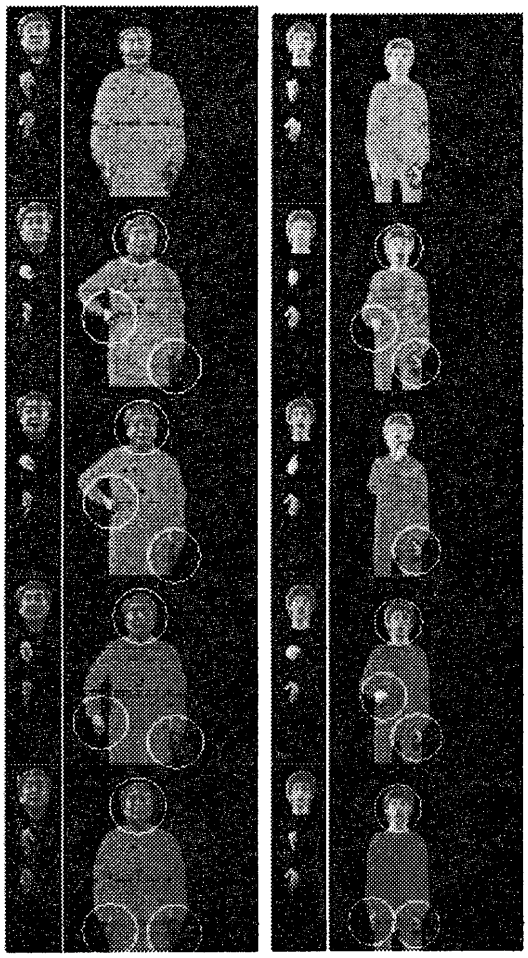
手話単語の認識は，認識したいデータの出力ベクトル列が，前節で単語ごとに学習された HMM に対して最も尤度が高い HMM を認識結果とする．認識には Vitebi アルゴリズムを用いる．

5.2 ロバストな認識のための認識候補の限定

手話は片手のみで行う手話や両手で行う手話があり，また手の動きだけで十分判別可能なものも多い．動かしていない方の手は被験者も特に意識をしていないため，手の置き方などによって形状もバラバラになりやすい．前章で述べた両手の特徴量すべてを用いて認識するよりも，手話ごとに必要な特徴量のみを選んで認識したほうが認識精度が上がると考える．また動きによって候補を限定した方が，画像処理のエラーなどで形状がうまく抽出できなかった場合にも認識できる可能性がある．このように候補を限定して認識をする方法について述べる．

片手の手話であるか，両手の手話であるかを判定する方法は，片手の手話の場合，動かしていない手（たいてい左手）は，初期位置を中心としたある半径内から出ることはない．このことより，左手領域の重心がある範囲内に手話実演内に収まっていれば片手の手話であると判定する．それ以外の手話は両手の手話であると判定する．片手の手話である場合前章で述べた手話特徴量の両手 16 次元分のうち片手分 8 次元のみを用いて HMM を学習，認識する．両手の手話と判定されたものについては，16 次元分を用いる．

また動きによる候補の限定であるが，この場合，手の動きの特徴量（顔からの距離，向き，速度，動きの方向）を出力ベクトル列とする HMM と，手の形の特徴量（面積，突起数，慣性主軸方向，円形度）を出力ベクトルとする HMM の二つを別々に学習する．認識時にまず動きに関する特徴量の出力列から，



(a)「かばん」　　(b)「白」

図 14　画像処理の結果

その特徴量を出力しやすい上位の HMM をその単語の候補として限定する．その後手の形を出力ベクトルとする HMM の中から，手の形の特徴量を用いて，動きで限定した候補の中で最も形の特徴量を出力しやすい HMM を認識結果とする．

6. 実　　験

6.1　撮影に用いた手話画像

本研究で用いた手話画像は，手話熟練者に手話をしてもらったものと著者がその手話をまねして行ったものである．神田氏は手話に慣れた被験者であり，手話のスピードは速い．撮影は大学内の講義室で行い，背景や服装も特に指定せず，普段通りの服装で撮影している．画像の大きさは 640 × 480 である．

撮影した手話単語は，服屋における買い物を想定した単語を撮影した．これはこの手話認識システムを店頭にカメラを置いて実現することを想定していることからである．

なお特徴量抽出（画像処理）は現在 1 フレームあたり数秒かかるため，撮影した画像（30 フレーム/秒）を 1 フレームおきに処理する．

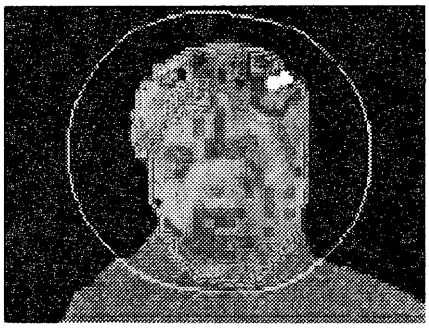


図 15　画像処理の失敗例

6.2　手指特徴量の抽出結果

手話画像を画像処理し，正しいな結果が得られているか確認する．図 14 に示すように，フレームごとに手領域と顔領域を抽出し，特徴を抽出できていることがわかる．図中の円が肌色探索範囲であり，赤い領域が右手，青い領域が左手，緑の領域が顔である．画像処理は４８単語中，４５単語を正確に処理することができた．

図 15 は，顔の中に手領域が完全に入ってしまったため，手首候補点を限定できず，追跡に失敗したものである．このような場合は手首基準にテンプレートマッチングを行わず平行移動でテンプレートマッチングを行うなどをしなくてはならない．ただ，顔に完全に手がオーバーラップしてしまうと，顔領域の面積の変化がなくなり，手と手が隠蔽しているのか，顔と手が隠蔽しているのか，ということを誤認識しやすくなる．顔領域と手領域の分離の問題については現在色情報と相関を用いて少しテクスチャを考慮しているが，今後は顔内のエッジなどを考慮したり，手の腕の長さや，肘と手首を結ぶ直線上に顔領域があるのはおかしいといったような，形状の情報や知識を用いて顔領域から手領域の分離をする必要がある．このような手が顔領域に入ってしまう以外の問題点は特になく，画像処理自体は現在のサンプルにおいては非常にうまくいっていると考えられ成功といえる．

6.3　手話単語の認識実験

画像処理がうまくできたもの 45 単語について手話単語の認識実験を行った．撮影した被験者は二人であり，手話経験者の神田さんのものが一単語につき 3 シーケンス，自分が手話を行ったものが一単語につき 6 シーケンスある．

6.3.1　一人の被験者による認識の実験

神田氏のデータから組合わせを変えながら 2 つを学習用に 1 つを認識用として実験した．著者のデータでも同様に実験を行った．

認識結果を表 1 に示す．自分で撮影したデータについてはほぼ完全に認識することが出来た．神田氏のデータで認識率が低いのは，学習データが 2 例と少ないことに加え，神田氏の手話を撮影する際，同じ単語でも微妙な違いが出るように，一通りの単語を撮影してからもう一度そのリストを撮影するということを 3 回繰り返したため，同一単語でもかなり違う動きになっていることなどが原因である（自分で撮影したものは同じ単語

表1 予備実験の認識結果

熟練者			初心者		
	単語数	正解数		単語数	正解数
1回目	45	34	1回目	45	33
2回目	45	34	2回目	45	45
3回目	45	40	3回目	45	45
合計	135	108	4回目	45	44
	認識率	0.80	5回目	45	45
			6回目	45	43
			合計	270	255
				認識率	0.94

を6回連続で撮影したため，6回ともほぼ同じような動きをしている）．また，動きのにている手話（「肩」と「胸」など）を混同したり，頭が白髪であるので，頭付近に手を持ってくる手話（「「頭」,「黒」など」）で混同が起った．自分のデータにおいては，一度目に撮影したデータを認識にまわしたとき以外はほぼ１００%である．これは，一度目は初期状態（手を初期姿勢に置いている状態）が長いものなどが原因であると考えられる．これについては，手話を学習する際の初期状態／終了状態の時間を固定したりする工夫が必要である．自分で撮影したデータ内での認識率がかなり良いので，これを学習データに用いて，神田さんのデータも個人の癖を吸収する目的で学習データに追加して次節で学習，認識を行ってみる．

6.3.2 複数の被験者による認識の実験

次に，著者の手話データ2つと神田氏のデータの3つ中2つを学習として計4つの学習データからモデルを構成し，学習に用いるデータを変えながら神田さんの残り1つのデータで認識実験を行った．候補を限定せず認識した場合と，候補を限定して認識を行った場合の認識結果の違いと，認識できなかった単語について考察する．なお手の動きで候補を限定する際，動きのHMMで認識した際の上位3位までを候補として残し，そのうちで手の形のHMMによる認識結果が最も上位に来るものを認識結果とした．実験結果を表2に示す．候補を限定して認識した結果の◎は，動きのHMMで認識した結果1位になり，形のHMMで認識した結果も1位だったものであり，○は動きで動きで3位以内に入り，形で1位に入ったものであり，●は，動きで3位以内に入り，形では候補内で最も順位が高くなったため認識成功としたものである．候補限定をしたものの方が，3シーケンスとも認識できた単語数がかなり向上し，3シーケンスとも認識に失敗する手話は無くなった．候補を限定することで，画像処理のエラーで形状が正確に抽出しづらい場合も動きによって候補を限定しているのでその中で形状が似ているものを選べばよいので有効である．

6.3.3 考　　察

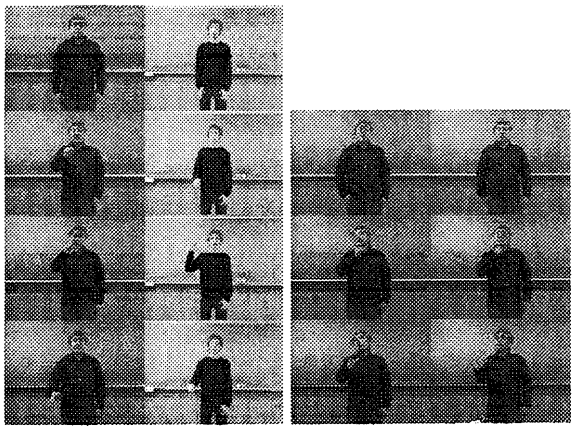
候補の限定をしない場合に比べて認識率自体は向上したが，この実験で使用した3位という候補の限定の仕方は暫定的なものであり，動きが非常に似た手話（数字など）がたくさん登録されている場合，上位のものを固定的に候補とすると，本来の正解が候補にあがらないことが有ると考えられる．今後は尤度の差や，動きの似た単語の数も考慮して候補の数の絞り方を考える必要がある．

表2 実験結果

単語No	単語名	候補 I	限定 II	無し III	候補 I	限定 II	有り III
1	スカート	○	○	○	○	◎	×
2	ズボン	○	○	○	◎	◎	◎
3	シャツ	×	○	×	○	×	●
4	靴	○	○	○	×	◎	◎
5	カバン	○	○	○	◎	◎	◎
6	ネクタイ	○	○	○	●	●	○
7	セーター	○	○	○	◎	◎	◎
8	メガネ	○	○	○	◎	◎	◎
9	色	○	○	○	○	◎	◎
10	赤	×	×	×	◎	×	●
11	青	○	×	○	●	◎	◎
12	黒	○	×	○	◎	×	◎
13	白	○	×	○	◎	◎	×
14	シルク	○	×	×	◎	●	◎
15	皮	○	○	○	◎	◎	◎
16	薄い	○	○	○	◎	◎	◎
17	厚い	○	○	×	●	×	○
18	小さい	○	×	○	●	●	●
19	大きい	○	○	○	○	◎	◎
20	長い	○	○	○	◎	◎	◎
21	短い	○	○	○	◎	◎	◎
22	センチメートル	○	○	○	◎	◎	◎
23	円	○	○	×	×	◎	○
24	夏物	○	○	○	◎	◎	◎
25	秋物	○	○	○	◎	◎	◎
26	冬物	○	○	○	○	◎	◎
27	涼しい	○	○	○	◎	◎	◎
28	暖かい	×	○	○	○	◎	◎
29	暑い	○	○	○	◎	◎	◎
30	寒い	○	○	○	◎	◎	◎
31	肩	○	○	×	◎	◎	◎
32	胸	○	○	○	◎	◎	◎
33	頭	○	×	×	◎	×	◎
34	安い	×	×	○	○	×	◎
35	高い	○	○	×	◎	◎	○
36	合う	○	○	○	●	●	◎
37	流行	○	○	○	◎	○	◎
38	背が高い	○	○	○	◎	×	◎
39	背が低い	×	○	○	●	◎	×
40	好き	○	×	○	×	◎	◎
41	嫌い	×	○	×	×	○	×
42	値上げ	○	○	○	●	◎	◎
43	～がありますか	×	○	○	◎	○	○
44	～はどこですか	○	×	○	●	◎	◎
45	～していいですか	×	×	○	○	◎	◎

また，手話のなかには，手話を行う場所自体はそれほど重要ではなく，その場所で行われる手の形状変化が重要な手話もある（例：「厚い」「小さい」）など．このような単語は，動きで候補を限定使用とした場合，正解以外の単語の動きに入力が似ていた場合，正解が候補に上がらないことが考えられる．これにたいしては，学習データ内で，位置の分散を大きくしたり，位置は重要ではないという事前知識のもと，認識候補を限定するなどの工夫が今後必要であると考えられる．

また似たような問題であるが，同一被験者の場合でも，例え



(a) 厚い　　(b) 嫌い

図 16　認識が難しい単語

ば「嫌い」という単語については，手をあごあたりから指を広げながら顔の前方方向に下げる手話であるが，現在は奥行き方向を考慮していないため，顔が正面を向いている場合は画面上での位置があまり変化せず，顔を少し横に向けた場合は横方向の動きが大きく画面上で現れて動きでうまく学習・認識できなかった．このような場合についても，顔の方向を検出する，ないしは奥行き方向の動きを検出するといったことをする必要があると考える．

実験結果で以上のような問題点が出ることがわかったが，それ以外の手話については認識できたので，候補を限定して認識するということは有効な手法であることがわかった．また，片手の手話は動かしている手の方の特徴量しか用いず学習することや，候補を限定するため，学習データが少なく単語データベースの信頼性が低い時にも認識できるということと，認識の際にかかる時間も減るというメリットがある．

7. おわりに

本論文では，複雑な背景下において，手話画像の処理方法と，それを学習・認識するための適切な手話特徴量，学習・認識する際の状態数の決定および候補の限定方法について述べた．

手話画像処理の方法については，被験者の肌色の情報をあらかじめ取得することによって，固定的な肌色範囲の決定ではできなかった，様々な肌色に対応することができるようになった．

手話の学習と認識においては，音声認識用の HMM ツールキットである HTK(Hidden Markov model Took Kit) を用いた学習と認識について述べた．HMM は時間的に伸縮性があり，多少のエラーにも対応できる．手話の学習時には，適切に HMM を学習できるような状態数の決め方を検討した．手話の特徴として，「手が動いているときは移動中で場所が重要であり，手が止っている時には手の形状が重要であり形状の変化をよく見なくてはならない」という方針によって単語ごとに状態数をきめる方法について述べた．

また手話の認識候補を動きを用いたり，どちらの手による手話かを判定することによりあらかじめ候補を限定する方法についても述べた．このように候補を限定することで，誤認識が減らすことができた．

また，手話の日本語変換システムにおける今後の課題は以下のようなものである．

- 難しい隠蔽状況（手が顔領域に完全に含まれている等），画像処理の難しい状況での特徴量抽出
- 撮影した手話のサンプルが少ないため，確率・統計的手法である HMM の有効性や問題点が見えにくい．サンプル数を増加して実験を行う必要がある．
- 手話単語認識における問題点の改善を重視したため，手話文の学習・認識の実験には至らなかった．今後は手話文の学習・認識方法について考える必要がある．

8. 謝　辞

手話撮影および研究に協力していただいた，中京大学教授の神田和幸氏に感謝いたします．

文　献

[1] H.Sagawa,M.Takeuchi, "A Method for Recognizing a Sequence of Sign language Words Represented in Japanese Sign Language Sentence",*Face and Gesture*,pp. 434–439,2000.

[2] N. Shimada, K. Kimura and Y. Shirai, "Real-time 3-D Hand Posture Estimation based on 2-D Appearance Retrieval Using Monocular Camera", *Proc. Int. WS. on RATFG-RTS (satellite WS of ICCV2001)*, pp. 23–30, 2001.

[3] K.Imagawa, "Color-Based Hands Tracking System for Sign Language Recognition", *Face and Gesture (FG1998)*, pp. 462–467,1998.

[4] K.Imagawa,H.Matsuo,R.Taniguch, and D.Arita, "Recognition of Local Features for Camera-based Sign Language Recognition System", *Face and Gesture (FG2000)*,pp. 849–0853,2000.

[5] Jurgen Kinscher,Holger Trebbe, "The Munster Taging Project - Mathematical Background", *Arbeitabereich Linguistik, University of Munster*,D-58149 Munster,May 19 1995.

[6] 坂口，大谷，岸野："隠れマルコフモデルによる顔画像からの表情認識"，テレビジョン学会誌，Vol.49 NO.8 1995.

[7] Takahiro Otsuka,Jun Ohya: "Spotting Segments Displaying Facial Expression from Image Sequences Using HMM", 0-8186-8344-9/98,1998 IEEE

[8] Takio Kurita, Satoru Hayamizu, "Gesture Recognition using HLAC Features of PARCOR Images and HMM based Recognizer", *Face and Gesture (FG1998)*,pp. 422–427,1998.

[9] 大和，大谷，石井，" 隠れマルコフモデルを用いた動画像からの人物の行動認識"，電子情報通信学会論文誌,vol.J76-D-II NO.12 1993.

[10] Christopher R. Wren, Brian P. Clarkson, Alex P. Penland, "Understanding Purposeful Human Motion", *Face and Gesture (FG2000)*,pp. 378–383,2000.

[11] NPO 手話技能検定協会，"ひと目でわかる実用手話辞典"，新星出版社