手の機能的な構造特徴量に基づく指文字認識

正会員 長 嶋 祐 二[†],

藤 井 昌 紀[†],

長 嶋 秀 世[†]

1. ま え が き

手話は、聾者（聴覚障害者）のコミュニケーション手段のひとつであり、一種の視覚的言語である。手話は、3次元空間上の手の形（手型）、位置および動きの形態要素（手指動作）が同時結合し、時間軸方向に逐次結合することにより語が形成される。手話の調動は、これらの手指動作の他に、うなづき・表情・口形に代表される非手指動作をつけることにより文法要素が加わり、統語構造をなし、視覚系によって受信者へ伝達される。

手話の調動のうち手指動作を認識するには、初期の手型とその位置および軌跡を抽出しなければならない。また、手話の会話中に使用される指文字は、日本語の五十音に相当するものであり、手型と掌方向（一部に動き要素を含む）により決められている。これら指文字には、手指動作を構成している多くの手型要素を含んでいる。

そこで本報告では、実情に即した手話の調動認識を行う手段として、既存の方法とは異なり、以下の方針で画像処理による手型（指文字）認識について検討する。

- ・手話の自由な動きを束縛するデータグローブなどの装着型デバイス^{1)~3)}ではなく、非装着かつ非接触型の画像処理により認識を行う。
- ・指先や関節などに色などのマーカを付け、それらを特徴点として捉えるのではなく、手のシルエットから手型自体が持つ普遍的な特徴量を抽出する。

- ・用いる画像は、手領域のみ接写したり⁴⁾、複数台のカメラにより左右の手掌の動きを追跡する⁵⁾のではなく、通常の手話の調動で用いる空間（上半身以上）を含む画像を用いる。

まず2章では、RGB データから肌色情報を用いて手領域を抽出する方法について述べる。次に3章では、2章で得られた手領域から指文字認識に必要な特徴量を抽出する。4章では、抽出された特徴量から指文字を認識する方法について述べ、実験とその結果について検討する。最後に5章では、本報告のまとめと今後の課題について述べる。

2. カラー画像からの手領域の抽出

手話の調動画像は、RGB データのカラー画像を用いる。カラー画像では、入力画像から“肌色”という色情報を用いて、手領域を容易に抽出することが可能となり、背景や服装の色、提示位置などに対する制約条件が少なくてすむ。

2.1 手領域の抽出

ここでは、入力画像から手領域を抽出する手順を示す。

(1) 表色系の変換

入力画像はRGB データである。しかし、このRGB 表色系では、肌色という人間の感覚的な色を示すのは困難である。そこで表色系には、心理的に独立した色相 H 、明度 V 、彩度 C の3軸をもち、修正マンセル色空間をモデルとした HVC 表色系を用いる。RGB 表色系から HVC 表色系への変換には、MTM 法⁶⁾を用いた。図1に原画像の一例を示す。

(2) 肌色領域の抽出

肌色領域は、(1)で変換された HVC 系の色相 H 、明度 V 、および彩度 C の各データに対して、肌色の範囲を包含するものとして抽出される。実験に用いた入力画像の肌色は、色相 H の値を5 Y から5 YR ま

[†] 工学院大学

“Recognition of Manual Alphabets Based on the Feature of Functional Structure of Hand” by Yuji Nagashima, Masanori Fujii and Hideyo Nagashima (Kogakuin University, Tokyo)

図 1 原画像（白枠内は手領域候補）
Original image.

図 2 2 値化した手領域
Binary image of hand region.

での範囲とし、彩度 C の値を 0.6 から 2.8 までの範囲とする。また明度 V は、撮影時の照明条件によって変動しやすいので、判別分析法を用いて自動的に閾値を求め、得られた閾値以上を明度 V の範囲とする。

(3) 手領域候補の抽出

(2) で抽出された肌色領域には、顔の領域と手の領域が含まれている。そこで、肌色領域の各孤立領域ごとに面積を求め、最大領域を顔の領域、2 番目に広い領域を手の領域とする。この手の領域が含まれる範囲のみ切り出すことで、手領域候補を得る。図 1 の白枠内は、原画像から抽出された手領域候補を示す。

(4) 手領域候補の修復

(3) で得られた手領域候補には、雑音により穴や途切れがあり手領域として不完全である。そこで、 5×5 画素の正方領域から 4 隅の画素を除いた要素を用いて手領域候補画像に対し拡大/縮退処理を行う。これにより 4×4 画素以下の穴や途切れ領域の除去が可能と

なる。ただし、これら画像サイズは、画像中の指サイズにより決定した。

(5) 手領域の抽出

(4) で修復した手領域候補から輪郭線を求める。この中で最大の長さを持つものを求め、その輪郭線に対して閉じた領域を手領域、開いた領域を背景として 2 値化する。このようにして手領域を抽出することができる。

以上の処理により、入力画像から自動的に手領域を抽出することが可能となる。また、この 2 値化した手領域の抽出処理結果を図 2 に示す。

3. 特徴量の抽出

調動体としての手型には各指ごとの屈曲、伸展などの組合せにより多様な形状が存在する。

画像処理による手型認識を考慮すると、単純で抽出が容易な特徴量として、伸展している指の本数、部位および方向があげられる。指の伸展には、伸展した指どうしが離れている離開伸展と、伸展した指どうしが接触している密着伸展に分けられる。そこで、手型認識には、離開伸展あるいは密着伸展した指の本数、部位および方向を特徴量として用いる。

3.1 離開伸展した指領域

2 章で得られた手領域のシルエットを見ると、離開伸展した指領域は、掌領域から突起した領域と考えられる。そこで、縮退/拡大処理の突起の除去という性質を利用して、離開伸展した指領域と掌領域とを分離する。

以下で、それぞれを抽出する方法について述べる。

(1) 処理手順

(a) 縮退/拡大処理の適用 2 章で得られた手領域に対して、指の幅の画素数を直径とする円を要素として縮退処理を行い、同じ要素を用いて拡大処理を行う。これにより、指の幅の画素数より小さいものは突起と見なされ消滅し、掌領域が得られる。ここで指の幅の画素数は、参照の画像を用いて各個人ごとに求める。

(b) 伸展した指領域の抽出 指領域候補は、2 章から得られた手領域と (a) で得られた掌領域とを比較して、(a) で消滅した領域として抽出される。次に、抽出された指領域候補に対して 8 近傍連結のラベル付けを行い、一定面積以下の孤立領域を雑音成分として指領域候補から除去する。この処理で残った指領域候補が指領域となる。原画像からの指領域と掌領域の抽出結果を図 3 に示す。

図 3 指領域と掌領域
Finger region and palm region.

(2) 離開伸展した指の特徴量

離開伸展した指領域から、以下のような特徴量が求められる。

- ・離開伸展した指の本数 離開伸展した指の本数は、指領域のラベル付けされた領域の数とする。
- ・指ベクトル 指ベクトルは、伸展した各指ごとに抽出され、その位置と方向成分を持つ。指ベクトルは、各指の領域に対する外接矩形の短辺方向に見た時、各列の重心点を最小二乗近似することで抽出する。また、ベクトルの始点は掌重心（後述）に近い方とする。これらを基本特徴量として、各指の開き、掌に対する接続位置などの高次の特徴量を求める。

(3) 掌の特徴量

抽出された掌領域から得られる特徴量には、重心、外接矩形および方向がある。

- ・掌重心、掌の外接矩形 掌重心は、掌領域の重心点とし、掌の外接矩形は、掌領域の外接矩形とする。これら2つの特徴量は、実際に指文字を認識する処理過程で各指ベクトルの位置関係を記述するのに用いる。
- ・掌方向 掌方向は、上下・左右・無方向の3つの方向をもち、掌領域の外接矩形で長辺方向を掌方向とする。なお、無方向は外接矩形の長辺と短辺の比が1に近いときとする。

(2)で得られた指領域の特徴量と、(3)で得られた掌の特徴量を図4に示す。

3.2 密着伸展した指領域

密着伸展した指は、2値化前の手領域の明度画像において、2つの明度の高い領域（指の腹や背）と、それらに挟まれて1つの暗い領域（境界線）が存在するという特徴を用いて分離する。

しかしこの境界線は、撮影時の照明などの環境要因や指形状などの個人差の影響を受けやすい。明度値や微分強度による単純な2値化処理のみでは、境界線を安定的に抽出することが難しい。

図 4 離開伸展した指と掌の特徴量
Topological feature of finger and palm region.

原画像 境界線

図 5 境界線の抽出
Boundary extraction.

そこで、ここではソーベルオペレータにより得られる微分勾配の方向を用いて、密着伸展した指領域から境界線を安定的に抽出する方法について述べる。

(1) 処理手順

- ソーベルオペレータの適用 2値化処理前の手領域明度画像に、メディアンフィルタをかけ平滑化する。次に、ソーベルオペレータを用いて微分勾配の方向を求める。ここで、微分勾配の方向は、等間隔な8方向に量子化する。
- 境界線候補の抽出 すべての量子化された勾配方向のうち、掌領域に含まれる点に対し隣接した点のベクトルの方向が向いている（明度値 V では、 V 字の谷を成している）とき、この点を境界線の候補点とする。次に、各境界線候補点に対し8近傍のラベル付けを行う。
- 雑音の除去と誤抽出の補正 (b)で抽出した各境界線候補の角度と長さを求める。一定長以下のものは候補から削除する。また、候補に対し、隣接する2組ずつ角度の差と距離を調べる。角度の差が小さくかつ距離が近いものは、途切れが発生して2つに分けられたと考え、接続し1つの境界線候補とする。
- 境界線の抽出 以上の処理により残った境界線候補を、境界線とし抽出する。

本手法により抽出した境界線の例を図5に示す。

(2) 特徴量

境界線から得られる特徴量には、次のものがある。

- ・密着伸展した指の本数 密着伸展した指の本数は、抽出された境界線の数に1を加えた本数とする。
- ・境界線ベクトル 境界線ベクトルは、各境界線ごとに抽出する。境界線ベクトルの方向は、境界線の外接矩形において短辺方向に見た時の各列の重心点を最小二乗近似することで抽出する。

4. 指文字の認識

ここでは、3章で得られた特徴量を用いて指文字の認識を行う。各特徴量の照合を階層的に配置した決定木に基づいて行う。本認識システムの処理概要を図6に示す。

4.1 認識実験

各指文字の調動画像は、背景を青として左右から各4灯のライトを透過光として用い撮影した。撮影にはソニー製のカメラDCX-537Aを使用し、ソニー製の追記型ビデオディスクLVR-3000Nに記録した。マイクロ・テクニカ製のパーソナルコンピュータ用フレームメモリCVFM 01 (512×512画素,R,G,B各8bits)を用いて、パーソナルコンピュータPC-9821APに取込み、イーサネットを介してHP製のEWS HP 9000/735に転送し、すべての処理をEWS上で行った。入力画像サイズは、フレームメモリの有効撮影面積(512×512画素)に上半身が余裕をもって納まるように撮影した。認識に用いた指文字の調動画像は、上記の撮影条件により、ネイティブサイナ、NHK手話ニュースキャスタ2名、健聴者の手話通訳ボランティア2名の合計5名の話者により撮影した。

図6 本システムの処理概要
Processing of this method.

4.2 実験結果

今回対象とした指文字数は、動き要素を持たない39文字の指文字である。

本システムで個別に認識が可能である24文字の指文字を対象にした場合の個別認識率は、平均74%であった。

また、本手法では基本的に手のシルエットを用いるため、個別に認識できない指文字として図7(a)に示す「い」と「ち」、図7(b)に示す「ゆ」と「わ」など15文字がある。この原因は、「い」と「ち」のように手型が酷似している、「ゆ」と「わ」のように前腕の回外か回内の違い、などからシルエットが酷似してしまうためである。これら15文字の指文字は個別に識別できないが、決定木においてシルエットが酷似して

図7 手型の類似した画像例

Examples of similar handshapes in Manual Alphabets.
(a) "I" and "CHI", (b) "YU" and "WA".

表1 認識結果
Result of recognition.

手話者	個別認識率	総認識率
NHK キャスタ A	83%	79%
NHK キャスタ B	70%	76%
ネイティブサイナ	70%	76%
健聴者 A	79%	71%
健聴者 B	70%	74%
平均	74%	75%

いる数種類の指文字群と判定するものも正解に含めた、39文字の指文字の総認識率は、平均75%となった。認識結果を表1に示す。

4.3 考察

認識実験の結果、以下のような点が明らかとなった。

- ・標準手型位置からの手の機屈あるいは尺屈に対しロバストな認識結果が得られる。
- ・背景や服装の制限が緩和され、容易に手領域を抽出できる。

また、問題点として、以下のような点が挙げられる。

- ・手領域が顔に接触しハイライトが生じた時、接触領域の境界線抽出が困難となり、手領域を安定に抽出できない場合がある。
- ・予測した本数の境界線を抽出できない場合がある。

これらの原因は、照明状態や個人の手型の形状差による影響が大きい。

5. む す び

本報告では、画像処理による日本手話の指文字認識について検討を行った。ここでは、指文字認識を手話の調動構造のひとつである手型認識の基礎として位置づけ、上半身が余裕をもって入る画像を用いた。また、手領域は肌色という色情報を用いることで容易に抽出することができた。さらに、特徴量には手型自体の構造を反映した普遍的なものを採用した。その結果、手話に熟練した5名の話者に対して、個別認識率は平均74%となり、全体の認識率は平均75%となった。

今後は、シルエットが酷似する手型などの問題点を改善するため、手話調動のとき得られる各種の動き情報を用いることで、安定した認識手法を確立して行く。

く。

手話影像撮影にご協力いただきました中野佐世子、深海久美子、木村美鈴、中博一および吉永喜美子の諸氏に感謝いたします。

【参 考 文 献】

- 1) 阿部, 酒匂, 佐川: 構文意味解析に基づく手話・文章変換方法, 信学論 (D-II), **J76-D-II**, 9, pp.2023-2030 (1993)
- 2) 岸野, 高橋: 手振認識とその応用, 信学論 (D-II), **J73-D-II**, 12, pp.1985-1992 (1990)
- 3) 渡邊, 黒川: コンピュータ・デスクトップへの身振り対話の導入, 計測自動制御学会: HI 研究論文集, **2**, pp.65-72 (1993)
- 4) 落合, 鎌田: 画像特徴を用いた日本指文字の記述方法, 信学技報, IE89-116, pp.61-68 (1989)
- 5) 石淵, 竹村, 岸野: パイプライン型画像処理装置を用いた実時間手形状認識, 情処学 HI 研資, 92-HI-42, 14, pp.91-96 (1992)
- 6) 宮原, 吉田: 色データ (RGB) - (HVC) 数学的変換方法, テレビ誌, **43**, 10, pp.1129-1136 (1989)
- 7) 神田和幸: 指文字の研究, 光生館 (1986)
- 8) 日本色彩学会編: 新編色彩科学ハンドブック, 東京大学出版会 (1980)

ながしま ゆうじ
長嶋 祐二 昭和55年, 工学院大学大学院電気工学専攻修士課程修了。同年, 同大助手, 講師を経て, 平成6年より, 同大助教授。手話の認識, 生成, 電子化辞書の構築の研究に従事。工学博士。正会員。

ふじい まさのり
藤井 昌紀 平成6年, 工学院大学電子工学科卒業。現在, 同大学大学院情報学専攻修士課程在学中。手話の画像処理による認識の研究に従事。

ながしま ひでよ
長嶋 秀世 昭和41年, 工学院大学大学院電気工学専攻修士課程修了。同年, 東京工業大学助手, 昭和45年, 工学院大学講師, 助教授を経て, 昭和57年より同大教授。商標図形の類似検索の研究に従事。工学博士。