

指文字認識に関する基礎的検討

長嶋 祐二 小野寺 卓 長嶋 秀世 寺内 美奈 大和 玄一
工 学 院 大 学 職業訓練大学校

概要

日本語手話の指文字の認識について検討する。本手法では、上半身全てを用いた一般の手話認識への応用を考慮して、フレームメモリの有効画素(512×512[画素])の中心付近128×128[画素]のみを使って指文字画像を取得した。縮退/拡大処理により伸展させた指領域を抽出し、これから指ベクトルなどの基本特徴量を抽出した。また、微分勾配の方向を用いて指同士の境界線を抽出することにより、密着伸展させた指の状態を検出し、特徴量として用いた。これらの特徴量に基づいて決定木を構築し、指文字の認識を試みた。

A Study of Recognition Method of Japanese Finger Spelling

Yuji NAGASHIMA † , Taku ONODERA † , Hideyo NAGASHIMA † , Mina TERAUCHI † and
Genichi OHWA †

† Kogakuin University, 2665-1, Nakano-Machi, Hachioji-shi, Tokyo 192, Japan

† Institute of Vocational Training, 4-1-1, Hashimoto-dai, Sagamihara-Shi, Kanagawa
229, Japan

ABSTRACT

In the present paper, we propose a method to recognize Japanese finger spelling. In this method, finger spelling images are taken using only 128×128[pixels] region around the center of the frame memory(512×512[pixels]) in order to apply this method to recognition of general sign language expressed by the upper half of a body in the future. Our proposing method extracts regions of stretched fingers from images by erosion-dilation method, and from these regions extract fundamental features such as finger vector. And also the method detects the status of contacted and stretched fingers through extracting boundary between fingers using gradient vector, and use it as features. Based on these features, we have constructed decision tree and tried to recognize finger spellings.

1 まえがき

聴覚障害者のコミュニケーションの手段の一つとして用いられている手話を、画像処理により認識する研究が盛んに行われている。

手話はその名が示すとおり、手掌を中心とした両上肢を上半身の前で形や位置などを様々に変化させ、その一つ一つの形や動き、移動方向に特定の意味情報を結び付けることによって意思伝達を図る視覚言語である。画像処理による手話認識では、上半身に対する肘と手掌の位置、手掌の形（指型）と方向、およびこれらの動きを抽出する処理が重要となる [1]。

手掌の形状抽出に関する研究としては、対象を指文字に限定した指文字認識に関する様々な検討がなされている [1, 2, 3]。

日本語手話の一部である指文字は、日本語の 50 音に相当する表音記号であり、50 音の各文字に異なる指形状を割当てたものである [4]。指文字には屈曲、伸展など様々な指形状が含まれており、一般の手話単語において手掌がつくる形の多くの要素がこの中に内包されていると考えられる [5]。

本報告では、上半身画像全体を対象とした一般の手話認識における手話画像認識辞書部分 [6] への応用を前提とした指文字認識について述べる。すなわち、

- 1) 手掌領域の周辺分布 [2] といった、特定の指文字の形状に依存する画像特徴量を用いるのではなく、各指の状態を表わすような普遍的な特徴量を用いる。
- 2) 手掌部のみを接写するのではなく、上半身画像から手掌部を抽出した状態を想定して画像サイズを設定し、画像処理を行う。

という二項目に重点を置いた。

まず初めに、縮退／拡大処理を用いた伸展させた指領域の抽出方法と、微分勾配に注目した密着伸展させた指の境界線の抽出方法、及びこれらから手掌の状態を表わす特徴量を抽出する方法について述べる。次に、得られた特徴量を用いて決定木を構築し、指文字認識を行う方法について述べる。最後に、実際に指文字画像を取得して認識実験を行った結果について述べる。

2 伸展させた指領域の抽出

手話を構成する手型および指文字には様々な形態のものが存在する。画像処理の観点からこれらの分類／認識を考えたとき、各々の形態の差異が顕著に現れる状態として、伸展させた指型（部位、方向、本数）が挙げられる。これは、爪や関節のしわ、指の影や肌に生ずる陰影など、手掌に存在する様々な濃度情報を完全に取去って二値化したシルエットを用いても、大半の伸展した指型の識別が可能であることからわかる。

そこで本節では、二値化した手掌領域から伸展させた指領域を抽出し、これから特徴を抽出する方法について述べる。

2.1 縮退／拡大処理による伸展させた指領域の抽出

縮退処理と拡大処理は本来、二値化された領域を扱う問題において、穴や途切れ、突起や孤立雑音を除去して領域の形状を滑らかにする前処理として一般的に用いられている。拡大処理は領域を外側に向かって広げ、縮退処理は逆に領域を内側に狭める処理であり、以下のように定義される。

・拡大：

$$\text{if } d[i, j] = ON \text{ then} \\ \{ d[x, y] \mid d[x, y] \in \Omega[i, j] \} = \{ ON \}$$

・縮退：

$$\text{if } d[i, j] = OFF \text{ then} \\ \{ d[x, y] \mid d[x, y] \in \Omega[i, j] \} = \{ OFF \}$$

ここで、 $d[i, j]$ は任意の画素の濃度値を、 $\Omega[x, y]$ は $d[x, y]$ の近傍領域を表す。

[n 回の拡大 $\Rightarrow n$ 回の縮退] または [n 回の縮退 $\Rightarrow n$ 回の拡大] のように、これらの処理を組み合わせる。拡大によって一度消滅した穴や溝は、次の縮退において再び現れることはない。同様に、縮退によって一度消滅した突起や孤立雑音が、次の拡大において再び現れることもない。これらの性質を利用して、領域の概形を保存したまま、穴や溝、突起や孤立雑音を消滅させる。 $\Omega[x, y]$ に方向性を持たせることにより、特定方向の領域にのみ選択的に作用させることも可能となる。

本報告では、二値化した手掌画像における伸展させた指領域を、手掌からの突起と見なして [n 回の縮退 ⇒ n 回の拡大処理] (以後は縮退/拡大処理と略す) によりこれを抽出することを試みた。

ここで用いるのは、 $\Omega[x, y]$ を $d[x, y]$ の周囲八近傍とする無方向性の縮退/拡大処理である。これは、任意の方向に最大で $2n$ [画素] の幅をもつ突起を消滅させることができる処理であり、この $2n$ [画素] を手掌画像における伸展させた指領域の太さに対応させることにより、手掌の回転の影響を受けずに伸展させた指領域を抽出できる。

2.2 処理手順

具体的な指領域の抽出処理手順を以下に示す。

1) 前処理

手掌画像を平滑化した後、Sobelの微分フィルタによりエッジ強調処理を行い、この結果を基に手掌領域と背景領域とを二値化する(図1a,b)。

2) 縮退/拡大処理の適用

二値化された手掌画像に対して、n回の縮退処理とn回の拡大処理を適用する。この結果、伸展させた指領域および手掌領域境界の微細な凹凸が消滅した画像が得られる(図1c)。

3) 伸展させた指領域の検出

1)と2)の結果とを比較して、2)で消滅した領域を抽出し指領域候補とする。次に、抽出された指領域候補に対して八近傍連結のラベルづけを行い、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(1d)。この処理で残った指領域候補を指領域とする。

以上で、伸展させた指領域群が抽出され、個々の識別が可能となる。

2.3 特徴量の抽出

縮退/拡大処理によって抽出された伸展させた各指領域を基に、手掌の状態を表わす基本特徴量を抽出する。認識処理に用いる基本特徴量について述べる。

1) 伸展させた指の本数

抽出された領域数を、伸展させた指の本数とする。

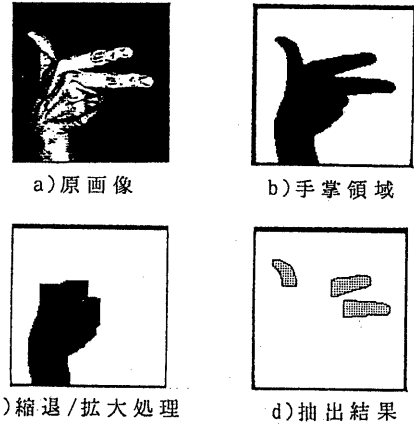


図1: 縮退/拡大処理による伸展させた指領域の抽出

2) 指ベクトル

指ベクトルは、伸展させた各指領域ごとに抽出され、その位置、方向、長さを表わす。指ベクトルは、指領域を外接矩形の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される(図2)。また、ベクトルの始点は、二端点のうちの手掌重心(後述)に近いほうとする。

3) 各指領域の重心点

伸展させた指の位置を一点で表わす場合に用いる。

4) 手掌重心、手掌の外接矩形

手領域から伸展させた指領域を除いた全ての部分の重心点、外接矩形である。これらは、各指ベクトルの位置関係を記述する際に用いる。

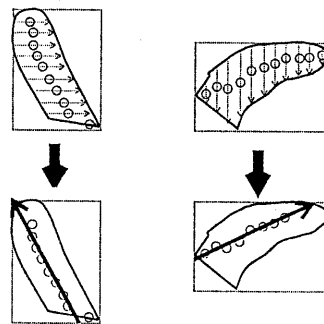


図2: 指ベクトルの抽出方法

これらの特徴量を抽出した例を図3に示す。

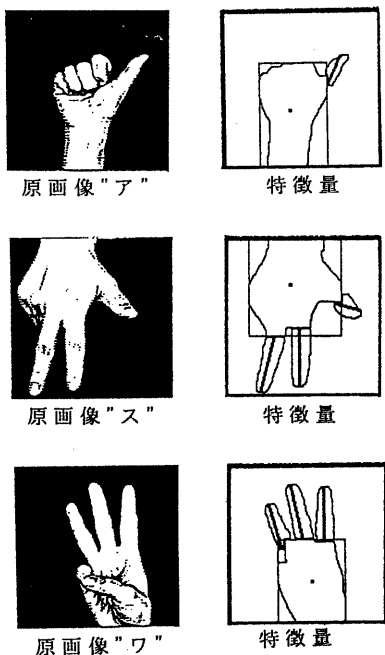


図 3: 抽出された基本特徴量

3 密着伸展させた指の境界線の抽出

前節では、二値化した手掌画像から伸展させた指領域を抽出し、これから特徴を抽出することを検討した。既に述べたように、伸展させた指については、二値化された手掌のシルエットを用いることにより多くの指文字の認識が可能となる。しかし、次にあげる形態の指文字は伸展させた指領域の抽出が不可能なため、指文字を認識することはできない。

- 1) 指を握った指文字
- 2) 二本以上の指を密着して伸展させた指文字

そこで本節では、これらの指文字の構造を認識できる特徴量として、原画像のエッジ勾配を用いて密着伸展させた指の境界線を抽出する方法について述べる。

3.1 エッジ勾配の方向を用いた指の境界線の抽出

密着させた指の境界線は、指文字画像において指と指の間に生ずる暗い筋としてその存在が容易に識別できる。しかしこの境界線は、照明強度や個人の指形状、肌色の違いなどの影響を受けやすく、濃度値や微分強度に対する単純な二値化処理だけで安定的に抽出するのは難しい。

そこで、微分フィルタにより得られるエッジ勾配の方向を用いて、安定にこれを抽出することを試みた。

密着させた指の境界線は、画像特徴として二つの明部（指の背や腹）に挟まれた暗部と見なされるので、勾配の方向が向い合った二つのエッジ領域の境界上に現れると考えられる。従って、微分フィルタにより得られた微分画像をエッジ勾配の方向に基づき領域分割し、条件に適合する領域境界を抽出する。この時、局所雑音の影響により境界線の連続性が損われるのを防ぐため、微分処理の後に境界線強調処理を施す。

3.2 処理手順

つぎに、密着伸展させた指の境界領域の抽出手順を述べる。

1) 前処理

Sobel の微分フィルタを用いて、原画像よりエッジ勾配の方向を求める。局所的雑音の影響を軽減するため、得られた各勾配値を等間隔 4 値に量子化し、4 通りのコード付けを行う（図 4）。

2) 境界線の強調

図 4 で "1" と "3"、"2" と "4" を指の境界線候補として整合する方向コードの組とする。注目点のコードとその周囲 8 近傍のコードとを比較し、整合 / 不整合状態で接しているコードの数を調べる。この数を基に、注目点のコードを整合するコードに修正していく。これにより、境界線の連続性が強調され、その抽出が容易となる。

3) 境界線の抽出

以上の結果から、隣接した整合コード領域の境界線候補を検出し、一定長以上のものを密着伸展させた指の境界線とする。

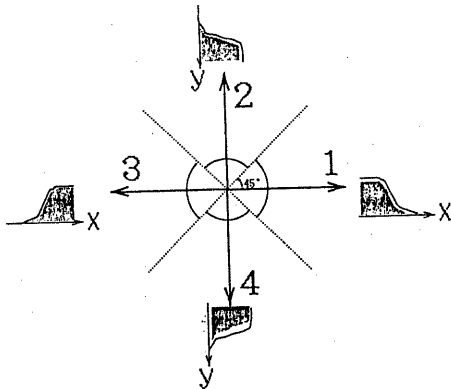


図 4: エッジ勾配の方向の量子化

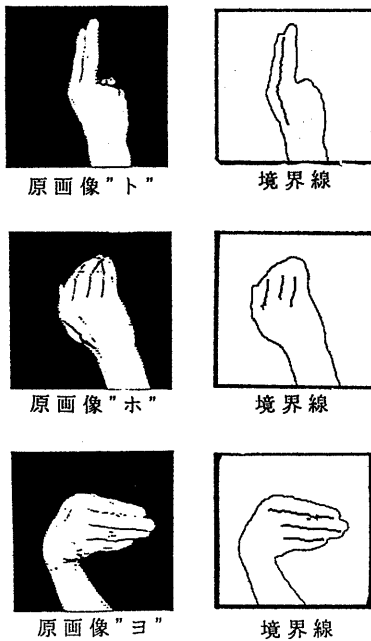


図 5: 密着伸展させた指の境界線の抽出

4) 特徴量の抽出

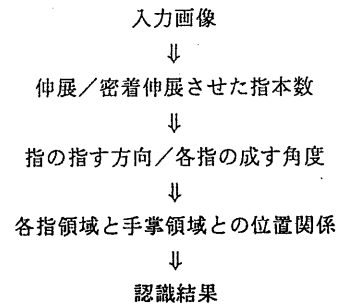
得られた境界線情報を基に、密着伸展させた指の本数および画像平面内で指す方向を抽出し、特徴量として用いる。

本手法により抽出した指境界の例を図 5 に示す。

4 指文字の認識

ここでは、2, 3 で得られた伸展させた指に関する特徴量を用いて指文字の認識を行う。本システムでは、入力された指文字画像より各特徴量を抽出し、構築しておいた特徴データベースと比較/照合を行い、認識結果を出力する。

認識方法は、各特徴量の照合ステージ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の階層構造の概要を次に示す。



この階層構造の特徴は、

- 1) 話者の癖による相違が少ない指文字の普遍的特徴を上階層に配置する。
- 2) 指の伸屈や指関節の角度、手首全体の捻りといった、話者ごとの個性差が多く現れる特徴を下階層に配置する。

ことによって、個体間の差異の特徴量への影響の度合いを過大評価することなく、誤認識を減少させるという考えによるものである。

5 認識実験

各指文字は、背景を暗幕として左右から 2 灯のライトを当て撮影し、影の発生を防いだ。撮影に使用した

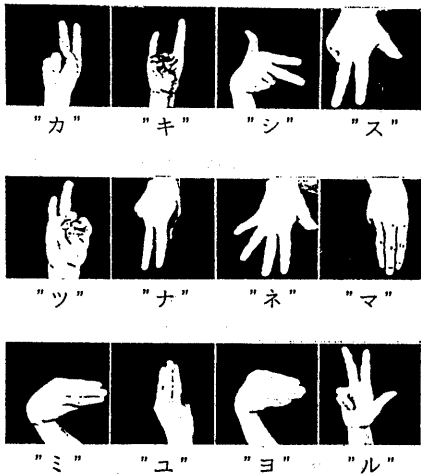


図 6: 個別識別が可能な指文字

ビデオカメラは1/2インチサイズの CCD 素子を用いた SONY 製白黒 CCD カメラ XC-75 である。ビデオ信号の入力には、パーソナルコンピュータ用フレームメモリ、マイクロ・テクニカ製 CVFM01 (512 × 512 画素, モノクロ 8 bits) を用いた。入力された画像データは、ワークステーション HP-APOLO425t に転送して全ての処理を行った。

撮影サイズは、フレームメモリの有効撮影面積を全て使用するのではなく、中心付近の 128 × 128 画素だけを用いて、この範囲内に手掌のみを納めた。これは 1. で述べたように、上半身全てを使った一般の手話認識への応用を考慮してのことであり、上半身をフレームメモリー一杯に納めた場合の手掌の占める面積にほぼ一致するためである。

指文字認識用の画像は、上記の画像入力条件により、4 名の話者（内 2 名は日時を変えて 2 回）をのべ 6 回の撮影により得られた合計 241 枚の指文字画像を用いた。

5.1 実験結果

現在のシステムで個別認識が可能な 12 の指文字を図 6 に示す。この 12 の指文字を対象にした場合、平均認識率は約 69% であった。

また、個別認識までは至らないが、途中経過の時点で正しく分類されているものを正解に含めた場合のシステム全体の認識率は約 65% であった。

5.2 問題点

認識実験の結果、以下のような点が明らかとなった。
・縮退／拡大処理による伸展させた指領域の抽出処理は、手掌全体の傾きや指の方向の影響を受けずに良好な抽出結果が得られる。

しかし、指文字“コ”における伸展させた母指では、抽出される領域が極端に小さい時に、雑音と見なされて除去される場合がある（図 7）。

また、中指と他の指を密着伸展させる指文字“ウ”や“ケ”では、中指の先端だけがわずかに突出するため、単独で伸展させた指領域として誤抽出される場合もある（図 7）。

これらの問題は、手掌領域からの突起部を単純に伸展させた指領域と見なす、本手法の適用限界によると考えられる。従って、手掌の構造を知識として用いて誤抽出を減少させることを検討する必要がある。



図 7: 伸展させた指領域の誤抽出が生じやすい指文字

・微分勾配を用いた密着伸展させた指の境界線の抽出処理は、画像の質に依存して決定されるような二値化パラメータ等を用いずに特徴抽出が安定的に行える。しかし、より安定した検出を行うためには境界線強調処理について、さらに検討する必要がある。

・抽出した特徴量を基に認識を行う処理では、各特徴量の照合ステージを階層的に配置する決定木の手法を用いた。この手法は指文字間の類似性を利用することにより、必要最小限の照合で認識結果が得られるという利点がある。しかし、入力からそれぞれの認識結果までの道筋が一つしかないことから、各分岐点における照合ミスや特徴量自体の誤抽出が直ちに誤認識に結び付くという問題点がある。これを改善するためには、特徴量の抽出精度を向上させると共に、全特徴量を並列に扱い、統合分析することによって最も有力な候補を出力するような認識処理についても検討する必

6 あとがき

本報告では、画像処理による日本語手話の指文字の認識方法について検討を行った。ここでは指文字認識処理を、上半身全てを用いた一般の手話認識における手掌形状認識部分の基礎データを得るための一段階として位置づけた。すなわち、フレームメモリの有効撮影面積の一部分だけを用いて指文字画像を取得することにより、話者の上半身画像から手掌部を切り出した状態に近づけた。また、画像から抽出する特徴量としては、特定の指文字の形状だけに依存するものではなく、手掌の普遍的な構造を反映するようなものを採用し、一般の手話認識への応用を考慮した。

本システムでは、縮退/拡大処理による伸展させた指領域の抽出処理、および微分勾配に注目した密着伸展させた指の境界線の抽出処理を用いて、伸展させた指の位置や方向を表わす指ベクトルや、密着伸展させた指の状態等を抽出した。さらに、現在認識可能な12の指文字について認識実験を行った結果、約69%の認識率を得た。

今後の課題は、各特徴量の抽出処理における問題点を改善すること、及び本手法を一般の手話認識へ応用し、その有用性を検証することが挙げられる。

参考文献

- [1] 落合, 鎌田: 画像特徴を用いた日本指文字の記述方法, 信学研資, IE89-116, pp.61-68(1989)
- [2] 秋原, 脇: 投影像を用いた指文字認識法, 信学秋期大会, pp.1536(1986)
- [3] 安達: 手話通訳システムの研究動向, 信学研資, NLC92-5, pp.33-40(1992)
- [4] 金田: 日本語手話辞典, 光書房(1980)
- [5] 長嶋, 寺内, 大和: 日本語手話の形態素分析とその記号化に関する検討, 日本ソフトウェア科学会第8回大会(1991-09).
- [6] 長嶋, 寺内, 佐藤: トータルコミュニケーション支援用辞書構築に関する検討, 情報処理学会ヒュー

