

手話認識技術の研究*

布施 泰史^{*1}・巢山 昭文^{*2}

Study on Recognition Method for Sign Language

Yasufumi FUSE and Akifumi SUYAMA

本研究は、全国で約35万人、当県でも約5千人いると言われている聴覚障害者と健聴者との対話を支援するための手話を自動認識するシステムを構築することを目標としている。手話の一種で会話と併用される文字の表現法の一つに指文字がある。今回、近赤外光カメラを用いて認識するシステムについて検討を行い、手、指を静止させて表現する指文字（静止指文字）と手、指を動作させて表現する指文字（動作指文字）との全指文字を対象とし、ほぼ良好に認識できることを確認した。

キーワード：聴覚障害者、指文字、認識、画像処理、近赤外光

1 はじめに

聴覚障害者の3割程度は、手話を日常の会話手段としている。これに対して、聴覚障害者でなく（健聴者ともいう）かつ手話ができる人は少なく、聴覚障害者と健聴者の会話がスムーズに行われない場合が多い。そこで、今回、聴覚障害者と健聴者との対話を支援することを目標とし、最終的に手話を認識するシステムの開発を実施した。その第一歩として手話の一種である指文字を認識する手法の検討を行った。指文字は手の形を使った全国共通の文字の表現法である。日本語は50音46文字あるが、この日本語一つ一つに対して、すべて指文字が備わっている。聴覚障害者は手話の合間で固有名詞を表現する場合、手話がわからない場合等にこの指文字を補助的に用いて会話を行っている。

従来、手話認識の研究ではセンサー付手袋（データグローブ）などが多用されている。しかしこの方法では、ケーブルの繋がった専用手袋を装着するため運用上非常に不便である。我々は研究に着手する上でシンプルなシステムを提案するため画像処理技術による研究を進め、最終的に素手の

まま認識可能な手法を検討した。

また、手話の身振り動作についても、この手法の延長を別の画像処理技術を使い検討したので報告する。



2 実験方法

2-1 実験構成

近赤外光の反射光を捉える画像入力デバイスを用い、特定のタイミングで得られた画像について各種の画像処理を行い、指文字の認識や簡単な手話動作の認識について実験を行った。

(1) ハードウェア構成

本装置は、モーションプロセッサ（㈱東芝）と呼ばれる画像入力デバイス（専用カメラ1台、近

* 手話認識技術の研究(第3報)

*1 機械電子・デザイン部

*2 現 企画調整部情報政策課

赤外光を発光するLED、反射光を捉えるレンズと結像用イメージセンサ及び画像処理ボード）と画像処理用パソコンで構成される（図1）。このデバイスは、反射光の強さを取得することで距離情報を濃淡画像として表現できる機能を持つ。その能力は表1のとおりである。

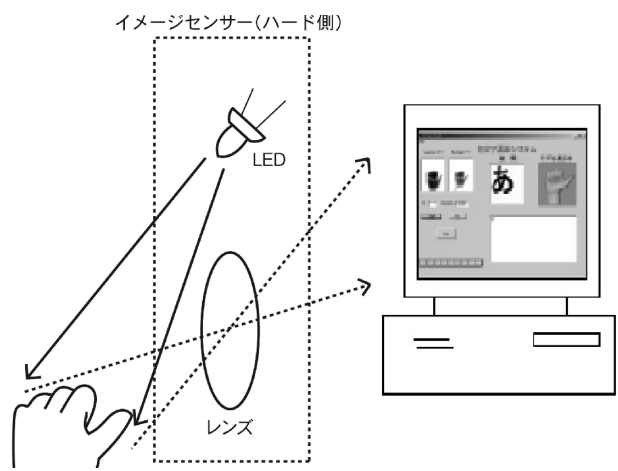


図1 実験の構成

表1 性能表

分解能	64 × 64 pixels
奥行き深度	256 階調
動作レート	25 ～ 50 フレーム/sec
撮影距離	30 ～ 90 cm

(2) 画像処理ソフト

画像入力デバイスのシャッタースピード等を制御する部分、二値化、正規化、パターンマッチング処理等を行う各種の処理部分、その結果を認識する部分をそれぞれ作成し認識の検討を行う。

2-2 認識処理の流れ

認識処理は図2のようなフローに従って行った。まず、専用のカメラから取得された三次元画像を画像処理により、手領域のみの抽出を行う。処理装置から得られた検出画像をセグメンテーション部で、指文字として静止指文字部と動作指文字部に分ける。次に、この指文字部で検出された画像を認識処理部に送る。認識処理部では判別基準を計算し、あらかじめ登録した各指文字データと判別基準値とのデータベース照合（パターン照合）を行い、どの指文字かを最終判定する。

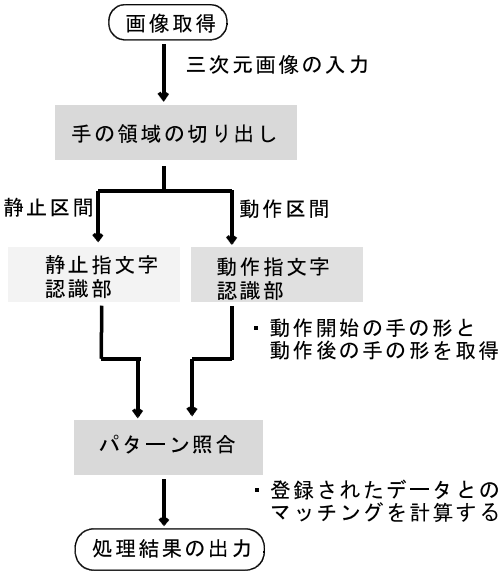


図2 認識処理の流れ

2-3 実験条件

対象指文字は50音46文字（静止指文字41文字＋動作指文字5文字）。被験者は手話の経験のある1名と未経験者2名。被験者には、「あ」～「ん」までの指文字を連続動作で実施してもらった。

3 結果及び考察

3-1 指文字の認識について

処理結果の一例を図3に示す。ディスプレイ上には、予め用意した「指文字画像データ」とキャプチャーしたリアルタイム画像とが「指文字（あ）」一致した結果を、文字表記として表している。被験者は筆者自身が行ったもので、映し出された2枚画像の指の大きさには若干の相違があるが、正規化により認識の精度をあげることができた。現在、指文字46字の内、静止指文字41文字、動作指文字5文字全てリアルタイム認識が可能となった。未経験者の認識結果に関しては、指の開き具合、カメラへの手の向ける角度のズレ等により多少の誤認識は見られたが、指文字の練習を繰り返すごとに結果は良好となり、手話経験者と同じような結果が得られた。

今回、認識処理において画像間（取得画像とデータベース画像）の類似度をX,Y,Z上での重心点補正、さらに面積値を用いた正規化を行った後パ

ターン照合を行った結果、より精度の高い認識率
が得られた。また、実際の状況を想定しPC上には
文字による表記と音声発話による機能を持たせた。
この音声発話機能に関しては、指文字（静止
指文字、動作指文字）間の連続した動作（文章を
伴うもの）も発話可能のように音素を連続発話と
して表現した。この一連のシステムにより指文字
話者から健常者への伝達が実現した。

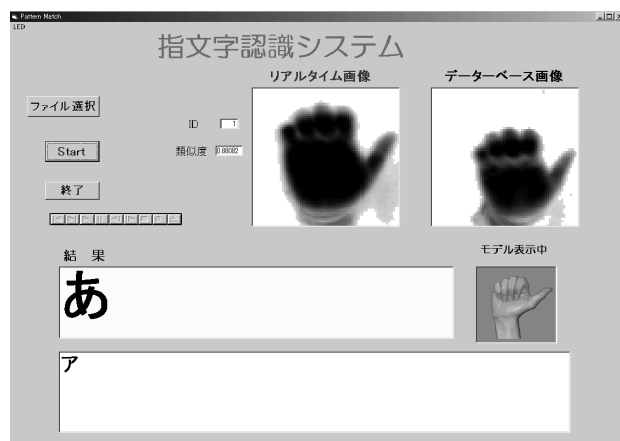


図3 指文字認識結果

3-2 動きのある指文字、簡単な挨拶への展開

動作を伴う指文字については、開始はじめての手の形と終了時の手の形をデータベースとの照合により判別することが可能となったが、この手法では語彙も限られる。手話動作への展開を考える上で、手の動作軌跡を捉えることが必須となり、移動ベクトル分布のパターンを照合させる手法で検討を行った。この結果、自己紹介、簡単な挨拶等の認識が一部可能となった。

3-3 上半身画像処理についての考察

上記システムでは、システム構成上、近距離しか対応できず、上半身すべてを使う手話動作を捉えることは困難である。そこで、通常カメラで人物の上半身部の画像を得て背景除去を行った後、前述の手法による認識処理を適用することで、両手での手話動作等まで対応できる手法を検討した。320×240画素のカラー画像を処理するソフトを試作し、以下の手法で手の動作部分のみの抽出実験を行った。

手法

(1) 動画より1/10秒程度の画像の切り出し

- (2) 近接画像による差分処理&孤立点除去
- (3) エッジ抽出&孤立点除去
- (4) 線画処理による外形輪郭の抽出
- (5) 元画像より人物のみ画像復元
- (6) 差分処理&(5)の画像とのAND処理
- (7) 肌色のみ抽出



図4 (2)の処理（差分&孤立点処理状況）

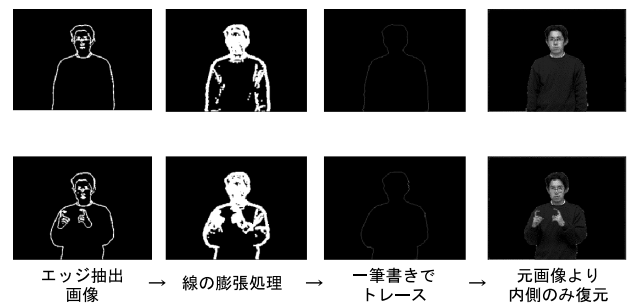


図5 (3)～(5)の処理

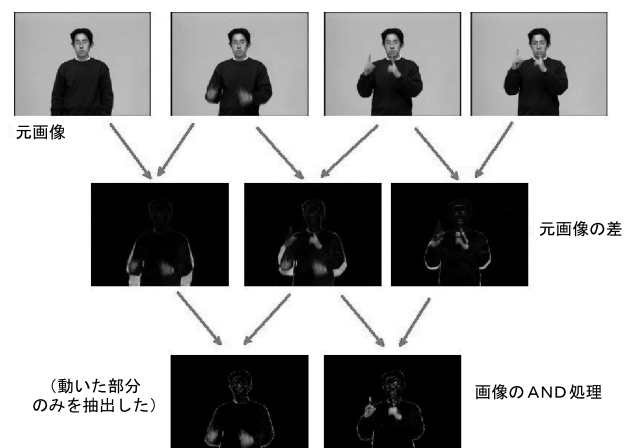


図6 (6)の処理

これらの処理を連続して行うことで、前述の動きのある指文字と同格に扱う画像を得られたが、処理時間が大幅にかかり、リアルタイム処理が不可能なこと、わずかな画像の差で最終画像が変化し、パターン照合しにくく誤認識しやすいこととなるため、そのままの適用は難しいことが分かった。

技法の単純化、処理速度の向上、最終画像安定のための検討が不可欠といえる。

4 まとめ

- 1) 指文字の認識において、手袋（センサ付き）等を装着しない非接触による認識が可能であることを示した。
- 2) パターンマッチング等を用いた画像処理による手法が有効であることを確認した。
- 3) 対象指文字50音46文字全て認識が可能となり、認識率95%と良好な結果が得られた。

5 参考文献

- 1) 肥後一彦, 森田秀樹, 手話認識技術の研究 (第1報), 宮崎県科学技術交流会 (2000)
- 2) 肥後一彦, 布施泰史, 手話認識技術の研究 (第2報), 宮崎県工業技術センター成果発表会 (2001)
- 3) 手話教室入門: 財団法人全日本ろうあ連盟
- 4) 米川明彦 他: “日本語手話辞典” 日本手話研究所, 財団法人全日本ろうあ連盟 (1997)
- 5) 沼崎俊一, 森下明, 梅木直子, 土井美和子: “ジェスチャ入力に適した画像入力装置の提案とその3次元情報検出性能の検討” 情報処理学会論文誌, Vol.41 (2000)
- 6) 布施泰史, 巢山昭文, 第18回リハ工学カンファレンス講演論文集, 139 (2003)
- 7) 布施泰史, 巢山昭文, 手話認識技術の研究, 第5回福祉技術シンポジウム, 31 (2003)