

2つのシルエット画像を用いた指文字認識手法とカメラ配置の検討

木元 文章 舟川 政博 平山 亮

金沢工業大学

1. はじめに

ろう者と健常者のコミュニケーション手段として欠かせない技法の 1 つに指文字がある。これは、固有名詞や手話にない単語の表現が可能で、手話において使用される局面が多々ある。そのため、コミュニケーション支援として指文字の認識システムが求められている。最近では、1 台のカメラで撮影し、カメラの性能に低依存なシルエット画像から特徴を抽出して、認識する研究がある[1][2]。しかし、1 方向のシルエット画像だけでは類似した形状の指文字は判別できず、高い精度の認識は困難であった。

本研究は、2 つのシルエット画像を用いて指文字を高い精度で認識するために、2 台のカメラを用いる。著者らは今まで、静止指文字を研究対象としてきた[3]。しかし、本研究では、濁音、半濁音、拗音の動作指文字にも対象を拡大し評価を行う。また、カメラの配置パターンを増やすことで、どの配置が指文字認識に有用であるかをその認識精度から検討した。

2. 認識対象と撮影方法

2.1. 認識対象とする指文字

静止指文字は 41 種類、動作指文字は濁音の 20 種類、半濁音の 5 種類、拗音の 9 種類の計 34 種類を対象とする。動作指文字の「の」、「も」、「り」、「ん」と長音は対象としない。

2.2. 撮影を行うカメラ配置

図 1 に撮影を行うカメラ配置を示す。カメラは撮影対象の右側の配置 1 を 0° とした場合に、配置 9 の 180° まで 22.5° の間隔で均等な位置に配置する。カメラから撮影対象までの距離は約 60cm とする。すべての組合せ 36 パターン ($= {}_9C_2$) について認識精度を評価する。以降、配置 2 と 1 を用いる場合を[2,1]と表すことにする。

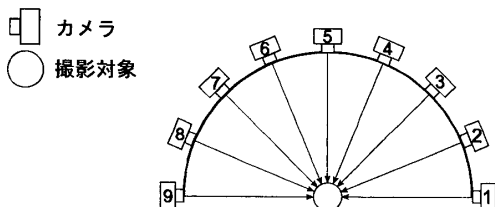
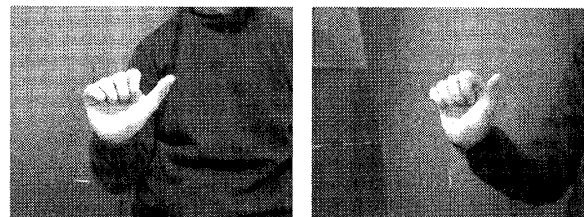


図 1. カメラを配置する位置

A study of recognition method and camera position for finger spelling using two silhouette images
Fumiaki Kimoto, Masahiro Funakawa, Makoto J. Hirayama
Kanazawa Institute of Technology

2.3. 指文字の撮影

指文字を撮影するにあたり、シルエット画像を容易に得るため、図 2 のように背景と着衣に青色を用いる。



配置 5

配置 7

図 2. 撮影画像

3. 静止指文字の認識方法

3.1. シルエット画像からの特徴量抽出

1 つのシルエット画像から計 638 種類の以下の特徴量を抽出する。

- ・円形度 (1 種類)
- ・縦横比 (1 種類)
- ・シルエットと外接長方形との面積比 (1 種類)

さらに、以下の特徴量は外接長方形を 25×25 画素に縮小したシルエット画像から抽出する。

- ・縦を 5 分割し、その矩形内にあるシルエットと矩形との面積比 (5 種類)
- ・横を 5 分割し、その矩形内にあるシルエットと矩形との面積比 (5 種類)
- ・縮小シルエット画像の画素値 (625 種類)

3.2. ニューラルネットワークによる認識方法

ニューラルネットワークを用いて、静止指文字 41 種類の認識を行う。図 3 のように入力層のニューロン数は、2 つのシルエット画像から特徴量を抽出するため、1276 個とし、中間層は 300 個、出力層は 41 個とする。学習データは被験者 10 人から撮影した指文字画像を使用し、学習にはバックプロパゲーション法を用いる。

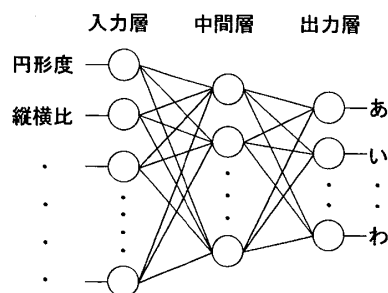


図 3. 認識ニューラルネットワーク

4. 動作指文字の認識方法

4.1. 重心の座標による閾値の決定

動作指文字は静止指文字の形状を保持しながらの移動であり、濁音は右、半濁音は上、拗音は手前に直線動作をさせて表現する。そこで、動作したシルエットの開始と終了時の重心座標を求め、縦方向と横方向に移動した距離から認識を行う。その際の閾値は、配置パターンで静止指文字を 41 種類行った場合に、シルエットの重心間の縦方向と横方向の最大画素数である。図 4 のように配置パターンにつき 4 つの閾値を求めることができる。

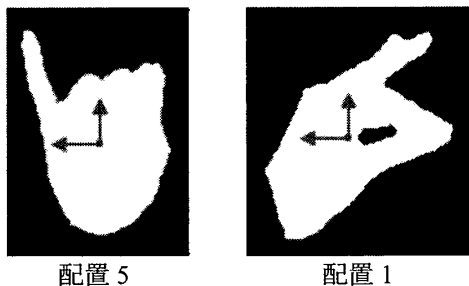


図 4. 閾値の例

4.2. 配置パターンによる認識方法

指文字が動作指文字か否かを判定する閾値は、カメラの配置パターンごとに異なる。以下に各動作指文字の認識に用いる閾値を示す。

- ・濁音：配置 5 に最も近い配置における横方向
- ・半濁音：2 つの配置における縦方向の平均
- ・拗音：配置 1 または 9 に最も近い配置における横方向

5. 認識実験

5.1. 静止指文字の認識実験

未学習データとして、学習データに用いた被験者とは異なる被験者 8 人の静止指文字 41 種類を用いる。その結果の上位 5 配置パターンを表 1 に示す。これにより、配置[9,6]で最も高い認識精度 88.11% が得られた。

表 1. 静止指文字の認識精度の上位 5 配置

認識精度の順位	配置パターン	認識精度(%)
1	[9,6]	88.11
2	[8,6] [6,4]	87.80
4	[7,4]	87.20
5	[6,2]	86.59

認識精度の上位 5 配置の内 4 つに配置 6 が含まれた。正面の配置 5 のカメラに向けて指文字を行うと手首に負担がかかる。そのため被験者が負担のかからない、配置 6 が正面になったと考えられる。なお、1 つのシルエット画像だけで同様の特徴量による認識を行った場合も配置 6 の認識精度が最も高い 81.10% となり、次いで配置 5 が 78.35% という

結果が得られている。

また、類似した形状の指文字である‘い’と‘ち’については、1 つのシルエット画像では約 50% の認識精度に対して、2 つのシルエット画像により 100% の認識を行うことができた。

5.2. 動作指文字の認識実験

静止指文字画像の認識結果から、最も高い認識精度の配置[9,6]により本実験を行った。閾値は、濁音が配置 6 で横方向に 59 画素、半濁音は配置 6 と 9 が縦方向に 44 画素、拗音は配置 9 で横方向に 47 画素とした。実験データは被験者 8 人の動作指文字 34 種類である。表 2 にそれぞれの認識結果を示す。

表 2. 動作指文字の認識結果

動作指文字	認識精度(%)
濁音	88.75
半濁音	95.00
拗音	90.28

撮影開始時の静止指文字で認識に誤りがあったが、濁音、半濁音、拗音の動作 3 種類の区別については 100% の識別ができた。

6. まとめ

本研究では、2 つのシルエット画像を用いて、特徴量抽出を行うことで、類似した静止指文字の誤認識を減らし、高い精度で静止指文字の認識を行うことができた。また、重心座標による動作指文字の認識も行うことができた。異なる被験者でも高い精度で認識できることから、不特定の被験者にも対応できる手法だといえる。さらに、カメラの配置パターンの 1 つに配置 6 を用いることが指文字認識に有用であるとわかった。

今後の課題として、‘の’、‘も’、‘り’、‘ん’と長音の動作指文字の認識と連続した指文字の認識をするためにフレームの抽出方法について検討する必要がある。

謝辞

本研究の一部は、文部科学省のハイテク事業による私学助成を得て行われた。

参考文献

- [1] 舟川政博, 平山亮, “指文字画像からの手指形状特徴量抽出方法の検討,” 第 5 回情報科学フォーラム 2006, pp. 561-562, 2006.
- [2] 岩崎聡, 朝倉俊行, 広瀬謙治, “ニューラルネットワークを用いた指文字認識,” 日本機械学会講演論文集, Vol. 2002, pp. 239-240, 2002.
- [3] 木元文章, 舟川政博, “2 方向のシルエット画像による指文字認識精度の向上と有用なカメラ配置の検討,” 電気関係学会北陸支部連合大会, F-36, 2008.