

指関節の距離と角度による指文字認識手法の検討

山口 陽功[†]増田 健志[†]澤野 弘明[†][†] 愛知工業大学

1 はじめに

聴覚障害者の会話において手話や指文字が日常的な場面で使われている。指文字は日本語の仮名文字それぞれに対応した手の形や動きにより一文字ずつ表現される手話言語の一つである。指文字を知らない人や経験が少ない人は、指文字を会話の中で認識し、理解することは難しいため、本研究では指文字の自動認識手法に着目する。指文字を自動認識するために、小松らがセンサグローブを用いた認識手法 [1] を提案している。この手法ではセンサグローブで5本の指関節の屈折や指の位置関係を取得して指文字を認識している。センサグローブを扱うことにより認識精度は高いが、特徴的なセンサを扱うことで日常的に普及することは難しい。そこで、センサグローブを用いない比較的普及しやすい Web カメラを扱った指文字認識手法に着目する。

渡辺らは、機械学習を用いて Web カメラから指文字を自動認識する手法 [2] を提案している。この手法では Web カメラと手の距離や指文字を自動認識する処理において 21 種類の指関節の三次元座標を指文字ごとにすべて扱う。渡辺らの手法は指文字の認識精度が実験数全体の 90% であり、指文字の認識精度の向上が課題として挙げられる。そこで本研究では機械学習に使用するデータを指関節の三次元座標から、距離と角度に変更し、渡辺らの手法に対して認識精度の向上を実現する指文字認識手法を提案する。本稿では、提案手法に対して評価実験を行い、今後の課題について述べる。

2 提案手法

本節では、機械学習を用いて指文字を 1 文字ずつ認識する手法について述べる。認識手法で扱う指文字は、静止した状態で表現する指文字に限定する。Web カメラから得た画像に対して、MediaPipe*により指関節の骨格座標を三次元座標で取得する。取得された指関節の骨格座標を図 1 に示す。空間の位置に依存させないために、基準点を手首に位置する図 1 の点 0 に設定して、21 個の指関節の骨格座標の相対座標を扱う。相対的な指関節の骨格座標から、距離と角度を 1 フレーム



図 1: 検出された指関節の骨格座標

ごとに計算して、計算された距離と角度から機械学習用の学習データを構築する。構築する機械学習データの概要について表 1 に示す。表 1 のデータ番号 1-5 は、基準点から指先の距離を示し、それぞれの指ごとに番号を付与する。データ番号 6-10 についても指ごとに番号が付与されており、基準点から指先の距離と基準点から第一関節の距離の差を示す。データ番号 11-14 は、親指の先を基準点としたそれぞれの指先までの距離を示す。データ番号 15-17 は、隣り合う指先同士の間の距離を示す。データ番号 18 は、基準点から点 9 の中指の付け根と x 軸との成す角を示す。データ番号 19, 20 は、基準点から点 4, 20 による線分と z 軸の成す角それぞれを示す。データ番号 21-25 は、点 9 と各指先による線分と x 軸との成す角を示す。データ番号 26-30 は、指の第二関節と指先の線分と x 軸との成すそれぞれの角を示す。取得したデータを教師あり機械学習 Support Vector Machine により学習モデルを作成して認識する。

3 実験と考察

本節では提案手法を用いた評価実験について述べる。本実験では、指文字が表現されている動画を 1 フレームごとの画像を、学習データとテストデータに分けて実験を行う。実験に使用する学習データの一部には、静止した状態で表現する指文字 41 種類を指文字の知識のない 1 人が真似た画像を用いた。静止した状態で表現

A Study on a Fingerspelling Recognition Method with Finger Joint Distances and Angles

[†]Aichi Institute of Technology

*MediaPipe: <https://google.github.io/mediapipe/>

表 1: 機械学習データ概要

データ番号	データ概要
1-5	基準点から 5 本の指先の座標
6-10	基準点から 5 本の指先と第一関節の差
11-14	親指の先と残りの指先の距離
15-17	それぞれの指先同士の距離
18	基準点と図 1 の点 9 と x 軸との成す角
19	基準点と図 1 の点 4 と z 軸との成す角
20	基準点と図 1 の点 20 と z 軸との成す角
21-25	図 1 の点 9 と指先の線分と x 軸との成す角
26-30	指の第二関節と指先の線分と x 軸との成す角



図 2: 入力画像 手話の指文字 50 音「あ〜ん」
<https://youtu.be/IRIUPHpAzmQ>

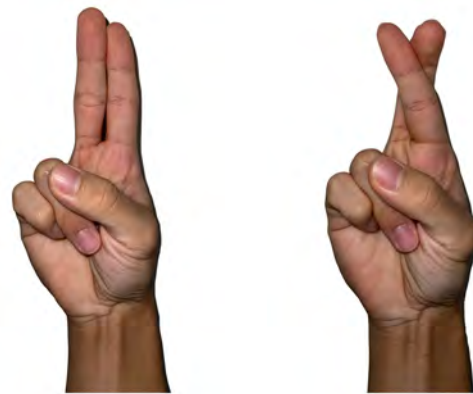
する指文字 41 種類に対してそれぞれ 200 種類の指関節間の距離や角度のデータを取得して学習モデルを作成する。動画配信サイトに投稿されている「あ」から「ん」までの指文字 50 音を表現している動画*を用いて、指文字が検出されたフレームの合計 3,690 フレームをテストデータとする。ここで評価実験の入力画像は図 2 に示すように、指文字と日本語の仮名文字が提示されており、YouTube で一般公開されている。また指文字を切り替えるタイミングのフレームと手の形と手の動きで表現される動く指文字は評価実験の対象外とする。

提案手法により認識された文字と目視確認による指文字が示した文字が一致した場合を正検出、一致しない場合を誤検出とする。実験結果を表 1 に示す。実験の結果、全体の正検出は 99.0%、誤検出は 1.0%であった。また、誤検出のうち図 3 に示す「う」「ら」の文字が誤検出の中で多く確認された。「う」「ら」の文字の表現には、人差し指と中指の奥行きにおける前後の相対的な位置関係が重要であり、2 本の指関節の前後の相対的な位置関係を正しく判別しきれない場合に誤検出が確認された。指関節の前後の相対的な位置関係に関して時系列前後のフレームの位置関係を考慮することで誤検出を削減できると考えられる。渡辺らの手法 [2] では静止した状態で表現する指文字 41 種類に対して、

*YouTube: 手話の指文字 50 音「あ〜ん」: <https://youtu.be/IRIUPHpAzmQ>

表 2: 評価実験の結果 (正検出率が低い順)

指文字	正検出 (%)	誤検出 (%)	計
う	74 (82.6)	16 (17.4)	90
ら	82 (91.1)	8 (8.9)	90
へ	85 (94.4)	5 (5.6)	90
ふ	88 (97.8)	2 (2.2)	90
す	87 (96.7)	3 (3.3)	90
せ	89 (98.9)	1 (1.1)	90
た	89 (98.9)	1 (1.1)	90
み	89 (98.9)	1 (1.1)	90
その他	2,970 (100)	0 (0)	2,970
計	3,653 (99.0)	37 (1.0)	3,690



(a) う (b) ら

図 3: 誤検出の多い文字

認識精度が 90%の精度だった結果に対し、今回の提案手法を用いて構築した学習データは認識精度が 99.0%になり直接的な比較はできないが精度が 9%増加した。

4 おわりに

本稿では、指関節の距離と角度による指文字認識手法について述べた。評価実験では、全体の正検出が 99.0% 誤検出数が 1%であった。渡辺らの手法 [2] と直接的な比較はできないが認識精度が 9%増加した。表現する形が似ている文字に対して、指関節のカメラに対する前後関係について正しい判別ができないことで誤検出となった。今後の課題として、誤検出の対応を検討、動いた状態で表現する指文字の判別方法が挙げられる。

参考文献

- [1] 小松周生, 白石優旗: “深層学習を用いたセンサグループによる指文字認識の検討”, 情処研報, Vol. 2018-AAC-6, No. 4, pp. 1-5 (2018)
- [2] 渡邊聡, 五十嵐悠紀: “Web カメラを用いた指文字自動認識システム”, WISS 2021, No. 2-B04 (2021)