

## 深層学習による多次元時系列データを用いた連続指文字認識手法の検討

白石優旗<sup>1)</sup>, 土屋智彦<sup>2)</sup>, 加藤伸子<sup>1)</sup>, 米山文雄<sup>1)</sup>, 設楽明寿<sup>3)</sup>

筑波技術大学 産業技術学部 産業情報学科<sup>1)</sup>

筑波技術大学大学院 技術科学研究科 産業技術学専攻<sup>2)</sup>

筑波大学大学院 図書館情報メディア研究科 図書館情報メディア専攻<sup>3)</sup>

キーワード: 手話言語, センサグローブ, 機械学習, 畳み込みニューラルネットワーク

### 1. はじめに

昨今のダイバーシティ推進により, ろう・難聴者と聴者が共に生活する機会のさらなる増加が期待され, 両者の円滑なコミュニケーションの実現は急務である。したがって, 代替手段を用いて情報獲得の支援を行う「情報保障」は重要である。これまでの情報保障は, 聴者の音声を文字(手話)に変換(通訳)して, ろう・難聴者に提示することが主流であった。しかしながら, コミュニケーションは双方向的であることが本質であるため, ろう・難聴者の発話(手話・指文字)を聴者に文字(音声)で提示することも必要となる。

1対1の局面では, 双方が文字を使う(筆談)等の工夫でコミュニケーション自体は可能であるものの, そのような対応がそもそも困難な状況がある。例えば, 講演・会議・グループでの話し合い等, ろう・難聴者が多数の聴者に対し手話を用いて発話するケースである。本研究ではこれらの状況を第一義的に想定する。

本研究では, 高精度の手話認識の実現により, 本課題の解決を目指す。本稿では, 動的指文字(手指を動かしながら提示する指文字)を含む全指文字(76文字)に対し, それら指文字が連続的に提示された状態(連続指文字)における認識手法について検討する。具体的には, センサグローブの改良, データ収集実験に最適な単語の選定, 実際のデータ収集実験, 並びに, 連続指文字認識への足がかりとして, 単一(つまり連続的に提示しない)動的指文字の認識精度の向上のための認識アルゴリズムの改良について述べる。

認識手法には, 近年高い認識精度の報告が多数なされている深層学習を用いる。また, センサグローブを採用することにより, カメラと異なり環境光やオクルージョン(手と手の重なり)の影響を受けずに, またカメラの存在を意識せずに, 聴講者に対して発話することができる(図1)。開発システムの詳細については, 参考文献[1, 2]を参照されたし。

### 2. 関連研究

現在の手話認識研究は, カメラ(RGBカメラ・深度センサ)を用いたものが主流であるものの, 約1,000単語に対する単語誤り率22.9%(訓練データと同一手話者)であり, 実用的なレベルで「文」を認識するまでには至っていない[3]。センサグローブとカメラを共に利用したものについて Ongらの報告[4]があるが, 隠れマルコフモデルを活用した認識により, 英数字の指文字37文字と手話30単語の識別率80%程度に留まっている。センサグローブのみを用いた手法について Mummadiらの報告[5]があるが, 様々な機械学習による手法(深層学習は含まれていない)を比較・検討することで識別率92%を達成したと報告されているものの, 識別対象はフランス手話の内, 手指を動かさずに提示する静的指文字22文字に限定されている。

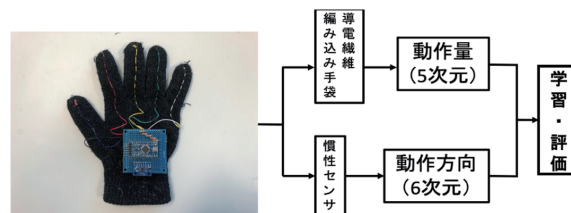


図1 システム構成

### 3. 連続指文字データ採取実験

連続指文字データ採取実験に先立ち, 最初にセンサグローブの改良を行った。単一文字認識データ採取実験[2]のデータを解析したところ, センサや制御基板の物理的な不安定さに起因するノイズが混入したデータを一部で確認した。そこで, それらをセンサグローブに確実に固定できるように3Dプリンタを用いて治具を作成した。同時に, 基盤回路の見直しを行い, 予備のセンサグローブを含めて複数個制作した(図2)。

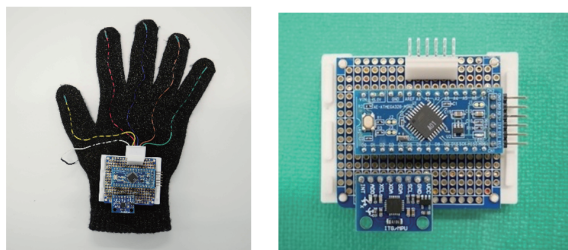


図2 改良されたセンサグローブ

機械学習においては、使用する学習データが非常に重要となる [3]。そこで、実験で採取する手話単語（連続指文字での提示）について参考文献 [6] を参考に選定した。今回の目的は単語の認識ではなく、単語に含まれる指文字の認識のため、日本語の指文字（76 文字）がすべて含まれるようにし、単語の文字数も適切な長さに調整した（2 文字から 5 文字）。その際、連続的な指文字の提示動作に起因する他の指文字への認識ミスの発生のしやすさを考慮し、64 単語を選定した。同時に、想定される様々な認識ミスを類型化し、整理した。これにより、今回選定した手話単語が、今後の連続指文字データ採取実験の指針になることが期待される。

その後、実際に連続指文字データ採取実験を行った。具体的には、日常的に手話を使用している、ろう・難聴の実験協力者 40 名（20～24 歳）に対し、先に説明した 64 単語を対象としたデータ採取実験を行った。その際、センサグローブを着用し、指文字の動作を一単語ずつ 10 秒以内で行ってもらった（各単語最大 5 文字のため、十分にゆっくりと提示可能である）。

なお、サンプリングレートは 200 個 / 秒で、指の動作量（5 次元）と手の動作方向（加速度 3 次元、角速度 3 次元）について計測データ（多次元時系列データ）を取得した。また、データの収集と同時に手動でラベリングを行った。この一連の動作を 5 回繰り返した。

なお、本実験は筑波技術大学倫理審査委員会の承認を得て実施した。現在、採取したデータの解析と認識アルゴリズムの検討を行っており、実験の詳細と識別実験の結果は、本年度中に学術論文に投稿予定である。

#### 4. 単一指文字識別アルゴリズムの改良

動的指文字を含む 76 文字（20 名×5 回分のデータ）に対して平均識別率 70.0%（20-CV 法）の結果を得ていた [2] が、十分な精度とは言えない。そこで、識別精度向上のため、データクリーニング、動的指文字における動作範囲の特定、データ不足解消のためのデータ拡張の導入、認識深層学習ネットワークの改良を行って認識実験を行った。結果については、本年度中に学術論文に投稿予定である。

#### 5. まとめと今後の課題

本研究では、深層学習を用いて、動的指文字を含む全指文字に対し、連続指文字の認識手法についての検討を行った。具体的には、データ収集実験に最適な単語の選定、データ収集実験、単一動的指文字の認識精度向上のための解析を行い、次年度の認識システム開発の足がかりとした。今後は、センサグローブの改良、前処理の改良を含む識別アルゴリズムの改良、並びに、学習データの追加収集を行い、識別率の向上と連続的に提示された指文字の認識に取り組む予定である。

#### 謝辞

本研究は筑波技術大学教育研究等高度化推進事業並びに JSPS 科研費 19K11411 の助成を受けたものです

#### 参考文献

- [1] 土屋智彦, 白石優旗, 深層学習を用いたセンサグローブによる指文字認識の改良. 情報処理学会アクセシビリティ研究会 (IPSJ SIG AAC) 第 9 回研究会; 2019-3.
- [2] T. Tsuchiya, et. al., Sensor Glove Approach for Japanese Fingerspelling Recognition System Using Convolutional Neural Networks, Proc. of the 13th Int. Conf. on Advances in Computer-Human Interactions (ACHI 2020), pp.152-157, Valencia, Spain, Mar. 2020.
- [3] D. Bragg, et. al., Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective, pp.16-31, ASSETS '19, 2019.
- [4] C. Ong, et. al., Sign-Language Recognition Through Gesture & Movement Analysis (SIGMA), In: Billingsley J., Brett P. (eds) Mechatronics and Machine Vision in Practice 3. Springer, Cham, 2018.
- [5] C. K. Mummadi, et. al., Real-time Embedded Recognition of Sign Language Alphabet Fingerspelling in an IMU-Based Glove, Proc. iWOAR 2017, Rostock, Germany, Sep. 2017.
- [6] 風間康介他, 言語情報を用いた Kinect による連続指文字認識手法の検討. 信学技報 WIT2017-74, 2018-3.