

配信先：大阪科学・大学記者クラブ、文部科学記者会、科学記者会

2024 年 12 月 19 日

大阪公立大学

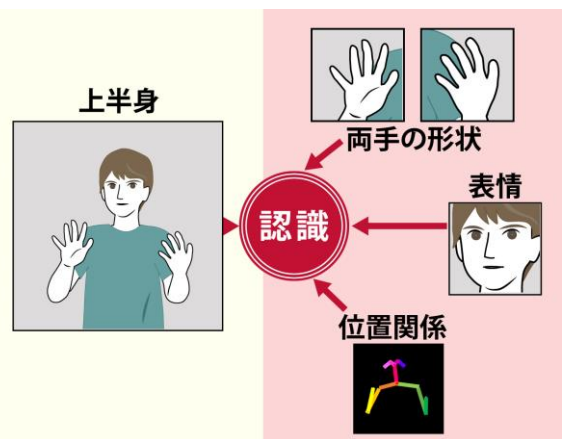
AI による手話単語の新しい認識手法を提案

<ポイント>

- ◇AI を用いて、動画から手話単語を認識する研究が活発に行われており、さまざまな手法が開発されている。
- ◇手話における上半身の大まかな動きの情報に、手や顔の表情などの局所情報と人物の骨格情報を加えることで、手話単語の認識精度が従来手法と比較して 10～15% 向上。

<概要>

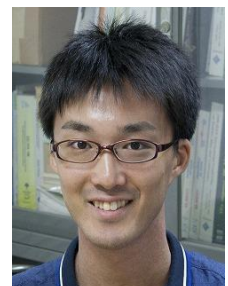
手話は、手や腕の動き、手指の形、顔の表情などの情報で単語を作り、その単語を組み合わせで会話をする視覚言語です。手話単語の多さが手話の習得を難しくしており、AI を用いて動画から手話単語を自動で認識する研究が活発に行われています。これまでの研究で使用されていた手法では、手話を大まかな動きの情報として捉えていました。しかし手話は、手指の形の細かな違いや、手と体の位置関係によっても意味が異なるため、従来の手法では手話単語を高精度に認識できないという問題がありました。



大阪公立大学大学院情報学研究科の井上 勝文准教授、岩村 雅一准教授、大阪公立大学 研究推進機構 協創研究センター 文書解析・知識科学研究所の Partha Pratim Roy 客員研究員らの研究グループは、アメリカの手話単語の AI を用いた認識手法について研究を進めています。本研究では、従来の認識手法である上半身の大まかな動きの情報に、手や顔の表情などの局所情報と、手と体の位置関係を表す骨格情報を加えました。これにより、手話単語の認識精度が従来手法と比較して 10～15% 向上しました。

本研究成果は、2024 年 11 月 11 日に国際学術誌「IEEE Access」に掲載されました。

健聴者の多くが手話を未習得なため、聴覚障害者の方々とのコミュニケーションにおいて、ある種の障壁が存在しています。本研究では、この障壁を取り除く第一歩として、手話単語の認識精度を向上させることに成功しました。この技術が聴覚障害者の方々とのコミュニケーション活性化に繋がる一助となることを期待します。



井上 勝文准教授

<研究の背景>

聴覚障害者とコミュニケーションを取る方法の一つとして、手話があります。手話は、手や腕の動き、表情、手の形状、手と体の位置関係などの情報を組み合わせることで表現される単語（手話単語）を、さらに組み合わせることで会話を実現する視覚言語の一種です。世界各国独自の手話が体系化されており、どの国の手話も数千個以上の手話単語が存在しています。習得すべき手話単語の多さが、手話の習得を難しくしている要因の一つと考えられており、手話単語を表現する話者を撮影した動画から、AI を用いて手話単語を自動で認識する研究が盛んに行われています。

これまでの手話単語認識に関する研究では、手話を「お辞儀をする」「手を振る」などの一般的な動作の一種と見なし、話者の大まかな動きの情報を捉える手法を用いてきました。しかし実際には、手の形の細かな違いや、手と体の位置関係の違いによって意味が異なるため、大まかな動きの情報だけでは手話単語を高精度に認識できないという問題がありました。

<研究の内容>

本研究では、別の研究グループによる顔認識研究において、顔全体の情報に加えて目や口などの局所領域の情報も組み合わせることにより、顔認識精度が向上したという知見を参考にしました。本研究グループは、手話を表現している話者の上半身の大まかな動きの情報に、手や顔の表情などの局所情報と、手と体の位置関係として人物の骨格情報を加えることで、手話単語の認識精度を向上できるのではないかと考えました。これを実現するために、上半身と局所領域の情報には深層学習モデル I3D^{*1}、骨格情報には ST-GCN^{*2} を用いて、情報ごとに手話単語を認識した結果を統合しました。そして、アメリカの手話単語認識データセットで検証した結果、話者の大まかな動きの情報のみを用いた従来手法と比較して、認識精度を 10～15%向上できることが判明しました。

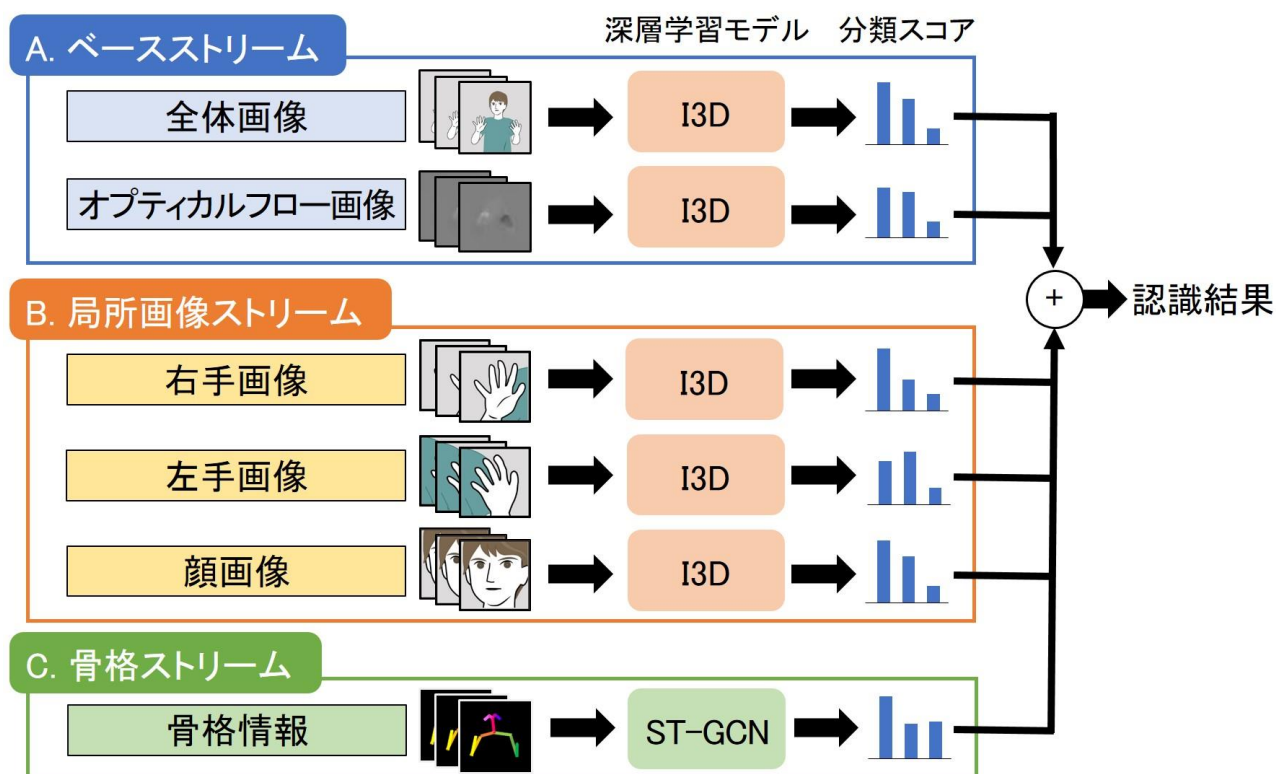


図 1 手話単語認識に用いた深層学習モデルの概略図

<期待される効果・今後の展開>

本研究によって手話単語認識の精度を高めることができました。また今回提案した手法は、どの国の手話にも適用できるため、さまざまな国で聴覚障害者の人々とのコミュニケーション活性化に繋がることが期待できます。今後は手話単語認識のさらなる精度向上や、本手法を用いた手話翻訳の構築と精度向上などの課題解決を目指します。

<資金情報>

本研究は、JSPS 科研費 19K12023 の支援を受けて実施されました。

<用語解説>

※1 I3D : 時間情報と空間情報を同時に取得し、行動認識に応用されている深層学習モデルの一種。

※2 ST-GCN : 点と点を線でつなげたようなもので表現されるグラフ情報を入力とする、深層学習モデルの一種。本研究では、人物の骨格情報を図 1 に示す棒人間のようなグラフ情報を入力として、人物行動を認識するために用いている。

<掲載誌情報>

【発表雑誌】 IEEE Access

【論文名】 Word-Level Sign Language Recognition With Multi-Stream Neural Networks Focusing on Local Regions and Skeletal Information

【著者】 Mizuki Maruyama, Shrey Singh, Katsufumi Inoue, Partha Pratim Roy, Masakazu Iwamura, and Michifumi Yoshioka

【掲載 URL】 <https://doi.org/10.1109/ACCESS.2024.3494878>

【研究内容に関する問い合わせ先】

大阪公立大学大学院情報学研究科
准教授 井上 勝文 (いのうえ かつふみ)
TEL : 072-254-7279
E-mail : inoue@omu.ac.jp

【報道に関する問い合わせ先】

大阪公立大学 広報課
担当 : 谷
TEL : 06-6605-3411
E-mail : koho-list@ml.omu.ac.jp