



HOKKAIDO UNIVERSITY

Title	Enhancing Recognition and Improved Processing Speed for Isolated Sign Language Recognition [an abstract of dissertation and a summary of dissertation review]
Author(s)	堀, 紀章
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第16000号
Issue Date	2024-03-25
Doc URL	https://hdl.handle.net/2115/91903
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Noriaki_Hori_abstract.pdf, 論文内容の要旨



学位論文内容の要旨

博士の専攻分野の名称 博士(情報科学) 氏名 堀 紀章

学位論文題名

Enhancing Recognition and Improved Processing Speed for Isolated Sign Language Recognition (手話単語認識における認識率の向上と処理速度の改善)

手話認識の研究は、2つに分けられ、ニュース番組などの同時手話通訳のように連続した手話を認識するものと、単語単位の手話を認識するもので、これを手話単語認識と呼んでいる。本研究では手話単語認識について、認識率の向上と処理速度の改善処理に注目する。手話単語認識の具体的な応用例として、会話型手話認識への応用や聴覚障がい者や手話学習者の支援などが挙げられる。例えば、手話を使わない人々が手話を理解するために、手話者を撮影し自動的に言葉や音声に変換することでスムーズにコミュニケーションができるのではないかと考えた。近年、さまざまな研究者が手話認識のモデルを提案し、手話認識の研究は非常に前進しているがまだ完全ではない。230程度の単語単位のデータセットでは98%を超え人間の認識レベルに近づきつつある。最近では、GPUの性能が高まり、リアルタイムで処理することができるようになってきた。本論文では、先行研究の手話単語認識モデル SAM-SLR(Skeleton Aware Multi-model Sign Language Recognition) をベースに、さらに認識率を向上させるために、モダリティの1つ Multi-stream で使われる Joint と Bone に、学習の際にエポックごとの推定結果を再利用する手法と、認識結果の Top-1 と Top-2 の評価値の差に着目した再評価による手法を提案する。また、手話認識の実用化のために認識率と評価値を維持しつつ処理速度の改善処理の手法を提案する。

先行研究 SAM-SLR は、トルコ語手話の手話単語データセット AUTSL(Ankara University Turkish Sign Language) を使った2021年のコンペティションで優勝したモデルで、97.64%と高い認識率を達成したモデルである。データセット AUTSL の認識率は99%弱まで高められる可能性があるとの研究者の報告もあり、さらに認識率を高める研究を行った。1つ目の提案は、学習の際にエポックごとの推定結果を再利用する手法で、Top-1 の評価値をクラスごとに集計し、最高値のクラスを採用し、最新のエポックの Top-1 がそのクラスの推定結果を採用するという手法を提案する。2つ目の提案は、すでに認識率が高められる既存モデルに対し、認識結果で Top-1 と Top-2 の評価値の差が小さいものの認識率が低いことに着目し、既存のモデルで使用されていない新たな手法で再評価する手法である。この手法は既存のモデルよりも認識率を高められる可能性があると考えた。処理速度の改善については、データセット AUTSL の認識率が人間のレベルに達しつつあり、実用化を実現するために、主に認識率の維持、現行の GPU で動作しない部分の修正、1つの動画から独立した4つのモダリティをシームレスにつなぎ、認識応答時間の短縮の提案を行った。

本論文は全5章で構成される。

第1章では、研究背景と目的を述べている。特に、先行研究の手話単語認識モデル SAM-SLR について、認識率の改善、処理速度の改善の実現のための課題について述べている。

第2章では、手話単語認識モデルについて関連研究について述べる。特に、本研究のベースとなる先行研究のモデル SAM-SLR について、詳しく述べている。

第3章では、手話単語認識モデル SAM-SLR をベースに、認識率向上の手法について述べる。認識

率の高いストリーム Joint と Bone に着目し, 学習の際にエポックごとの推定結果を再利用し, 推定結果の Top-1 の評価値をクラスごとに集計し, 最高値のクラスを採用し, Top-1 がそのクラスのもの を推定結果とするアルゴリズムを提案した. 2つのストリーム Joint, Bone のモダリティをこの手法 を適用して認識率を向上させ, さらに, この結果を Multi-stream, SAM-SLR モデルにも反映させる ことで認識率の向上が期待される.

第 4 章では, 手話単語認識モデル SAM-SLR をベースに, さらに認識率向上の手法について述べる. 再評価の実施の判断は, SAM-SLR の認識結果から Top-1 と Top-2 の評価値の差, 片手などの条件 で決定し, 目や鼻, 口, 顎, 頬などの顔パーツの位置で三角形メッシュを作り, 人差し指の停留し ている位置をそのメッシュ上で捉えることで, 異なる顔の形状でも顔パーツ基準の位置で比較する 手法である. この手法により, 口や鼻などに触れる手話認識をより正確に捉えられ, クラスごとの 全 Train データの位置, Test データの位置を比較し評価する. 認識率は先行研究と比べ 97.94% から 98.24% に向上した.

第 5 章では, 処理速度の改善の手法について述べる. SAM-SLR モデルでの認識率を維持させ, 認 識応答時間を短縮することを目的とする. まず, 1 つの動画から独立した 4 つのモダリティをシー ムレスにつなぎ, 認識処理をさせる. SAM-SLR モデルのモダリティ RGB-Flow で使われる Optical Flow が現在の GPU で動作しないため, 同等の結果が得られるようにコードを修正する必要がある. 次に, 認識応答時間を短縮するために, 処理の削減, 並列化, Python から C++ 化, 内部メモリの活用 などを行った. 認識率を維持させるために, 処理の削減結果としては, Optical Flow の処理ではオリ ジナルの手法と提案手法との差は, Optical 画像は見た目上はほぼ同じとなり, 数値上は多少の違い があるが, 軽微な範囲に収まった. また, 姿勢推定処理 MMPose では, オリジナルの手法と提案手法 ではほぼ同等の結果が得られた. 認識評価値は概ね同等の値となり, 認識率も同じ結果となった. 平均認識応答時間は約 4.4 秒から 0.72 秒に短縮し, シリアル処理したものと提案手法では, 約 6 倍高 速化された.

第 6 章においては, 手話単語認識の認識率向上と処理速度の改善について考察し, 本論文の結論 を述べている. 今後の GPU の発展により, 手話認識の認識率向上と処理速度の改善に関する研究に もたらした知見をまとめ, 今後の展開についても述べている.