



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
Citation	北海道大学. 博士(情報科学) 甲第16000号
Issue Date	2024-03-25
Doc URL	<a href="http://hdl.handle.net/2115/91903">http://hdl.handle.net/2115/91903</a>
Rights(URL)	<a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
Type	theses (doctoral - abstract and summary of review)
Additional Information	There are other files related to this item in HUSCAP. Check the above URL.
File Information	Noriaki_Hori_review.pdf (審査の要旨)



[Instructions for use](#)

## 学位論文審査の要旨

博士の専攻分野の名称 博士 (情報科学) 氏名 堀 紀章

審査担当者 主査 教授 山本 雅人  
副査 教授 野田 五十樹  
副査 教授 川村 秀憲  
副査 特任教授 小野 哲雄

### 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)

手話認識の研究は大きく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連続手話認識と、単語単位の手話を認識する手話単語認識がある。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。近年、さまざまな研究者が手話単語認識のモデルを提案し、手話単語認識の研究は進んでいるが未だ発展途上である。最近では、GPUの性能が高まり、画像の集まりである動画においてもリアルタイムで処理することができるようになってきている。本論文は、手話単語認識の認識率の向上と処理速度の向上を目的とし、先行研究の手話単語認識モデル SAM-SLR (Skeleton Aware Multi-model Sign Language Recognition) をベースに研究を行っている。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。本論文は、さらに認識率を高めるため、大きく2つの提案を行っている。1つ目の提案では、SAM-SLRのモダリティの1つ Multi-Stream で使われる Joint と Bone に対して、学習の際、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 におけるそのクラスの推定結果を採用するという手法の構築を試みている。2つ目の提案は、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の構築を試みている。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の構築を行った。

本論文は全6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、手話単語の認識モデルに関する先行研究の問題点を述べ、本論文で提案される認識率の改善、処理速度の改良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ており、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た認識率向上の手法を導入している。再利用の手法では、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である。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の向上を示した。

第4章では、手話単語認識モデル SAM-SLR をベースに、認識結果を再評価することによりさらに認識率向上の手法を導入している。すでに認識率が高い状態のモデルに対し、既存のモデルの認識率を損なわずに、さらに認識率を高める手法であり、再評価の実施の判断は、SAM-SLR の認識結果から Top-1 と Top-2 の評価値の差が小さいものの認識率が低い点に着目した手法である。この手法では、口や鼻や額などに触れる手話認識をより正確に捉えようとするもので、評価方法は、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人差し指の位置を比較する方法である。これによって、クラスごとの Train データのすべての位置と Test データの位置を比較し評価することで、既存のモデルと比べ認識率の向上が示されている。

第5章では、処理速度の改善の手法について述べる。SAM-SLR モデルでの認識結果を維持させ、認識応答時間を短縮することを目的としている。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ために、SAM-SLR モデルのモダリティ RGB-Flow で紹介されていた 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している。また、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、修正された Optical Flow の処理ではオリジナルの手法と提案手法との差は、軽微な範囲に収まり、高速化のために1枚の画像のみで姿勢推定処理 MMPose を行うなどの工夫によって認識率をほぼ保ったまま平均認識応答時間を約4.4秒から0.72秒に短縮している。

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、手話認識の認識率の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。

これを要するに、著者は、手話単語認識において、認識率の向上と処理速度の改善の手法の構築を行った。特に、認識率の向上を行うために、学習過程での評価結果を再利用する方法と評価値が拮抗するデータの評価について、顔パーツ基準による人差し指の相対的位置を比較する方法を提案した。これは手話単語認識における精度向上に有益な知見を得たものであり、情報科学の分野に対して貢献するところ大なるものがある。よって著者は、北海道大学博士(情報科学)の学位を授与される資格あるものと認める。