

骨格推定と機械学習を利用した日本指文字の分類

Classification of Signed Japanese Character with Pose Estimation and Machine Learning

小林大起[†] 石川孝明[‡] 渡辺裕[‡]
 Hiroki Kobayashi[†] Takaaki Ishikawa[‡] Hiroshi Watanabe[‡]

[†] 早稲田大学大学院基幹理工学研究科 [‡] 早稲田大学国際情報通信センター
[†] Graduate School of Fundamental Science and Engineering, Waseda University
[‡] Global Information and Telecommunication Institute, Waseda University

1. まえがき

近年、ネットワークの進歩により多様なユーザ間でのコミュニケーションを容易にする目的で、自然言語と手話との間の自動翻訳が研究されている。また、人体で最も複雑な骨格を持つ手の正確な識別は、手話翻訳に限らずコンピュータビジョンにおいて重視されている。しかし、現状では手話と日本語間のコーパスが不足しており、機械学習に用いるサンプル数が不十分である。

この原因として、健聴者における手話理解者が少ないことが挙げられる。一般に、手話習得には多大な労力が必要であるという問題がある。また、細部の相異しかない単語群が存在し、その分別は困難である。手話翻訳の実現には、一般的なカメラ画像から効率的に学習データを構築し、手話の可読性とコーパス量の両面を向上させる必要がある。

本研究では、手話翻訳技術の一例として日本語対応手話の指文字を OpenPose[1] により骨格推定し、対応する文字へ分類する手法を提案する。

2. 関連研究

手話翻訳技術として、デプスセンサなどを利用した特定単語の翻訳が行われている。しかし、このようなアプローチは導入に特殊な環境を要求する場合が多い。一方で画像の RGB 値を直接用いる形式では、背景等のノイズが推定精度を低下させる。

一方、OpenPose は映像内の人体の各関節位置の検出と、それらの適切な接合とを行うことで単一 RGB 画像から骨格検出を可能とする手法である。深層学習により身体各部位間の接合を行うことでボトムアップリアルタイム処理を実現している。

3. 提案手法

3.1. 静的指文字と動的指文字

日本語対応手話の指文字は文字を表す際に手のポーズだけで表現されるものと動作を伴って初めて意味を成すものとの二つに大別できる。そこで、本研究では前者を静的、後者を動的と定義する。動的指文字の多くは静的指文字と共通するポーズを動作の起点とする。そのため、本研究では静的指文字の骨格構造を分類した後、その結果と一定方向への移動の検出とを組み合わせることで動的指文字を検出する。

3.2. 概略

OpenPose によって RGB 画像から関節座標を得ることで入力ノイズを吸収する。その後、座標を相対化し、

(1)により隣接する関節 (x_i, y_i) , (x_{i+1}, y_{i+1}) 間の角度情報 θ_i を得ることでスケール差を吸収する。この値を多重化したベクトル X 内の各要素 X_j について(2)式によって正規化した後、SVMに入力することで静的指文字を分類する。

$$\theta_i = \text{Arccos} \frac{x_i x_{i+1} + y_i y_{i+1}}{\sqrt{x_i^2 + y_i^2} \sqrt{x_{i+1}^2 + y_{i+1}^2}} \quad (1)$$

$$X_j^{\text{norm}} = \frac{X_j - \min X}{\max X - \min X} \quad (2)$$

4. 実験

被験者を用いて作成したデータセット内の静的指文字を提案手法により分類した。映像は被験者正面上半身が遮蔽なく収まり、各文字 60~150 フレーム程度を撮影したものである。Leave-One-Out 交差検定での正解率の全文字の平均を表 1 に、1 文字ごとの平均 F 値、適合率、再現率を図 1 に示す。

表 1 静的指文字の平均正解率

正解率	0.63633
-----	---------

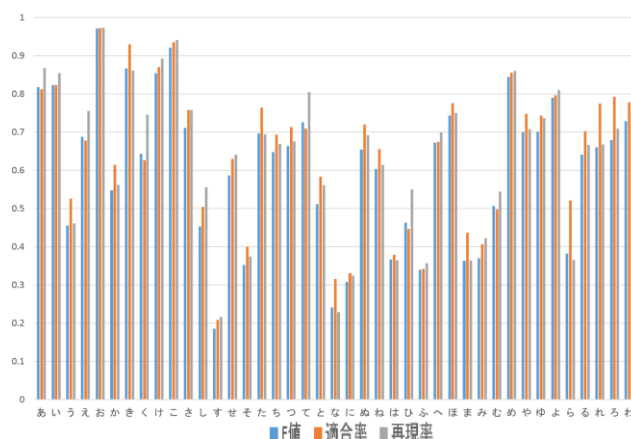


図 1 静的指文字各種別の平均データ

5. むすび

本研究では OpenPose による骨格情報から算出した角度ベクトルを利用して指文字を分類する手法を提案し、静的指文字に対して正解率 63.6%を得た。

参考文献

- [1] Z. Cao, T. Simon, S.E. Wei, and Y. Sheikh : “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, CVPR2017 Oral, <https://arxiv.org/abs/1611.08050>.