

MediaPipe Hands と CNN を用いた空中手書き文字認識

1. はじめに

近年、クラウドコンピューティングや SNS の普及により、ふと思いついたアイデアを保存する機会が増加している。従来の入力方法であるタッチパネルやキーボードは、スマートフォンでは画面スペースが限られるため文字の入力がしづらく、汚れや破損のリスクも存在する。また、音声認識は公共施設や騒音の多い環境では使用が困難である。そこで、新たな入力手法として空中手書き文字認識が提案されている。加速度を取得する装置やセンサ付きカメラなどの特殊な装置を使わない手法[1]も提案されているが、色判定の欠点として、光の反射や背景によるノイズの発生が挙げられている。また、筆記に色のついたデバイスが必要となる。そこで、本研究では MediaPipe Hands を用いて手の姿勢を推定し、カメラのみで動作可能な手書き文字入力の手法を提案する。

2. 提案手法

2.1 提案システムの概要

本研究ではディスプレイ上部にカメラを固定し、図 1 のようにカメラに向かって空中に文字の筆記を行う。筆記した文字部分を画像として切り出し、CNN を用いて入力された文字の推定を行う。MediaPipe Hands を用いて手の姿勢を推定し、取得した座標から文字の筆記動作やジェスチャー動作による文字の取り消しを行う。

2.2 手書き文字の筆記

本システムでは、文字の筆記は親指と人差し指でペンをつまむような動作で行う。親指から人差し指までの距離を計算し、指同士の開き具合で変化する距離で線描画の ON/OFF を切り替える。また、1 文字の書き終わり判定は描画が OFF の状態で 2 秒経過した場合とする。また、誤って入力したい文字とは別の文字が分類結果として出力された場合には、ピースの動作を行うことで文字の取り消しを行う。

2.3 CNN を用いた文字分類

筆記された文字を切り出す際、文字を分類しやすくするために白背景のキャンバスに黒線で筆記したデータに変換する。また、切り出した文字領域が正方形となるように領域の外側に均等に余白を加え、リサイズを行うことで文字の大きさによって画像データが大きく変わらないようにしている。そのため、今回は拗音の分類は行わず、ひらがな 46 文字に濁音、半濁音の 25 文字を加えた合計 71 文字で分類を行う。また、学習データには各文字 900 枚を用いた。文字画像を CNN に入力し、認識率が一番高い文字を分類結果とする。



図 1 MediaPipe Hands を用いた文字入力の様子

表 1 評価実験の結果

	被験者 A	被験者 B	被験者 C
認識率	100%	100%	97%
入力時間	215 秒	277 秒	259 秒

2.4 類似文字の分類

2.3 節の分類方法では線が 1 本増えただけのような形が似ている文字の誤認識が多かったため、画像全体ではなく、画像の上部分に注目した別の学習モデルを用いて分類結果の修正を行う。作成した追加モデルは「あ」と「め」、「さ」と「き」、「う」と「ろ」、「ね」と「ぬ」、「は」と「ほ」、「ち」と「ら」の 6 通りである。分類結果に類似文字が出力された場合には、入力された画像の上部分を使用し、別モデルでの分類を行うことで認識率の向上を図っている。

3. 評価実験

提案システムの評価実験を行った。学習済 1 名、未学習 2 名を被験者とし、46 文字を筆記し、認識率と全ての文字が出力されるまでの入力時間を計測した。実験結果を表 1 に示す。学習済の人物に対しては十分な認識率が得られ、未学習の人物に対しても 1 文字の誤認識と良い結果が得られた。また、システムに対する慣れや丁寧に筆記する意識などによって入力時間に差ができることが確認された。

4. おわりに

本研究では USB カメラと MediaPipe Hands を用いた空中手書き文字認識システムを提案した。評価実験では学習済の人物、未学習の人物どちらに対しても十分な認識率が得られることが確認できた。今後の課題として、拗音の分類やカタカナの分類などが挙げられる。

[参考文献]

- [1] 花房, 西田: 機械学習を用いた空中手書き文字認識の検討, 福井工業大学研究紀要, vol.49, pp.12-18, 2019