

MediaPipe Hands と CNN を用いた手軽な指文字認識システムの提案

1. はじめに

我が国における聴覚障害者の総数は約 34 万人とされている。その中には後天的に障害を患った中途失聴者も含まれ、そのような方々の多くは手話を習得していないことが多く、コミュニケーションをとることが困難であるため、手話や指文字を認識するシステムも提案されている。田中ら[1]は Kinect を用いた指文字分類手法を提案している。また、他の研究ではより高価な機材を用いることで認識率の向上を試みるものもある。これに対し、本研究では安価な USB カメラのみを用いて指文字の認識率の向上を試み、ユーザが指文字を習得するためのシステムを提案する。

2. 手軽な指文字認識システムの提案

2.1 システム概要

本システムでは USB カメラの画像を用いて指文字を認識し、ユーザが指文字を学習するシステムの開発を目的とする。まず、撮影した画像から手領域を切り出し、CNN を用いて指文字の清音を認識する。また、MediaPipe Hands で手の座標を取得して動作の有無を検知し、必要に応じて濁点・半濁点・拗音・促音の認識を行う。さらに、認識率が低い指文字は指先などの距離情報から分類する。

2.2 手領域画像の取得

CNN の学習に用いるための手領域画像の取得方法について述べる。まず、肌色抽出によって手領域のみを抽出し、背景を除去する。また、MediaPipe Hands で求めたランドマーク(図 1)の座標から x, y 座標の最大値, 最小値を取得し、手の中心座標と手の外接矩形の幅と高さを計算する。幅と高さの大きな方を基準に正方形を作成し、その 1.4 倍のサイズで手領域画像を切り出す(図 2)。

2.3 CNN を用いた清音分類

2.2 節の方法で取得した手領域画像を CNN に学習させ、動きの無い指文字を認識する。今回は清音 41 種類を認識対象として各 1,000 枚、総数 41,000 枚の学習データを用いた。分類時にはフレームごとに切り出した手領域を CNN に入力し、認識確率が一番高い指文字を分類結果とする。

2.4 認識率が低い指文字の判別方法

2.3 節の分類方法のみでは正面からの形が酷似している指文字の認識率が著しく低くなるという問題点があるため、MediaPipe Hands のランドマークの座標を用いて分類結果の修正を行う。指先同士や指先と手の中心までの距離情報を取得し、それらにしきい値を設定することで形が酷似している指文字の認識率の向上を図っている。

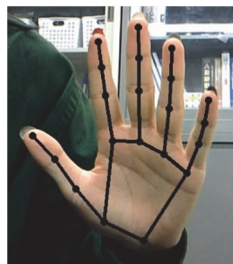


図 1 手のランドマーク



図 2 手領域の取得

表 1 実験結果

被験者	清音	清音以外
学習済の被験者	91.7%	80.0%
未学習の被験者	67.8%	60.0%

2.5 濁点・半濁点などの付与

清音の指文字の形を維持したまま、右方向に動かすと濁音、上方向に動かすと半濁音、手前方向に動かすと拗音・促音となる。なお、動作の途中で正面からの形が変化する清音の指文字が 4 種類存在するが本研究では対応していない。

濁音・半濁音の認識には MediaPipe Hands で取得した手の中心座標を用い、清音を認識したフレームの座標との差分がしきい値を超えると濁音・半濁音を付与する。手前方向の動きの検知には手領域面積の差分を利用する。

3. 評価実験

提案システムの認識性能を評価するため AM12 時頃、室内で自然光のもと学習済、未学習各一名ずつの被験者で実験を行った。清音 41 種類と清音以外(濁音・半濁音など) 33 種類を 1 種類につき 5 回ずつ行い、認識率の差を調べた。

実験結果を表 1 に示す。学習済の被験者ではかなり良い結果が得られたが、未学習の被験者ではあまり良い結果が得られなかった。これは指の太さや手の縦横比などの違いが誤認識の原因と思われる。また、光源の強さや色味、方向による影の有無、被験者の肌の色の違いの影響もあると考えられる。

4. おわりに

本研究では MediaPipe Hands と CNN を用いた手軽な指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に対して十分な認識率が得られた。今後の課題として、今回は対象外とした形が変化する清音 4 文字の実装が挙げられる。

[参考文献]

- [1] 田中成典, 塚田義典, 田熊美沙: 距離画像センサを用いた指文字の認識に関する研究, 情報処理学会第 76 回全国大会 5Q-7, 2014