

深層学習を用いた単眼カメラによる空書の自動認識

Automatic recognition of Kusyo by a monocular camera using deep learning

藤本 一文^{*1}, 長谷川 忍^{*2}

Takafumi FUJIMOTO^{*1}, Shinobu HASEGAWA^{*2}

^{*1*} 北陸先端科学技術大学院大学

^{*1*} Japan Advanced Institute of Science and Technology

Email: s1810165@jaist.ac.jp

あらまし：深層学習を用いた空書の自動認識の開発を行った。色手袋の特徴量から重心の座標を取得し、軌跡を描画することで空書を可視化した。また学習データには空書データのほかに手書き文字データを追加した実験を行った。実験結果では5つのモデルによる比較を行い、補正を加えることや手書き文字データを追加することにより良い結果を得られた。

キーワード：空書, 単眼カメラ, 深層学習

1. はじめに

手話とは、ろう者が健聴者とコミュニケーションを行うために用いられる方法の一つである。だが、健聴者の中で手話ができる、もしくはわかる人はごくわずかである。日本のろう者の数が34万人であるのに対し、手話通訳士や要約筆記者といったサポートができる人材は5万人弱しかないのが現状である。また、私たちが使う口語で日々新しい言葉が生まれるように、手話言語も新しい言葉に対応しなければならない。これらの問題を解決するために手話の自動認識の研究も進んできている。近年はAI研究分野における深層学習が急速に進歩していることにもなっており、画像や動画から時系列的に認識する研究⁽¹⁾⁽²⁾も進んでおり、認識精度は大きく向上している。しかし、新しい言葉にも対応しなくてはならないためその都度データを集める必要があることは課題の一つである。

そこで本研究では、コミュニケーションをとる方法として次に多い筆談や空書という文字に注目した。空書とは筆や紙を使わず、空中に文字を書くことであり、主に指一つで書くことができる。空書は手話単語や指文字で表すことができないときや「形」に意味があることを伝えたいときに用いられることが多い。また、手話とは違い、健聴者との簡単なやりとりにはこちらの方が早い。そのため手話と同じように、空書の自動認識できる環境も要求されるが、空書までサポートされた研究はあまりない。

本研究では、深層学習を利用して単眼カメラの前で書かれたひらがなによる空書を認識するための手法を開発し、手話の自動認識の範囲を拡張することを目的とする。単眼カメラはデュアルカメラよりも安価に導入できるため環境を整えやすい。また、カメラ以外には特別なセンサを必要としない非接触型の環境開発を目指す。

本研究で用いるひらがなは50音(ゐ, ゑを除く)、濁点、半濁点の71文字である。空書の検出・認識を行う上で課題として3つのことが挙げられる。1)空

書は空中文字ということもあり、どこから書き出しであり、どこが書き終わりなのかの識別ができないこと、2)一筆書きとなるため、線の切れ目(1画目と2画目の間など)がわからないこと、3)学習を行う上で、ひらがなによる空書のデータを用意しなくてはならないこと、である。

2. 空書の課題

2.1 書きはじめ・書き終わり

色手袋を利用して手の領域の検出、特徴量を抽出する。画面内の特徴量の量で書きはじめ・書き終わりの設定を行った。図1に赤色手袋が検出されている様子を示す。

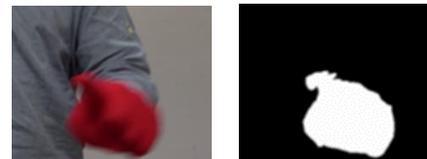


図1 赤色手袋が検出される様子

2.2 線の切れ目

重心の軌跡を表示することで空書の再現を行った。ひらがなは基本的に上から下へと向かって書かれるため、左上方向の範囲に向かう線を消し、線の切れ目の再現も行った。左上方向の線は重心の座標の移動から求める。座標の方向ベクトルを図3に示す。左上に向かう線は重心の移動前の点 P_{i-1} から移動先の点 P_i の角度パラメータ θ_i で求めることができる。角度パラメータ θ_i を求める方法として式(1)に示す。 θ_i の範囲を(2)に示す。

これにより、できた文字を学習データとして用いる。図4に取得した空書画像を示す。

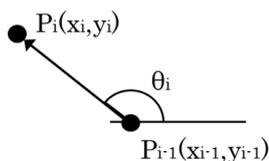


図3 方向ベクトル

$$\theta_i = \tan^{-1} \left(\frac{y_i - y_{i-1}}{x_i - x_{i-1}} \right) \quad (1)$$

$$90 < \theta_i < 180 \quad (2)$$

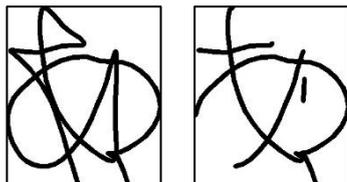


図4 空書画像(左:補正なし, 右:補正あり)

2.3 データセット・トレーニング

学習するデータはビデオカメラで被験者によるひらがなの空書の様子を集めた。空書を行うとき、対面で見ることが多いため、本研究ではカメラを対象者の対面に設置し、ひらがなの空書の様子を画像で取得する。29人分のデータから各クラス約58枚の画像を取得し、データ加工(アフィン変換でx, z軸疑似回転、y軸回転)により各クラス約870枚に増やした。図5に加工例を示す。

また、空書は手書き文字の一環である。手書き文字には草書と呼ばれる線が流れるように書かれた文字がある。そのため学習データに動画から集めた空書データだけでなく公開されている手書き文字データ³⁾を追加した実験も行った。

少ない学習データで行うため、学習はファインチューニングを用いる。ネットワークには ResNet50⁴⁾を使用する。

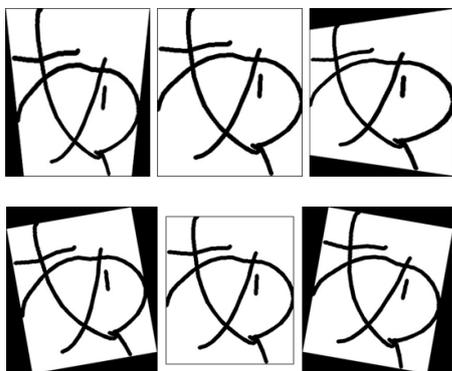


図5 加工例

3. 実験

3.1 評価方法

手書き文字データは空書データとは違うため、データ量の差によっては認識精度が落ちる可能性があ

る。そのため3パターンに分けて追加を行った。これによって比較実験では補正なしの学習モデル、補正ありの学習モデル、補正ありに手書き文字データを追加した学習モデルの3つのモデルで比較を行った。また、5分割交差検証を行うことで汎化性能のある結果が得られた。

実験では、5人の被験者に2回ずつ50音(ゐ, ゑを除く)と濁音・半濁音のひらがな71文字の空書の認識精度を求める。

認識精度を求める際には(3)式のように、識別回数(Answer)を識別回数と誤検出(Miss)の合計で割ると識別率(Ave.)が求めることができる。

$$\text{Ave.} = \frac{\text{Answer}}{\text{Answer} + \text{Miss}} \quad (3)$$

3.2 実験結果

実験結果では、各モデルで5分割検証を行うためそれらの平均値を求めた。その結果、model1の補正なしの場合で96.7[%]の識別率であった。model2の補正ありの場合では97.7[%], そこへmodel3の手書き文字データを追加した場合は98.1[%]の識別率であった。補正や手書き文字データの効果はあったと言える。表1に各モデルの識別率を示す。

表1 各モデルの識別率

	model		
	1	2	3
Answer	3434	3468	3482
Miss	116	82	68
Ave.	0.967	0.977	0.981

4. まとめ

上方向の線を削除することや手書き文字データを追加したことにより識別率向上の効果が見られた。

だが、ひらがなの中には補正を加えたことにより識別率が下がったものや識別率が低いままのものもあった。実用性を見る場合、更なる改良が必要だと思われる。

参考文献

- (1) 橋村佳祐, 齊藤剛史, “Light-HMM を用いた手話認識,” 電気関係学会九州支部連合大会講演論文集, p. 511(2015)
- (2) 松田啓佑, 飯塚博幸, 山本雅人, “手話動作分類における RCNN モデルの性能評価と内部状態解析,” 人工知能学会全国大会論文集, pp. 1-4(2018)
- (3) 文字画像データセット(平仮名 73 文字版)を試験公開しました, <https://lab.ndl.go.jp/cms/hiragana73/>, (参照 2020.1.16)
- (4) K. He, X. Zhang, S. Ren, & et al. “Deep Residual Learning for Image Recognition”, arXiv:1512.03385 (2015)