

# 手指形状・動作特徴を用いた LSTM による 連続指文字 3 クラスインデクシング手法の精度向上

横山和明<sup>†</sup> 森本正志<sup>‡</sup>  
愛知工業大学<sup>†‡</sup>

## 1 はじめに

平成 25 年度における厚生労働省の調査によると日本国内における聴覚・言語障害者数は 34 万人にのぼる [1]. このような人々との豊かなコミュニケーションを支援する手段として、手話が挙げられる。手話とは手や腕の動きや位置、形によって意思を伝える視覚言語である。このうち指文字は手の形と動きを日本語の五十音と対応させた手話であり、固有名詞や新語などを表す際に用いられる。

指文字を学習することは、手話普及促進のための有効な第一歩であると言える。我々の研究室ではモーションセンサを用いて指文字動作を倣い学習できる指文字学習支援アプリケーションを開発してきた [2]. ここで重要となるのは指文字認識性能である。指文字は連続して用いられること、また形状だけでなく動作を伴う指文字（動的指文字）があることから、我々は指文字認識の前処理として手指動作系列を非指文字/静的指文字/動的指文字区間に分類する指文字インデクシング手法を提案した [3].

[3] では 4 種類の指文字動作特徴および SVM を用いて分類を行っていたが、本研究では入力特徴量および学習データ数を増やし、学習手法として時系列特徴に強い LSTM を用いることで分類精度の向上を図る。

## 2 指文字インデクシング

上述したように、連続した指文字認識のためには手指動作系列が指文字なのか指文字間の遷移（非指文字）なのかを識別し区間を定める必要がある。本研究ではこの処理を指文字インデクシング（索引づけ）と呼ぶ。指文字認識も含めた処理の流れを図 1 に示す。計測された手指動作系列データを、あらかじめ学習を行った識別器によりフレームごとの 3 クラス分類を行う。指文字は動きを伴わない静的指文字と動きを伴う動的指文字に分類することができれば、静的指文字認識では手指形状特徴を、動的指文字では形状のみではなくフレーム間の時系列特徴を用いて効率的に認識処理を行うことができる。なお、本研究ではこの指文字インデクシング手法のみに言及する。

A study on improving 3-class indexing method of continuous JSL fingerspelling using fingers' shape and motion features with LSTM model

<sup>†</sup>Kazuaki Yokoyama, Aichi Institute of Technology

<sup>‡</sup>Masashi Morimoto, Aichi Institute of Technology

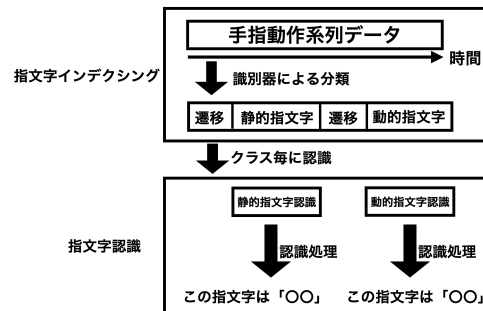


図 1 処理の流れ

## 3 手指形状・動作特徴と LSTM

[3] ではモーションセンサである LeapMotion を用いて掌および各指先の 3 次元座標データを計測し、ある対象フレームから前後 15 フレーム間での座標移動量を用いて、対象フレームにおける掌の移動距離、指先の移動距離、手の向きのばらつき、掌の移動距離の 4 種類の動作特徴を算出した。その後、SVM を用いて対象フレームに対する 3 クラス分類の学習を行い識別器を作成した。評価実験により非指文字/静的指文字/動的指文字の F 値平均はそれぞれ 0.74/0.73/0.21 となり、主に動的指文字の識別率が低くなった。その原因として、入力特徴の不足および SVM による時系列特徴の学習不足が考えられる。

そこで本研究では、時系列特徴を学習することができる LSTM を識別器として用い、その一方で LeapMotion で取得した 3 次元座標データをそのまま手指形状・動作特徴として用いる。LSTM とは RNN の拡張モデルであり、中間層のユニットを LSTM ブロックに置き換えることで長期依存の特徴を学習することができる深層学習モデルの一つである [4].

入力データとしては上述した掌および各指先の 3 次元座標データ 6 種類に加え、5 指における中手骨・基節骨・中節骨・末節骨の下端・中心・上端の各 3 次元座標データ、掌および各指先の向き、掌の法線ベクトル、掌の位置の変化率の 68 種類を合わせた 74 種類を入力特徴として用いる。これらのデータは、[2] の指文字認識においても用いられている。

## 4 データセット

本研究では被験者 6 人を対象に 6 種類の指文字系列パターンを各 4 回ずつ計 144 セットの手指動作系列データを計測した。計測した 6 種類の指文字系列パターンを表 1 に示す。各々の指文字系列パターンにおける平均フレーム数は 8842F であり、このフレーム内に 12 文字の指文字および遷移区間が撮影されている。これは [3] で用いられていたものと同じであり、各系

列パターンにおいて静的および動的指文字の出現順序を揃えている。データ計測後に座標データに基づいて Unity アプリケーションで手指形状を復元し、目視で各フレームの正解クラスを付与し、データセットを作成する。

表1 指文字系列パターン

	静	静	動	静	動	動	動	静	静	静	動	動
1	や	よ	ん	る	ぞ	が	の	か	あ	す	じ	ぬ
2	ち	と	ぢ	お	ぎ	も	ぼ	ゆ	ふ	に	ぐ	こ
3	ほ	そ	ぜ	た	げ	も	ど	て	む	そ	ぎ	め
4	な	し	で	く	ず	づ	よ	れ	ろ	ら	ゆ	へ
5	ま	け	ご	え	り	だ	を	う	わ	は	べ	ひ
6	き	い	ぼ	み	び	っ	ぶ	ね	せ	つ	や	さ

## 5 評価実験

本実験では、4章で述べた平均フレーム数より1パッチを2210Fとした。また最適化手法を adam, 入力および出力層を全結合層, 中間層を100ユニットとした3層のLSTMネットワークを構成し、活性化関数にReLU, 損失関数にソフトマックスクロスエントロピーを用いて学習を行う。

### 5.1 交差検証

はじめに [3] との比較を行うため、入力に用いる手指動作系列データを [3] で特徴量算出に用いたものと同じ掌および各指先の3次元座標データのみとして実験を行なった。また交差検証におけるデータ分割方法として、被験者毎(6分割)・指文字系列パターン毎(6分割)および計測回毎(4分割)の3種類を実施した。評価結果として、正解率の平均および標準偏差はそれぞれ、 $0.53 \pm 0.06 / 0.57 \pm 0.02 / 0.70 \pm 0.01$  となった。被験者および指文字系列パターンに比べ計測回毎に対する正解率が高い結果となった。また、どの対象においても標準偏差は十分に低くデータごとのばらつきはほぼないことがわかった。

次に、計測回毎の分割方法で得られた4種類の識別器のうち、最も高い正解率を示した識別器における再現率、適合率、F値を表2に示す。[3]で述べたSVMでのF値平均と比べ、動的指文字の分類性能が大きく向上したことがわかった。

表2 再現率, 適合率, F値

正解率	0.72		
	遷移	静的	動的
再現率	0.81	0.67	0.59
適合率	0.78	0.75	0.54
F値	0.80	0.71	0.57

### 5.2 追加特徴を用いた評価

特徴量の追加による識別性能の検証のため、3章で述べた追加特徴を加えた手指動作系列データで識別器の学習を行なった。学習用データに被験者6における4回目以外の撮影データ, 検証およびテストデータに学習用データから除いた被験者6における4回目の撮影データを用い65回の学習を行なった。表3にその結果を示す。遷移クラスにおけるF値は0.66と低くなったが、静的および動的指文字クラスにおけるF値は向上していることがわかった。

表3 追加特徴を用いた識別器による再現率, 適合率, F値

正解率	0.72		
	遷移	静的	動的
再現率	0.57	0.80	0.77
適合率	0.79	0.74	0.68
F値	0.66	0.77	0.72

## 6 考察

### 6.1 交差検証

交差検証の結果、データ分割方法を計測回毎にした場合と比べ被験者および指文字系列パターンは低い数値となった。これは、分割方法によっては十分に汎化ができていないことを表し、手指形状・動作パターンや遷移組み合わせの不足が要因であると考察する。一方、[3]におけるF値平均との比較結果、静的指文字クラスを除いた2クラスにおいて向上が見られた。遷移クラスおよび動的指文字クラスの認識においてはフレーム間の時系列特徴が重要となるため、LSTMの学習能力が有効に機能したと考えられる。

### 6.2 追加特徴を用いた評価

追加特徴により、特に動的指文字クラスのF値において精度向上が見られた。これは動作に関する情報を多く追加特徴として用いた事によるものと考えられる。また後段における指文字認識処理を想定した場合、静的、動的指文字クラスにおいては再現率、遷移クラスにおいては適合率が分類の際に重要であると考察される。表3からは表2に比べてそれぞれの向上が確認されることから、追加特徴量の有効性が示されていると考察する。

## 7 まとめ

本研究では指文字インデクシングの精度向上に関する手法を提案した。LSTMを用いて識別器を作成することで、先行研究[3]と同種類のデータを用いた場合でも動きのあるクラスの分類精度が向上した。また、特徴量を追加することで静的および動的指文字クラスにおいてより高いF値を示すことができた。今後の課題として、データセットの増強、入力特徴の有効性評価・取捨選択、指文字認識機能との連携等が挙げられる。

## 謝辞

本研究の一部はJSPS科研費JP18K11364の助成による。

## 参考文献

- [1] 厚生労働省, 平成25年版障害者白書(全体版)付録8章, 厚生労働省(オンライン).
- [2] 森本他, "モーションセンサを用いた学習支援アプリケーション", 情処論 DCON, Vol.8, No.2, pp.28-40 (2020).
- [3] 加藤他, "手指動作特徴を用いた連続指文字に対する3クラスインデクシング手法に関する検討", 情処全, 2R-03, No.2, pp.161-162 (2019).
- [4] S. Hochreiter et al., "Long Short-Term Memory". Neural Computation, Vol.9, No.8, pp.1735-1780 (1997).