

写真画像のスケルトン化を用いた 指文字認識の正答率の向上

知念次郎†1 當摩拓也†2 築地立家†3

東京電機大学理工学部†1 東京電機大学理工学部†2 東京電機大学理工学部†3

1. はじめに

WHO (World Health Organization) によると、2018年時点では世界人口の約5%である4億6600万人が聴覚障害者であるとされている。

彼らの意思疎通の手段として筆談や手話、指文字などが挙げられる。

筆談はメモ帳などに字を書いて相手に伝える方法であり、手話は手の形や動き、顔の表情で挨拶や物などを表し、指文字は手の形や動きで文字や数字を表すことで会話する方法である。

筆談では相手に伝えるためにメモ帳やペンが必要となるので使いにくい場面が多い。手話では一般的なものを指し示すことは容易だが固有名詞のような手話では表せないものがある。そのような場面では文字を一文字ずつ表すことのできる指文字を使う。

手話や指文字で会話を行うためには会話を行う双方が指文字に関する知識が必要である。しかし、多くの人はそれらに関する知識がなくこれらを用いた会話が困難である。

指文字の認識については、2012年には筑波技術大学の三宅、若月、内藤らは Kinect を使用して^[3]、2016年には東京電機大学の飯森は Leap Motion を使用して^[4]指文字の画像認識の研究を行なった。これらの研究の問題点として、深度カメラやセンサ等を用意する必要があるため普及しにくい。そこで深度カメラなどの特殊な機材を使用せずに一般的なカメラで撮影した画像だけで指文字認識を行い高い正答率を出すことを目的とした。

2. 研究目的

過去の指文字関連の研究では指文字認識を Kinect や Leap Motion で撮影した深度画像などを使用して研究が行われてきた。これらの研究は、深度カメラやセンサ等を用意する必要があるため普及しにくいという問題点がある。そこで深度カメラなどの特殊な機材を使用せずに一般的なカメラで撮影した画像だけで指文字認識を行い高い正答率を出すことを目的とした。

3. 指文字認識の方法

指文字とは、手の形や動きに文字を対応させた視覚言語の一種である。指文字には国や地域によって同じ文字でも表し方が異なるため本研究では、日本で使用されている指文字ではなく広く知られているアルファベットの指文字を認識する研

究を行った。

アルファベットの指文字には2種類があり両手を使用して表現するイギリス指文字と片手を使用して表現するアメリカ指文字がある。この2種類の中で多く使われていて画像を取得する時に片手で表現でき安易なため、本研究では上記理由により、アメリカ指文字の American manual alphabet を採用した。

アメリカ指文字の手の形状は図1に示す。

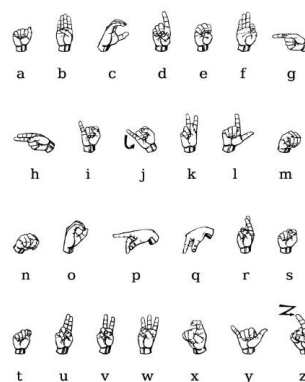


図1. American manual alphabet ^[9]

今まで一般的なカメラを用いた指文字の研究では撮影した画像のグレースケール化などを行い二値化してから特徴量の抽出を行ってきた。本研究では撮影した指文字画像から OpenPose を使用してスケルトン画像にすることで特徴量を減らすことにした。

OpenPose は Carnegie Mellon 大学が発表した深層学習によって単一画像上における人体の関節などの特徴点をリアルタイムで検出し、画像上の人体の姿勢を可視化できる技術である。本研究では指文字の手の形状を検出するために深度カメラやセンサ等を使用する代用として OpenPose の技術を使用した。

指文字を判定するためのモデルを作るため、指文字一文字ごとに600枚の画像を用意し、用意した画像を OpenPose でスケルトン画像に変換し、その画像を使用して指文字を判定するモデルを作成した。本研究ではこのモデルを使用して指文字の判定を行った。

本研究では Web カメラで取得した画像を OpenPose にてスケルトン画像に変換し、作成したモデルが取得したスケルトン画像をどの指文字であるかを判別する。

本研究で目的としているのは指文字の正答率の向上であるため、1文字ごとに40枚の画像を用意しその正答率を測定し記録した。

今回用いる文字は動きのある「J」と「Z」以外の A～Y の文字である。データの取得については2人で行い、手の撮影位置はカメラから約50cmの地点とする。

Improvement on Fingerspelling Recognition with Skeletonization of Photographic Images

†1 JIRO CHINEN, Tokyo Denki University Science and Engineering

†2 TAKUYA TOMA, Tokyo Denki University Science and Engineering

†3 TATSUIE TSUKIJI, Tokyo Denki University Science and Engineering

4. 開発環境

本研究で使用したハードウェア、ソフトウェア、PCをそれぞれ表1, 表2, 表3, 表4に示す。PC1は指文字を判定するためのモデルの製作に使用し、PC2は指文字画像のOpenPoseを使用したスケルトン画像への変換と、実際に判定する時に使用した。

表1. 開発環境(ハードウェア)

名称	メーカー名	型番	シリアル番号
webカメラ	iBUFFALO	BSW20K M14BK	A40825

名称	メーカー名	型番	シリアル番号
PC 1	LENOVO	0896-CTO	PB08NA0
PC 2	HP Pavilion	500-430jp	JPA4514TJ3

表2. 開発環境(PC)

名称	CPU	メモリ	GPU
PC 1	Intel(R)Core(TM) I5-3470, 3.20GHz	16.0 GB	NVIDIA GeForce GTX 1060 6GB
PC 2	AMD A6-7400K Radeon R5, 6 Compute Cores 2C +4G, 3.50 GHz	16.0GB	NVIDIA GeForce GTX 1060 6GB

表3. 開発環境(ソフトウェア PC1)

名称	メーカー名	バージョン
Windows 10 Pro	Microsoft	1709
Python	Python ソフトウェア財団	3.6.6
tensorflow-gpu	Google	1.10.0
Keras		2.2.2
OpenCV	Intel, Itseez, Willow Garage	3.4.1
NumPy	Travis Oliphant	1.14.5
Cuda	NVIDIA	9.0.176
NVIDIA グラフィックドライバ	NVIDIA	385.54

表4. 開発環境(ソフトウェア PC2)

名称	メーカー名	バージョン
Windows 10 Pro	Microsoft	1803
Python	Python ソフトウェア財団	3.6.6
tensorflow	Google	1.12.0
Keras		2.2.4
NumPy	Travis Oliphant	1.15.4
chainer	Preferred Networks	5.0.0
Cupy-cuda90	Preferred Networks	5.0.0
scipy		1.1.0
Cuda	NVIDIA	9.0.176
NVIDIA グラフィックドライバ	NVIDIA	385.54

5. 指文字認識の実験結果

文字ごとの正答率の測定は表5の正答率より、すべての指文字の正答率が85%を超えており、高い精度で認識しているといえる。「T」は「S」と誤認識されることが多く、「R」は「K」と誤

認識されることが多かった。

表5. 文字ごとの正答率の測定実験の正答率(%)

文字	A	B	C	D	E	F
正答率	100	100	100	100	95	100

文字	G	H	I	K	L	M
正答率	100	100	97.5	100	100	100

文字	N	O	P	Q	R	S
正答率	92.5	100	87.5	100	87.5	100

文字	T	U	V	W	X	Y
正答率	85	97.5	90	100	100	100

6. おわりに

本研究で行った手法を用いることでカメラ画像からの指文字認識の正答率を上昇させることができた。このような結果を得られた理由として、スケルトン画像にする事によって指の形状が明確になったから^[10]と考えられる。

本研究では指文字の画像から判定していた為、画像上から動きのある指文字を判定することはできなかった。今回対応のできなかった動きのある指文字認識も判別できるように、動きのある指の動きを追えるような画像の変化を読み取る指文字認識の手法を検討する必要がある。

参考文献

- [1] 太田満久, 須藤広大, 黒澤匠雅, 小田大輔. 現場で使える! TensorFlow 開発入門 Keras による深層学習モデル構築手法. 初版. 株式会社翔泳社. 2018.277p.
- [2] 斎藤康毅.ゼロから作るDeepLearning Python で学ぶディープラーニングの理論と実装. 初版. 株式会社オライリー・ジャパン 2016. 298p.
- [3] 三宅太一, 若月大輔, 内藤一郎. 距離画像を用いた動きをとらえる指文字認識に関する基礎的検討. 筑波技術大学テクノロジーレポート. 筑波技術大学学術・社会貢献推進委員会. 2012. Vol.20(1). P7-13.
- [4] 飯森義貴. LeapMotion を用いたリアルタイム指文字認識の検討. 東京電機大学理工学部情報システムデザイン学系. 2016. 卒業論文.
- [5] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, Yaser Sheikh. realtime multi-person 2D pose estimation using Part Affinity Fields. CVPR. 2017.
- [6] TensorFlow. <https://www.tensorflow.org>. (参照 2018-8-27).
- [7] World Health Organization. Deafness and hearing loss. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>. (参照 2018-12-11).
- [8] Chainer_Realttime_Multi-Person_Pose_Estimation. https://github.com/DeNA/Chainer_Realttime_Multi-Person_Pose_Estimation. (参照 2018-9-18).
- [9] 「American manual alphabet」. https://en.wikipedia.org/wiki/American_manual_alphabet. (参照 2018-11-29).
- [10] 安藤博士, 佐々木守, 岩田穆. 見え方に基づいた顔認識手法のための画像前処理. FIT2005, 2005, p21-22, (第4回情報科学技術フォーラム).