

上方に設置したセンサによる指文字認識方法の研究

大塚友章[†] 山崎雄太[†] 内山健斗[†] 水谷晃三[†] 荒井正之[†]

帝京大学理工学部ヒューマン情報システム学科[†]

1. はじめに

本研究では手話行為者の使う指文字を上方に設置したセンサによって認識することで、コミュニケーションを支援する方法を検討する。指文字とは聾者が手話だけでは表せない単語などを手で一文字ずつ表現する方法である。

既存の指文字認識の研究[1-3]では深度センサや RGB カメラなどを指文字使用者の正面に設置するため、対話者同士の立ち位置や設置場所の制約などが生じやすいと考えられる。そこで本研究では上方にセンサを設置することにより設置場所の制約を改善した指文字認識手法の実現を検討する。

2. 研究手法

2.1. 研究概要

本研究では深度センサと RGB カメラを床から 2m の高さに設置し、センサを真下に向けて認識を行う。認識する範囲はセンサから 500~900mm (床から 1100~1500mm) とする。上方からのセンサから取得した手指の画像をニューラルネットワークにより認識する。

まず上方にセンサを設置して指文字の認識を行う場合についての影響を検証した。その結果、図 1 に例示する「は」を表す指文字のように指を正面に向けている指文字や、正面からでは指が隠れてしまう指文字などの 7 文字は、上方にセンサを設置することで認識精度が向上する可能性がある。一方、指を下へ向けている指文字など 7 文字の認識が困難になる可能性がある。

2.2. 指画像の取得と前処理

センサからの RGB 画像については、深度情報を活用して前処理を行う。まず、画像中の認識範囲外の影響を除外するために、深度情報に基づいて前述の範囲以外を黒に塗りつぶす。コントラストや明るさが偏っている画像の全体的なバランスを調節するために画像のヒストグラムにおける画素値の分布を均等化する(図 2 左)。次に深度情報を可視化した画像(以下、深度画像と記す)を作成する。深度画像は認識範囲内

A study for a method of fingerspelling recognition using a mounted sensor on the ceiling.

[†]Tomoaki Ohtsuka, Yuuta Yamazaki, Kento Uchiyama, Kozo Mizutani, Masayuki Arai : Department of Human Information Systems, Faculty of Science and Engineering, Teikyo University.

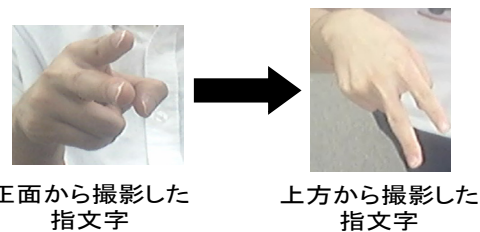


図 1 センサの位置を変えて撮影した指文字「は」



図 2 RGB 画像と深度画像の例

の深度情報にグレースケールを割り当て作成したものである(図 2 右)。ニューラルネットワークの学習と後述の指文字会話支援システムでは図 2 の様な RGB 画像と深度画像を用いる。

2.3. 指文字認識の方法

指文字認識の既存研究では、手指形状を認識する手法としてテンプレートマッチングやニューラルネットワークなどが用いられている。本研究では、主に画像認識に応用される畳み込みニューラルネットワーク (Convolutional Neural Network, 以下 CNN と記す) による認識を行う[4]。本研究ではニューラルネットワークを扱うオープンソースのフレームワークである Caffe を利用する[5]。

2.4. 指文字会話支援システム

Caffe を使用した指文字会話支援システムの試作を行う。図 3 に試作するシステムの概要を示す。①, ②では Kinect から取得した RGB データ、深度情報を基に 2.2 で述べた方法により RGB 画像、深度画像を生成する。③と④では Caffe を利用し、生成された画像を認識する。⑦では CNNServer の 2 つの認識プログラムが得た結果(⑤, ⑥)のうち、それぞれ最上位のものを 1 つずつ選択し、認識精度の高い方の結果を画面に出力する。

3. 実験

3.1. 学習モデルの作成

CNN を使って動作の伴わない指文字 41 種類

を分類する学習モデルの作成を行った。学習用のサンプル画像として、指や手の角度、高さを変えながら撮影した RGB 画像と深度画像を、それぞれ各文字につき約 4500 枚を用意した。サンプル画像は、学習用に用いる学習画像群と学習後の学習モデルの精度を確認するためのテスト画像群に 1 対 6 に振り分けた。図 4 は、テスト画像群に対する学習モデルの識別正解率の推移を示したものである。RGB 画像の学習モデルでは 600 回目の学習以降、識別正解率が 100%に近い数値となり、深度画像も 800 回目の学習以降、識別正解率が 90%以上の数値となった。

3.2. 支援システムの動作結果

3.1の実験で得られた学習モデルを用いる指文字会話支援システムを試作した。図5に動作例を示す。RGBセンサ画像処理画面にはセンサが取得した上方からの指文字が表示されている。認識結果表示画面にはその指文字を認識した結果が表示されている。

4. 考察

学習モデルの作成では深度画像の識別正解率の方が低い。深度画像はグレースケールのため、認識が困難になっている可能性があり改善方法の検討が必要である。

試作システムの認識結果表示では2つの認識プログラムの最上位の結果を取得し、高い認識精度を持つ方を出力しているが、最上位の結果が誤っている場合もある。2つの認識プログラムのすべての結果を活用し、各文字に対する認識精度の平均値を求め、その値が最も大きい結果を出力するなど他の手法の検討が必要である。

5. おわりに

本研究では上方に設置したセンサからの指文字認識の方法の研究と指文字認識の支援システムの試作を行った。上方にセンサを設置する方式では、他の技術と組み合わせて手の位置の検出を行うことで、立ち位置の制約を更に改善できる可能性もある。今後は認識精度の向上や実用性の向上を目指したい。

参考文献

[1]Otiniano Rodriguez : Finger Spelling Recognition from RGB-D Information Using Kernel Descriptor, 26th Conference on Graphics Patterns and Images (SIBGRAPI 2013), pp.5-8, 2013.
 [2]Hee-Deok Yang : Sign Language Recognition with the Kinect Sensor Based on Conditional Random Fields, Sensors, Vol.15, pp.135-147, 2014.
 [3]島田 貢明 : 指文字の認識に関する基礎的研究,

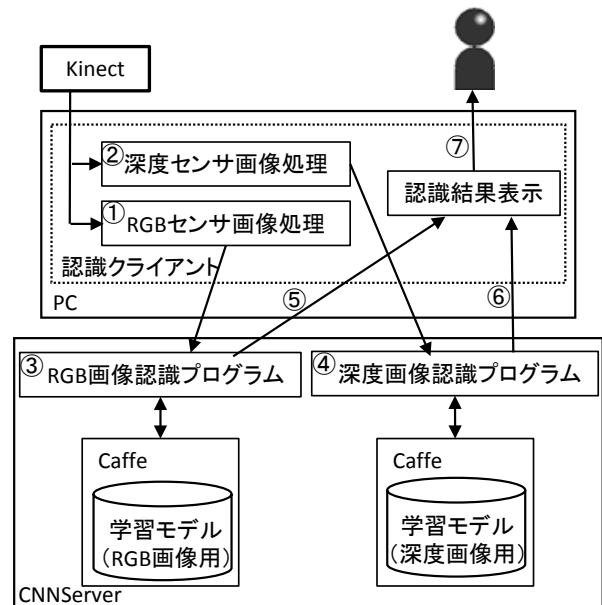


図3 指文字会話支援システムの概要

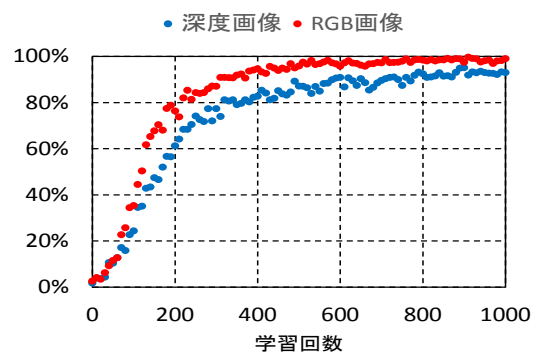


図4 学習中のモデルの識別正解率の推移



図5 試作システムの動作例

仁愛女子短期大学研究紀要, Vol.36, pp.9-14, 2004.

[4]岡谷貴之 : 深層学習, 講談社, 2015.

[5]Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor : Caffe: Convolutional Architecture for Fast Feature Embedding, arXiv preprint arXiv:1408.5093, 2014.