

# 複数の視覚センサを用いた指文字計測と認識に関する研究

岩本 一希<sup>1</sup> 浜野 遼太郎<sup>1</sup> 本田 久平<sup>1</sup> 長谷川 由美<sup>2</sup> 田中 省作<sup>3</sup>

大分工業高等専門学校<sup>1</sup> 近畿大学<sup>2</sup> 立命館大学<sup>3</sup>

## 1.はじめに

手話は手の形や動きにより話者の意図を伝えることができ、聴覚障がい者とのコミュニケーションに使用される。近年では日本でも手話を学ぶ人口が増えてきているが、英語などの第二言語と比較して、教材や支援システムは十分整備されているとは言えない。今回、手話のうち指文字を研究対象とし、学習者を支援するシステムを目指し、指文字の計測や認識について試みた。

本研究では、画像を取得し骨格検出アルゴリズムである OpenPose<sup>1)2)</sup>を用いて指文字データを収集する。複数の視覚センサからの情報を組み合わせて学習者の手指の三次元位置計測を行い、指文字を認識するシステムの提案を行う。

## 2.システム概要

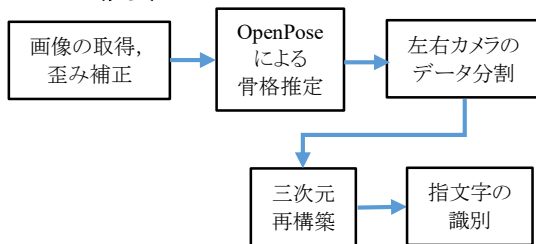


図1 システムの概要

図1に三次元データを用いた指文字認識のシステムの概要図を示す。

まず、ステレオカメラから取得した画像の歪みを補正する。補正された画像を OpenPose で処理し、骨格推定情報を取得する。得られたデータを左右のカメラそれぞれのものに分割し、三次元座標の復元を行った後、得られた三次元座標からどの指文字なのかを SVM(Support Vector Machine)で識別する。

## 3.カメラキャリブレーション

今回、ステレオカメラモジュール (ELP 960P2CAM-V90-J) を用いた。このモジュールは、2つのカメラの同期をハードウェア的にとっており、PC の1つの USB ポートに挿して同時刻の画像を取得できる。

このカメラで画像を取得すると、画像には魚眼レンズのような歪みが発生する。そこで、カメラキャリブレーションを行いカメラの内部パラメータを予め求めておき、指文字計測の際は、求めた内部パラメータを用いて画像の歪みを補正する。

## 4.OpenPose

OpenPose は深層学習を用いた単眼カメラでのスケルトン検出アルゴリズムで、静止画や動画、リアルタイム映像から全身や手の骨格位置推定を行うことができる。



図2 手の骨格推定

図2に OpenPose による手の骨格位置推定の例を示す。指ごとに色分けして推定結果が表示されている。手については、手首位置を含めた 21 個の関節座標の推定値が得られる。各関節の位置はピクセル単位で Json ファイルとして書き出すことができるため、今回はそれを取得データとして用いた。

## 5.三角測量

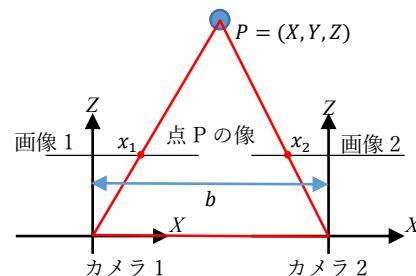


図3 三角測量

人間が左右の目を使って物体の奥行きを感じるように、カメラを2台用いることで同シーンに取得した2枚の画像とそれぞれの画像の点PのX座標の差から点Pの三次元座標の推定ができる。この方法を用いるためには、光軸が平行、光学中心だけが X 方向に距離 b だけ離れているなどの諸条件を満たす、図3のような理想的なカメラ配置に変換することが必要になる。よって実際に行う際は、カメラの回転や距離を測定し、補正する必要があるため、今回は OpenCV のクラス関数を用いてその推定と変換を行った。

## 6.指文字の識別

指文字の識別には、パターン認識モデルの一つである SVM を用いる。今回は SVM 中の RBF カーネルを用いる。RBF カーネルは学習の際にパラメータ C,  $\gamma$  を設定することで、誤りに対する厳しさを境界の複雑さを変えることができる。SVM の学習と識別の際に用いる特徴量として、OpenPose で

Fingerspelling Measurement Using Several Visual Sensors  
Kazuma Iwamoto<sup>1</sup>, Ryotaro Hamano<sup>1</sup>, Kyuuhei Honda<sup>1</sup>,  
Yumi Hasegawa<sup>2</sup>, Shosaku Tanaka<sup>3</sup>  
National Institute of Technology, Oita College<sup>1</sup>, Kindai  
University<sup>2</sup>, Ritsumeikan University<sup>3</sup>

得られた手の関節座標を用いる。

## 7.実験



図4 処理の流れ

まず、構築したシステムにて、指文字の三次元再構築を行った。取得画像から三次元再構築までの流れを『あ』の指文字を例にして図4に示す。三次元で表現された指文字データは、いろんな視点からの見え方を表示することができる。

次に、認識率について調査した。今回は指文字のうち、手の動きによって表現するものや指が隠れるなどして OpenPose で認識できなかったものを除いた 30 種を認識対象とし、それぞれについて 15 データを収集した。そのうち 12 データを学習データとし残り 3 データをテストデータとする交差検証法と、 $C$ ,  $\gamma$  に対するグリッドサーチを行って SVM による認識率を計算した。

SVM の特徴量として、手の関節 21 関節のうち、手首以外の 20 関節の座標値から手首の座標値を引いて算出した値を用いた。これは、提示する指文字の位置の変化による影響を軽減するためである。三次元再構築を行って手指関節の三次元座標を使った場合には、60 次元 (20 関節  $\times$  3 次元) の特徴ベクトルを用いる。比較のため、1 つのカメラの画像内の二次元座標を用いた場合の 40 次元 (20 関節  $\times$  2 次元) の特徴ベクトルおよび両方のカメラの画像内の二次元座標を用いた場合の 80 次元 (20 関節  $\times$  2  $\times$  2 次元) の特徴ベクトルを用いた認識も行った。

認識率は、三次元座標の 60 次元の特徴ベクトルを用いた

場合が 98.7%、二次元座標の 40 次元と 80 次元を用いた場合が 99.7%と、三次元座標を用いた方が若干低いという結果となった。この原因としては、今回用いたステレオカメラのカメラ間距離が小さかったため三角測量の際に奥行方向の距離計算の誤差が大きくなったことが考えられる。

誤認識が多かった指文字は、図 5~7 に示す『い』と『や』、『う』と『る』、『ひ』と『ぬ』などであった。原因としては、対になる指文字間の形状の違いが 1 つまたは 2 つの関節の曲げのみであり、違いが小さいために両者の区別が困難であったことが考えられる。



図5 『い』の指文字(左)と『や』の指文字(右)

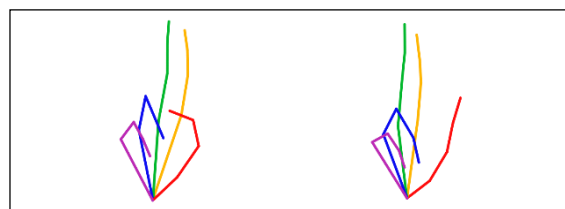


図6 『う』の指文字(左)と『る』の指文字(右)

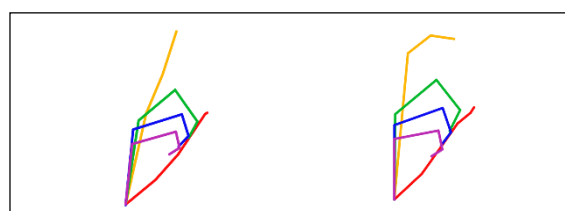


図7 『ひ』の指文字(左)と『ぬ』の指文字(右)

## 8.まとめ

本研究ではステレオカメラと OpenPose を用いた三次元再構築による指文字認識を行った。認識については、二次元座標による認識と比較して三次元座標による認識の優位性は見られなかった。

今後は多くの学習データを収集するとともに、カメラ間の距離や角度を変えた距離計算や指文字の認識率の変化を調べることが必要だと思われる。また、指文字学習者支援のための本格的なシステム構築に向け、三次元計測結果と認識結果を分かりやすく提示できるよう取り組んでいく。

## 参考文献

- 1) Tomas Simon, Hanbyul Joo, Iain Matthews, Yaser Sheikh, “Hand Keypoint Detection in Single Images using Multiview Bootstrapping”, CVPR, pp.1145-1153, 2017.
- 2) Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, CVPR, pp.7291-7299, 2017.