

UbiScript: スマートフォンを用いた手指ジェスチャー認識

生野優輝^{†1} 外村佳伸^{†2}

概要: 我々は手指ジェスチャー認識を用いた指示（コマンドイング）により、音声に依らない新しい空間インタフェースの創成をめざしている。特に聴覚障がいを持つ人が指文字を用いて文字や数字を豊富に表せることに着目し、従来のジェスチャー認識より深いコマンドイングをシンプルに実現することを検討している。本報告では手指ジェスチャー認識のプラットフォームとして、近年登場した MediaPipe をスマートフォン上に搭載し、その上に実装した指文字認識、数字認識手法、ならびにそれを用いた基本的な応用例について述べる。

1. はじめに

手指ジェスチャー認識は、近年普及が進む音声認識とともに、今後身近な操作インタフェースとして有望であると考えている。特に筆者を含む聴覚障がい者にとっては、音声に依らないインタフェースへの期待は非常に大きい。

本研究では、聴覚障がいを持つ人が指文字を用いて文字や数字を豊富に表現できることに着目し、従来の手のおおまかな動きによるジェスチャー入力に加え、指文字を使用することでより豊富なコマンドイングができることをめざしている。本報告ではそのための基本となる手指ジェスチャー認識手法とその簡単な応用例について述べる。

2. 手指ジェスチャー認識

コンピュータに対する空間インタフェースとしての手指ジェスチャー認識についてはこれまでも多くの研究がされている。近年では Leap Motion [1] や Kinect などにおいて、赤外線を用いたセンサと骨格モデル認識を組み合わせることで手指や腕の動きを簡単に取得できるようになり、多様なジェスチャー認識が行われるようになった。

さらに近頃では、機械学習技術および計算能力の向上により、1台のカメラで撮影された画像から、手指ジェスチャーを3次元でリアルタイム認識することが現実的になってきた。そのことにより、スマートフォンやスマートウォッチなどのモバイル機器のカメラを用いた手指ジェスチャー認識の研究も盛んになりつつある。処理のために使われる認識基本環境のひとつに OpenPose [2] がある。利用するにはある程度の計算能力が必要だが、単眼カメラから骨格座標の推定が可能である。MediaPipe [3] では同じく単眼カメラから骨格座標の推定ができるが、OpenPose より軽量で、最近のスペックの高いモバイル機器でなら十分利用可能になってきた。例えば、Erwin Wu et al. は竜頭の部分に小型カメラを搭載したスマートウォッチを用いて、利用者の手の甲のみを撮影し、この画像を入力として、手指の3次元姿勢を出力とする深層学習ネットワークを設計・開発した[4]。

3. MediaPipe

3.1 MediaPipe

MediaPipe は機械学習のパイプライン処理を構成できるクロスプラットフォームとして Google 社が提供するオープンソースである。様々な機械学習ソリューションとして、顔検出、手検出、ボディ検出、物体検出など機械学習モデルを持ち、Android, iOS, デスクトップ/クラウド, ウェブ, IoT と、多様なプラットフォーム上で機能する。

3.2 MediaPipe Hands

本研究では数ある機械学習ソリューションのうち、MediaPipe Hands [5] を使用している。

MediaPipe Hands は機械学習により単眼カメラから3Dの手指ランドマーク（図1）を推測し、ハンドトラッキングを提供する。また、機械学習の方法に工夫をしており、他のハンドトラッキングライブラリにない軽量さを合わせ持つことから、デスクトップ環境はもちろん、モバイル環境でも動くという強みを持つ。

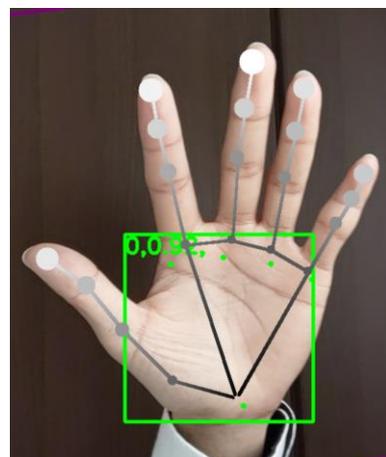


図1 MediaPipe Hands で取得できる手指ランドマーク

取得できるランドマーク座標は、図1にて示されている点で、手首、各関節、指先の合計21個の3次元座標が取得できる。

^{†1} 龍谷大学大学院理工学研究科

^{†2} 龍谷大学先端理工学部

4. MediaPipe Hands を用いた手指ジェスチャー認識

4.1 手指ジェスチャー認識

坂坂らは Leap Motion を用いた指文字認識にて、取得したランドマーク情報を使用し、手が正面を向いているかどうか、各指が曲がっているか、伸びているか、など指の形状に注目した説明変数を用いた分類木分析によって、74.7%の認識結果を得ている [6]。本研究では、Leap Motion と取得できるランドマークの形状が似ている MediaPipe を用い、手の向き、指の形状に注目した指文字認識を試みる。

4.2 手指ジェスチャー認識の下処理

MediaPipe により 21 個の手指ランドマークを取得できるが、座標情報をそのままジェスチャー認識に使用するのではデータ量が多く処理に時間がかかってしまう。そこで、21 個のうち指先と手首を除く 15 個の手指ランドマークについては、角度を求めることができ、その角度により曲げ伸びを識別することとした。本研究ではこうした情報を用いて以下、数字とひらがなの指文字認識を試みた。

4.3 数字指文字認識

0~9 までの数字指文字について、角度データと、ランドマークの同士の距離を用いた表 1 の条件変数を使用した決定木分析を行う。

表 1 数字指文字認識の説明変数

	説明変数
A	親指が伸びている
B	人差し指が伸びている
C	中指が伸びている
D	薬指が伸びている
E	小指が伸びている
F	親指と人差し指で輪を作っている

4.4 平仮名指文字認識

～ 指文字表 ～

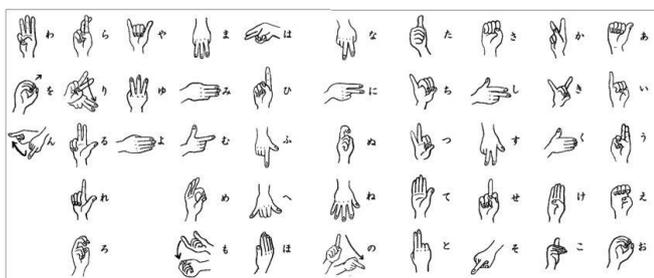


図 2 平仮名指文字表

図 2 は平仮名の指文字を説明したもので [6]、動きのない静的指文字と、動きを持つ動的指文字の二つに分類できる。本研究では静的指文字全てと、動的指文字から「も」の終

端動作を認識の対応とした。他の動的指文字である「の」、「り」、「を」、「ん」については終端動作に類似する他の指文字があったため今回は対象としていない。

決定木分析を行うにあたって、手全体の向きの真偽値を返す A~D と、指の曲げ伸びの真偽値を返す E~R、手指ランドマークの位置関係の真偽値を返す S~U、計 21 個の説明変数を用意した。

認識は「の」、「り」、「を」、「ん」を除いた平仮名指文字、各 20 個を訓練データとしてランダムフォレストを生成し、各決定木の多数決により行う。

表 2 平仮名指文字認識の説明変数

	説明変数
A	相手に向いている面が手の平
B	相手に向いている面が手の甲
C	手の向きが上
D	手の向きが下
E	親指が伸びている
F	人差し指が伸びている
G	中指が伸びている
H	薬指が伸びている
I	小指が伸びている
J	親指以外が伸びている
K	曲がっている指がある
L	全部の指が曲がっている
M	人差し指以外が曲がっている
N	人差し指と中指だけが伸びている
O	中指の第三関節が伸びている
P	人差し指の第三関節が伸びている
Q	親指以外で第一と第二関節が伸びている指がある
R	人差し指と中指の第二関節が伸びている
S	人差し指と中指が交差している
T	親指と人差し指で輪を作っている
U	親指が中指についている

5. 指文字認識評価実験

5.1 数字指文字認識

本実験では、被験者一人を対象に、スマートフォンをインカメラで自分を撮影するように片手で持ってもらい、もう一方の手で連続数字指文字をしてもらった。被験者が表現する指文字は、ランダム生成された 0~9 の数字が 1 回ずつ出現する 5 桁の数字列である。この実験を 20 回繰り返した結果得られた正解率は、94%であった。図 3 に 2021 と入力した例を示す。

5.2 平仮名指文字

本実験では、被験者一人を対象に、スマートフォンをインカメラで自分を撮影するように片手で持ってもらい、もう片方の手で連続平仮名指文字をしてもらった。被験者が表現する指文字は完全ランダム生成された 5 文字の平仮名である。100 回繰り返した結果、全体の認識率は 79.2%であった。

「う」を「ら」とよく誤るように、特定の平仮名同士で、片方の平仮名が、もう片方の平仮名とよく誤って認識され

るケースがいくつかあった。そのため、正解率が低い平仮名指文字と、適合率が低い平仮名指文字のペアが幾つか出た。図 4 に「インタラクション」から、本システムにて対応している指文字である「いたらくしよ」を表現した例を示す。

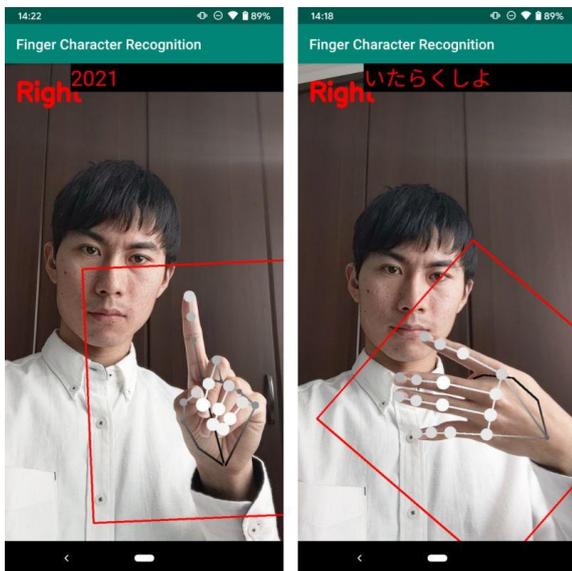


図 3 数字指文字認識

図 4 平仮名指文字認識

6. UbiScript

上述の指文字認識を活用して、シンプルな手指ジェスチャーによる操作指示（コマンドィング）について検討する。

手指ジェスチャーによる実用的なコマンドィングを考えると、指文字認識率の高い文字を選択的に用い、それと併せて数字を効果的に使うことが考えられる。

6.1 UbiScript Number

既に示している図 3 は、数字の認識のみを出力する基本アプリケーションとして様々な使い方ができる。セキュリティー等数字の組み合わせを必要とする場面や、機器制御で量的なセットを行う場面の基本となる。図 5 にパスワードロックを数字ジェスチャーで解除する例を示す。

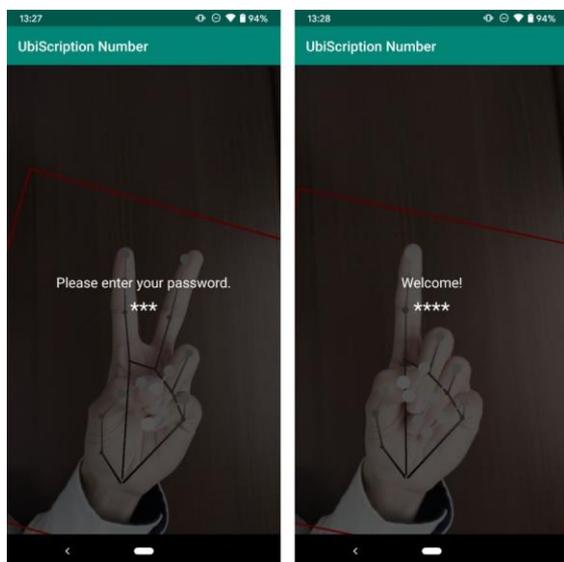


図 5 : UbiScript Number (セキュリティー)

6.2 UbiScript Color

UbiScript Color は平仮名指文字によって入力された単語に対応した色を表示したものである。一見、任意の色を選択するだけに見えるが、単語入力ができることにより、直感的な制御対象の選択が可能になる。図 6 に色を単語入力で選択する例を示す。

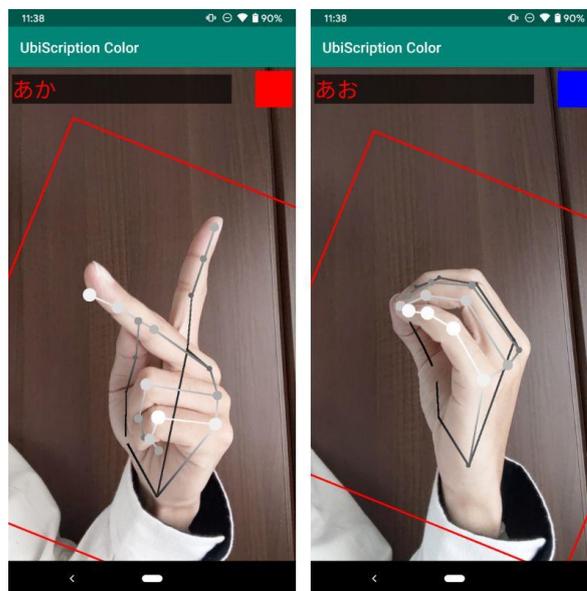


図 6 : UbiScript Color

6.3 UbiScript Chorus

量的なセットを担当する UbiScript Number と、色を選択する UbiScript Color を組み合わせることで、スマートフォンを用いてそれぞれに表示したものを一つの大きな画面に並べることによって、聴覚障がいを持つ人による「色のコーラス」が実現する。その場合数字シーケンスを各自が順に表現していく。

7. まとめ

本研究では、MediaPipe Hands をトラッキングシステムとして用いた、スマートフォン上での数字指文字認識と平仮名指文字認識を検討・実現し、さらにそれを用いた簡単な応用系を示した。数字指文字認識は単純な決定木分析で高い認識率を得られたが、平仮名指文字はランダムフォレストを使用しても 70~80%程となった。平仮名指文字については説明変数の工夫や、他の機械学習手法を使用するなど認識率の向上をめざしていきたい。さらに、数字指文字、平仮名指文字に加えて、アルファベット指文字の 3 種類の指文字を使用してのコマンドィングも今後、検討していく。場合によってはコマンドィングとしてはアルファベットの方が汎用性が高いであろう。

今回はまだ応用システム以前の基本部分を示したに過ぎないが、将来的には、UbiScript の実用的な展開系を模索するとともに、日常的には手指ジェスチャーを家電等の操作インターフェースとして、また聴覚障がいを持つ人向けには本ツールを使った新しい表現手段の提供を検討していきたい。

参考文献

- [1] “Leap Motion Developer”, <https://developer.leapmotion.com/>
(参照 2020/21/21)
- [2] “Open Pose”, <https://github.com/CMU-Perceptual-Computing-Lab/openpose> (参照 2020/21/21)
- [3] “Media Pipe”, <https://github.com/google/mediapipe> (参照 2020/21/21)
- [4] Erwin Wu, Ye Yuan¹, Hui-Shyong Yeo, Aaron Quigley, Hideki Koike, Kris M. Kitani. “Back-Hand-Pose: 3D Hand Pose Estimation for a Wrist-worn Camera via Dorsum Deformation Network.” UIST '20, October 20–23, 2020, Virtual Event, USA
- [5] “MediaPipe Hands” <https://google.github.io/mediapipe/solutions/hands.html> (参照 2020/21/21)
- [6] 船坂真生子, 石川由羽, 高田雅美, 城和貴. Leap motion controller を用いた指文字認識. 情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告, Vol.2015,No.8, pp.1-6, 2015