

ニューラニューラルネットワークを用いたハンドジェスチャ認識による ゲームユーザインターフェースの研究

A Study on Game User Interface Utilizing Hand Gesture Recognition with Neural Networks

郭 翔[†] 山富 龍[†] 齋藤 友彦[†] 二宮 洋[†]
Guo Xiang Yamatomi Ryou Saito Tomohiko Ninomiya Hiroshi

1. はじめに

ハンドジェスチャは、人間とコンピュータのインタラクションにおいて、直感的かつ自然な方法を提供する。特に、ゲームや仮想現実の分野において、ユーザが物理的なデバイスに依存しないインタラクションを行える点が重要である。しかし、ハンドジェスチャ認識は依然として多くの技術的課題を含んでいる。

一方、深層学習はコンピューターサイエンスの分野において革命的な進展を遂げている。ハンドジェスチャにおいても人間の自然な動作をコンピュータに理解させる可能性を秘めている。

石黒らは Media Pipe を用いたハンドジェスチャ認識を提案した[1]。Media Pipe は顔や手、ポーズなどのリアルタイム検出の機能を提供するライブラリである[2]。さらに文献[1]では、ハンドジェスチャにより操作可能な地図アプリを作成し、その操作性について検証を行った。

本研究では、Media Pipe を用いたハンドジェスチャ認識のゲームへの応用について検討を行う。特に、FPS (First-Person Shooter) ゲームなど、銃のような武器を用いたシューティングゲームを想定する。本研究では、まずハンドジェスチャにより操作できる簡易的な射的ゲームを開発する。射的ゲームに有用と考えられるジェスチャを考案し、Media Pipe とニューラルネットワークを用いたそれらのハンドジェスチャ認識アルゴリズムを提案する。そして、提案したアルゴリズムを射的ゲームに組み込み、認識精度や操作性に関して実験、検証を行う。

2. 準備

カメラ入力画像から手を検出するために、Google 社が提供するオープンソース機械学習ライブラリ Media Pipe を用いる[2]。Media Pipe はカメラ入力画像から手を検出し、21 個ある手指の各関節の 3D 座標を推測する。Media Pipe を用いて、手を検出した様子を図 1 に示す。また、21 個の関節と識別番号 (0~20) の対応関係を図 2 に示す。

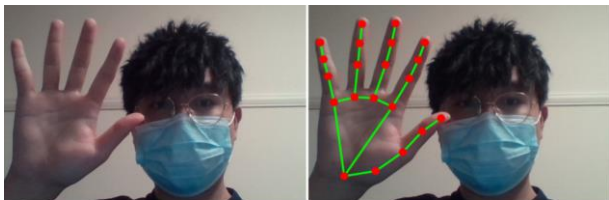


図 1 MediaPipe を用いて手を検出する様子



図 2 手の各関節の座標[1]

3. 従来研究

文献[1]では、Media Pipe を用いて、ジェスチャ認識を行っている。利用する座標は図 2 における点 8、点 1 2、点 1 6 である。これらの座標データを深層学習モデルの入力として扱い、ジェスチャの学習・推論を行う。なお、深層学習モデルとして多層パーセプトロン、畳み込みニューラルネットワーク、LSTM (Long Short-Term Memory) を用いている。

認識を行う具体的なジェスチャパターンは、人差し指の上下左右への移動、人差し指と中指の 2 本指によるプッシュ動作、それに薬指を加えた 3 本指によるプッシュ動作である。さらに、何も操作をしていないときの反応を防ぐため、上下左右動作時の XY 座標上の静止状態、プッシュ動作時の Z 座標上の静止状態もジェスチャとして認識している。

文献[1]では、さらに上記ジェスチャを用いた地図アプリを開発している。具体的な操作として、ブラウザ上の地図に対して、人差し指による上下左右への移動、2 本指のプッシュ動作による地図の拡大、3 本指のプッシュ動作による地図の縮小を実装している。

4. ハンドジェスチャを用いた射的ゲーム

4.1 射的ゲームの概要

本研究で開発した、射的ゲームの概要を説明する。射的ゲームの様子は図 3 の通りである。図 3 の左はゲームをプレイする様子であり、右はゲーム画面である。図 3 左のようにプレイヤーは手で銃のような形を作り、そのジェスチャに基づいて、ゲームを操作する。図 3 右において、青い円は射的ゲームの的を表している。プレイヤーは照準 (図 3 右の白い円) を的に合わせ、銃を発射させる。

銃の照準 (図 3 右の白い円) の座標は次のように定める。Media Pipe により検出された点 8 (人差し指の先) の座標を求める。この座標データはカメラ画像のフレーム内の相対座標として、0~1 の値で与えられる。与えられた座標に対して、ゲーム画面の幅と高さを掛けて得られた座標を銃の照準座標とする。なお、本研究ではノート PC 及び内臓の Web カメラ (ノート PC の中央上部) を使って実装しており、ゲーム画面は幅 800、高さ 600 ピクセルである。

[†]湘南工科大学 Shonan Institute of Technology

また、弾の発射に関するジェスチャに関して、本研究ではルールベースによるもの、及び機械学習によるものの二つを提案する。詳細は 4.2 節で述べる。

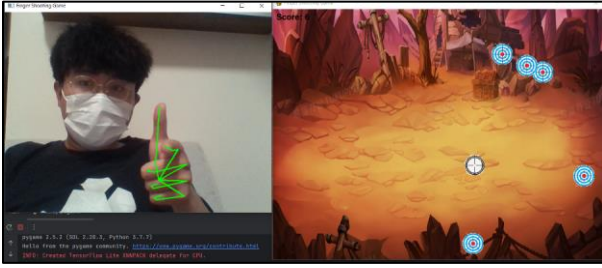


図 3 射的ゲーム

4.2 弾の発射の認識

4.2.1 ルールベースによる認識

弾の発射に関して、本研究では、親指を人差し指に近づけ、親指でボタンを押すようなジェスチャを行うことで、発射するものとする。このジェスチャ認識として二つのアルゴリズムを提案する。一つ目はルールベースによるものである。Media Pipe で得られた点 4 (親指の先) の座標を (x_4, y_4) 、点 5 (人差し指の付け根) の座標を (x_5, y_5) とする。点 4 と点 5 の距離 $d = \sqrt{(x_5 - x_4)^2 + (y_5 - y_4)^2}$ が閾値以下のとき、弾が発射されたと認識する。なお、本研究では閾値を 0.1 とし、実装を行った。

4.2.2 機械学習による認識

二つ目はニューラルネットワークを用いた機械学習によるものである。弾を発射した場合の手の画像、及び発射していない場合の画像をそれぞれ 1600 枚用意し、これらを訓練データとしてニューラルネットワークを構築する。本研究で用いたニューラルネットワークのモデルは多層パーセプトロン (MLP: Multi-Layer Perceptron) である。MLP はドロップアウト層と全結合層を含み、入力層は Media Pipe で得られた関節の座標データを XY 座標として入力する。入力層は平坦化され、128 ユニットを持つ全結合層を 2 つ持ち、それぞれに Dropout 層を挿入する。中間層の活性化関数には ReLU 関数を用い、出力層の活性化関数には Sigmoid 関数を用いて 2 クラス分類を行う。学習時のハイパーパラメータについて、バッチサイズ 128、エポック数 500、最適化手法は Adam、学習率は 0.001 とする。

なお、本研究では、4.2.1 節で述べたルールベースによるジェスチャ認識と比較するため、ニューラルネットワークの入力として、点 4 (親指の先) と点 5 (人差し指の付け根) の XY 座標を用いる。その他、点 0～点 8 の 9 点、及び点 0～点 20 の 21 点を入力としたものを合わせて提案し、それぞれの認識精度を 5.2 節で検証する。

5. 評価実験

5.1 照準に関する実験

4.1 節で述べた照準 (図 3 白丸) を定める手法に関して実験を行う。大学生 10 名に本研究で開発したゲームを各 10 回プレイしてもらう。なお、ゲームの的の位置は毎回異なり、ランダムに出現する。毎回のゲームにおいて照準を全ての的に当てるために要した時間を計測し、その平均を計算する。その結果を表 1 に示す。また、表 1 には比較とし

て、マウスを用いて照準を合わせるために要した時間も計測し、その平均を記している。

表 1 照準に関する実験の結果

	ジェスチャ	マウス
平均時間	2.11 秒	0.65 秒

5.2 弾の発射認識に関する実験

4.2 節で述べた弾の発射に関するジェスチャの認識精度に関して実験を行う。弾を発射した場合の手の画像、及び発射していない場合の画像、それぞれ 400 枚をテストデータとして用意し、各手法の認識率を計測する。その結果は表 2 に示す。表 2 において、ルールベースとは、4.2.1 で述べた手法であり、2, 9, 21 点モデルとは、4.2.2 で述べたニューラルネットワークによる手法において、入力として、2, 9, 21 点を用いたものである。

表 2 弾の発射認識に関する実験の結果

	認識率
ルールベース	76%
2 点モデル	85%
9 点モデル	95%
21 点モデル	98%

6. 考察

照準に関する実験では、ジェスチャ操作とマウス操作を比較すると、ジェスチャ操作の方が照準を合わせるための時間がかかり、操作性に関して課題が残っていると言える。この原因の一つとして、ユーザの慣れ不足が考えられる。今後はユーザのトレーニングを十分に行った上での検証が必要である。

弾の発射認識に関する実験では、ルールベースとニューラルネットワークを用いた手法を比較すると、2 点モデルにおいても認識精度が高く、ニューラルネットワークを用いた手法の方が優れていると言える。また、利用する関節の点の数を増やすほど、認識率は高くなり、弾の発射のようなジェスチャにおいても、手の全ての関節の点の情報が重要であることがこの実験から分かる。

7. まとめ

本研究では、ハンドジェスチャ認識のゲームへの応用について検討を行った。簡易的な射的ゲームを開発し、それに適したジェスチャを考案し、その認識アルゴリズムを提案した。今後の課題は様々なゲームとそれに適したハンドジェスチャを提案し、その操作性とエンターテインメント性の検証を行うことである。

謝辞

本研究は JSPS 科研費 22K02811 の助成を受けたものです。

参考文献

- [1] 石黒大樹, 尾関智子, “深層学習を用いたハンドジェスチャ分類による非接触ユーザインターフェースの開発,” 人工知能学会全国大会論文集, Vol. 36, ROMBUNNO. 3F3-GS-9-01, 2022.
- [2] Google, 「Media Pipe Solutions guide」, <https://ai.google.dev/edge/media-pipe/solutions/guide> (2024-06-14 閲覧)