4ZE-04

指差しポインティングシステムにおける指先ジェスチャの検出方法の研究

野田雄希† WONG CHING JIN‡ 水谷晃三†‡
帝京大学大学院理工学研究科† 帝京大学理工学部情報電子工学科‡

1. はじめに

人が行う指差しジェスチャをセンサで検出し，その指差し方向にポインタを表示したり，ジェスチャに対応した操作を実行したりする研究が行われている．筆者らは天井に下方に向けて設置した RGB-D センサを用いてこれを実現する方法を検討し，指差しポインティングシステムの開発を行った[1]．本研究では実用性の向上のため，指を動かすジェスチャによりマウスのクリックに相当する操作ができるユーザインタフェース（以下クリックジェスチャと呼ぶ）の実現を目指す．

2. 天井に設置した RGB-D センサによる指差しポインティングシステムの概要

図 1 に筆者らの方式による指差しポインティングシステムの概要を示す．RGB-D センサを天井に下方に向けて設置する．こうすることで，複数ユーザが重なるオクルージョンの発生が軽減され，ユーザの正面や側方などに設置したセンサを用いる既存研究の方式[2,3]に比べて，指差しの方向に対する柔軟性や複数人の指差しジェスチャを同時に捉えることが容易になっている（図 2）．将来的には複数のセンサを用いて室内全体で複数人が同時に使用できる指差しポインティングシステムの実現を目指している．

3. 深層学習による指差しジェスチャの検出

指を伸ばした状態から指を曲げた状態に変化し，一定時間指を曲げた状態が継続した後，再び指を伸ばした状態に戻る動作したとき，クリックジェスチャを行ったとみなす．このとき手の甲は動かさず，指だけを動かすものとする．筆者らの本方式では上方に設置した RGB-D センサを用いるため，既存手法と比べ，前述の特徴

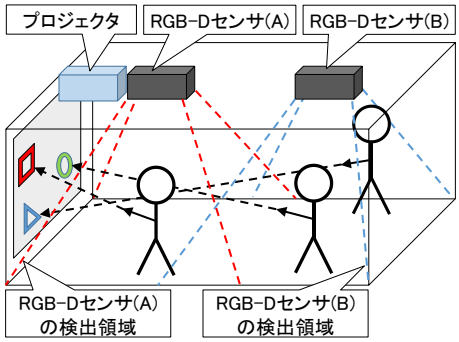

図 1．システム概要図

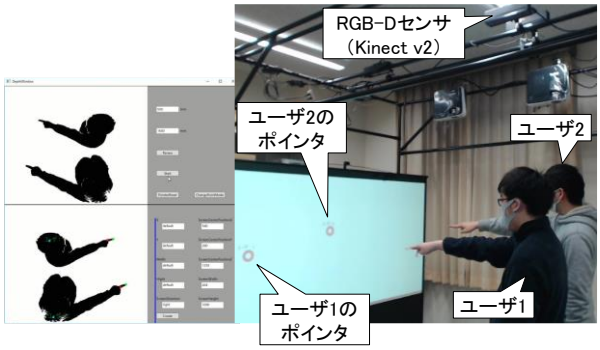

図 2．複数人でのシステム使用例
（左：システム画面，右：使用の様子）

を維持しながら手の甲と指の関係を捉えることができ，クリックジェスチャの判定が可能になると考えられる．

ジェスチャの検出方法として，センサから取得した手や指の深度値・座標の値の変化から判定する方法が考えられる．しかし，指などの細い領域の深度値はセンサの特性上不安定になりやすい傾向があり，誤検出の原因になりうる．そこで，指部分だけではなく手全体を含む深度画像を用いて深層学習により指先の状態を識別することでクリックジェスチャの検出を行う方式を試みる．深度画像とは，センサから取得した深度データを任意の深度範囲でグレースケール画像化した画像である．ニューラルネットワークに指を伸ばした状態と指を曲げた状態の 2 種類の深度画像を与えることで，指先の状態を識別しモデルを生成する．生成したモデルを用い

A Study of the Detection Method for Finger Tip Gesture on Finger Pointing System
Yuki Noda†, WONG CHING JIN‡, Kozo Mizutani†‡
†Graduate School of Science and Engineering, Teikyo University.
‡Department of Information and Electronic Engineering, Faculty of Science and Engineering, Teikyo University.

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.

てクリックジェスチャを検出するには，まずモデルを用いて深度画像に写っている手の位置とその手の状態を一定フレームごとに識別し，記憶しておく．そして，指を伸ばした状態から指を曲げた状態が一定フレーム継続し，再び指を伸ばした状態というパターンが検出されたとき，クリックジェスチャを行ったと判定する．

4. 検出モデルの作成と評価

指先ジェスチャの検出モデルの作成には TensorFlow Object Detection API を用いた．同 API 上で公開されている学習済みの物体検出モデル EfficientDet D0 512x512 を使用した転移学習を行う．バッチサイズは 4，ステップ数 10000 で学習を行った．

学習画像として使用する深度画像は床から 2.5m の高さに設置した Kinect v2 により取得した深度データから作成した．画像化する深度範囲が狭いほど物体の凹凸が濃淡で表現されやすいため，手の甲の深度値を基準に上限 100mm，下限 100mm の深度範囲を画像化した．学習画像には，指を伸ばした状態の深度画像 734 枚，指を曲げた状態の深度画像 722 枚を用意した．2 種類の状態の深度画像は，右手・左手，指差し方向が水平・上・下の組み合わせ全 6 通りの画像となっている．図 2 に学習画像の例を示す．学習の際，(a)水平反転，(b)垂直反転，(c)反時計回りに 90° 回転，(d)拡大・縮小し，一部を切り抜き正方形にリサイズ，これら 4 つの中からランダムな組み合わせのデータ拡張を深度画像に対して一定の確率で行い，学習画像として使用している．作成したモデルに対し，585 枚の深度画像を別途用意して評価☆した結果，IoU＝0.75 で AP＝0.736，IoU＝0.5-0.95 で AR が 0.683 となった．

作成したモデルを用いてクリックジェスチャの判定を行うプログラムの試作を行った．図 3 にプログラムの実行例を示す．クリックジェスチャを行ったとき，指先の状態を識別し，クリックジェスチャを検出できることを確認した．しかし，指先状態の誤認識によりジェスチャの誤検出が発生することも確認した．

5. 考察

指先の状態の誤認識の主因として学習画像の不足があると考えられる．各ジェスチャの学習画像を増やすことで識別精度がより向上すると考えられる．また，複数人同時に写っている画像や個人ごとの体格の違いを含む画像の追加など，学習画像のバリエーションを増やすことで精度が改善すると考えられる

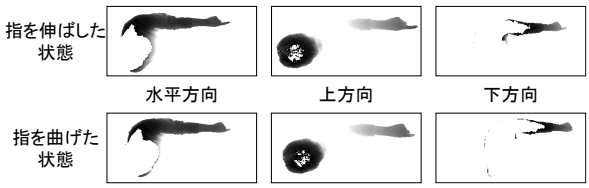

図 2. 学習に用いた深度画像の例（一部分）

図 3. 試作プログラムの実行例

また，深度画像を用いる方法はグレースケール化により深度情報が 255 段階に丸められてしまうため欠落が発生する．検出精度の向上のために，深度値を直接与えるニューラルネットワークを用いる方法が考えられる．

6. おわりに

本研究では，天井に下方に向けて設置した RGB-D センサを用いた指差しポインティングシステムにおいて，指先ジェスチャによるユーザインタフェースを実現するため，深層学習により指先の状態を識別し，ジェスチャを検出する方法を検討した．指先の状態を識別するモデルを用いることで，指差しポインティングシステムでクリックジェスチャを利用したインタラクションの実現が期待できる．

謝辞

本研究の一部は JSPS 科研費 JP18K11580，21K12163 の助成を受けた．

参考文献

[1] 野田雄希, 水谷晃三, 天井から下方に向けて設置した RGB-D センサによる指差しポインティングの研究, 情報処理学会第 83 回全国大会講演論文集, 5ZB-08, 2021 .

[2] Dai Fujita, Takashi Komuro: Real-time 3D Hand Pointing Recognition using Appearance Difference between Two Camera Images, The 3rd IAPR Asian Conference on Pattern Recognition (ACPR 2015) Program Booklet, pp.36-37, 2015.

[3] K. Hu, S. Canavan, L. Yin: Hand Pointing Estimation for Human Computer Interaction Based on Two Orthogonal-Views, Proceedings of International Conference on Pattern Recognition, pp. 3760–3763, 2010.

☆COCO Object Detection Challenge の評価指標．画像あたりの最大検出数を 100 とした．

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.