

手の動きに追従する UI のための深度センサを用いた タップジェスチャの認識方法

田村 律起[†] 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

筆者らは、天井から下方に向けて設置した RGB-D センサとプロジェクタにより、ユーザの手のひらに UI を投影し、操作するコンピューティング環境の研究を行っている[1]。本稿では、本研究の試作システムにおいてタップジェスチャを認識するための方法について述べる。

2. 手の動きに追従する UI の概要

図 1 に手の動きに追従する UI を実現するための試作システムの概要、図 2 にシステムの処理の流れを示す。天井に設置したセンサとプロジェクタを用いて利用者の手に追従するように UI を投影する。先行研究では、センサから取得した深度値を二次元の深度画像に変換し、これに対して手の形を学習させた物体検出器により手の領域を推定する。推定された手に対して手の領域の位置や大きさなどの情報を含むタグを付与し、前後のフレームで類似するタグを同一のものと判別しながらタグの情報を更新することで同じユーザの手を識別して UI を継続的に投影する。これにより、複数人のユーザにそれぞれ別の UI を投影できるようにしている。本研究では、投影された UI の操作を実現するために、一方の手のひらを他方の手の指でタップするジェスチャを認識する方法について検討する。

3. タップジェスチャ認識方法

3.1. 関連研究

Sun らは、タップ時の指先の色の変化により平面上へのタップを認識する EigenNail 法を提案した[2]。筆者らは、これを参考に手のひらへのタップ時の指先の色の変化を RGB センサで捉えることで、タップを認識する方法を検討した[3]。この方法では、被験者の指の色の個人差や投影光の影響などいくつかの課題が残っている。また、深度センサからの値を用いる方法として、

A Method for Tap Gesture Recognition of the projected UI on User's Hands Using a Depth Sensor

Ritsuki Tamura[†], Kozo Mizutani^{†‡}

[†]Graduate School of Science and Engineering, Teikyo University.

[‡]Department of Information and Electronic Engineering, Faculty of Science and Engineering, Teikyo University

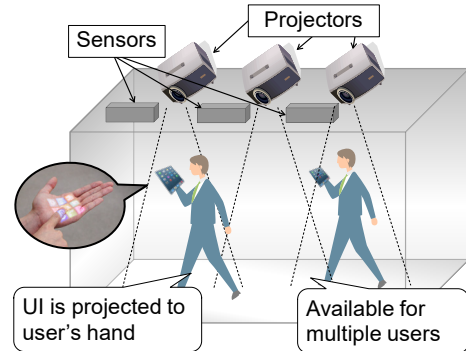


図 1 試作システムの概要図

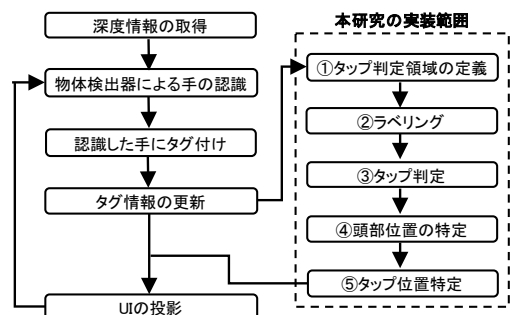


図 2 試作システムの処理の流れ

Harrison らはユーザの肩に装着したセンサとプロジェクタにより手のひらや腕に UI を投影する OmniTouch を提案した[4]。この方法では、本研究のように遠距離からの認識や複数のユーザを対象にした同時認識への対応が困難である。天井に配置したセンサと手の間に 1m 程度の距離があり、センサの精度を踏まえて指先の数 cm の動きの変化を捉えることが課題となる。

3.2. 認識方法

本研究におけるタップ認識方法では、画像処理としてよく用いられるラベリング手法と深度値による判別方法を組み合わせた方法を用いる。図 3 は、試作システムのモニタ画面での判定状態の表示にタップ判定領域などの説明を加えた認識方法のイメージ図である。

①まず、手が複数認識されているとき手の領域同士が重なっているかどうかの判別を行う。重なっていた場合、物体検出器が手を推定できなくなりタグ情報が更新されなくなる。そこで 2 つのタグ領域を含む矩形範囲を「タップ判定領域」として新たに定義し、それを対象に画像ラベリングを行う。②ラベリングでは、深度画像

上の画素値だけでなく対応するピクセルの深度値も参照し、両方の手が重なって隣接しており、なおかつ深度値の差が閾値 D_{th} 以下だった場合、同一のラベルを付与する。このとき、両方の手が接触している状態ではタップ判定領域内にラベルが1つしか存在していないことになり、手が離れている場合には、ラベルが2つ存在していることになる。これにより、③前者の状態をタップ状態と判定する。実際には深度センサのノイズなどにより領域内に3つ以上のラベルが存在することがあるため、ピクセル数が一定以下のラベルについては除外する。

タップ位置の決定は、まず、④タップ判定領域内のラベリングされたそれぞれの領域から、深度画像上で隣り合っているピクセル、つまり人体部分のピクセルを走査し、その中で深度値が小さい（センサに近い）位置に存在する一定数のピクセルの重心を頭部中心位置とする。次に、⑤この中心位置から最も離れた手のピクセルを指先と判定する。指先はラベリングされた手ごとに存在するため、上側に重なっている手の指先をタップ位置特定対象とする。そしてタップを認識したとき、その直前フレームで指先が存在していた位置をタップ位置とする（図4）。

4. 実験と考察

前述の方法を試作システムに実装して簡易な評価実験を行った。実験では Kinect V2 センサを手のひらから約 1.3m 上方の位置に下方へ向けて設置した。センサの真下で、一方の手の指で他方の手のひらを 1 秒間タップし、1 秒間離す動作を 10 回繰り返す。それを、センサを基準にして立ち位置を変えて上下左右の 4 方向で行い、それぞれの状態の出現回数を算出した。ただし、指の上下移動にかかる時間もあるため、それぞれの状態が継続した時間基準を 0.7 秒以上または 0.8 秒以上としたうえで回数を算出した。その結果を表 1 に示す。時間基準 0.7 秒以上では、 $D_{th}=10\text{mm}$ のときに、タップ状態が 40 回、非タップ状態が 34 回となり、92.5% の正確性でタップ状態が認識されたことを確認した。時間基準の違いで比較すると、 $D_{th}=5\text{mm}$ に対して 10mm のときは回数が減り悪化している傾向がある。これは、閾値が大きい分、非タップ状態の継続時間が短くなっているためであると考えられる。

なお、本手法ではスワイプやフリックなどのタップ後の操作の認識が原理的に困難である。また、現状ではタップジェスチャによる UI 操作の仕組みが未実装である。これらの解決が今後の課題である。

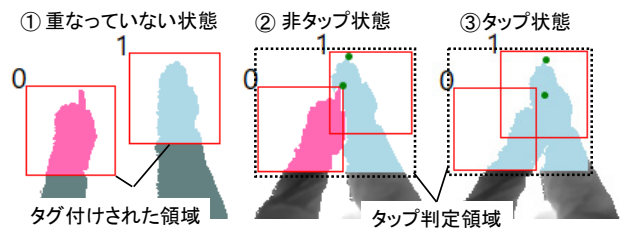


図3 ラベリングによるセグメンテーションとタップ認識

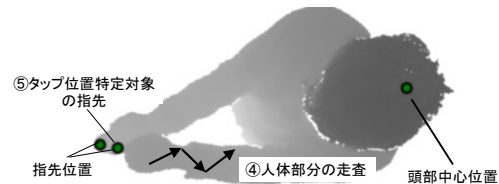


図4 タップ位置の特定

表 1 試作システムのタップ判定結果

時間基準(秒)	判定結果 (回数)				
	>0.7		>0.8		
状態	タップ	非タップ	タップ	非タップ	
閾値 D_{th} (mm)	5	35	38	32	35
	10	40	34	40	14

5. おわりに

本研究では、手に追従する UI のためのタップジェスチャの認識方法として、深度センサと画像ラベリングを応用した方法を検討し、システムを試作した。今後は正確な評価や改良などを行い実用的なシステムの実現を目指す。

謝辞

本研究の一部は JSPS 科研費 21K12163 の助成を受けた。

参考文献

- [1] 岩崎真大, 松岡雄介, 土屋博雅, 水谷晃三, 荒井正之. 手の動きに追従する UI における手の認識とタグ付けによる判別の研究, 情報処理学会 第 80 回全国大会, 6ZB-05, 2018.
- [2] Y. Sun, J. M. Hollerbach and S. A. Mascaró. EigenNail for Finger Force Direction Recognition, Proceedings 2007 IEEE International Conference on Robotics and Automation, pp.3251-3256, 2007.
- [3] 田村律起, 野田雄希, 水谷晃三. 手の動きに追従する UI におけるタップジェスチャの認識方法の研究, 情報処理学会第 85 回全国大会, 5ZD-06, 2023.
- [4] Harrison, Chris and Benko, Hrvoje and Wilson, Andrew D.. OmniTouch: Wearable Multitouch Interaction Everywhere, Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, pp.441-450, 2011.