

## 仮想現実環境における入力のための陰影に基づく指と皮膚の接触判定

関谷 悠冬<sup>†</sup> 梅澤 猛<sup>‡</sup> 大澤 範高<sup>‡</sup>千葉大学工学部総合工学科情報工学コース<sup>†</sup> 千葉大学大学院工学研究院<sup>‡</sup>

## 1. はじめに

バーチャルリアリティ (VR) 技術の発展に伴い、VR 環境に適した入力操作手法が研究されている。中でも、皮膚に指で触れることで入力を行う手法は、触覚フィードバックを得られる点で優れている。川口らは掌上にフリックキーボードを投影する手法を提案した [1]。また、Harrison らは皮膚を含む様々な表面に入力 UI を投影できる OmniTouch を開発した [2]。しかしながら、これらの手法では手や指の位置認識に深度センサが用いられており、その精度が不十分であるために、誤入力が多い。

本研究では、指が皮膚に触れた際の影やテクスチャの変化に着目し、指先周辺の RGB 画像から Convolutional Neural Network (CNN) を用いて接触を判定する手法を提案する。

## 2. 陰影に基づく接触判定

前腕に対して指で触れることによる入力動作と、その様子をヘッドマウントディスプレイ (HMD) に搭載した RGB カメラで撮影して認識するシステムを想定する (図 1)。



図1 入力動作

接触判定は指先抽出、画像補正、判定の 3 フェーズによって行う。指先抽出フェーズでは、撮影した手の映像をトリミングし、指先周辺のみを生成する。画像補正フェーズでは、指先周辺画像のコントラストや色味の補正を行う (図 2 右)。判定フェーズでは、補正した指先周辺画像を入

力とする CNN によって接触の判定を行う。



図2 指先周辺画像例 (左: 補正前 右: 補正後)

## 3. 実験条件

画像補正手法とデータ拡張の有効性の調査と判定精度の評価のため、実験を行った。

## 3.1. 入力動作の撮影

机から垂直方向に約 20cm 離れた位置にカメラを設置し、机においた腕に対する入力動作を撮影した。撮影した映像は 24bitRGB, 1080×1920 ピクセル, 30fps である。指で触れる、離すを 1 秒程度の間隔で繰り返し、10 秒程度の映像を撮影した。被験者は 3 人について、それぞれ光源等の環境を変えて 3 回撮影し、合計 9 回分の入力動作映像を得た。

## 3.2. 指先抽出フェーズ

まず、手の映像から骨格情報を推定する。本実験では Google の提供する手指位置推定モデル MediaPipe Hands [3] を使用し、手首、各関節、各指先の計 21 点の XYZ 座標を推定した。

次に、骨格情報をもとに画像上における指先位置、指の向く方向、指の長さの 3 つを算出する。指先周辺画像の一辺の大きさ  $S$  は、指の幅のおおよそ 2 倍となるように、産業技術総合研究所の統計データ [4] をもとに、 $S = \text{指の長さ} \times 0.355$  として算出した。

最後に、指先位置を中心として指の方向が上向きになるように回転した後に、 $S \times S$  の正方形を切り出し、これを指先周辺画像とする。

## 3.3. データセットとデータ拡張

指先周辺画像には、非接触時を 0、接触時を 1 として手動でラベル付けを行った。

1 回分の入力動作映像から、入力動作に用いない親指を除いた各指で、接触時・非接触時それぞれ 95 枚、計 760 枚の指先周辺画像を生成した。これを 1 ブロックとし、9 ブロック合わせてオリジナルデータセットとした。さらに、指先抽出

Detection of Contact between Finger and Skin based on Shadow for Input in Virtual Reality Environment

<sup>†</sup> Yuto Sekiya, Department of Information Engineering, Faculty of Engineering, Chiba University

<sup>‡</sup> Takeshi Umezawa, Noritaka Osawa, Graduate School of Engineering, Chiba University

フェーズにおいて切り出し時の指先位置に変位を加えることでデータ拡張を行い、位置シフトデータセットとした。変位は、XY方向それぞれ独立に、指先周辺画像の一辺の大きさ  $S$  に対して  $-S/4 \sim S/4$  の範囲で一様乱数を設定した。比較のため、オリジナルデータセットと位置シフトデータセットを結合して複合データセットとした。各データセットのサイズを表1に示す。

表1 各データセットに含まれる画像枚数

データセット	接触	非接触
オリジナル	3420	3420
位置シフト	3420	3420
複合	6840	6840

### 3.4. 画像補正フェーズ

以下の6つの補正手法それぞれを適用した場合を比較し、判定精度への影響を検証する。

- 補正なし
- MinMax正規化
- 標準化
- 標準化 + MinMax正規化
- グレースケール変換
- MinMax正規化 + グレースケール変換

### 3.5. 判定フェーズ

50×50ピクセルにリサイズした補正後の指先周辺画像を入力として、接触/非接触の2値判定を行うCNNモデルを構築した(表2)。最適化関数はAdamax, 損失関数はbinary cross entropyを使用し、50エポックで学習を行った。

表2 接触判定モデルの構造

Layer	OutputShape	Activation
Input	(50, 50, 3)	-
Conv2D	(50, 50, 32)	relu
Conv2D	(48, 48, 32)	relu
MaxPooling2D	(24, 24, 32)	-
Conv2D	(24, 24, 64)	relu
Conv2D	(22, 22, 64)	relu
MaxPooling2D	(11, 11, 64)	-
Flatten	(7744)	-
FC	(512)	relu
FC	(1)	sigmoid

オリジナルデータセットの評価の際には、各ブロックを分割されたデータセットとして9分割交差検証を行った。位置シフトデータセット及び複合データセットの評価の際には、そのうち8ブロックを学習に用い、残りの1ブロックに対応するオリジナルデータセットのブロックをテストデータとして使用して9分割交差検証を行った。

## 4. 実験結果

画像補正とデータ拡張の組み合わせごとの正解率を図3に示す。複合データセットでグレースケール変換を行った場合に最も正解率が高くなった。

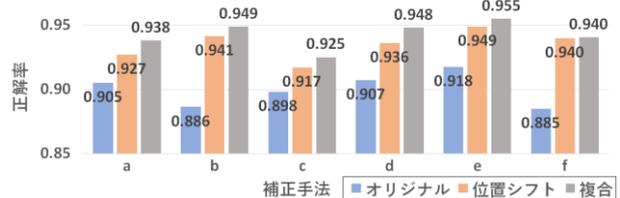


図3 各データセットと画像補正手法での正解率

## 5. 考察

位置シフトのデータ拡張を行った場合、指先位置が画像中心からずれているデータに対する判定精度が向上したため、正解率が高くなった。

標準化は、他の補正手法とは異なり、補正後の画素値の範囲が定まらないことが精度の低下を招いたと考えられる。MinMax正規化はコントラストを強調するような効果が得られるため、画像全体の明度が低いデータに関して判定精度の向上が見られたが、腕面の微妙な明暗を強調してしまうことによる誤判別が起こった。また、指先周辺画像はほぼ全面が肌の色であり、光源環境以外での色の変化が少ないことから、光源色等の差を吸収できるグレースケール変換が有効であったと考えられる。

## 6. おわりに

本研究では、影やテクスチャの変化をもとに、指と皮膚の接触をCNNによって判定する手法を提案した。5つの画像補正手法とデータ拡張の組み合わせを比較し、その効果と有効性を検証した。また、正解率より、本手法によってRGB画像から指と皮膚の接触判定が可能であることが示唆された。

### 【参考文献】

- [1] 川口航平ら. VR向けの掌上における日本語フリック入力手法の提案. ヒューマンインターフェースシンポジウム2019, NO.5B1-3, (2019).
- [2] Chris Harrison *et al.* OmniTouch: Wearable Multitouch Interaction Everywhere. Proceedings of the 24th annual ACM symposium on user interface software and technology, pp.441-450, (2011).
- [3] Fan Zhang *et al.* MediaPipe Hands: On-device Real-time Hand Tracking. arXiv:2006.10214, (2020).
- [4] 河内まき子. 2012: AIST 日本人の手の寸法データ. <https://www.airc.aist.go.jp/dhrt/hand/index.html>.