

陰影変化に基づく指先と柔らかい表面の接触判定精度向上

新井 祥悟[†] 梅澤 猛[‡] 大澤 範高[‡]千葉大学工学部総合工学科情報工学コース[†] 千葉大学大学院工学研究院[‡]

1. はじめに

仮想現実感技術、拡張現実感技術の発展に伴い、仮想空間内、拡張現実空間内でのさまざまな文字入力手法が検討されている。仮想キーボードを用いたキー入力のために、川口ら[1]はヘッドマウントディスプレイに Leap Motion を取り付け、1人称視点による掌上における文字入力手法を提案した。1人称視点による接触判定は表示装置と一体化できるため、手にデバイスを装着せずに済むというメリットがある反面、横方向から空間検知をすることができないため、判定精度の向上が課題となる。

先行研究[2]ではヘッドマウントディスプレイ前部に取り付けた RGB カメラを用いて、前腕部と指先の接触判定を行なっている。前腕部にもう一方の手の指先で触れる様子を撮影し、接触の際に生じる指先周辺の陰影変化を基に深層学習モデルを用いて接触判定を行なっている。その後の研究で時系列データを扱う深層学習モデルを適用することで、高精度な接触判定が可能であることが示唆されている。しかし照明条件による判定への影響が大きく、接触面の指先の陰影変化が小さい場合に接触判定精度が低いことが課題として挙げられる。また学習データに用いた指先と皮膚の接触映像が、全て同一人物の映像であったため、学習データに用いていない未知の人物に対する接触判定精度が低いことも課題であった。

本研究では陰影変化が小さい環境における複数人のデータを多数収集して、深層学習の訓練データとすることで、陰影変化が少ない環境および、未知の人物における接触判定精度の向上を図る。また、明度変換によるデータ拡張の効果も評価する。4つの深層学習モデルを用いて F1 Score を算出し、評価する。

2. アプローチ

2.1. データ収集

片方の手の前腕部に対し、他方の手で人差し指・中指・薬指・小指の順に1秒間隔で1回ずつ接触する動作を1本の動画として、14人の参加者、10箇所の撮影環境下で計84本の動画を撮影した。次に、撮影した動画に Mediapipe Hands を適用して手指の位置情報を取得した。その情報を基に全フレームに対して、指先が上向きになるように指先画像(人差し指・中指・薬指・小指)を作成した。1本の動画から指先画像系列を4つ、計 $84 \times 4 = 336$ 系列を作成し、この内12人分の320系列をデータセット T1、2人分の16系列をデータセット V1 とした。なお、すべての指先画像には接触(1)、非接触(0)としてラベル付けを行った。

2.2. データ拡張

照度差が小さい環境における接触判定精度向上のため、データセット T1 に明度変換を適用することでデータ拡張を行いデータセット T2 とした。同様にデータセット V1 からデータセット V2 を作成した。明度変換には、OpenCV 4.5.5.64 の `convertScaleAbs` 関数を用い、コントラストの倍率は 0.5, 1.5 とした。したがって、データセット T2 のデータ数は明度変換を適用しない元の画像を含め、 $336 \times 3 = 1,008$ 系列である。

2.3. モデルの検討

深層学習モデルとして次の4つを比較評価した。

- LSTM (Long Short-Term Memory)
- GRU (Gated Recurrent Unit)
- ConvLSTM (Convolutional Long Short-Term Memory)
- ViT (Vision Transformer)

LSTM, GRU, ConvLSTM のそれぞれを用いたモデルでは特徴抽出モデルと判定モデルを組み合わせしており、特徴抽出モデルでは共通して CNN (Convolutional Neural Network) を、判定モデルではそれぞれ LSTM, GRU, ConvLSTM を用いて時系列データを処理し接触判定を行った。入力画像サイズは 50×50 pixels とした。ViT では事前学習済みの ViT-B/16 を用いており、CNN を用いずに ViT で特徴抽出と判定を行う。パッチサイズが 16 であるため、入力画像サイズは 16 の倍数である 64×64 pixels とした。

Accuracy improvement in contact detection between fingertips and soft surface based on shadow variation

[†]Shogo Arai, Department of Information Engineering, Faculty of Engineering, Chiba University.

[‡]Takeshi Umezawa, Noritaka Osawa, Graduate School of Engineering, Chiba University

3. 接触判定

3.1. データ収集とデータ拡張の効果検証

データ量の増加が推定結果に与える影響を検証するため同じ学習モデルを使った比較を行なった。先行研究[2]を基にした LSTM を用いる深層学習モデルにより 1 人分のデータを訓練データとして学習したモデル(A)を比較対象とし、同じ構造でデータセット T1 を訓練データとしたモデル(B)、データセット T2 を訓練データとしたモデル(C)を使ってデータセット V1 と V2 を推定したときの F1 Score を図 1 に示す。データセット V1 を用いた評価の F1 Score が A の 0.2003 に比べ B では 0.8897 に改善したことより、複数人によるデータを深層学習の訓練データとすることで未知の人物に対する接触判定精度が向上したことを確認した。

データセット V2 を用いた評価の F1 Score が B の 0.8570 から C では 0.9481 に改善したことより、明度変換によるデータ拡張の効果も確認した。図 2 は、A と B では実際は接触しているが非接触と判定され、C では正しく接触と判定された指先画像の例である。収集データの拡充と明度変換によるデータ拡張によって、図 2 に示すような陰影変化が小さい環境においても正しく判定できる場合が増加したことを確認した。

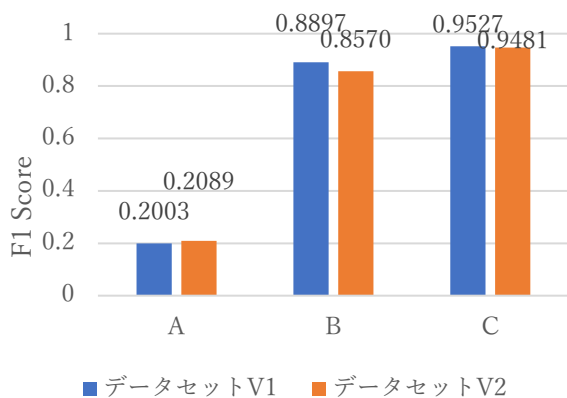


図 1 データ収集とデータ拡張の効果

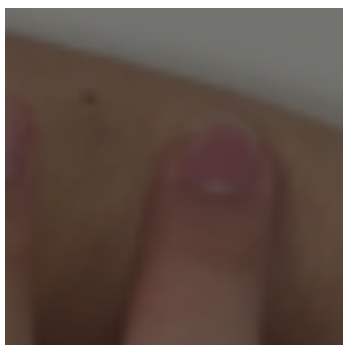


図 2 C でのみ正しく判定できた画像

3.2. モデルの比較

データセット T2 を訓練データとして学習した 4 つの深層学習モデルに、データセット V2 を用いて評価した結果を表 1 に示す。F1 Score が高い順に GRU の 0.9667, LSTM の 0.9481, ConvLSTM の 0.9299, ViT の 0.7772 であった。ViT の F1 Score が低かった原因として、適合率が 0.7894, 再現率が 0.7655 で両方とも低いことより、接触しているが非接触、非接触だが接触と推定されていることが他のモデルよりも多いことが挙げられる。またすべてのモデルにおいて、推定結果における接触の始まりと終わりに対し正解ラベルの切り替わりタイミングにずれがあった。ラベル付けは手動で行ったが、接触の始まりと終わりを目視で判断することが困難であったため、ずれの発生はやむを得ない。しかし実用的な文字入力のためには、違和感を生じない時間的許容範囲内で接触、非接触が正しく検知できることが重要であり、キー入力のための適切な評価指標の検討が今後必要である。

表 1 モデルの評価結果

モデル	F1 Score	適合率	再現率
LSTM	0.9481	0.9929	0.9071
GRU	0.9667	0.9619	0.9715
ConvLSTM	0.9299	0.8895	0.9740
ViT	0.7772	0.7894	0.7655

4. まとめ

本研究では訓練データの量および多様性を増やすことと明度変換によるデータ拡張の効果を確認した。4 つの深層学習モデルの比較を行い、GRU を用いたモデルが最も高い F1 Score であった。

参考文献

- [1] 川口 航平, 磯本 俊弥, 志築 文太郎, 高橋 伸, "VR 向けの掌上における日本語フリック入力手法の提案", ヒューマンインタフェースシンポジウム論文集, pp.676-682, 2019
- [2] Y. Sekiya, T. Umezawa and N. Osawa, "Detection of Finger Contact with Skin based on Shadows and texture around Fingertips", *Human-Computer Interaction: Interaction Techniques and Novel Applications*, vol.12763, pp.109-122, 2021