

USB カメラを用いた AR 環境における入力システムの改善

1. はじめに

近年, AR グラスやスマートグラスといった装着型のデバイスが増加している. 現在は深度センサ内蔵デバイスや付属のコントローラを用いて操作するのが主流だが, デバイスが高額な価格設定や直感的な操作ではない問題がある. 例えば, 指文字の手の形をパターンマッチングして文字を入力する手法[1]が提案されているが, 指文字を覚える必要があり, 学習にコストがかかる. また, 先行研究として指先と掌を使った直感的な入力システム[2]も提案されているが, 二次元画像からの手の大きさの差を用いて入力判定を行っているため, 個人差の影響を受けやすい. そこで, 本研究では接触判定に機械学習を用い, 認識精度を向上させる方法を提案する.

2. 提案手法

2.1 入力システムの概要

先行研究[2]と同様に, 本研究では図 1 のように左手の掌をキーボードとし, 右手の人差し指の接触を検知する. レイアウトはスマートフォンで日本語文字入力できる最低限の 3×4 を採用した. また, 2 次元画像から手の位置を 3 次元座標として予測できる MediaPipe を使用した. ただし, 奥行き情報は必ずしも正しいものが取得できないため, 掌と人差し指が接触しているか否かを別の機械学習によって 2 クラス分類する. 誤入力を避けるため, 入力のタイミングは掌から人差し指が離れたときとした.

2.2 システム構成

本研究では手全体が画角に収まるように USB カメラを左肩に固定した. 画像サイズは 640×480 ピクセル, 処理速度は 20 fps 程度である. 現状の動作中の画面出力はディスプレイにて行ったが, 将来的にはユーザが装着した AR 機器上で自分の手を見て使用することを想定している.

2.3 接触判定

MediaPipe で予測された座標点を使用し, SVM を用いて, 左手の掌と右手の人差し指が接触しているか否かを 2 クラス分類で判定する. 最初に, 左手の掌 7 点と右手の人差し指 4 点を選択し, 計 11 点の 3 次元座標を取得する. 次に, 各特徴点について前フレームとの差分を計算する. これにより, 各特徴点ごとに速度を取り入れ, 合計で 66 次元の特徴ベクトルを作成する. これを SVM で学習させ, 接触判定のモデルを作成する. モデルが「接触している」と予測した後の一定時間経過後に, 「接触していない」と予測した場合, 前 12 フレームから現在フレームの中で人差し指の先端と最も重なっている文字が入力される.

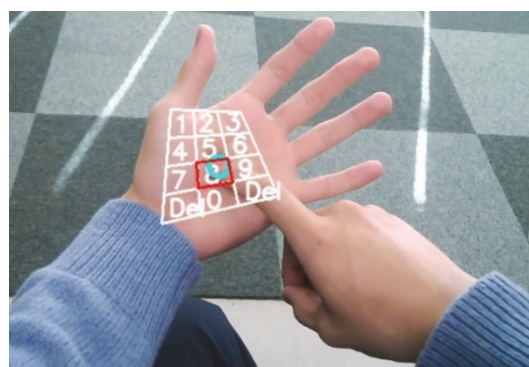


図 1 本システムの動作様子

表 1 性能比較実験の結果

	先行研究	本研究
KSPS	0.42	0.45
ER	34%	22%

3. 評価実験

先行研究と提案システムの性能比較実験を行った. 学習データには, 掌に接触しているか否かの画像を計 1000 枚×2 人分用意した. また, 抽出した特徴ベクトルにノイズを加えて 11 倍にデータ拡張を行った. 被験者は 4 人とし, キーボードに割り当てられた文字を 3 分間入力してもらい, KSPS (1 秒当たりの文字数) と ER (エラー率) を計測した. 表 1 は 4 人の平均値であり, いずれも先行研究よりも高い結果が得られた. なお, 1 文字入力した際に複数の文字が誤入力される状況もあり, KSPS はそれもカウントしている. そのため, 本研究は先行研究に比べて, 正しく入力できた文字が 1.27 倍多く入力できていることが確認できた.

4. おわりに

本研究では USB カメラを用いた入力システムの改善方法を提案した. 機械学習を用いて接触判定を行うことで認識精度を向上させ, 先行研究に比べて間違いを抑えて早く入力できることが確認した. 今後の課題として, 入力時の姿勢に影響を受けずに入力可能にすること等があげられる.

[参考文献]

- [1] 久保田千尋, 湯村翼: ヘッドマウントディスプレイにおける指文字を用いた文字入力手法の提案, 情報処理学会インタラクション 2022 論文集, pp. 180-183, 2022
- [2] 溝上凱都: USB カメラを用いた AR 環境向け指先と掌の接触判定, 福岡工業大学情報工学部情報工学科卒業論文, 2023