



簡易型 HMD に用いられるスマートフォンの 内蔵センサを入力とした運動識別

松井良太¹⁾, 加藤花歩²⁾, 杉浦裕太¹⁾²⁾

1) 慶應義塾大学 理工学部 情報工学科 (〒 223-8522 神奈川県横浜市港北区日吉 3-14-1)

2) 慶應義塾大学大学院 理工学研究科

概要: スマートフォンと VR ゴーグルを用いた VR 環境が普及している. 人間と VR 環境との新たなインタラクションへの応用を見据え, 本研究ではこのような環境の下, スマートフォンの内蔵センサを用いて人間の運動を検知, 識別する手法を提案する. 運動を行った際の各センサのデータをサポートベクタマシンによる学習に用いることで, ジャンプやターンなどの運動の識別を実現している. 4 名の実験協力者からデータを収集し, ユーザごとの識別精度検証を行った結果, 16 種類の運動に対して平均 92.03% の精度が得られた.

キーワード: インタラクション, スマートフォン, センサ, 運動識別

1. はじめに

近年, VR ゴーグルの内部にスマートフォンを挿入することで VR 環境を構築する簡易型ヘッドマウントディスプレイ (以下, HMD) が普及してきている. PC やスマートフォンなどの外部接続を必要としないスタンドアロン型の HMD は, 購入に数万円を要することも多く, 高価である. これに対して, 簡易型 HMD 環境の構築で用いられる VR ゴーグルは, 安価なものであれば数百円程度で購入でき, 導入コストが低いという利点がある. ただしこのような VR 環境では, スマートフォンの画面がゴーグルで覆われてしまい, タッチパネルを用いた操作が困難になる. そのため, この環境を利用する際には, スマートフォンとは別に手持ちのリモコンを使用して操作することも多い.

一方で, これまで既存のタッチパネルなどに頼らない, モバイル環境における新たなインタラクションの手法を探る研究が数多く取り組まれてきた. その一例として, GestEar はスマートウォッチの内蔵センサとマイクを入力としてノックや拍手などの手の動きを識別し, オーディオデバイスなどの操作への利用を目指すものである [1]. これによって, 特殊なデバイスなどを使わずに, より直感的なインタラクションを実現している.

本研究では, 簡易型 HMD 環境において, このようなインタラクションの手法を, 身体全体の運動をセンシングすることで実現することを目指す. その手法として, 簡易型 HMD で用いられるスマートフォンに内蔵された複数のセンサを利用し, ジャンプやターンなどの多様な運動を検知, 識別することを提案する (図 1). 加速度センサ値の変化によってゴーグル装着者の姿勢変化を検知し, その前後のセンサ値を記録する. これらのデータは周波数成分への変換などの前処理が施された後, サポートベクタマシン (以下,

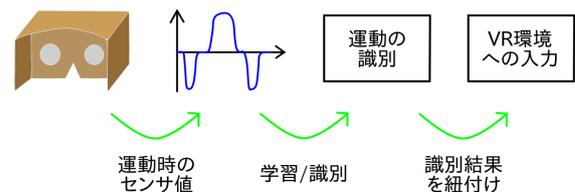


図 1: 提案するインタラクションのモデル

SVM) による学習器の生成, 運動の識別に用いられる. 4 名の実験協力者から運動時のデータを収集し, ユーザごとに識別精度の検証を行った結果, 16 種類の運動に対して平均 92.03% の精度が得られた.

2. 関連研究

2.1 スマートフォンの内蔵カメラによる動作の識別

スマートフォンを用いた新たなインタラクション手法として, 内蔵カメラによる動作の識別が多く研究されてきている. Lv [2] は, 手首や膝に固定されたスマートフォンで装着者の手足を撮影することにより, 手や足の動作を識別する手法を提案した. 上倉ら [3] は, 簡易型 HMD 環境で用いるスマートフォンの背面カメラを使用することで, ゴーグル装着者が前方に伸ばした手の動作を識別する手法を提案した.

これに対して本研究は, 身体全体の運動を身体に近接したスマートフォンで識別するため, 内蔵カメラではなくセンサを用いる手法を採用する.

2.2 センサによる動作の識別

モバイル端末を用いた動作の識別における他のアプローチとして, 加速度センサやジャイロセンサなどの入力を利用する手法が挙げられる. Becker ら [1] や Shi ら [4] は, スマ

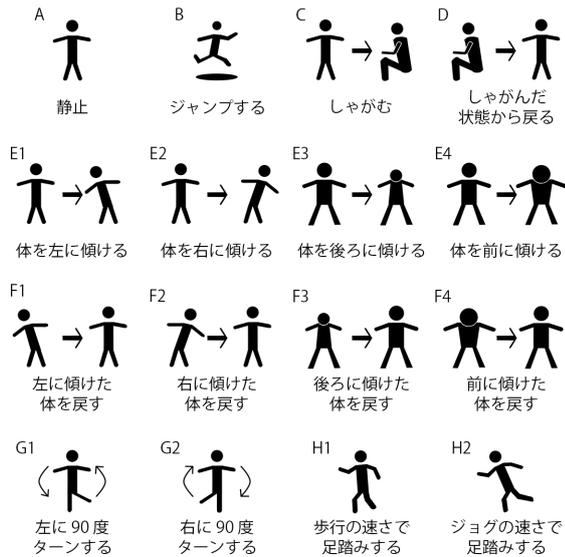


図 2: 16 種類の運動

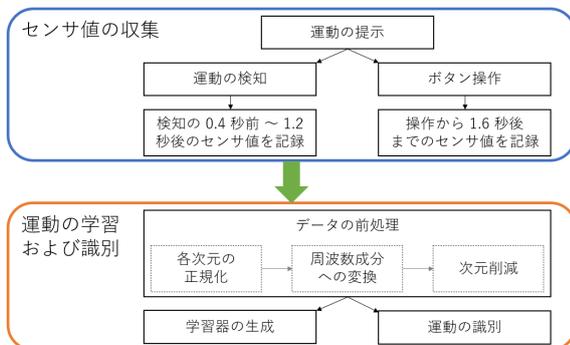


図 3: 本システムにおける処理の流れ

トウォッチの内蔵センサを用いて手の動作を識別し、IoT デバイスの操作などに利用することを提案した。また、スマートフォンの内蔵センサに着目した研究も存在する。Chen ら [5] は、スマートフォンの内蔵センサを入力として人間の動作を識別する際に、最も精度が高くなるスマートフォンの位置や機械学習手法の条件を検討した。Jain ら [6] や Barna ら [7] は、健康管理アプリケーションなどへの応用を見据え、スマートフォンの内蔵センサを用いて、歩行や階段の昇り降りなどの運動を識別する手法を提案した。

これに対して本研究は、スマートフォンの使用環境を簡易型 HMD に限定した上で、運動識別の結果を VR 環境への入力と直接的に紐付けることを提案する。これにより、センサを用いた、より密接なインタラクションの実現を目指す。

3. 実装

本研究では、図 2 で示した 16 種類の運動を行った際の各センサのデータを収集するアプリケーションを作成し、収集したセンサ値から学習器を生成して運動の識別を行った。このシステムの流れを、図 3 に示す。

3.1 センサ値の収集

運動中の各センサのデータを収集するアプリケーションを、Android スマートフォンでの使用を前提に、Unity を用いて作成した。アプリケーションは実行中、ランダムな順番で 16 種類の運動を提示する。ただし、動きが対となる運動 (図 2 の C と D, E1-4 と F1-4, G1 と G2) については、それぞれ連続して提示される。ユーザは簡易型 HMD 使用時のスマートフォンの位置を想定し、スマートフォンを目線の高さに構えた状態で、提示された運動をそれぞれ 1 回ずつ行う。これにより、各センサのデータはスマートフォンの内部ストレージに、運動の種類と紐付けられて記録される。

アプリケーションは実行中、20 [Hz] のサンプリングレートで 3 軸加速度センサ値 \mathbf{a} 、3 軸ジャイロセンサ値 \mathbf{b} を取得する。また加速度センサ値は、手ブレなどの影響を軽減するため、指数移動平均 $\bar{\mathbf{a}}$ の計算を行う。これは新たな加速度 \mathbf{a} を取得する度に、式 1 を用いて更新される。ただし、 α は平滑化係数である。

$$\bar{\mathbf{a}}_{\text{new}} = \alpha \cdot \mathbf{a} + (1 - \alpha) \cdot \bar{\mathbf{a}}_{\text{old}} \quad (1)$$

また、センサ値を取得する度に、新たな加速度 \mathbf{a} と 1 つ前の指数移動平均 $\bar{\mathbf{a}}_{\text{old}}$ との差分におけるユークリッド距離 $\|\mathbf{a} - \bar{\mathbf{a}}_{\text{old}}\|$ を計算する。この値が任意の閾値を超えた場合、ユーザが運動したと判定し、その 0.4 秒前から 1.2 秒後までのセンサ値を記録する。ただし、加速度の変動が小さい運動 (図 2 の A, H1, H2) については、適切な閾値の設定が困難であったため、スマートフォン画面上の測定開始ボタンをユーザが押してから 1.6 秒間のセンサ値を記録する方法を採用した。

3.2 運動の学習および識別

収集したセンサ値を用いた学習器生成と運動識別には、Python を用いた。

まず、データに対する前処理として、各次元の正規化処理、高速フーリエ変換 (以下、FFT) による周波数成分への変換、主成分分析 (以下、PCA) による次元削減の 3 つの処理を施す。正規化処理では、6 次元データ (3 次元加速度センサ値、3 次元ジャイロセンサ値) の各次元について、変換後の平均、分散がそれぞれ 0, 1 となるように線形変換を行う。これにより、単位の違いなどによる次元間での寄与度の差を除去できる。次に FFT による周波数成分への変換により、センサ値の変動に関する周波数特性やその位相特性を抽出する。このとき、データの次元数は $6 \times 20 \text{ [Hz]} \times 1.6 \text{ [sec]} = 192$ である。最後に PCA によって、32 次元に削減する。上述の前処理を施した後、SVM による学習器を生成することで運動を識別する。

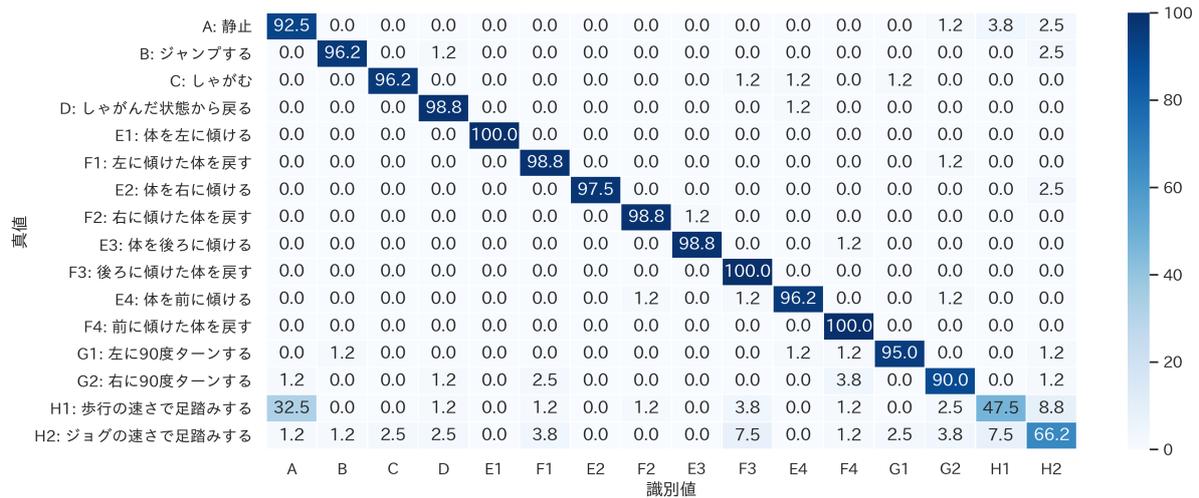


図 4: 1.6 秒分全てのデータを用いて個人毎学習により識別精度を検証した際の混同行列

表 1: 学習データとテストデータの分割方法

分割方法	説明
個人毎学習	実験参加者ごとのデータに対して、K-fold 交差検証で分割する。
全ユーザ学習	全実験参加者のデータに対して、K-fold 交差検証で分割する。
他ユーザ学習	全実験参加者のデータに対して、Leave-one-subject-out 交差検証で分割する。

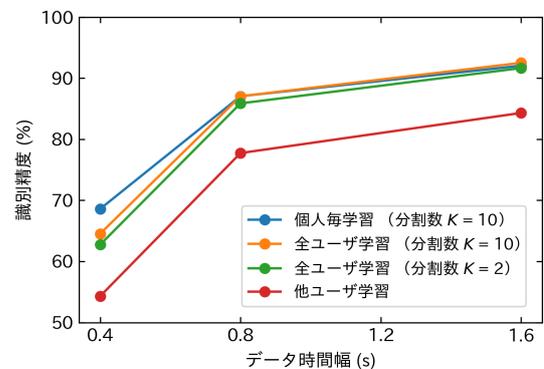


図 5: データの時間幅および分割方法と識別精度

4. 実験

4.1 概要

本実験には 4 名の実験協力者が参加した。参加者はそれぞれ、各センサのデータを収集するアプリケーションを使用しながら、図 2 に示した 16 種類の運動を 20 回ずつ行った。なお、COVID-19 の感染拡大により、参加者から対面でデータを収集することが困難になったため、上述のアプリケーションを参加者に配布し、実験方法に関するオンラインの講習を行うことでデータの収集を行った。これにより、4 名 × 16 種類 × 20 回 = 1,280 個のデータが収集された。このデータセットを用いて識別精度を検証したが、その際には Chen ら [5] の手法を参考に、学習データとテストデータの分割方法を 3 種類用意し、分割方法による精度の変化にも着目した。この分割方法についての説明を、表 1 に示す。ただし、K-fold 交差検証における分割数 K は、個人毎学習では $K = 10$ に設定し、全ユーザ学習では $K = 2, 10$ の 2 通りの設定を用意した。これは、全ユーザ学習、分割数 $K = 10$ の条件が最も学習データ数の多い条件となる中、学習データ数がより個人毎学習と近い条件を用意するためである。また、データの時間幅と精度の関係についても考えるため、収集した 1.6 秒分のデータを全て使用した場合に加え、前半 0.4 秒分または 0.8 秒分のデータのみを使用した

場合についても精度の評価を行った。この時間幅は、データの前処理として FFT を施す際に、各次元のセンサ値の長さが 2 のべき乗である必要があることから選定された。

4.2 結果

1.6 秒分のデータを全て使用し、個人毎学習により精度を検証した際の混同行列を図 4 に示す。この場合の識別精度は 92.03% となった。なお、図 4 より、2 種類の足踏みの運動に対する精度が低いことが分かる。そこで、この 2 つの運動を除外し、残りの 14 種類の運動のみで同様の精度検証を行った結果、識別精度は 97.32% に向上した。

また、使用するデータの時間幅および学習、テストデータの分割方法と識別精度の関係を表したグラフを図 5 に示す。まず、データの分割方法による識別精度の違いに着目すると、他ユーザ学習は個人毎学習や全ユーザ学習と比べて精度が低かった。次に、データの時間幅と識別精度との関係に着目すると、いずれの分割手法においても時間幅が小さくなるにつれて精度が低下する結果が得られた。

5. 議論

5.1 実験結果に関する考察

個人毎学習による精度検証の結果から、足踏みの運動に対する識別精度が比較的低いことが明らかとなった。特に図4より、「歩行の速さで足踏みする」が、動きのない「静止」と激しく混同されていた。これは、足踏みの運動の実施時は、他の運動と比べて頭部の動きが小さいことに起因すると考えられる。

次に、データの分割方法に着目すると、他ユーザ学習が他の方法と比べて精度が低い結果となった。一般的に、運動の振る舞いには対象者の身体的特性などによって個人差が存在するが、このような運動時の特性を識別器に反映できず、他ユーザ学習で精度が低下したと考えられる。

最後に、データの時間幅に着目すると、0.4秒にまで短縮した際に精度が大きく低下した。これは、ほとんどの運動において、データの始点から0.4秒後にセンサ値が大きく変動したためであると考えられる。0.4秒付近でセンサ値が大きく変動することは、身体が大きく動いていることを意味するため、その部分のデータに運動の特徴が多く含まれていることを示唆する。時間幅を0.4秒に短縮するとそのデータの多くが損失してしまうため、運動の識別が困難になったと考察できる。

5.2 制約および今後の課題

個人毎学習や全ユーザ学習では、収集した1.6秒分のデータを全て使用することで90%を超える精度が得られた。一方で、運動を検知してから運動終了まで、1.6秒-0.4秒=1.2秒のタイムラグが存在する。これは運動を始めてから運動が識別されるまでの遅延時間が最短でも1.2秒を上回ることを意味し、リアルタイム推定の場合、ユーザが反応の大きな遅れを覚悟することとなる。この遅延時間はデータの時間幅を短縮することで改善できるが、それとともに識別精度が低下する結果も得られており、トレードオフである。そのため、VRゲームなどの迅速な応答が求められるアプリケーションの場合、本システムでの識別速度では不十分である。今後は、データの時間幅の短縮後も高い精度が維持できるよう、より効果的な特徴量の選定や機械学習手法の工夫などを行っていく。

また、VR環境での本システムの利用においては、センサを利用した他の機能との競合を考慮する必要がある。例えばVR映像の視聴時に、首を振って顔の向きを変え、それに応じて映像の視点が変化する機能がしばしば用いられる。そのため、本手法におけるターンなどの運動をすると、必然的に映像の視点が変わってしまう。このような機能の競合を許容しない場合には、競合を回避できる運動の選定などが必要であると考えられる。

操作に利用する運動については、ユーザ側の使いやすさも考慮する必要がある。特にジャンプなどの激しい運動は、頻繁に行くと疲れてしまうなどの観点から、アプリケーションの満足度を低下させる要因になりかねない。今後は識別

精度のみならず、アプリケーションの機能や使いやすさなどの様々な観点から各運動を評価し、使用環境ごとに最適な運動の選定を行っていく。

6. おわりに

本研究では、簡易型HMDで用いられるスマートフォンの内蔵センサを利用して、ジャンプやターンなどの16種類の運動を識別する手法を提案した。ユーザの運動は加速度センサ値の変化によって検知され、その前後のセンサ値が取得される。取得されたセンサ値は前処理が施された後に、SVMによる学習器の生成、運動の識別に用いられる。ユーザごとに識別精度を評価した結果、平均92.03%の精度が得られた。

謝辞 本研究は、JST AIP-PRISM 課題番号 JPMJCR18Y2 の支援を受けたものです。

参考文献

- [1] Vincent Becker, Linus Fessler and Gábor Sörös. “GestEar: Combining Audio and Motion Sensing for Gesture Recognition on Smartwatches”. *Proceedings of the 23rd International Symposium on Wearable Computers (ISWC '19)*. pp. 10-19, 2019.
- [2] Zhihan Lv. “Wearable Smartphone: Wearable Hybrid Framework for Hand and Foot Gesture Interaction on Smartphone”. *2013 IEEE International Conference on Computer Vision Workshops*. pp. 436-443, 2013.
- [3] 上倉裕大, 赤池英夫. “簡易 VR システムにおける内蔵スマートフォンのみを用いた入力手法の提案と実装”. 情報処理学会第81回全国大会講演論文集. pp. 79-80, 2019.
- [4] Lei Shi *et al.* “Knock Knock, What’s There: Converting Passive Objects into Customizable Smart Controllers”. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18)*. article 31, 13 pages, 2018.
- [5] Yufei Chen and Chao Shen. “Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition”. *IEEE Access*. vol. 5, pp. 3095-3110, 2017.
- [6] Ankita Jain and Vivek Kanhangad. “Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors”. *IEEE Sensors Journal*. vol. 18, no. 3, pp. 1169-1177, 2018.
- [7] Arnab Barna *et al.* “A study on Human Activity Recognition Using Gyroscope, Accelerometer, Temperature and Humidity data”. *2019 International Conference on Electrical, Computer and Communication Engineering*. 6 pages, 2019.