

モーションキャプチャーと AI を用いた万引き行動早期検出に関する研究 －姿勢推定による行動識別の基礎的検討－

日大生産工(院)○LIU XIAOQING
日大生産工 豊谷 純

1. まえがき

近年、小売業界では万引き行為による損失が深刻な社会的課題となっている。経済産業省の報告によると、国内の小売店における年間の万引き被害額は数千億円に達するとされ、特に人手不足が顕著なコンビニエンスストアや無人店舗などでは、防犯対策の強化が急務となっている。

従来の監視手法では、防犯カメラと店員による目視確認が中心であったが、人的監視には限界があり、見逃しも多い。また、AIを用いた映像分析技術も導入されつつあるが、顔認識など外見情報を用いる手法はプライバシーや法的制約の問題がある。

そこで本研究では、個人を特定しない形で行動解析を行う姿勢推定（Pose Estimation）に注目する。人体の関節位置を基に行動を解析することで、プライバシーを保ちつつ「通常とは異なる動き＝万引き行動の兆候」を検出できる可能性がある。

2. 提案手法

本システムは、MoveNetによる人体骨格推定を基盤として、静止画像またはカメラ映像から人物の姿勢を検出し、当該動作が「万引き（盗窃）行為」であるかを識別するシステムである（Fig.1）。従来研究では映像や顔特徴を利用した識別が多かったが、本研究では骨格情報に基づく動作特徴に着目し、よりプライバシーに配慮した検出を試みる点に特徴がある。

入力された映像はフレーム単位でMoveNetモデルに入力され、各フレームから17箇所の関節点（キーポイント）座標が抽出される。抽出された座標は座標系の違いや被写体の大きさ差を補正するために正規化処理を行い、欠損点は補間・平滑化処理により補完する。

次に、各フレームあるいは短時間ウィンドウ（例：0.5～1.0秒）内のキーポイント列から特徴量を抽出する。これらを分類器の入力とする。また、必要に応じて画像パッチから抽出した補助的特徴（手の開閉推定や物体検出など）を骨

格特徴に統合するマルチモーダル手法も検討する予定である。

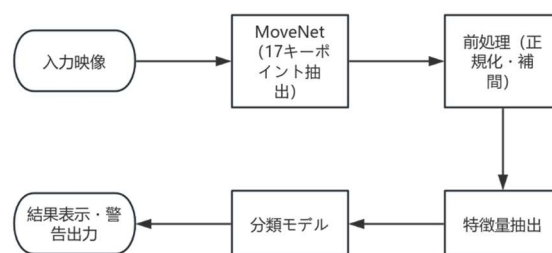


Fig.1 本研究の処理手順（現段階では既往研究の方法に基づき基礎的な検討を進めている）

分類モデルは教師あり学習に基づく二値分類器（「通常行動」／「万引き行動」）として設計する。初期段階では学習データ量や計算資源を考慮し、ランダムフォレストやXGBoost、小規模な全結合ニューラルネットワークを用いてベースラインを構築する。その後、必要に応じてCNNによる画像特徴抽出やLSTMなどの時系列モデルへ拡張を検討する。学習時にはデータ不均衡への対応として、オーバーサンプリングまたは重み付き損失関数を導入する予定である。

本研究の開発方針として、まずは教員から提供されたMoveNetの骨格認識コードを動作確認し、当該コードにより抽出された17関節の座標をCSV等の汎用フォーマットで出力・蓄積するパイプラインを構築する。具体的な作業手順は次のとおりである：

- (1) MoveNetコードの動作確認と実験環境への組み込み、
- (2) 被験者による模擬動作映像の撮影とラベリング（万引き／非万引き）、
- (3) フレームごとのキーポイントをCSV出力してデータベースを整備、
- (4) 特徴量抽出パイプラインの実装とベースライン分類器の学習・評価、

(5) リアルタイム表示および検出時の警告出力機能の実装。

システムはリアルタイム処理を目標とし、検出結果は画面上に可視化されるとともに、万引き行為が検出された場合には即時に警告メッセージを出力することで店舗側の迅速な対応を支援する。

3. 実験方法および測定方法

本研究では、提案手法の有効性を検証するため、骨格認識プログラム (MoveNet) を用いて取得した関節座標データを基に、AIモデルの学習および識別性能評価を行う。以下に、実験環境、データ作成手順、および測定方法を示す。

(1) 実験環境

実験はノートPC上で実施し、PythonおよびTensorFlow, OpenCVライブラリを用いてプログラムを作成した。現段階ではスマートフォンで撮影した人物画像を入力とし、MoveNetモデルによる骨格推定処理を適用予定である。

(2) データ作成

実験用データとしては、(a) 公開されている人物行動データセットの一部 (例: Human3.6M, MPII Human Pose Dataset など) および、(b) 研究室で撮影した模擬万引き動作映像を使用する。

MoveNetにより抽出された17関節点の(x, y)座標をCSV形式で保存し、それぞれに「通常」または「万引き」のラベルを付与した。

(3) 学習および評価方法

データの80%を学習用, 20%を評価用として分割し、教師あり学習により分類モデルを学習させる。分類モデルとしては、初期段階ではランダムフォレストおよび小規模な全結合ニューラルネットワーク (3層程度) を用い、後にCNNによる画像特徴との統合モデルを検討する。学習は300エポックを基本とし、損失関数には交差エントロピー、最適化アルゴリズムにはAdamを使用する。

(4) 測定および評価指標

検出性能は、Accuracy (正解率), Precision (適合率), Recall (再現率), および F1スコアを用いて評価する。また、検出の早期性を確認するため、動作開始から警告出力までの平均遅延時間を測定し、リアルタイム性を評価する。

さらに、誤検出 (False Positive) の発生状況を分析し、誤警報の原因を特定することで

モデル改良の方向性を検討する。

本研究は現在、システム開発およびデータ整備の初期段階にある。現時点では、MoveNetを用いた骨格推定の動作確認を行い、人物の肩・肘・腰など主要関節点の抽出が可能であることを確認している。

今後は、万引き行動に特徴的な身体姿勢として、肩をすくめる・上体を前に傾けるといった動作に着目し、これらの特徴を数値化して分類モデルの学習データとして利用する予定である。また、動作の個人差や撮影環境による誤検出の影響を検討し、特徴量の選定および前処理手法を改良していく。

4. まとめ

本研究では、モーションキャプチャー技術とAIを組み合わせ、万引き行動に特徴的な動作を識別する手法の基礎的検討を行っている。現在は、MoveNetを用いた骨格推定の動作確認および関節点データの取得を進めており、今後は肩や体幹の姿勢変化に着目した特徴量の抽出と、それを用いた動作分類モデルの構築を行う予定である。

最終的には、AIによって不審な姿勢変化を自動検出し、店舗などの防犯監視を支援できるシステムの実現を目指す。

参考文献

- 1) Julieta Martinez, Rayat Hossain, Javier Romero, and James J. Little “A simple yet effective baseline for 3D human pose estimation”, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2640–2649.
- 2) Narges Rashvand, Ghazal Alinezhad Noghreh, Armin Danesh Pazho, Babak Rahimi Ardabili, Hamed Tabkhi “Shopformer: Transformer-Based Framework for Detecting Shoplifting via Human Pose”, *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024.
- 3) TensorFlow Blog (Google Research) : “Next Generation Pose Detection with MoveNet and TensorFlow.js”, *TensorFlow Blog*, May 2021.
<https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html>