

関節動作の特徴量を用いた日常動作からの個人識別システムに関する一検討

石元 道哉[†] 武 博[†] 佐藤 公則[†]

東京工科大学 コンピュータサイエンス学部[†]

1. はじめに

カメラ映像のみを用いて歩容から個人識別をする技術として, Teepe ら (2021) ^[1] により提唱された, 人体関節の位置の時間変化から特徴量を抽出する, グラフ畳み込みネットワーク (GCN) が高い精度を示している. この技術を応用して, 歩行のみでなく, さまざまな日常動作から信頼性の高い個人識別を行うシステムを作成することが本研究の目的である. この論文では, GCN を日常動作の特徴量抽出に利用したときの精度を確かめるために行った実験について述べる.

2. 実験目的

本実験の目的は, 複数の日常動作データに対して GCN モデルを評価し, 精度を比較することである. また, 学習データに, 関節の位置情報と, 位置情報もとに計算した相対的な関節角度情報の 2 つのパターンを用意し, それぞれを用いてモデルを学習させた場合に精度に差が出るかを調べることである. 関節角度を用いたのは, 被験者の空間的な位置に影響されにくく, 精度向上に繋がると予測したためである.

3. 実験内容

本実験では, 日常動作の公開データセットである MMAct trimmed cross-scene dataset ^[2] を骨格推定 API の MediaPipe ^[3] で処理し, 関節情報を抽出した後, GCN モデルの学習と評価に使用した. モデルは, Teepe による GCN の PyTorch 実装 ^[4] を利用した.

3.1. データセット

MMAct trimmed cross-scene dataset は, 20 人の被験者による, 35 種類の日常動作を行う様子を撮影した動画データセットである. データセットは, 部屋の四隅に設置されたカメラで被験者が日常動作を行う様子を同時に撮影した動画ファイル群である. 同一の動作に対し, 部屋

の環境を変えた最大 3 つのシーンが用意されており, 各シーンに対し 1 個以上の撮影セッションで構成されている. 本実験では, 35 の動作の内, carrying, fall, jumping, kicking, loitering, pushing, running の 7 つを使用した. これらの動作を選択した理由は, 歩行などの持続的な動作が特徴量の効率的な抽出に向いていると考えたからである. Fall, jumping, kicking は, 持続的でない動作と結果を比較するために含めた. これらの動作は全て, シーン 1 にセッション 4 つ, シーン 4 にセッション 2 つの計 2 シーン 6 セッションで構成されている. シーン 1 は何も置かれていない部屋で, シーン 4 は大人の腰程の高さの 3 つの固定障害物が部屋に置かれている環境である (図 3.1).

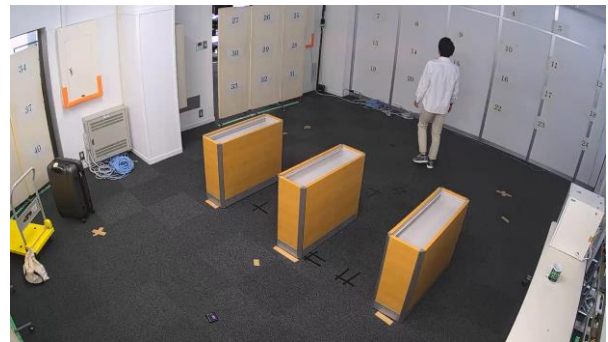


図 3.1 MMAct trimmed cross-scene dataset ^[2] シーン 4, 障害物が置かれた部屋の例.

3.2. データ処理

各動作について, 全てのセッションから, 60 フレームを 1 スライスとし, 30 フレームおきに動画から切り出し, 最大 10 スライスに分割した. それぞれのスライスから関節の位置情報を抽出した学習・評価データと, 位置情報を関節の相対的な角度情報に変換した学習・評価データを作成した.

関節位置から関節角度への変換には, 体の末端にある関節を除いて, 前後の関節を含めた 3 つの関節がなす角度を関節の x 座標として格納した. 関節の y 座標には, 前後の関節までの距

A Study on Personal Identification System Based on Features Extracted from Joint Movements of Everyday Actions

[†]Michiya Ishimoto, Bo Wu, Kiminori Sato, Department of Computer Science, Tokyo University of Technology

離の和を格納した。末端の関節については, x, y 座標共に 0.0 を格納した。

3.3. 学習と評価

それぞれの学習・評価データに対して, GCN を学習させ, その後学習したモデルの個人推定性能を評価した。

各学習・評価データに対するモデル評価には, $k=4$ の k -分割交差検証を用いた。学習・評価データの分割は, 被験者 ID をシャッフルした後, 被験者 ID で分割した。精度の計算は, まず評価データを学習済みモデルに入力して特徴量ベクトルを計算させた。シーン 1 のセッション 1 に含まれるスライス全てを Gallery セットとし, それ以外のスライスを Probe セットとした。それぞれのスライスに対し, 4 つのカメラ角度それぞれのデータを Gallery セットと比較し, 最も近い Gallery スライスをカメラごとの推定候補 ID とした。Gallery と Probe の近似度は, それぞれの特徴量ベクトルのユークリッド距離で計算された。最終的な推定 ID は, 4 つのカメラ角度から得られた推定候補 ID の最頻値とした。最頻値が存在しないときは, 最も近似度の高い推定候補 ID を推定 ID とした。Probe セットのスライス全てに対し被験者 ID を推定し, その正解数をスライスの総数で割った正解率を計算し, これをモデルの精度指標とした。

4. 実験結果

実験した結果を, 関節の位置情報を用いた学習・評価データを normal, 位置情報を関節の角度に変換した学習・評価データを angles とし, 表 4.1 に示す。

学習・評価データごとの精度のうち, 持続的な動作 (carrying, loitering, pushing, running) の平均値は, normal が 65.03%, angles が 65.20%であった。持続的でない動作では, normal が 42.33%, angles が 43.93%であった。

表 4.1 動作, シーンごとの平均精度. (N=Normal, A=Angles, N1=Normal Scene 1) .

Action	N1 (%)	A1 (%)	N4 (%)	A4 (%)
carrying	70.75	72.65	61.11	62.06
fall	33.85	36.92	30.00	37.42
jumping	46.39	46.11	43.06	54.17
kicking	46.20	47.87	53.13	36.87
loitering	68.61	60.47	64.86	60.48
pushing	66.91	72.13	53.95	67.21
running	67.92	64.49	60.65	54.27
mean	57.23	57.24	52.39	53.21

5. まとめ

持続的な動作と比べ, 持続的でない動作ではモデルの個人推定精度が落ちており, 動画から連続的なスライスを切り出す学習データでは, 持続的動作のほうが特徴をうまく抽出しやすかったことが分かった。持続的動作間での精度には大きな差は無かった。

関節の位置情報を用いた場合の結果と, 位置情報を相対的な関節角度に変換した場合の結果の差は, 平均して 1%以下と小さく, 誤差の範囲であると考えられる。よって, このモデルは入力データが関節の位置でも角度でも同様の精度が得られるため, あえてデータを角度に変換する意義は, 精度向上のためには無いことが示された。

シーン 1 とシーン 4 では平均して 5%程度シーン 4 での予測精度が低く, 障害物がある環境では, 4 つのカメラからの情報を複合しても, 推定精度が落ちることが分かった。

6. 今後の予定

Teepe ら (2021) の論文では, 同じモデルで最高 8 割近い推定精度が示されており, それと比較すると本実験で得られた推定精度は低い傾向にある。これはデータセットの処理の違いや, 学習時のハイパーパラメータの差が影響していると考えられる。よって, 今後はより高い推定精度の得られるデータセットの処理方法と, 適切なハイパーパラメータの探求を試みる。

また, システムとして運用する際, 持続的でない動作も登録して個人識別に使えたほうが都合がよいと考えられるため, 持続的動作と同様の精度が得られるデータセット処理や学習の工夫について模索する。

7. 参考文献

- [1] T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, "GaitGraph: Graph Convolutional Network for Skeleton-Based Gait Recognition," in 2021 IEEE International Conference on Image Processing (ICIP), 2021, pp. 2314-2318. doi: 10.1109/ICIP42928.2021.9506717.
- [2] Hitachi, Ltd., "MMAct: A Large-Scale Dataset for Cross Modal Learning on Human Action Understanding", <https://mmact19.github.io/2019/> (2023/01/10 閲覧)
- [3] C. Lugaresi et al., MediaPipe: A Framework for Building Perception Pipelines. arXiv, 2019. doi: 10.48550/ARXIV.1906.08172.
- [4] T. Teepe, "GaitGraph", <https://github.com/tteepe/GaitGraph> (2023/01/10 閲覧)