

グラフニューラルネットワークを用いた 製造工程における作業者の個人識別手法の提案

工藤 諒太[‡] 堀川 三好[†] 小村 皓大[†] 岡本 東[†]

岩手県立大学大学院 ソフトウェア情報学研究科[†]

岩手県立大学 ソフトウェア情報学部[‡]

1. はじめに

近年、作業者の動線分析や習熟度の可視化など、製造現場において作業者の行動分析が取り組まれている。特に、動画による行動分析を行う際には、複数人いる作業者が誰であるかを把握する個人識別が必要となる。しかしながら、作業者は、工場着、帽子やマスクを着用する機会が多く、顔認証や衣服の特徴から個人識別をすることが困難な場合が多い。そのため、身体的特徴や動作の特徴から個人識別を行う技術が求められている。

本研究は、作業者の動画から骨格データを取得し、グラフニューラルネットワーク（以下、GNN）を用いて個人識別を行う手法を提案する。提案手法では、カメラと作業者の距離（以下、撮影距離）の変化に対応するため、骨格データに対して座標変換を行う。その上で、骨格データをグラフ構造化し、身体や動作の特徴量を考慮した個人識別モデルを提案する。

本稿では、複数人が個別に工場作業を想定した動作をしている場合に、提案手法により個人識別が可能なことを報告する。これにより、外観が似ている作業者が複数人いる動画の個人識別を可能とし、工場における行動分析のための一助を得ることを目的としている。

2. 関連研究

動画解析の活用例として、歩行者の人物再識別を行った事例が挙げられる¹⁾。この課題に対しては、取得された骨格データをもとに識別を行う手法や、人物同士の位置関係をグラフとして捉

えて識別を行う手法が多く見られる。これにより周辺環境の影響を受けずに、複数のカメラに渡る人物の追跡が可能となる。また、製造現場などにおいては、作業者が帽子につけるマーカーや、腕章等を身につけ、物体検出による個人識別が行われている。しかしながら、教師データの収集にコストがかかることや識別精度に課題を抱えている。

3. 身体的特徴による個人識別

動画から得られる骨格データから、身体および動作の特徴量を取得して個人識別モデルを生成する。個人識別モデルの作成手順を述べる。

3.1. 骨格データの取得

作業者を撮影した動画から、姿勢推定を用いて骨格データを取得する。本稿では OpenPose を用いて首や腰等の 25 箇所の関節点の x, y 座標を取得する。

3.2. 骨格座標の座標変換

骨格データのスケールは、撮影距離によって変化する。本稿では、スケール変換の方法として、実測値および深度値を用いた変換方法を比較する。深度値の取得方法として MiDaS²⁾を用いる。予備実験の結果、骨格データの取得で最も安定して取得可能かつ標準偏差が小さい首から腰の長さを基準とし、実測値および深度値を用いてスケール変換を行うこととした。

3.3. 個人識別モデルの生成

座標変換を行った骨格データをグラフ構造化し、GNN 系列の機械学習による個人識別モデルを生成する。本稿では、3 つの GNN 系列の機械学習モデルを比較する。

(1) GCN (Graph Convolutional Networks)

グラフデータにおけるノード特徴量とエッジによるノード同士の隣接関係を示すデータを用いて、エッジで繋がれた関連度の高いノードを畳み込み計算する手法である。

(2) GAT (Graph Attention Networks)

グラフ構造の畳み込み層と Self-Attention 層を併せた手法である。Attention 機構では、グラフ頂点

Proposal of Individual Identification Method for Workers in Manufacturing Process Using Graph Neural Network

Ryota Kudo[‡], Mitsuyoshi Horikawa[†], Kodai Komura[†], Azuma Okamoto[†]

[†] Software and Information Science, Graduate School of Iwate Prefectural University

[‡] Software and Information Science, Iwate Prefectural University

を表す特徴量の更新と頂点同士の「つながり」の重要度計算を行う。

(3) GC-LSTM (Graph Convolutional LSTM)

グラフの空間情報と時間情報を考慮した手法であり、時間 t の特徴量の学習に対して直前の特徴量を加えることで、時系列を考慮した学習が可能である。

4. 検証実験

4.1. 実験目的・方法

提案する個人識別手法の有効性を検証するため、セル生産を想定した学内実験を行う。被験者3名(身長 169, 168, 158cm)が、3つの作業台を移動しながら、「組み立て・梱包・検査」を行う。予備実験において骨格データが信頼度高く取得できるのが 14m 付近までであることから、撮影距離を 3, 8, 13m の3水準とする。各被験者が各距離 21セット、合計 189回実施する。

4.2. データセットおよび学習方法

学習に利用する骨格データ件数は約 785,000 件となる。座標変換は、実測値または深度値ごとに、各被験者の首から腰の長さの実寸と骨格データの距離の比率で座標をスケール変換し、グラフデータを生成する。バッチサイズは 4096 であり、リーブワンアウトによる検証を行う。

4.3. 実験結果

個人識別モデルの精度評価の指標として、適合率と再現率の調和平均である F 値を用いる。

(実験 1) 機械学習モデルの精度比較

個人識別に最も適した GNN 系統の機械学習モデルを探るために、実測値を基にした座標変換を行った撮影距離 8m のデータのみを用いて比較を行った。各モデルの精度を表 1 に、GCN と GC-LSTM による学習モデルの混合行列を図 1 に示す。

(実験 2) 撮影距離の影響

実験 1 において最も精度が良い GC-LSTM を用いて、撮影距離の影響を比較する。各々の撮影距離を学習データとして用いた場合 (3m, 8m, 13m) の精度と、全て撮影距離を学習データとして用いた場合 (実測値による座標変換有り無し) の精度を表 2 に示す。今回の実験では、縦方向の移動が少ないため、撮影距離による影響は小さいのがわかる。併せて、座標変換の処理の有無による影響も小さい。

(実験 3) 座標変換手法の影響

撮影距離を考慮した個人識別モデルを構築するために、実測値と深度値を用いた手法で座標変換を行う手法を比較した (表 3)。精度に大きな差は見られないため、縦方向の移動が多い作

表 1 GNN 系列のモデルによる精度比較

	GCN	GAT	GC-LSTM
F 値	0.61	0.88	0.95

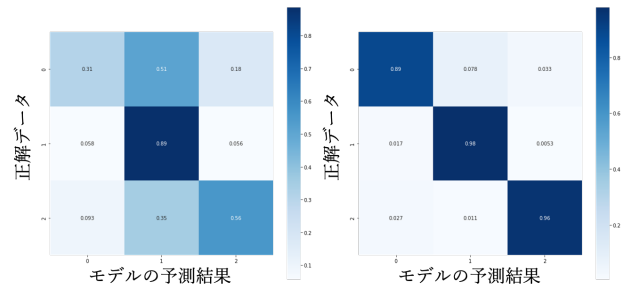


図 1 混合行列 (左: GCN 右: GC-LSTM)

表 2 撮影距離と精度の比較

	3m	8m	13m	全体 (変換有)	全体 (変換無)
F 値	0.96	0.92	0.94	0.95	0.93

表 3 座標変換手法による精度の比較

	実測値	深度値
F 値	0.95	0.92

業の場合、深度値を考慮した変換方法の適用が期待できる。

5. おわりに

本稿では、動画解析から得られる骨格データを用いた個人識別モデルの提案をした。実験結果から、身体的特徴と動作的特徴を考慮することで、高い精度で個人識別可能であることを示した。また、座標変換を行うことで、作業者とカメラとの距離の変動に対応可能だと思われる。今後は、同時に複数人の作業者の識別が可能な機械学習モデルの考案と同時に、実際の製造現場での実証実験を行い、実用性を高めていきたい。また、他のセンサとのマルチモーダル学習でさらに精度を向上させる手法についても検討していく予定である。

参考文献

- 1) Yan, Yichao, et al. "Learning context graph for person search." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
- 2) Ranftl, René, et al. "Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer." IEEE transactions on pattern analysis and machine intelligence 2020.