

## 天井設置カメラによる俯瞰動画を用いた人物検出と再同定

河 竜星<sup>†</sup>梅澤 猛<sup>‡</sup> 大澤 範高<sup>‡</sup>千葉大学工学部総合工学科情報工学コース<sup>†</sup>千葉大学大学院工学研究院<sup>‡</sup>

## 1 はじめに

カメラ画像による個人再同定では、側面から撮影する場合よりも天井からの方が遮蔽が生じにくい。カメラを天井に設置することで対象者の視界に入らない位置から画像を取得することができるため、撮影されているという心理的負担を軽減することができる。天井に設置したカメラで深度情報を取得し個人識別を行う先行研究 [2] では、頭部や肩周りのサイズなどの体格情報や、髪や服の色情報を用いている。しかし、この手法は服装の変化に影響受けやすく、日をまたぐような大きな時間差がある場合には再同定が困難になる。

本研究では、歩行時の人の特徴点の時間的位置変化を、天井に設置した深度カメラで得られる深度情報から推定し、その連続した特徴点の座標の時系列データを用いた個人の再同定を検討する。図1に人物の特徴抽出と再同定の方法を示す。深度画像から頭頂の特徴点の位置を推定する深層学習モデルを構築し、天井に設置した深度カメラで撮影した動画の頭頂を示す特徴点の位置を推定する。推定した特徴点の座標の時系列データの類似度を動的時間短縮法 (DTW)[1] を用いて計算することで個人の再同定を行う。この方法では、体格情報や色情報を用いないため、服装の影響を受けないと考えられる。高精度な特徴点の位置推定ができるように EfficientNetV2[3] を基に、特徴点推定モデルを構築する。推定された特徴点の時間的位置変化の波形同士を DTW を用いて類似度を計算し、最大類似度に基づいて人物の再同定を行う。

## 2 実験

EfficientNetV2 を用いた特徴点推定モデルによる頭頂の特徴点の推定精度の評価する。特徴点推定モデルに

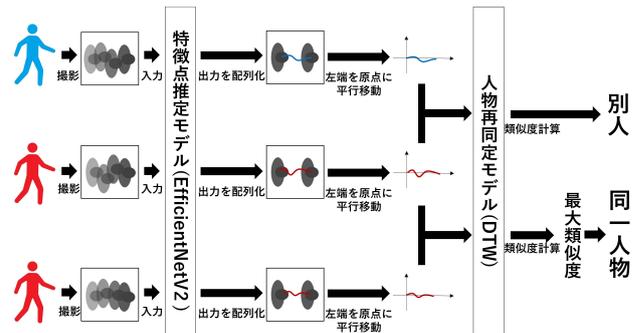


図1 人物の特徴抽出と再同定の方法

より推定された特徴点の時間的位置変化の時系列データを用いた個人再同定の有用性を確認する。パーカーのフードを着用し頭部を大きく見せた際と、上着を着用し肩周りを大きく見せた際の再同定への影響を調べる。

データセット作成時の撮影には Intel 社の RealSense D435i の深度センサと RGB センサを使用する。床から約 2.5m の位置にカメラを設置し撮影を行った。

## 3 特徴点推定モデル

本研究では ImageNet によって事前学習済みの EfficientNetV2 を基に、データセットを使用し転移学習することで深度画像の入力を受け、頭頂の座標を推定するモデルを構築する。事前学習済みの EfficientNetV2 の入力サイズは 3 チャンネル  $\times$  640pixel  $\times$  480pixel とし、出力は 1,792 次元となっているため、特徴点の 2 次元座標が出力として得られるように、全結合層を用いて 2 次元に線形変換するモデルを構築した。

転移学習のためのデータ収集では、頭頂に赤いマーカーをつけた 1 人の被験者が上着やフードを着用せずにカメラの下を歩く様子を撮影した。深度情報を RGB 画像に変換した深度画像をデータセットに用いる。深度画像と RGB 画像の位置合わせを行い、RGB 画像から検出した頭頂の赤いマーカーの座標を正解ラベルとする。データセットに含まれる画像は 7,329 件である。

Person detection and re-identification using top-view video from ceiling-mounted cameras

<sup>†</sup> Ryusei Kawa, Department of Information Engineering, Faculty of Engineering, Chiba University

<sup>‡</sup> Takeshi Umezawa, Noritaka Osawa, Graduate School of Engineering, Chiba University

転移学習では、データセットに対し上下反転、左右反転、180°回転の操作を行い、データセットを4倍にデータ拡張して構築した学習用データセット(29,316件)を9:1に分割し、訓練データ(26,385件)とテストデータ(2,931件)とした。

転移学習のEpoch数を50、学習率を0.0001、損失関数には二乗平均平方誤差RMSEを用いた。

構築したモデルをテストデータによって評価したところ、特徴点の位置推定誤差RMSEは2.75となった。テストデータに対しての出力例を図2に示す。図は人物領域のみを切り抜いた深度画像であり、図の上方向を向いている。カメラに近いほど黒く、遠いほど白く表現され、測定できなかった箇所は黒で表現される。中央の濃い灰色が頭部、周りの薄い灰色が肩と胸である。赤の十字の交点が正解座標であり、青の十字の交点が推定座標である。

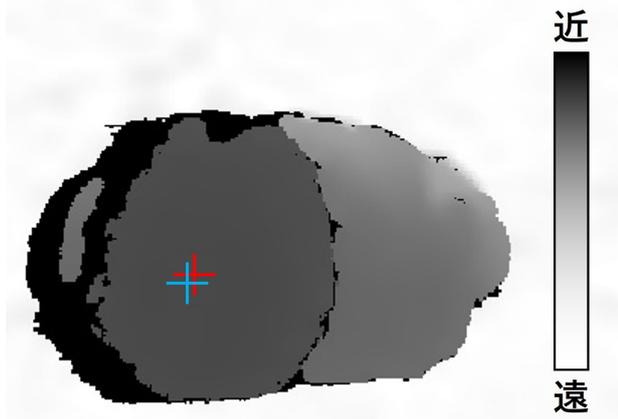


図2 テストデータに対する推定結果

#### 4 人物再同定モデル

入力画像に頭頂の時間的位置変化の時系列データを受け、DTWで求めた波形の類似度が最大となる人物と同じ人物として再同定を行うモデルを構築する。

人物再同定のためのデータ収集には、天井に設置した深度センサのみを用いた。被験者3人について、上着またはフードの着用有無条件を変えそれぞれ10回ずつ、計60件の深度動画を撮影した。このデータから特徴点推定モデルで推定した頭頂の特徴点の連続した時系列データを作成した。

データセットが60件と小さいため、データを最大限活用するために歩容データセットを一つ抜き交差検証(leave-one-out cross-validation)に基づき評価した。

DTWによって再同定を行った結果を表1に示す。表1中のA',B',C'は被験者A,Bが上着を、Cがフードを着用している状態を表す。上着やフードを着用していなかった場合の結果は、30回の内27回同定に成功し、成功率は90.0%となった。上着やフードを着用していた場合の結果は、30回の内21回同定に成功し、成功率は70.0%となった。

表1 各被験者の同定結果

		予測値			
		A	B	C	
正解値	A	A	10	0	0
		A'	10	0	0
	B	B	0	10	0
		B'	4	6	0
	C	C	0	3	7
		C'	0	5	5

#### 5 おわりに

上着やフードを着用している場合の同定成功率は着用していない場合と比べ20%パーセントポイント低くなり、服装の影響を受けないという予想に反する結果となった。特徴点推定モデルの精度向上のために訓練データを充実させ、同定成功率の向上のために俯瞰画像からどのような歩容特徴を抽出し、その特徴に基づいたどのような同定方法が適しているかの検討が今後の課題である。被験者数を増やした際の再同定成功率への影響についても調査する予定である。

#### 参考文献

- [1] D.J. Berndt and J Clifford. Using dynamic time warping to find patterns in time series. *KDD Workshop*, 1994.
- [2] D. Liciotti, M. Paolanti, E. Frontoni, and A. Mancini. Person re-identification dataset with rgb-d camera in a top-view configuration. *International Workshop on Face and Facial Expression Recognition from Real World Videos*, pp. 1–11, 2016.
- [3] M. Tan and Quoc V. Le. Efficientnetv2: Smaller models and faster training. *International Conference on Machine Learning*, 2021.