

## 被写体のプライバシーに配慮した人物再同定手法の提案

大谷 紀子 研究室

2272064 土井 博暉

### 1. 背景と目的

各団体がブースを設けて展示する形式のイベントにおいて、効率的に来場者を勧誘するためには、既訪問者と未訪問者の識別が不可欠である。しかし、勧誘担当者が単独で遂行できる業務には限界があり、複数人が交代で勧誘業務にあたる場合には、来場者情報の共有が困難となる。また、勧誘担当者の誤記憶により、既訪問者を再勧誘する可能性がある。

解決策として人物再同定 (Re-ID; Person Re-identification) 技術の活用が考えられる。既存の人物再同定手法の多くは深層学習に基づいているが、深層学習は内部処理に不透明性を有するため、同定の根拠が被撮影者にとって理解しにくく、心理的な抵抗感を生じさせる恐れがある。また、イベント会場では計算資源が限られるため、実運用においては、計算コストを抑えつつリアルタイムな人物再同定を可能とする手法が求められる。

本研究では、展示ブースにおける来場者勧誘支援を目的とし、内部処理が明瞭でリアルタイムに処理可能である軽量な人物再同定手法を提案する。

### 2. 提案手法

提案手法は、人物検出、属性抽出、人物再同定の3フェーズで構成される。第1フェーズである人物検出は常時実行される処理であり、人物が検出された場合に第2フェーズの属性抽出が実行される。第2フェーズの終了後、第3フェーズとして人物再同定を実行する。各処理は、身長推定用のARマーカーが映るように設置されたカメラで、展示ブース周辺における来場者の往來を撮影しな

がら実行される。訪問者情報を保持するデータベースは、初期状態を空として処理を開始する。

第1フェーズでは、映像から人物を検出する。カメラで撮影した映像から、軽量な物体検出モデルである You Only Look Once (以下 YOLO) [1]により人オブジェクトを検出し、DeepSortにより画面外に出るまで追跡する。また、MediaPipe Poseを用いて検出した人オブジェクトの骨格情報を取得し、全身が明瞭に撮影された人物のみを解析対象とする。

第2フェーズでは、検出された人物から、服装の種類と代表色、身長、および髪色を属性情報として抽出する。服装の種類は、ファッション解析に特化した物体検出モデルである YOLOS-Fashionpediaにより特定する。服装の代表色は、RGB色空間における3次元ヒストグラムを用いた量子化処理により決定する。髪色は、頭部検出用データセット SCUT-HEADの再学習モデルで検出した頭部領域に対し、3次元ヒストグラムを用いた量子化処理を適用して決定する。身長は、ARマーカーを幾何学的基準とし、画像上の画素数を実空間の長さに変換することで算出する。

第3フェーズでは、抽出された人物がデータベースに登録済みか否かを判定する。抽出された人物とデータベース内の人物について、同種の服が一定数以上あり、身長差、服の代表色の差、および髪色の差がそれぞれ設定された閾値以下である場合に同一人物と判定する。服の代表色と髪色の差は、3次元RGBヒストグラムの距離とする。データベース内に同一人物が確認された場合は対応

表 1 人物再同定精度と処理速度

実験	処理手法	純度	ペアワイズ適合率	ペアワイズ再現率	ペナルティ付き F1 スコア	平均 FPS
A	OSNet	0.321	0.061	0.553	0.108	—
	提案手法	0.372	0.048	0.129	0.069	—
B	OSNet	0.420	0.032	0.353	0.059	9.819
	OSNet+人物検出	0.445	0.054	0.286	0.091	7.245
	提案手法	0.582	0.021	0.037	0.027	7.909

する人物 ID を出力し、確認されなかった場合は新規人物をデータベースへ登録する。

### 3. 評価実験

評価実験では、提案手法の人物再同定精度と処理速度を検証するため、実験 A および実験 B を実施した。評価用データセットの構築にあたって、本学横浜キャンパスの 3 号館および 4 号館の計 3 箇所にカメラを設置し、画角内に AR マーカーを配置して、学生や教職員が廊下を往来する様子を撮影した。

実験 A では、再同定精度のみを検証するため、動画から抽出した 26 人分 70 枚の静止画像に対し、第 2 フェーズおよび第 3 フェーズの処理により人物再同定を行った。実験 B では、提案手法全体の性能を検証するため、複数台のカメラの前を通過した計 42 人の人物が映る 3 本の動画に対し、第 1 フェーズから第 3 フェーズまでの処理により人物再同定を行った。

いずれの実験においても OSNet を比較対象とした。OSNet は人物再同定専用に設計された、軽量かつ高精度な CNN アーキテクチャである。

精度評価指標には、純度、ペアワイズ適合率、ペアワイズ再現率、ペナルティ付き F1 スコアを用いた。純度は、予測 ID ごとに求めた最頻出 ID 数の総和を全データ数で正規化した値を表す。ペアワイズ適合率は同一人物と判定された ID ペアの正解率、ペアワイズ再現率は同一人物として判定されるべき ID ペアの検出率を表す。ペナルティ付き F1 スコアは F1 スコアと網羅率の積である。網羅率は、動画中から検出されたすべての人オブジェクトのうち、解析対象とした人オブジェ

クトの割合である。処理速度評価指標には、1 秒当たりの処理フレーム数を示す FPS を用いた。

実験 A, B における OSNet と提案手法の人物再同定精度と処理速度を表 1 に示す。実験 B については、提案手法の人物検出をしたうえで OSNet により人物再同定を行った結果についても併記する。

実験全体を通じ、提案手法は既存手法と比べ、純度が高く、ペアワイズ再現率とペアワイズ再現率が低い。提案手法では、解析対象が 42 人にもかかわらず、56 人がデータベースに登録されたことから、同一人物を複数の人物と判定する傾向が強いと見える。また、OSNet の前処理として提案手法の人物検出を行うと、再同定に適さない画像が事前に除外され、再同定精度向上に寄与することが確認された。

### 4. 考察

提案手法による誤検知事例を確認したところ、服装の外観特徴が正しく抽出されていない事例が散見された。特に、背景色と類似した色合いの服を身に付けた人物同士を誤認するケースが多く、背景との混同が精度低下の一因であることが示唆された。動画から検出した人物画像は、一般的な人物画像に比べ解像度が低く、識別に必要な服の細部特徴が失われ、撮影環境変化の影響を強く受けたと考えられる。

### 参考文献

- [1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.779-788, 2016.