

画像による歩行者 OD 自動観測手法の検討

松下 貴哉¹・小泉 鴻一²・塚井 誠人³・平野 雅也⁴・宮地 岳志⁵

¹学生会員 広島大学大学院 先進理工系科学専攻 (〒739-8527 広島県東広島市鏡山 1-4-1)
E-mail: m222901@hiroshima-u.ac.jp

²学生会員 広島大学大学院 先進理工系科学専攻 (〒739-8527 広島県東広島市鏡山 1-4-1)
E-mail: m222901@hiroshima-u.ac.jp

³学生会員 広島大学大学院 先進理工系科学専攻 (〒739-8527 広島県東広島市鏡山 1-4-1)
E-mail: b194869@hiroshima-u.ac.jp

⁴正会員 広島大学大学院 先進理工系科学専攻 (〒739-8527 広島県東広島市鏡山 1-4-1)
E-mail: mtukai@hiroshima-u.ac.jp

⁵非会員 株式会社ミオシステム 代表取締役社長 (〒611-0021 京都宇治市宇治妙楽 24-1 ミツダビル 3F)
Email: mhirano@miosys.co.jp

⁶正会員 Bridge over 代表取締役社長 (〒733-0022 広島市西区天満町 12-9-702)
Email: takeshi.miyaji@bridgeover.jp

OD 調査は自動車や公共交通機関ばかりではなく、鉄道駅や商業施設の賑わい評価を目的に歩行者流動にも用いられる。しかし歩行者流動は、自動車のようなナンバープレートによる個体の識別ができないため、目視による流動カウントと目的地調査を組み合わせるなど、観測コストがかかる。本研究では歩行者 OD の自動観測に関して、歩き方の特徴を映像から抽出する生体認証技術を応用する。このために本論文では 3 つの課題に取り組む。まず、画像解析分野の物体追跡と複数台のカメラから撮像された人物の識別を目的とする Person Re-Identification の研究を説明する。また、深層学習に基づく歩容認証によって個人を同定し、歩行者 OD 自動観測アルゴリズムについて、歩行者 OD 自動観測手法への応用可能性を検討する。

Key Words: 歩行者 OD, 歩容認証, Person Re-Identification

1. 背景と目的

中心市街地の活性化を狙った駅や駅前広場の改装や、駅周辺地区の再開発、および新たな短距離交通手段導入等の政策効果の把握では、アンケートやステークホルダーインタビューなどの定性的な情報に加えて、駅利用者、主要施設の入れ込み客数、または交通旅客数などの定量的指標が欠かせない。最近では、地点別時刻別の滞在者数ならば携帯電話位置情報によって、また来訪者の移動軌跡であれば GPS 軌跡や重要地点に設置した Wi-Fi 感知器が活用できる。まちの賑わいの計測では、周辺地域を含む面的な効果を把握することが望ましい。つまり、ある地点から流入した歩行者の立ち回り先を明らかにできる歩行者 OD の潜在的価値は高いと考えられる¹⁾。

歩行者 OD の計測には、従来は断面交通量の人手観測と、アンケートによる目的地分配係数の実地調査が必要だった。前者は比較的低コストで実施できるものの、調査員が疲労するため、長時間にわたる連続観測が困難であり、後者は調査内容が複雑になるため、信頼性が担保

できる程度のサンプル数の確保に難がある。

Wi-Fi パケットセンサ²⁾や接続ログ³⁾、スマートフォンの加速度センサや地磁気センサ⁴⁾⁵⁾を活用した歩行者通行量の自動観測には、調査員疲労の問題はない。さらに近年では、カメラ映像による歩容認証、骨格認証に基づく歩行者観測が注目されている。歩容認証は歩行者の歩き方の特徴に基づいて、また骨格認証は人体の関節点を結んだ骨格情報に基づいて、それぞれ映像に映る個人を識別する生体認証技術である。これらは、カメラ映像があれば計測可能であって個人の協力が必要ないこと、顔が判別できない場合でも個人を認証可能なことに加え、顔認証よりも相対的に個人プロフィールと結び付けにくいという利点がある。これらの認証技術によって抽出した特徴を別地点の撮像と突合すれば、歩行者 OD 表が得られる。

近年は機械学習の研究が急速に蓄積されている。その応用の一分野である画像解析に関して、情報分野では Person Re-Identification (人物再同定または人物再識別) と呼ばれる研究分野がある。これは、動画から検出された

人物の外見に基づいて、画角内に既に現れた人物かどうかを識別する手法であり、こちらも歩行者 OD を観測可能な基礎技術として、注目される。これまでに、深層学習を用いた画像認識に基づく交通量観測の研究^{6,7)}はいくつか発表されているが、同手法の歩行者 OD への応用例はまだない。本研究では骨格認証について、既存の機械学習モデルを用いて歩行者 OD を計測する際の課題を整理する。

2. 既往研究

本章では、画像に基づく歩行者 OD の観測を念頭に、情報分野で研究が進められている物体追跡(トラッキング)と Person Re-Identification の既往研究をレビューして、この分野を概説する。

(1) 物体追跡

アルゴリズムを用いて動画から情報を得るタスクは、機械学習の代表的な応用分野として知られているが、その難易度は決して低くはない。難易度を左右するのは入力画像の品質であり、たとえば対象と観測者の位置関係(視点の多様性)、照明条件の多様性、人の全体または一部が一時的に視界から隠れること(Occlusion/オクルージョン)、ならびに低解像度などが挙げられる。ここで視点の多様性は複数のカメラから得られる人物の向きが一定ではないために生じる外観情報の変化、照明条件はカメラの設置場所や時間帯による画角の明るさと RGB 素子に記録される色の見え方、オクルージョンはすれ違い等によって体の一部または全体が隠れることで生じる情報の欠落、低解像度は街中に設置された監視カメラなどで生じる低画質(画素の密度が低い)画像であり、いずれも状況によっては、対象の識別を困難にする。

物体追跡は、対象を継続して画角内で追跡する技術である。追跡対象の数、オクルージョンの有無、カメラの台数、撮影領域の重複の有無、オンライン/オフライン処理等によって技術が分類され、それぞれ難易度が異なる(図 1)。追跡対象が 1 つの場合は最も難易度が低く、たとえば 0.1 秒単位で取得する素動画(以下、フレーム)をつなぎ合わせれば実現可能である。一方で追跡対象が複数の場合は、あるフレームで検出された全ての物体が、別フレームで検出された全物体の同一性を確かめる突合(以下、照合)を行う必要がある。なお照合には、追跡対象の位置情報や外観情報が用いられることもある。複数カメラの場合は、それらのカメラ間で同一地点が不明または存在しないため、照合に位置情報は利用できず、外観情報に頼ることになる、ただし外観は撮影条件が変われば対象物の背景が変化し、それが外観の識別に

影響するため、タスクの難易度があがる。

Bewley が一台のカメラで複数の物体を追跡対象とする Multi Object Tracking (MOT) の基礎的なアルゴリズムとなっている SORT⁸⁾を発表した。SORT が動画内で物体検出を行うと、捕捉した物体は bounding box (bbox :バウンディングボックス)と呼ばれる短径領域で表示される。物体追跡、つまりトラッキングとは、bbox をフレーム間で関連付けることに等しい。その処理手順をやや簡単に述べると、時刻 t-1 までの追跡情報から予測された bbox と時刻 t において検出された bbox の位置の類似度が高い、という条件を満たすように最初のフレームから最後のフレームまで連続して画像処理することであり、その bbox を連ねた結果がトラッキングである。SORT は Convolutional Neural Network (CNN) の進歩によって当時の最先端の性能とオンライン処理を両立した。トラッキングに用いるデータは、以下でモデル化される。

$$x = [u, v, s, r, \dot{u}, \dot{v}, \dot{s}]^T \quad (1)$$

u, v は bbox の中心の水平と垂直のピクセル位置、 s, r は対象の bbox の面積と縦横比であり、上付 $\dot{\cdot}$ は各変数の時間微分を表す。なお、縦横比 r は一定とする。bbox の関連付けには、評価指標は画像の重なり割合をあらわす Intersection over Union (IoU) を用いる。bbox の割り当てにはハンガリアン法を用い、所定の IoU の最小値を下回る場合は、bbox を割り当てない。

SORT は、bbox の関連付けに位置情報のみを扱っていたため、オクルージョンによって ID の消失や入れ替わるなどの課題があった。そこで視覚的な特徴量を考慮して SORT を拡張した Deep SORT が発表された⁹⁾。これは物体(人物)の特徴量を取得する際に、後述する Person Re-Identification データセットを用いて学習したモデルを適用する。Deep SORT において、位置に基づくトラッキングを行う際のデータは以下でモデル化される。

$$x = [u, v, \gamma, h, \dot{u}, \dot{v}, \dot{\gamma}, \dot{h}] \quad (2)$$

SORT との相違点は、縦横比 γ を変数扱いとして、bbox の高さ h を加えた点である。Deep SORT は SORT よりも処理手順が複雑なため計算に時間がかかるが、長期的なオクルージョンからの回復に有用な外観情報を用いることで、トラッキング性能が向上し、画角の出入りを伴う物体のトラッキングもできるようになった。

(2) Person Re-Identification

Person Re-Identification (以下 ReID) は、異なるカメラから撮影された人物の画像間の対応関係を確立するアル

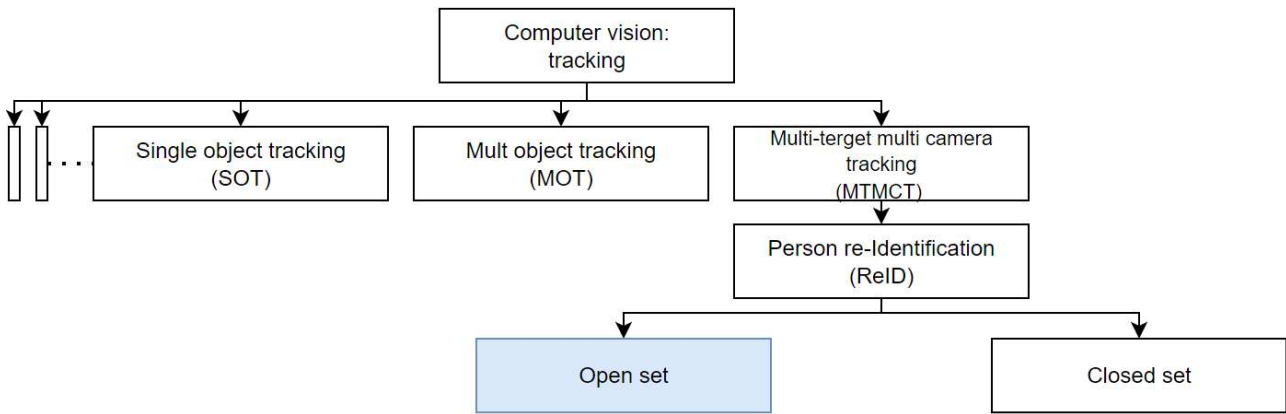


図1 画像解析分野の分類図

ゴリズムである。その手法は、もともと **Multi-Target Multi-Camera Tracking (MTMCT)** のサブタスクとして研究されていた¹⁴⁾が、2006年に Gheissari ら¹⁵⁾が、**ReID**を「重複しない複数のカメラから撮影された人物とその外観を照合することを目的とした研究テーマ」と定義したことにより、独立した研究分野として認知されるようになった(図1)。

ReID のタスクは非常にシンプルであり、「人間の目と脳が物体や人を検出、追跡、識別する日常的な認知プロセスを画像解析によって自動化すること」である。**ReID** の手順について簡単に説明する。まず動画から人物を検出して **bbox** で切り出した人物画像(ギャラリー)に **ID** を割り当て、**ReID** モデルを学習する。その後、新たに取得した動画の識別対象画像(プローブ)を学習した **ReID** モデルに入力してプローブとギャラリーの類似度を計算し、同一人物の確率が高い順にギャラリーがランキングされたリストを取得することで、同定精度を検証する。

図1下部右に現れる **Closed set** は、プローブがギャラリーに必ず存在する状況である。一方、**Open set** は、ギャラリーの一部がプローブで欠損する状況である。そのため **Open set** では、類似度を計算した後でプローブに無い人物として識別するタスクが必要である。そのうえで該当の人物は、ギャラリーに追加しなくてはならない。**Closed set** の例は、科学捜査などの特定の人物 **ID** を識別するなどの一対多間の照合タスクである。一方、**Open set** の例は不特定多数の人物を複数のカメラで撮影した動画内で一貫して追跡するような、多対多間の照合タスクである。市中に設置したカメラ画像から歩行者 **OD** を観測するタスクは **Open set** であり、画像解析分野の中でも難易度の高いことがわかる。

ReID モデルの開発では様々な課題が検討されている。以下ではそのうち代表的な課題について簡単に述べておく。特徴量抽出は、画像内の人物の色や形などの外観情報から信頼性の高い特徴量を得るためのネットワーク設

計に焦点を置いた研究である。人物同定の精度は取得する画像に大きく影響を受けるが、撮影状況によっては全身の画像が常に得られる保証がない。そのため、身体を部位ごとに分けて外観を捉えることでオクルージョンに対応する手法¹⁶⁾や、その簡略法として単純に画像を横方向に等分(この研究では7等分)して一部欠損に備える手法¹⁷⁾、同じく一部欠損に備えるために全身と身体部位を総合的に処理する手法^{18),19)}がある。

メトリック学習は、取得したサンプル画像に対応する特徴量の性質を改善する手法群である。具体的には、特徴量空間内において、類似または同一グループ間では近く、異なるグループ間では遠くに配置されるような学習手法である。つまり学習の結果として得られる特徴量ベクトルを識別性の高い特徴量空間に変換することで信頼性の高い照合を実現するため手法である。この分野では **ReID** モデルの学習時に設計する損失関数に対する研究が多くなされており、なかでもトリプレットロス関数を用いる手法^{20),21)}がよく使用される。トリプレットロス関数では、3つの画像を用いる。それらは、**Anchor** と呼ばれる画像 x_A と、**Positive** (正例) と呼ばれる同一人物の画像 x_P と、**Negative** (負例) と呼ばれる画像 x_N とは異なる人物の画像 x_N である。これらを用いてトリプレットロス関数は、以下のように定義される。

$$L_{\text{triplet loss}} = [d(z_A, z_P) - d(z_A, z_N) + m]_+ \quad (3)$$

ここで $z_i = \phi(x_i)$ は、学習された x_i の特徴量、 $d(_)$ は、二つのサンプルの距離、 m は予め設定するマージンであり、 $[c]_+$ は、 $\max(0, c)$ を表す。式(3)を最小化すると、 x_A と x_P の距離の最小化と、 x_A と x_N の距離の最大化を同時に行う特徴量空間の学習が進むことがわかる。さらに特徴量抽出とメトリック学習にまたがる手法として、重要な特徴量の絞り込みや、人物同定に重要な特徴量に重み付ける手法^{22),23),24),25)}などが検討されている。

ReID モデルの学習時に用いるデータセットでは事前

に教師ラベル，すなわち正解ラベルを手作業などの何らかの手法によって，付与する必要がある．たとえば，前述したトリプレットロス関数では，Anchor に対して正例，負例の準備が必要だが，これを膨大なデータに対して付与することは困難である．そこで正解ラベルが完全には付与されていないデータを一部に含む弱教師つき学習や，正解データの全くない画像からでも学習を進める教師なし学習に焦点を当てた研究が注目を集めるようになった．前者の例としてたとえば以下の手順が提案されている．まず画像から得た特徴量を特徴量空間にプロットする．次にプロットされた全ての画像に対して k の値（近傍の範囲をあらわす）を変えながら k 近傍法を用いると，それぞれの k に対して任意の 2 画像の近傍にプロットされた共通画像の割合が取得できる．このとき，ある画像の k の値を横軸，共通する画像が含まれる割合を縦軸にプロットすると，その画像が同一人物の場合と異なる人物の場合ではグラフの形状が異なることを利用して，同一人物を推定する²⁹⁾．このほか k -means 法を繰り返して適用して，特徴量が類似する人物画像集合をクラスタリングする手法²⁷⁾や，同一人物を使用しないで異なる人物画像の差を学習させる手法²⁸⁾がある．最後の手法は，厳密に言えば「負例であるという正解」が設定されていないが，たとえば離れた 2 地点間で同時に同一に人物が存在できない等の条件を加味できる．つまり，画像以外の何らかの外部情報を用いれば，正例よりも負例のペアが取得しやすい点に着目した手法である．

そのほかにも，出力するランキングリストを最適化する研究^{29), 30)}，低画質に対応した研究³¹⁾，服装が変化した場合の研究³²⁾などがある．

(3) 本研究の特徴

既往研究を概観すると，ReID として人物同定の研究が進んでいることがわかる．しかしこれらの一連の研究は全ての人物の外観に基づく特徴量取得を試みており，本研究のような歩容／骨格に基づく同定を行っていないわけではない．歩容／骨格は外観（服装や荷物など）の変化に強いと考えられるが，他方でフレームから取得できる情報が少なければ，同定精度が低下することも懸念される．しかしこれらの人物同定における得失には，実証的な実験の蓄積が欠かせない．そこで本研究では，これまでの検証の蓄積が乏しい骨格認証に着目して，その性能を明らかにする．

3 章で説明する本研究の骨格認証手法では，姿勢推定ライブラリである OpenPose¹⁰⁾の TensorFlow¹¹⁾に移植された tf-pose-estimation¹²⁾を適用する．tf-pose-estimation には，1) 複数人を画角に含むフレームを処理したときに人物 ID が入れ替わる問題と，2) 人間以外を人物と検出する誤検出問題がある．1) については，tf-pose-estimation の

処理はフレームごとであって，正確な背より手順としてはトラッキングを行っていないため，検出した人物に割り振られる ID は，同一人物が移動したときに一定になるとは限らない．仮に「ID は画角内の原点に近い人物から割り振る」というルールを加えても，2) の誤検出が起こったときや，オクルージョンによって人物が検出されなかったときはうまく機能しない．そのため，データ整形を行う前に，正しい人物 ID を振りなおす必要がある．中島らは，首座標の取得率が全関節 18 箇所のうち 90%と最も高いことを利用した骨格追跡¹³⁾を提案した．本分析でも同手法を適用して，上述した課題の緩和に努める．

3. 歩容認証手法の性能検証実験

(1) 実験内容

実環境を想定して，屋外で複数人の動画を取得するために 2021 年 7 月 14 日に学内で撮影実験を行った．撮影には SONY のビデオカメラ HDR-CX680 を使用した．撮影に当たっては，混雑の少ない状況と，動画内の歩行者数が多い状況を設定した．また，監視カメラなどの俯瞰の画角での撮像を想定して建物 2 階から対象領域全体を画角に収めるような設定で撮影を行なった．

地上からは，図 3 に示すようにエリア A と B に対して，正面と側面から撮影した．歩行状況は被験者 4 名が 1 人ずつ歩くパターン 4 つと，4 名が横並びで歩くパターンの 5 パターンである．また，動画は人物を正面と側面からそれぞれ撮影した．なお横並びで歩くパターンの側面撮影は，一番手前の人物のみ関節座標が得られる状況だったため，以下の検証では正面から撮影した動画のみを使用する．

画面内で，1 人または 4 人の歩行者が一定の方向に歩行する 3 秒程度の映像を用いる．映像は 4 名の被験者に対して，地点 A と地点 B の 2 種類，正面の撮影と側面



図 2 tf-pose-estimation 適用例



図3 実験エリア

の撮影の 2 種類、「1 人ずつ順番に歩く」と 4 人同時に歩く」の 2 種類の、合計 6 パターンである。動画の命名規則は、「地点 撮影方向 歩行人数」とする。たとえば、A 地点の一人ずつ歩く歩行者を正面から撮影した動画は「A 正面一人」である。撮影した動画のフレームレートは 30fps、解像度は 1920×1080 ピクセルである。tf-pose-estimation に入力するため、解像度を「640×480 ピクセル」に変換する。以下、サンプルはフレーム単位の静止画を指し、また動画はそれらサンプルの集合を指すこととする。tf-pose-estimation によって得られた 4 人の合計サンプル数は、30fps×3 秒強×4 人=約 400 件である。なお、被験者間のサンプル数のばらつきによる精度への影響を防ぐため、被験者 1 人あたり 100 件以上のサンプルが同一人について得られた場合は、時刻の早い順から 100 件を抽出した。

(2) 検証の手順

tf-pose-estimation は、1 フレームごとに鼻、首、両肩、両肘、両手首、両尻、両膝、両足首、両目の合計 18 箇所 の x, y 座標を推定する。なお画面内で該当箇所が検出できない場合は、欠損値として扱う。得られた結果を CSV ファイルで保存する。本研究内では、画面左上の端点を原点、右方向を x 軸、下方向を y 軸、カメラに対する奥行き方向を z 軸と定義する。

tf-pose-estimation の出力のうち、鼻や両目、両耳は歩容の特徴量への影響は小さいと考えて削除し、残り 13 箇所 の関節座標について 26 次元のベクトルを出力する。歩行者が z 軸方向に移動した場合、たとえば奥から手前方向 (z 軸にマイナス方向) に移動すれば撮像は徐々に大きく、反対に奥行き方向 (z 軸にプラス方向) に移動すれば撮像は徐々に小さくなる。フレームごとの見た目の大きさの影響を除くため、大腿部の長さが 1 となるように座標を変換して、画像のスケールを正規化する。

歩行者が画面内で x 軸または y 軸方向に移動した場合、得られた座標群は移動方向に関する情報を含んでいるが、それは歩容認証には不要である。そこで全フレームで首

の位置を原点とする移動座標系を採して、1 フレームあたり 12 箇所 の 24 次元ベクトルを得る。

本手順の最後に用いる Keras による深層学習モデルは、人物の骨格情報の同定を行うためのメトリック学習ステップである。ただし Keras はサンプルの一部の要素に欠損がある場合は動作しない。そこで Pandas の interpolate 関数を用いて、それぞれの関節座標の欠損値を時系列で線形補完する。

撮影期間全体でランダムに起こる欠損とは異なり、オクルージョンによる欠損は、ある一定期間継続してデータ欠損するため、人物同定精度を著しく低下させる。そこでオクルージョンと思われる期間の不完全なデータは、分析者が除外して、使用しないこととした。

以上の手順によって得た 24 次元ベクトルを入力とする多クラス分類の深層学習モデルを構築して、人物を同定する。この部分では Keras の Sequential モデル³³⁾を使用し、最適化アルゴリズムは Adam (Adaptive Moment Estimation)³⁴⁾を用いる。このアルゴリズムに識別対象の歩行者の 1 フレーム分のデータを入力すると、学習した歩行者の歩容に合致する確率が得られる。

深層学習モデルにおいて学習できないハイパーパラメータは、値を変えて複数回の精度検証実験を繰り返した結果、最も精度が高い値を採用した。実験的に設定するパラメータの中でも、エポック数 (1 つの訓練データを何回繰り返して学習させるか) は、実行時間に特に大きく影響する。学習時間の負荷を考慮して本実験ではこれを 100 回とした。

本実験における学習データと検証データを定義する。深層学習の分野では、学習データのうちの 2 割程度を抽出して、検証用データにすることが多い。なおクロス・バリデーションは一般のモデル推定でも行われる。その際には学習データの一部を抽出して検証データとする方法がとられることがあり、その方が検証時の精度は高くなる。しかし深層学習には、推定できないハイパーパラメータなどが存在するため、モデルの学習用データへのオーバーフィットを防ぎ、汎化性能の高いモデルを選択する必要がある。以上の手順をまとめると、離れた地点間の人物の同定を想定した本分析では、A 地点の撮像からギャラリーを、B 地点の撮像からプローブを構成する。さらにそれぞれから 2 割程度の画像を抽出して検証用として、残りの 8 割の画像で学習を行うことになる。

(3) 実験結果

学習データと検証データの組み合わせを変えて、5 種類の実験を行った。それぞれを比較して考察を行う。表中の精度は、深層学習モデルの重みの初期値による影響を減らすため、同様の実験を 100 回行った平均値である。実験 1 と実験 2 を比較すると、撮影方向によって同定

表 1 骨格認証実験の結果*

	学習	検証	精度(%)
実験 1	A 正面一人 (400)	B 正面一人 (400)	68.3
実験 2	A 側面一人 (400)	B 側面一人 (400)	64.7
実験 3	A 側面一人 (400)	B 正面一人 (400)	38.8
実験 4	A 正面一人+A 側面一人 (800)	B 正面一人 (400)	60.4
実験 5	A 正面一人+A 正面四人 (800)	B 正面一人 (400)	79.8

*表中は、学習データ、検証データに用いた実験動画の設定条件。()内は、それぞれのサンプル数を示す。

精度に大きな差がないことが分かる。実験 1 と実験 5 の比較してサンプル数の精度への影響を明らかにする。実験 5 では学習データに A 正面四人のデータを追加した。正面からの撮影の場合、4 人が横並びで歩いてもオクルージョンが発生しないため、tf-pose-estimation を適用した際に、一人ずつ歩いた場合と同様の信頼度で関節座標が得られる。つまり、被験者一人当たりのサンプル数が 2 倍になったことで、実験 1 に比べて精度が 11.5 ポイント向上したと考えられる。

実験 2 と実験 3 を比較して、学習データと検証用データの撮影方向が異なる場合の同定精度について考察すると、その場合は同定精度が低下することが示された。実験 3 と実験 4 の比較から入力データの撮影方向の一部が合致している場合について検討する。学習データに正面の動画と側面の動画の双方を用いた場合の精度は、異なる場合 (実験 1) より高く、同じ場合 (実験 3) より低い結果が得られた。

本実験全体を通して、サンプル数が少ないことが同定精度に大きく影響している可能性が考えられる。プローブのサンプル数が多ければ、実験 4 のように異なる方向から撮影されたデータが混在しても、同定精度が、さらに検証が必要である。また、ギャラリーの画像数が多い場合の精度への影響も、検証する必要がある。

4. 今後の開発課題の検討

ここまで物体追跡と ReID の研究動向を踏まえて、骨格認証手法の実証実験について示した。本章で本手法の課題をまとめた後、歩行者 OD 自動観測アルゴリズムの実用化に向けて開発課題を検討をする。

(1) 課題

本手法の課題として、同定手順における入力データの形式が挙げられる。骨格情報に基づいた人物同定を行う

本手法は、歩容情報のような (歩行時の腕の振り周期などの) 歩き方の時間情報を扱っていない。本手法で歩容認証ではなく骨格認証を行なった理由は、主に実環境でのデータ取得の困難さとデータ処理上の課題の 2 点である。歩容を学習するには、左右で一步分などの特定の人物の数秒のまとまったデータが必要だが、これに好適な画像の切り出しアルゴリズムの構成は困難だった。さらに人物のトラッキングを行っていない tf-pose-estimation では、ID の振り替わりへの対処ができなかった。歩容認証を行うには、その一連の期間について同一の ID を割り当てられる前処理アルゴリズムを組む必要がある。

3(3)の検証結果より、骨格認証は、撮影方向によって得られる情報に大きな違いが出るため、歩行者の方向変化に脆弱なことが明らかとなった。ただしこの結果は、本論の検証手順では歩行者の歩行方向が一定となるように管理したことが原因とも考えられる。すなわち、検出対象の人物像の撮影方向が偏ったことにより、Re-ID の入力データに必要なばらつきが失われた可能性がある。

一方で、現場ではトラッキング対象人物の歩行方向は、カメラの撮影方向に対して、必ずしも一定とは限らない。以上の考察より、ギャラリー(学習用データ)に同一人物に関して撮像方向が異なるケースが含まれるとプローブ照合時の頑健性が確保できる、という仮説が成立する。上述の、方向変化に強い骨格情報が抽出できる入力画像の設定は、今後検証を深める必要がある。

(2) 新たな処理手順

上述した処理手順について、さらに考察を深める。ギャラリー内の同一人物性を担保することは、撮像方向の多様性の確保につながるるとともに、学習用データをより多く獲得することにつながる。よって本論の検証手順では目視処理をしていた画角内の同一人物トラッキングは、きわめて重要なことがわかる。そこでまず、前処理として適用する人物のトラッキング精度を高める。そのうえで、取得した bbox に正確な人物 ID を付与した状態で出力する。このようにして得た bbox 群 (人物の画像群) に tf-pose-estimation を適用すれば、人物 ID の正確性を担保しつつ、多様性のある骨格情報が抽出できる。

本論の手順では、設定した地点 A をギャラリー、地点 B をプローブとした。しかし、実世界では地点 B から地点 A への移動は当然起こり得るため、一方をギャラリーに固定することが困難である。また現実の歩行空間では、時間とともに出現する人物が増加する Open set に対する同定が必要である。ただし、歩行者 OD の推定では、トラッキングや ReID で注目を集めるオンライン処理は必ずしも重要ではない。もちろんある程度計算負荷を低減する努力は必要だが、歩行者 OD の取得ではオフライン処理が可能である。この点を加味すると、人物

の情報が時間的に蓄積される条件ではなく、一定期間内で取得した動画を一括で処理する手順についても検討する余地がある。例えば複数地点と複数時点で撮影した全動画から検出された人物の特徴量を抽出して、特徴量のクラスタリングを行う手順が考えられる。ただしその前提は、骨格情報が十分に人物の特定情報を含むことである。その上でクラスタリングされた人物の位置・時刻情報を参照すれば、結果として歩行者 OD 表が得られる。

以上の手順をまとめる。画角内でトラッキングを行うことで **tf-pose-estimation** から得る特徴量の質を向上させる。さらに動画の一括処理による人物のクラスタリングを行うことで個人を分類して、各個人の位置情報と時刻情報を参照することで OD 表を作成する手順について、検討する必要がある。

5. 結論

本研究では、画像解析分野のレビューと歩容認証手法による実験を通して、歩行者 OD 自動観測手法の新たな処理手順の検討をした。歩行者 OD 自動観測の実装する上で重要な点は、人々が交錯する空間において、動画から人物を検出し、個人を識別するために効果的な特徴量を取得することである。そのために特徴量を抽出する前処理として人物のトラッキングを行うことで誤検知や ID の振り替わりを抑えられる可能性を明らかとした。さらに歩行者 OD のタスクは、オフライン処理が可能という点でクラスタリングという手順の成立可能性を指摘した。

参考文献

- 1) 国土交通省：まちの活性化を測る歩行者量調査のガイドラインの策定について https://www.mlit.go.jp/toshi/tosiko/toshi_tosiko_tk_000034.html
- 2) 末木祐多・佐々木邦明：Wi-Fi パケットセンサから得られるデータを用いた市街地における歩行者 OD 交通量の推計，公益社団法人日本都市計画学会 都市計画論文集，Vol.54, No.3, p.718-725, 2019.
- 3) 岡平考司・川名義輝：神戸市における Wi-Fi データを活用した歩行者行動分析，土木計画学研究・論文集，Vol.27, No.1, pp.147-155, 2010.
- 4) 遠藤巖・藤田悟：複数センサを組み合わせた屋内歩行者位置推定，マルチメディア，分散協調とモバイルシンポジウム 2013 論文集，pp.188-195, 2013.
- 5) 武島知勲・梶克彦・廣井慧・河口信夫・神山剛・太田賢・稲村浩：通路に設置した回転する磁石による歩行者の通過検出手法，情報処理学会論文誌，Vol.58, No.1, pp.43-56, 2017.
- 6) 坪井務・石坂哲宏・関弘翔・西脇大輔・D. Roy・高橋聡：インドにおける深層学習を用いた混合交通車両検出技術，一般社団法人交通工学研究会，交通工

- 学，Vol.56, No.1, pp.54-57, 2021.
- 7) 松岡禎典・里内俊介・坂本一誠：AI を活用した画像認識型交通量両観測の導入に関する研究，一般社団法人交通工学研究会，交通工学，Vol.56, No.1, pp.24-27, 2021.
- 8) A. Bewley, Z. Ge, L. Ott, F. Ramos, B. Upcroft, "Simple online and realtime tracking", IEEE, pp. 3464 – 3468, 2016.
- 9) N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric", *Proc. - Int. Conf. Image Process. ICIP*, vol. 2017-Sept, pp. 3645-3649, 2018.
- 10) Z. Cao, T. Simon, S. Wei, Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", CVPR, pp1302-1310, 2017.
- 11) Google Brain : TensorFlow, <https://www.tensorflow.org/>
- 12) Ildoo Kim : tf-pose-estimation, <https://github.com/ildoonet/tf-pose-estimation>
- 13) 中島雅貴・小篠裕子・斎藤英雄：時系列情報を考慮した人体骨格追跡と評価，電子情報通信学会技術研究報告，117 巻，391 号，pp.267-270, 2018.
- 14) X. Wang, "Intelligent multi-camera video surveillance: A review", *Pattern Recognition Letters*, vol. 34, no. 1, pp. 3-19, 2013.
- 15) N. Gheissari, T. Sebastian, R. Hartley, "Person Re-Identification using spatiotemporal appearance", *Computer Vision and Pattern Recognition (CVPR), 2006 IEEE Computer Society Conference on*, pp. 1528-1535, 2006.
- 16) S. Dong, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for Re-Identification", *British Machine Vision Conference*, pp. 68.1-68.11, 2011.
- 17) Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling", *European Conference on Computer Vision (ECCV)*, pp. 480-496, 2018.
- 18) Y. Suh, J. Wang, S. Tang, T. Mei, and K. Mu Lee, "Part-aligned bilinear representations for person Re-Identification", *ECCV*, pp. 402-419, 2018.
- 19) Xiaokang Zhang; Yan Yan; Jing-Hao Xue, "Semantic-Aware Occlusion-Robust Network for Occluded Person Re-Identification", *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 2764 -2778.
- 20) A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person Re-Identification", *arXiv preprint arXiv:1703.07737*, 2017.
- 21) C. Yan, G. Pang, X. Bai, "Beyond Triplet Loss: Person Re-identification with Fine-grained Difference-aware Pairwise Loss", *IEEE Transactions on Multimedia* Vol-ume 242022 pp 1665–1677, 2022.
- 22) X. Ning, K. Gong, "Feature Refinement and Filter Network for Person Re-Identification", *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 3391 – 3402, 2020.
- 23) J. Miao, Y. Wu; Y. Yang, "Identifying Visible Parts via Pose Estimation for Occluded Person Re-Identification", *IEEE Transactions on Neural Networks and Learning Systems*, pp. 4624 - 4634 ,2021.

- 24) X. Ningabde, K. Gongde, W. Li, “JWSAA: Joint weak saliency and attention aware for person re-identification”, 2020.
- 25) D. Yi, Z. Lei, S. Liao, “Deep features for person re-identification on metric learning”, ICPR, 2014.
- 26) M. Ye, A. J. Ma, L. Zheng, J. Li, and P. C. Yuen, “Dynamic label graph matching for unsupervised video Re-Identification”, ICCV, pp. 5142-5150, 2017.
- 27) H. Fan, L. Zheng, and Y. Yang, “Unsupervised person Re-Identification: Clustering and fine-tuning”, arXiv preprint arXiv:1705.10444, 2017.
- 28) A. J. Ma, P. C. Yuen, and J. Li, “Domain transfer support vector ranking for person Re-Identification without target camera label information”, ICCV, pp. 3567-3574, 2013.
- 29) Z. Zhong, L. Zheng, D. Cao, and S. Li, “Re-ranking person Re-Identification with k-reciprocal encoding”, CVPR, pp. 1318-1327, 2017.
- 30) M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, and R. Hu, “Person Re-Identification via ranking aggregation of similarity pulling and dissimilarity pushing”, IEEE Transactions on Multimedia (TMM), vol. 18, no. 12, pp. 2553-2566, 2016.
- 31) X. Li, W.-S. Zheng, X. Wang, T. Xiang, and S. Gong, “Multi-scale learning for low-resolution person Re-Identification”, ICCV, pp. 3765-3773, 2015.
- 32) Person Re-identification by Contour Sketch under Moderate Clothing Change
- 33) Keras-team : Sequential モデル API, <https://keras.io/ja/models/sequential/>
- 34) Diederik Kingma • Jimmy Ba : Adam: A method for stochastic optimization, ICLR, 2015.

(2022.09.30 受付)