

# 人物再同定のための歩行画像からの 特徴量抽出に関する研究

山崎 健太<sup>1</sup>・小泉 鴻一<sup>2</sup>・松下 貴哉<sup>2</sup>・塚井 誠人<sup>3</sup>・平野 雅也<sup>4</sup>  
・宮地 岳志<sup>5</sup>

<sup>1</sup>学生会員 広島大学 工学部第4類 (〒733-8527 広島市鏡山 1-4-1)

E-mail: b194869@hiroshima-u.ac.jp

<sup>2</sup>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: m214438@hiroshima-u.ac.jp / m222901@hiroshima-u.ac.jp

<sup>3</sup>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

<sup>4</sup>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

<sup>5</sup>正会員 Bridgeover 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

本研究では、画像から人物を特定する手法である Person Re-Identification (ReID) を用いて、カメラで観測する地点間の歩行者 OD を自動観測するための前段階として、ReID 手法を用いた人物識別に有効な特徴量の検証を行った。MediaPipe Pose により骨格特徴を、OSNet により外見特徴を抽出した上で、これらの特徴を組み合わせて人物の特徴のデータセットを作成して、ロジスティックモデル、非線形サポートベクターマシン、randomforest の3つの既成の分類器を用いて各タイプの入力特徴量に対する分類精度を比較した。各特徴の種類に対する精度を比較したところ、骨格特徴量を用いて前処理された外観特徴を入力した場合に、最も高い精度が得られることがわかった。

**Key Words:** Multi class classifier, Features of person, Image processing

## 1. はじめに

コンパクトアンドネットワーク等を目的として掲げる都市計画マスタープランでは、交通結節点の重要性を再評価し、活性化を図る施策<sup>1)</sup>がとられる。にぎわい創出を狙った施策整備やソフト施策の効果測定では、歩行者 OD (Origin&Destination) 表を得る必要があり、そのためには歩行者を計測しなければならない<sup>2)</sup>。従来の歩行者 OD 表の推定は、断面交通量の人手調査とアンケートによる目的地分配係数の実地調査の組み合わせが基本だった<sup>3)</sup>。しかしこれらの調査には一定の人員が必要であり、長期にわたる調査は困難である。

近年では、GPS データ、Wi-Fi データ、レーザーカウンターなどの ICT 技術を用いた歩行者の自動観測により、大量の歩行者データを取得できるようになってきた。その中でも、カメラ映像による骨格認証により人物の同定を行う研究は、個人の協力を必要とせず、遠距離からの撮影でも個人の認証ができるという利点がある<sup>4)</sup>。すな

わち、骨格特徴量を動画から抽出し、別地点の映像から得られた骨格特徴量と突合できれば、歩行者それぞれの再同定によって歩行者 OD が得られる。この方法で歩行者 OD が推定できれば、コストの大幅削減につながる。

上述の研究テーマに関連して、情報分野には Person Re-Identification(人物再同定、以下 ReID)と呼ばれる研究分野がある。これは機械学習に基づく画像解析により、複数カメラ間に映った人物の同一性を判別するタスクの研究である。このタスクの精度を高めるために、ReID 研究では特徴量抽出部分に焦点が当てられることが多い。

歩行者の骨格と外観から得られる情報を用いて ReID を行う上では、ReID に入力する人物特徴量を適切に設定する必要がある。本研究では、この問題を検討するため、学内撮影データから抽出された特徴量を分類器に入力し、ReID に優れた特徴量と、その入力方法について検証する。これにより、ReID の精度を高める特徴量抽出と入力方法を検討する。また上記で判明した最も良い特徴量を使用して、訓練データとテストデータ間で撮像

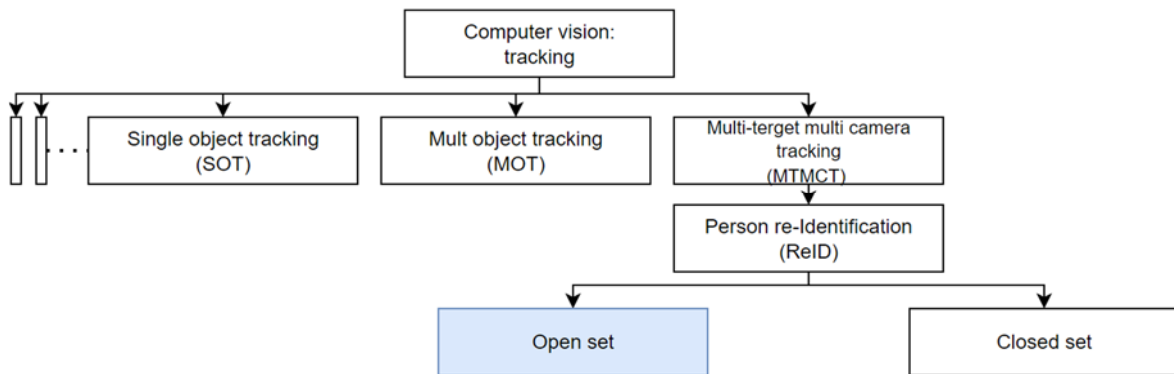


図 1 画像分野における ReID 手法の分類

方向が違う場合の精度も検証する。

## 2. 既往研究

本章では、ICT 技術を用いた人流把握に関する研究、および Person Re-Identification の分野の概説を踏まえて、本研究の位置づけについてまとめる。

### (1) ICT 技術を用いた人流把握に関する研究

葉ら<sup>5)</sup>は Wi-Fi パケットセンサを用いて来街者の回遊行動の把握を試みた。Wi-Fi パケットセンサの電波強度と、街路ネットワーク上の最短経路探索アルゴリズムを組み合わせると、少ない機材でも詳細な移動経路が推定できることを示した。

廣川ら<sup>6)</sup>は高山市の市街地において、Wi-Fi パケットセンサを使用した観光客の行動実態把握について考察した。1日当たりのユニーク ID 数と観光客数の比較や、駅前観光案内所の観測 ID と列車やバスの到着時刻の比較を行い、Wi-Fi パケットセンサによって観光客の動きが観測できることを示した。また取得 ID と取得日時の分析から、彼らの行動実態把握が可能なることも確認した。一方この方法では、一定期間データを取得すると個人が特定される可能性が高まること、分析内容に応じた移動の定義が必要なこと、さらに詳細な行動実態把握にはセンサーの設置範囲と密度の向上が必要なことを指摘した。

新階ら<sup>7)</sup>は阪神尼崎駅周辺において、サンプリング率は高いが空間解像度の粗さが課題の携帯電話基地局データ、計測は容易だが取得範囲の狭さとサンプル率の低さに課題のある Wi-Fi パケットセンサーデータ、最も詳細な移動経路が得られるがサンプルの一般性が低い GPS データといった、それぞれ特性の異なる課題のデータソースを組み合わせ、広域から狭域に至るまでの連続的な人流の把握を試みた。まず携帯基地局データと Wi-Fi パケットセンサーデータを組み合わせ、尼崎市内と首都圏居住者それぞれの滞在エリアを捕捉した。さらに、

Wi-Fi パケットセンサーデータと GPS データを組み合わせることで、詳細な経路選択について一定程度の妥当性を確認した。

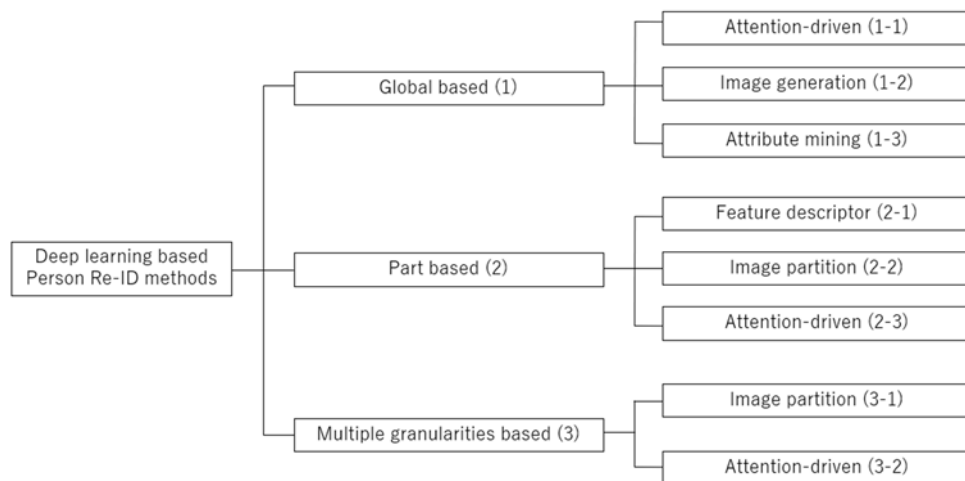
上述した方法は、スマートフォンのようなモバイル端末を所持していることが前提であり、さらにアプリケーションのインストールや Wi-Fi 接続などに関して、歩行者の協力が必要となる。このため、対象となる歩行者の網羅性に課題が残る。一方、カメラ画像による人流把握手法には、このような問題はない。

### (2) Person Re-Identification

Person Re-Identification (ReID) は、異なるカメラによって撮影された人物の対応関係を得るアルゴリズムを研究する分野である。この手法は、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<sup>8)</sup>が、2006年に Gheissari ら<sup>9)</sup>が「人物再同定 (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」と定義したことで、図 2-1 に示すような独立した研究分野として認知されるようになった。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。

図 1 に示すように、ReID には 2 種類の代表的なタスクがある。訓練データとテストデータに同じ人物が含まれる場合と、含まれない場合である。前者は、画像の多クラス分類の問題として扱える。一方で一般的な ReID モデルは、後者を想定してアルゴリズムを学習する。

以下、後者の一般的な ReID モデルの作成手順について、簡単に解説する。ReID モデルでは、人物の特徴量抽出とその特徴量空間内で類似度を比較することによって人物を同定する。まず、訓練データを ReID モデルに入力して特徴量を抽出する。その後、訓練データには含まれない人物の画像群で構成するテストデータを用意する。このとき、テストデータは 2 種類に分割される。ReID モデルに入力する識別対象の画像をクエリ、クエ

図2 特徴表現の分類<sup>10)</sup>

リとの類似度を計算して同一人物の確率が高い順にランキングする画像群を、ギャラリーと呼ぶ。そのうえでギャラリーについてランキングされたリストを用いて同定精度を検証する。ReIDにはクローズドセットとオープンセットという2種類の問題設定がある。前者はクエリとギャラリーに同じ人物が存在する状況であり、後者はクエリとギャラリーに同じ人物が存在しない可能性のある状況である。

オープンセットの場合、検出対象の人物=クエリとギャラリーの類似度を計算した後で、その人物がギャラリーに存在するか否かを識別する。存在すれば該当画像を出力してタスクが終了するが、存在しない場合はギャラリーに新規人物としてIDを割り当てて追加しなければならない。歩行者OD自動観測はオープンセットに分類されるタスクであり、難易度が高い。本論に先立って、以上のレビューを踏まえて論文検索を行った。結論から述べると2023年時点では、オープンセットのタスクに対して十分な精度で作動するReIDモデルは存在しなかった。つまり既存技術のシンプルな転用では、歩行者OD自動観測は実現できないことがわかった。

ReIDの難しさの一因は、あらゆる状況で画像から人物が識別できるような特徴量抽出が難しいためである。ReIDが困難となる典型的な状況・要因は、ギャラリープロンプト間の視点の変化、照明条件の変化であり、画像自体の状況としてオクルージョン、低解像度などが挙げられる。視点の変化とは、カメラごとの視点の違いにより、人体の見える部分が異なることや、カメラとの距離によって見え方が変化する問題である。照明条件の変化とは、画角内の明るさが時間や場所によって変わる問題である。オクルージョンとは、画角内で物や人が奥行き方向で重なる（つまり手前の物体が奥の物体を隠す／自分自身の手が体を隠す等の場合を含む）ことによって、体の一部、あるいは全部が隠れて全身の特徴を取得できない、また

は人物の検出自体が行われにくいという問題である。低解像度は、カメラの性能が不十分で低解像度画像しか得られない場合でもReIDの精度を保ちたいときに適用できるReID手法がないという問題である。これらの問題のため特徴量抽出には、同じ人物でも取得できる特徴が異なる場合や、人物の検出自体に失敗する場合がある。

ReIDでは、精度向上のために様々な技術的な検討が行われる。それらの中で特に本研究に関連する重要な研究テーマは、特徴表現である。これは、画像内の人物の色や形などの外観情報から、ReIDに優れる特徴量を抽出する方法や、その入力方法を検討する研究である。ReID研究の多くは深層学習に基づいて特徴量を抽出する。Liuら<sup>10)</sup>は、特徴を抽出する領域によって図2-2のようにグローバルベース（Global based）手法、部分ベース（Part based）手法、ならびに多重粒度ベース（Multiple granularities based）手法の3分類ができると主張した。以下、同図内に示したそれらの分類番号を参照しつつ、それぞれの具体的な手法を説明する。

(1)のグローバルベース手法では、画像全体から特徴を学習して特徴ベクトルを抽出し、このベクトルを用いてギャラリーを検索する。この手法は比較的単純なネットワーク構造を持っており、背景が単純でオクルージョンがない状況では有効だが、複雑な環境やオクルージョンが深刻な場合などには不向きである。

グローバルベース手法の下位分類には、注意駆動型手法（Attention-driven, 1-1）、画像生成手法（Image generation, 1-2）、属性マイニング手法（Attribute mining, 1-3）の3種類がある。

(1-1)の注意駆動型手法は、人が物を識別するときと同じ領域に、より多くの注意（モデル資源、すなわちモデルの表現力を集中的に割り当てること）を払わせることで、ReID性能を向上させる手法である。注意=Attentionは、入力されたデータのどこに注目すべきかを

特定する仕組みであり、うまく機能すれば精度向上と計算量の削減を同時に達成できる。しかし、注意の必要がないと判断した特徴を無視することにより、人物同定に必要な特徴を見逃す可能性もある。Zhang ら<sup>11)</sup>は、Attention を学習するときにグローバルな構造情報を補足する Relation-Aware Global Attention (RGA) モジュールを提案した。従来法は、Attention の学習に局所的な畳み込みを用いるが、この方法では人体の構造全体に関するグローバルな情報が無視される。そこで Zhang らは、局所的な特徴ベクトルとその他のベクトルとの関係ベクトル（類似度など）を積み重ねて、特徴ベクトルのグローバルな構造をコンパクトに表現する Relation-Aware 特徴量を提案した。これにより、Attention を決定するときにグローバルな情報も利用できるようになった。

(1-2) の画像生成手法は、姿勢、前景、背景などが異なる多様な学習サンプルの生成を目的としている。しかしこの手法には、生成された画像と実際の画像には差があること、複雑で莫大な計算リソースを要求するネットワーク構造になることなどの課題がある。Zhang ら<sup>12)</sup>は、データセット作成のための画像生成コストを低減する UnrealPerson という新しいパイプラインを提案した。この手法では、様々な仮想 3D シーンを 3D ゲームエンジンである Unreal Engine 4<sup>13)</sup> で作成し、ランダムに歩行者を生成して配置する。そして、その歩行者の歩行シーンを仮想空間内に設置した仮想カメラで撮影して、ID 付きの画像を作成する。この手法は、生成する歩行者の外観や周囲の環境を自由に設定した多様な画像を作成できるため、ReID 性能を高める学習データセットの作成に有効と報告された。

(1-3) の属性マイニング手法は、照明や視点の違いによる影響を受けにくい、ロバストな歩行者の属性特徴量を抽出する手法である。上記の通り、歩行者属性はロバストな特徴だが、そのクラス数は少ないので、データベース内で取得できたクラスが偏る問題が起こる。たとえば 2 種類しかない性別属性は、データセットによっては一方に偏る可能性がある。Su ら<sup>14)</sup>は、人物属性も考慮した ReID モデルである Multi-Task Learning algorithm with LOw Rank Attribute Embedding (MTL-LORAE) を提案した。Multi-Task Learning (マルチタスク学習) とは、単一のモデルで複数のタスクを解く機械学習手法である。Su らの手法では複数のカメラからの再同定をタスクとみなして、カメラ間で共有される特徴や情報を利用する。Low Rank Attribute Embedding とは、属性相関を学習して元の 2 値属性を連続関数に変換して、不正確または不完全な属性情報を修正/回復する手法である。Su らはマルチタスク学習に MTL-LORAE を適用することで、分類器の性能を向上させている。

(2) の部分ベース手法は、人体の局所領域に着目し

てきめ細かく識別性の高い特徴量を抽出する。これにより、グローバルベース手法で無視される細かい特徴を学習できる。この手法は、部分的な ReID やオクルージョンを含む ReID に有効だが、グローバルな識別情報を無視する可能性がある。部分ベース手法は、さらに特徴記述子手法 (Feature descriptor, 2-1)、画像分割手法 (Image partition, 2-2)、注意駆動型手法 (Attention-driven, 2-3) の 3 種類に分類できる。なお、(2-3) の注意駆動型手法は (1-1) とは異なり、局所領域に対して働く。

(2-1) の特徴記述子手法は、キーポイント周辺の特徴的な情報のみを含むベクトルの集合を求める手法である。この手法は回転や拡大縮小、視点変更、照明変化などに強いが、特徴マッチングに要する時間が長く、計算量も膨大である。Suh ら<sup>15)</sup>は、画角内に捉えた身体部位の違いに対応するため、部位ごとの特徴を学習するネットワークを提案した。このネットワークは、外観と身体部位の特徴マップを生成する 2 ストリームネットワークと、それらを 1 つに組み合わせるバイリニアプーリング層からなる。外観特徴マップは GoogLeNet<sup>16)</sup>、身体部分マップは OpenPose<sup>17)</sup> により出力する。バイリニアプーリング層によって 2 つの特徴マップが集約され、それらは外観記述子として出力される。Suh らはこれにより、身体的位置ずれによる影響を軽減できることを示した。

(2-2) の画像分割手法は、画角内の歩行者をいくつかの局所領域に分割して、それぞれ特徴抽出する。このとき、単純な分割を適用すると外れ値が多くなってあまり有効に機能しない。しかし、複雑なキーポイント推定やセマンティックセグメンテーションによる分割を行うと、その処理の計算コストが大幅に増加する。キーポイント推定は上述の難点を解消するために、関節などの特徴点を推定する方法である。またセマンティックセグメンテーションは、人物と背景をピクセル単位で区別するような、詳細なラベル付けを指す。Huang ら<sup>18)</sup>は、マルチスケール特徴学習、水平ストライプ分割、およびセマンティックセグメンテーションを 1 つのフレームワークに統合した MDRS (Multi-scale Discriminative network with Region Segmentation) を提案した。Huang らのモデルはマルチタスク学習をベースとしており、ID 損失、トリプレット損失、マルチセグメンテーション損失の 3 つをタスクとする。Huang らの手法では、まず、ResNet-50 によりマルチスケール特徴マップを抽出する。このとき、撮像の大きさによって分割数を変えながら、横方向に分割する。そして、分割前の特徴マップと分割された後の全ての特徴マップに対してプーリングと次元削減を行い、256 次元のベクトルとする。それらを足し合わせたストライプ特徴ベクトルを生成して、ID 損失を計算する。次に、ストライプ特徴ベクトルを結合した新しいベクトルを用いてトリプレット損失を計算する。そして、ピクセル単

位のクロスエントロピー損失を足し合わせてマルチセグメンテーション損失を求め、最後に、これらに重みをかけて各損失関数の和を求めて学習する。一連の処理によって Huang らは、姿勢の違いやオクルージョンなどに強いモデルとなった、と主張している。

オクルージョンに関連する問題は Partial ReID 問題と呼ばれていたが、この課題に正面から取り組んだ研究は乏しかった。つまり従来の研究では、基本的にすべてのギャラリー画像には全身が映っているという暗黙の仮定があった。(2-3)の注意駆動型手法として Miao ら<sup>19)</sup>は、オクルージョンの発生した領域を除外して識別性の高い特徴量を抽出する Pose-Guided Feature Alignment (PGFA) という手法を提案した。PGFA における特徴量抽出は、まず姿勢推定器により人体の 18 か所の座標を取得してヒートマップを作成する。これを特徴量マップと結合して、オクルージョン部分の特徴を無視した特徴量ベクトル；pose-guided global feature を抽出する。さらに特徴マップを水平方向に分割し、部分特徴量ベクトルを得る。ギャラリーとクエリのマッチングでは、両方に映った領域における部分特徴量ベクトルと、pose-guided global feature のそれぞれのコサイン類似度を計算してマッチングを行う。さらに彼らは、より実環境に近いオクルージョン画像で ReID を行う Occluded ReID 問題の定義と、この問題用のデータセット；OccludedDukeMTMC を作成した。Miao らの Occluded ReID 問題は、クエリとギャラリーの両方がオクルージョン画像となる場合の ReID 問題であり、実環境で取得できる画像で ReID を運用することを強く意識している。Miao らの手法は、3つのオクルージョンデータセット<sup>19), 20), 21)</sup>において既存の ReID 手法を大きく上回った。

(3)の多重粒度ベース手法は、グローバルベース手法と部分ベース手法の両方の利点を持つが、複雑なネットワークと膨大な計算リソースを必要とすることが多い。以下では、多重粒度ベース手法に広く用いられている画像分割手法 (Image partition, 3-1) と注意駆動型手法 (Attention-driven, 3-2) について述べる。

(3-1)の画像分割手法として Wang ら<sup>22)</sup>は、オクルージョンのある画像の ReID に強い Shrinking and Reweighting Network (SRNet) を提案した。オクルージョンの激しい画像からグローバル特徴を抽出するとき、抽出器は歩行者以外の部分にも注目してしまう。この問題を緩和するために、Wang らは Deep Residual Shrinkage Networks (DRS モジュール) を追加した。DRS モジュールは、0 に近い特徴量を 0 に置換するためにソフト閾値を学習して、オクルージョンに起因するノイズ混じりの特徴量を除く。この工夫により、集中的に歩行者部分に焦点を当てて識別性の高い特徴を学習できるようになった。さらに、オクルージョンが起きている身体部分の特徴が

ReID に及ぼす影響を弱めるために、部分ごとに特徴の重みを学習する Reweight Module for Part Matching (RMPM) を設計した。提案手法の性能を Partial-REID Dataset と Occluded REID Dataset を用いて検証したところ、当時の最先端手法よりも優れた精度を示した。

(3-2)として Liu ら<sup>23)</sup>は、多階層の Attention マップを異なる特徴層に多方向的に与える、HydraPlus-Net (HP-net) を提案した。これは、単純な CNN からなる MainNet (M-net) と、3つの multi-directional attention (MDA) モジュールを組み合わせた Attentive Feature Net (AF-net) によって構成される。MDA モジュールとは、CNN 内の複数の層から Attention マップを作成し、異なるレベルの特徴マップに掛け合わせて多レベルの Attention 特徴量を生成するモジュールである。なお、ここで“レベル”とは注目する部分の大きさであり、たとえば服装の柄のように画像全体から得られる情報は低レベル、髪の毛の長さのような局所的な情報は高レベルと呼ばれる。M-net と AF-net の出力は連結されて、最終的にグローバルな特徴と部分的な特徴を併せ持つ特徴ベクトルを出力する。

以上のように、特徴表現に関する研究は盛んに行われているものの、すべての課題を解決できる突出した性能の特徴量抽出手法はない。したがって、図 2 に列挙されたそれぞれの特徴量分類に対応する ReID アルゴリズムの優劣は、つけがたい状況にある。

ReID 分野では、基本的に外観特徴量を使用しており、本研究で使用する骨格特徴量を補助的に利用することが多い。その他の特徴量を使用した人物同定手法として歩容認証がある。歩容認証では、同一被験者を複数方向から撮影した Gait Energy Image (GEI: 歩容エネルギー画像) を使用する。GEI は、1 歩行周期を平均した画像であり、体型などの静的な成分と脚や腕の動きといった動的な成分を含む。守脇ら<sup>4)</sup>は人物属性を考慮した Parallel-GEINet という歩容認証手法を提案した。GEINet とは、GEI を入力する CNN であり、提案手法では性別や年齢といった属性と、人物 ID に関する特徴量抽出器として使用する。Parallel-GEINet は、人物 ID に関する GEINet に、抽出する属性数分の GEINet を並列に配置した特徴量抽出器である。ReID は、ギャラリーとクエリの抽出された特徴量のユークリッド距離を計算して行う。提案手法の性能を検証したところ、GEINet よりも精度が向上した。

Gul ら<sup>24)</sup>は、3D CNN による歩容認証手法を提案した。3D CNN は空間と時間に関する畳み込みを行うため、体型などの空間的特徴と歩行パターンの時間的特徴の両方を抽出できる。GEI を用いることにより、画像のノイズや照明の変動などを排除でき、歩行全体の情報をとらえた ReID ができる。また、ベイズ最適化によるハイパーパラメータ探索により、性能の向上が可能な点も注目される。

### (3) 本研究の位置づけ

以上で概観したように、これまで検討された候補の中で、他を圧倒する性能の特徴量は見つかっていない。歩容は多くの ReID で用いられる外観特徴量とは異なり、服装や持ち物の変化に左右されにくい一方で、算出の計算量が大きいことや、たとえば非常に重い物やかさばる物を抱えたときの歩容の違いやオクルージョンの問題も懸念される。つまり歩容まで含めて考えても、骨格、外観のいずれの特徴量が ReID に対して有効かは、今のところ結論は出ていない。

そこで本研究では、ReID に関して有効な特徴量に関する精度検証を行う。具体的には、昨年度に引き続き、骨格特徴量単体の性能を明らかにすると同時に、ReID で一般的に使用される外観特徴量単体での精度を確認する。また、既往研究と同様に外観特徴量を使用する際の、骨格特徴量の補助的な活用に関する検討も行う。

## 3. 人物再同定における特徴量抽出と分類方法

### (1) 同一人物画像群の作成

昨年度までの特徴量抽出手法では、歩行動画に `tf-pose-estimation` を適用して関節座標を取得していた。その際に、画角の中で複数の人物が歩行しているとき、人物のトラッキングが不十分なため、ある人物と別の人物の ID が、それらの人物のすれ違い時などに入れ替わることがあった。本研究ではこの問題を解決するために、特徴量抽出器に入力する前に `Simple Online and Realtime with a Deep Association (DeepSORT)` という物体追跡アルゴリズムを適用して、画角内の人物 ID が同一となるようにまとめた同一人物画像群を作成する。

Bewley は、2016 年に一台のカメラで複数の物体を追跡する `Multi Object Tracking (MOT)` の基礎的アルゴリズムである `Simple Online and Realtime Tracking31) (SORT)` を示した。SORT では、画角内の物体検出を行い、捉えた物体を `bounding box (bbox : バウンディングボックス)` と呼ばれる矩形領域で表示する。物体追跡、つまりトラッキングはこの `bbox` をフレーム間で結びつけることを指す。SORT は検出された `bbox` の位置情報を用いて移動先の `bbox` を予測し、検出された `bbox` と予測した移動先の `bbox` との重なり具合を表す `Intersection over Union (IoU)` を計算する。その上で、フレーム結合の候補となるいくつかの物体間の `IoU` 行列の擬対角成分を最大化するように `bbox` の割り当てを結合するハンガリアン法を行って、物体追跡を実現する。

位置情報のみを扱う SORT では、オクルージョンのとき、人物 ID の消失や振り替わりが起こる課題があった。そこで外観特徴量も考慮して物体追跡を行う

`DeepSORT32)` が開発された。`DeepSORT` では、外観特徴量に加えて、SORT では使用していなかった `bbox` の高さも考慮するようにした。これにより、画角の出入りを伴う物体のトラッキングが可能となり、トラッキング性能が向上した。本研究では `DeepSORT` を使用して動画内の全ての通行者を `bbox` で切り出し、画角内における同一人物の歩行画像集合を作成する。

### (2) 特徴量抽出手法

`DeepSORT` から得た同一人物画像群を `MediaPipe Pose33)` に入力して、関節座標を抽出する。なお、本研究では取得した関節座標を骨格特徴量と呼ぶ。`MediaPipe` は、Google 社が提供するオープンソースの機械学習ソリューションである。`MediaPipe` 内で提供されている機能の一つである `Mediapipe Pose` は、映像から人物の姿勢を推定できるライブラリである。

`Mediapipe Pose` は、検出器とトラッカーの 2 種類の機械学習パイプラインを使用する。まず、検出器により人物 / 姿勢関心領域 (`person/pose region-of-interest, ROI`) を特定し、トラッカーにより ROI 内の関節座標とセグメンテーションマスクを予測する。セグメンテーションマスクとは、背景と対象物体の領域の区別を指す。関節座標は、図 3-2 に示す 33 か所について取得する。取得方法は `POSE_LANDMARKS` と `POSE_WORLD_LANDMARKS` の 2 種類だが、本研究では後者で取得した。これは腰の中心を原点とするメートル単位の 3 次元座標を取得する方法であり、画像内の腰の中心から、x 軸は右方向、y 軸は下方向、z 座標はカメラから離れる奥行き方向が正となる。なお、`Mediapipe Pose` では、画像内に関節が映っていない場合も、それらの座標の推定値を取得できる。また各関節座標と共に、画角内にその関節が映る確率 (`visibility`) も同時に取得できる。

外観特徴量は、`DeepSORT` により作成した同一人物画像群を `Omni-Scale Network34) (OSNet)` に入力して抽出する。Zhou らが開発した `OSNet` は、異なる畳み込みフィルターのサイズを設定した複数のストリームから得られた特徴量を、一つに統合した特徴量であるオムニスケール特徴量を抽出して、人物の識別を行う `ReID` モデルである。彼らは人物を識別するためには全身の特徴だけではなく、局所的な特徴も同じく重要と考えて、これらのスケールの違う特徴を統合したオムニスケール特徴量を用いる手法を検討した。Zhou らはオムニスケールで表現した特徴量の組み合わせこそが、人物を同定する上で最も効果的でユニークな外観特徴となる、と主張している。`OSNet` は、オムニスケール特徴量を抽出して `ReID` を行う `deep convolutional neural networks (CNN, 深層畳み込みニューラルネットワーク)` である。本研究では、`OSNet` の `ReID` 機能は用いずにオムニスケール特徴量、



図3 学内のカメラ設置場所と画角

つまり人物の外観を表す 512 次元の特徴量を抽出するための特徴量抽出器として使用する。

### (3) 人物同定のための分類手法

本研究では、人物同定のための分類手法としてロジスティック回帰モデル<sup>35)</sup>、非線形サポートベクターマシン<sup>36)</sup>、およびランダムフォレスト<sup>37)</sup>を用いる。

## 4. 各特徴量による分類精度の検証

### (1) 歩行者画像取得実験の概要

本研究では、実際に歩行者 OD を推定する環境に近い状況を想定して、2022 年 10 月 24 日に広島大学東広島キャンパス内の工学部周辺にカメラを設置して歩行者を撮影した。この撮影には汎用の Web カメラとして普及している GoPro を使用した。設置場所と各画角は図 3 の通りである。この実験は様々な画角から撮影するために、地点 1, 2, 3, 4, 5 では三脚を使用し、地点 6 でカメラを街路灯に固定して撮影した。また地点 4 は、約 5.6 メートルの高さの渡り廊下から、地上を見下ろす俯角をつけて撮影した。実験は、6 名の被験者に協力の下で、1) 10:00~10:15 の間に、被験者に回る地点の順序を指示して歩行するパターン、2) 10:20~10:35 の間に、1) とは別の順序で指示を出して歩行するパターン、3) 10:45~11:00 の間に、指示を出さずに自由に歩行するパターンの、計 3 パターンを行った。

撮影に際して、実験内容とプライバシーポリシーを記載した A4 のチラシを各カメラ付近に掲示した。このチラシ内のリンクを辿ると、期間内に一般の歩行者から削

除の問い合わせがあった場合には、映像のうち本人の写っている箇所を削除することができる。これは、個人情報の実験データへの組み入れを望まない人物が、事後的にデータ削除を申し出られるオプトアウトという手続きであり、画像を撮影して開発を行う際の倫理コードである。本実験は広島大学が定める学術研究における個人情報保護ガイドラインに則っており、以上のプロトコルは社会基盤環境工学プログラムにおける研究倫理審査委員会の承認を受けている。

歩行経路の確認のため、被験者には「ジオグラフィカ | 登山用 GPS」<sup>38)</sup> というスマートフォンアプリをダウンロードさせて、GPS により 5 秒間隔で場所とその時間を記録した。記録したデータは被験者のスマートフォンに GPX 形式で保存させて、実験後に「カシミール 3D」<sup>39)</sup> というフリーソフトを用いて GPX ファイルからテキストデータを抽出できる。このファイルには被験者の歩行経路が時刻情報付きの点列として格納されている。この情報とカメラ画像の画角内に現れた時間を突合して、各被験者の移動形態を詳細に記録した。取得した GPS 経路の例を図 4 に示す。

本研究では、骨格特徴量、外観特徴量および、骨格特徴量と外観特徴量を結合した特徴量、骨格特徴量の visibility によって下処理をした外観特徴量の 4 種類の特徴量を抽出、作成する。本節ではその作成方法について述べる。まず DeepSORT を歩行動画に適用して同一人物画像群を作成した。撮影した動画のフレームレートは 30FPS だったが、画像の量が膨大なため 10FPS に変更して画像の切り出しを行った。さらに、実験 1,2 で被験者が指定した地点を回り終わる時間までとしてそれぞれ 10:00~10:15 の 15 分間および 10:20~10:31 の 11 分間とした。



図 4 カシミアール 3D による歩行経路の表示例

実験 3 に関しては 10:45~11:00 の 15 分間とした。また、画像ファイルの命名規則は「実験番号\_地点番号\_総フレーム数\_信頼度」とした。ここで信頼度とは、DeepSORT が人物を検出する際の信頼度を指す。DeepSORT による人物切り出しが終わったのち、それぞれの被験者と被験者以外のすべての一般通行者を分けるアノテーション作業を人手によって行って、同一人物画像群を作成した。

機械学習モデルの訓練に inputs する画像を増やすために、Data Augmentation (データオーギュメンテーション) を行う。これは、訓練データ画像に画像処理によるノイズを加えることにより、訓練に使用できる画像枚数を人為的に増やす手法のことである。学習の効率を高め、撮影条件に対する物体検出の頑健性を高めるために行われる。本研究では、機械学習におけるデータ拡張用 Python ライブラリ Albumentations<sup>40)</sup> を使用した。画像をぼかす Blur, ランダムに明るさを変える RandomBrightness, ランダムにコントラストを変える RandomContrast の 3 パターンを行い、訓練データを増やした。各方法における変化の範囲は全てデフォルト値で、Blur は (3, 7), RandomBrightness は (-0.2, 0.2), RandomContrast は (-0.2, 0.2) である。これらのオーギュメンテーションによって取得した画像の枚数を 4 倍にした。

上記の作業で作成したオーギュメンテーション済みの同一人物画像群を、骨格特徴量を抽出する MediaPipe Pose と、外観特徴量を抽出する OSNet にそれぞれ入力して、特徴量を抽出して CSV ファイルに保存した。各 CSV ファイルの 1 列目には人物ラベルを入力した。人物ラベル

は 0 から 6 までの番号で管理し、被験者以外の一般通行者は 0, 被験者はアルファベット順に 1 から 6 までの番号を付した。また 2 列目には、その他の一般通行者の場合、DeepSORT が自動的に付与される ID を入力した。被験者の場合、地点番号と DeepSORT が自動的に付与される ID を入力した。3 列目には画像名を記入し、4 列目から特徴量を入力した。

外観特徴量は、OSNet により抽出した。入力した画像の大きさを (128, 64) ピクセルにリサイズしたのち、512 次元の外観特徴量を抽出した。なお OSNet はどのような大きさの入力画像であっても特徴量抽出が可能だが、対象とする人物ができるだけ大きなサイズで歪みなく鮮明に撮影された画像を入力するほうが、外観特徴量の質は高くなると考えられる。骨格特徴量は、MediaPipe Pose により抽出した。MediaPipe Pose も OSNet と同様に特徴量抽出の前に、入力画像のリサイズを行った。画像の幅と高さのうち、大きい方を 480 ピクセルにして元画像とアスペクト比が変わらないように調整した。その後、33 か所の x, y, z 座標と visibility の合計 132 次元の特徴量を抽出した。骨格特徴量として使用する際には、人体の特徴に関係の無い visibility を削除した 99 次元を使用する。ただし、MediaPipe Pose は人物の検出に失敗した場合は関節座標が抽出されない。そのため、一連の処理によって出力された画像枚数を比較すると、外観特徴量に比べて骨格特徴量の画像枚数は少なかった。

骨格特徴量と外観特徴量を結合した特徴量とは、外観特徴量を骨格特徴量と同じ次元数をそろえて結合した 198 次元の特徴量である。異なる特徴量間の入力情報のばらつきを、次元数で制御してそろえるために外観特徴量を 512 次元から 99 次元に削減した。次元削減には Uniform Manifold Approximation and Projection (umap)<sup>41)</sup> を使用した。なお、両者の結合は骨格と外観の両方の特徴量が存在する画像のみ行われ、片方しか存在しないとき、その画像は削除される。そのため外観単体の特徴量よりも、取得できたデータ数は少なくなる。骨格特徴量によって下処理をした外観特徴量とは、骨格特徴量の visibility を参照して人体の各部分が十分に取得されていない画像を削除し、残った画像のみで構成した外観特徴量のデータセットである。

訓練データとテストデータの作成方法を述べる。訓練データとテストデータそれぞれにおいてサンプル数が最小の被験者に合わせて、その他の各被験者のサンプルをランダムに削減する。その後、テストデータからオーギュメンテーション済みの画像を削除する。特徴量を分類器に入力する際に各次元の値の範囲が異なる場合、モデルの学習に時間がかかったり、精度が低下したりする可能性がある。これを防ぐために、特徴量を正規化して入力する。

## (2) 検証結果

3 種類の分類器による的中率を平均して、分類器によらず的中率が高くなる特徴量の入力方法を明らかにする。さらに訓練データとテストデータ間で、歩行経路やカメラが異なる場合の的中率を検証する。分類対象は、被験者 6 人と被験者以外のその他の人物を 1 つにまとめた、合計 7 人分のラベルである。なお被験者に比べてその他の人物のデータ数が多いため、その他の人物のデータの中からランダムに抽出した人物の特徴量を使用した。以下の分析では、被験者のデータ数を踏まえてその他のデータ数は 20,000 個とした。

分類器の分類性能はデータセットの特性に左右されるため、試行錯誤によって値を定めなければならないハイパーパラメータが、分類器ごとにある。この値を適切に定めることを分類器のチューニングと呼ぶ。チューニングは、グリッドサーチと交差検証を組み合わせで行う。グリッドサーチとは、探索対象のハイパーパラメータの候補を列挙し、その中から最適な組み合わせを探索する手法である。交差検証とは、訓練データを分割して、片方を検証用、残りを訓練用としてモデルの学習を行うことである。検証用と訓練用の画像の組み合わせを分割数の分だけそれぞれ異なるデータセットを入力して学習を行い、それらの精度を平均してモデルを得る。グリッドサーチと交差検証の組み合わせとは、パラメータの組み合わせごとに交差検証を行い、モデルの精度が最良となるパラメータを探索する方法である。

本研究の分割数は 5 回として、分割方法は層化 K 分割法を採用した。これは、訓練用と検証用のデータ数、つまり本研究では人物ラベルの割合が等しくなるように分割する方法である。また、層化 K 分割法の引数 `shuffle` を `True` とした。これによりグリッドサーチのとき、元々地点別に並べて編成したデータの順序をランダムに並べ替える設定となるので、訓練用と検証用のデータで地点の偏りが無い設定とした。グリッドサーチでは、以下の表 4-1 の値を組み合わせで最適なパラメータの探索を行った。探索パラメータを増やすほど計算時間が増えるため、探索する組み合わせ総数は 10 個程度に抑えた。

グリッドサーチの対象となる分類器のハイパーパラメータは、ロジスティック回帰モデルでは正則化パラメータ `C`、SVM では正則化パラメータ `C` とカーネル関数内に現れる係数 `gamma`、ランダムフォレストでは不純度の評価指標 `criterion`、決定木の個数 `n_estimators`、決定木の深さ `max_depth` である。次に各分類器の引数の設定について解説する。ロジスティック回帰モデルでは、最適化関数 `solver` を大きいデータセット向けの `sag`、`multi_class` を多クラス分類用類用の `multinomial`、学習時の反復の最大回数を示す `max_iter` を 1000 とした。SVM では、`max_iter` を 1000 とした。ランダムフォレストの引数はデフォル

表 1 検証の仕様

データ	撮影時間	地点	人物ラベル
訓練	1,2,3	1,2,3,4,5	0,1,2,3,4,5,6
テスト	1,2,3	6	0,1,2,3,4,5,6

表 2 使用する合計のデータ数

特徴量	訓練データ数	テストデータ数
骨格特徴量	33,250	1,064
外観特徴量	88,200	2,765
骨格&外観特徴量	33,250	1,064
外観特徴量 (下処理)	20,349	833

表 3 特徴量ごとの的中率比較

特徴量	ロジスティック回帰	非線形 SVM	ランダムフォレスト	平均
骨格	22.7%	37.3%	23.4%	27.8%
外観量	51.0%	50.5%	42.5%	48.0%
骨格&外観	27.6%	22.5%	14.9%	21.7%
外観+下処理	71.4%	65.8%	56.7%	64.6%

ト値を採用し、各分類器の `random_state` を 0 とした。これは訓練データをシャッフルするときなどで使われる乱数生成のシード値である。この値を指定しておくと同様の分類を繰り返しても同一の乱数列が生成されるため、その乱数列の特性に左右されることなく、検証できる。これらのパラメータについて、グリッドサーチと交差検証を行った後に、最も優れたモデルを引き継いでテストデータで精度を求める。本分析の精度指標は、正解ラベルに対して最も分類確率が高いラベルを出力する場合の、両者の重複割合を示す的中率とした。

特徴量ごとの的中率を求め、`ReID` に優れる特徴量を探索する。歩行者 `OD` では複数地点間での人物同定を目指している。そこで地点 1 から 5 までを訓練データ、地点 6 をテストデータとした。この設定により、既知地点と未知地点間での精度検証を行う。的中率は特徴量ごとに 10 回求めて、それらを平均する。検証の仕様を表 1、検証に使用するデータの情報を表 2、分類器ごとの的中率、ならびに平均を表 3 にそれぞれ示す。なお、表 3 の訓練データ数は昨年行った先行研究での  $10^2$  程度から  $10^4$  に増加している。

表 3 に示した通り、特徴量ごとの的中率の平均値は 21.7% から 64.6% の範囲でばらついており、入力した特徴

量の種類または処理手順が分類の精度に影響することがわかる。また分類器ごとに特徴量の精度を降順に並べると、分類器間で大きな順序の違いはみられなかった。つまり、分類手法と特徴量の組み合わせに関する相互作用は、ほとんど見られないことが分かった。

全体として最も優れた特徴量の入力方法は、全身の主要な関節の visibility に基づいて下処理した外観特徴量であった。これは、下処理によって全身が映っていない画像や人物検出が難しい画像の外観特徴量が削除され、より特徴量を抽出しやすい画像を分類器に入力できたため、と考えられる。また下処理によって入力画像の枚数が 1/2 から 1/4 程度に絞られても、ReID 精度は同等以上となることが注目される。ただし、最も優れた分類器でも的中率は 71.4% に留まっており、訓練データに無い地点の画像を正確に人物同定することは難しいことも分かった。

次に被験者の撮像方向の違いに対する人物同定精度の頑健さの検証を行う。3 回の撮影のうち、被験者は各地点を少なくとも 2 回通過する。撮影時間で画像データセットを分割することにより、同一地点のカメラでもカメラとの距離、歩いてくる方向、オクルージョン、ならびに光の当たり方などの条件が異なる。つまり、訓練データとテストデータの間の撮影時間の違いは、被験者の撮像方向の違い、および付随する諸条件の違いとして、入力画像に埋め込まれている。使用する特徴量は、特徴量ごとの的中率比較で最も高い精度が得られた、全身の visibility を用いて外観特徴量の下処理を行った特徴量とする。先の的中率の検証と同様に、的中率は 10 回の平均値とする。検証の仕様を表 4、検証に使用するデータの情報を表 5、分類器ごとの的中率、およびその平均を表 6 に示す。

表 6 より、撮像方向が訓練とテストで違う場合、平均的中率は 70.5% となった。この結果を表 4 に示した同じ特徴量を使用した結果(64.6%)と比べると、こちらの的中率が高い結果となった。すなわち、学習と訓練で同じ画角を用いる方が、的中率は高くなる。これは、訓練データとテストデータに似た画像が含まれているためと考えられる。この検証では訓練データとテストデータで、撮影時刻のみ違う画像を使用した。特定地点の撮影時刻が違っていても、地点の歩行経路の制約のため同じ方向に歩いた被験者が生じて、訓練データとテストデータに似た画像が含まれることがある。つまり、訓練データとテストデータにおいて同じ画角ならば、個々の被験者の歩行経路が違っていても、カメラとの距離や人体の映る角度が似ている画像が一定程度含まれるため、人物同定精度を高める要因になると考えられる。

表 4 検証の仕様

データ	撮影時間	地点	人物ラベル
訓練データ	1,2	1,2,3,4,5,6	0,1,2,3,4,5,6
テストデータ	3	1,2,3,4,5,6	0,1,2,3,4,5,6

表 5 使用する合計のデータ数

特徴量	訓練データ数	テストデータ数
外観+下処理	12,622	2,722

表 6 撮像方向の違いによる的中率比較

特徴量	ランダム			平均
	ロジスティック回帰モデル	SVM	フォレスト	
外観+下処理	75.8%	75.7%	60.0%	70.5%

## 5. おわりに

本研究では歩行者 OD 自動観測を目指して、1) ReID の特徴量抽出手法に関するレビュー、2) 特徴量ごとの ReID の精度検証、および 3) 撮像方向が違う場合の ReID の精度検証に取り組んだ。1) について、ReID における特徴量抽出手法は、特徴量を抽出する領域によって 3 分類できると判明した。そして分類ごとに、代表的な手法をまとめた。

2), 3) への入力画像を得るため、学内 6 か所に設置したカメラで撮影した、実環境に近い歩行映像を使用した。映像からの特徴量抽出手順は、以下の通りである。まず撮影した歩行映像を DeepSORT に入力して、人物をフレームごとに切り出す。次に、それらの画像のアノテーションを行う。そしてそれらの画像にオーギュメンテーションを行って画像枚数を増やした上で、それらの特徴量抽出器に入力して特徴量を抽出する。本研究では、ReID で主に用いられている外観特徴量と、昨年度まで使用していた骨格特徴量、両者を結合した特徴量、および骨格特徴量により下処理した外観特徴量の 4 種類の特徴量を検討した。なお、外観特徴量は OSNet、骨格特徴量は MediaPipe Pose で抽出する。また ReID 精度は、ロジスティック回帰モデル、SVM、ランダムフォレストの 3 種類の分類器で求める。

2) について、訓練データとテストデータの地点が重ならないような設定とした。これは、歩行者 OD を推定する際にある地点のデータを用いて学習を行い、別地点のデータにその学習済みモデルを適用して ReID を行う状況を想定するためである。この設定の下で 3 種類の分類器の ReID 精度を平均して、分類器によらず的中率が

高くなる特徴量を検証した。これにより、どの ReID モデルに入力しても有効な特徴量を明らかにする。その結果、骨格特徴量により下処理した外観特徴量の ReID 精度が、64.6%と最も高くなった。これは、骨格を参照した下処理によって全身が映っていない画像が削除された結果、より特徴量を抽出しやすい画像を学習に使用できたためと考えられる。

3) では、訓練データとテストデータの撮影時刻を変えて、被験者が画角内に侵入する撮像方向の違いに関する ReID の頑健性の検証を行った。先の検証結果に基づいて、骨格特徴量により下処理した外観特徴量を使用した結果、平均の ReID 精度は 70.5%となった。先の結果と比べると、こちらの設定の方が精度が高くなった。したがって、訓練データとテストデータに同じ画角を用いる方が ReID 精度が高くなることが分かった。これは、訓練データとテストデータに似た画像が含まれているためと考えられる。3) では、訓練データとテストデータで同一地点において撮影時刻のみが違う画像を使用した。撮影時刻が違っていても、両方の撮影時刻のある地点において、同じ方向に歩いた被験者もいるため、訓練データとテストデータに似た画像が含まれていた。このことが ReID 精度を高めた要因と考えられる。

ここで「訓練データとテストデータが似ている」とは、人物ではなく背景であることに注意すべきである。ReID を人物特徴量にのみ基づいて行う方針ならば、この結果は、分析者に追加のデータ処理を要求する。つまり図 2 で示した ReID 手法(2-2)に関連するセマンティックセグメンテーションを適用することで、背景ノイズを除去する必要がある。しかしカメラ設置点の撮像条件を加味して ReID を行う方針ならば、背景情報はある程度積極的に活用すべきである。この場合、本研究の OSNet のように、既存の特徴量抽出アルゴリズムを活用する処理手順であれば、後段の分類器の学習において、汎化性能を高めるチューニング、すなわち転移学習を効率的に行う方策を検討しなければならない。今後は、特定の環境条下でいずれの ReID 性能が高くなるか、十分に注意を払った検討が必要である。このことはまた、既往研究の ReID 精度を比較する際に、検証に用いたデータの環境条件(背景の有無など)にも注意を払う必要があることも示すと思われる。

## REFERENCES

- 1) 国土交通省白書 2020, 第 2 部第 4 章 2.3 節 地域特性を活かしたまちづくり・基盤整備  
<https://www.mlit.go.jp/hakusyo/mlit/r01/hakusho/r02/html/n2423000.html> (2023 年 2 月 6 日閲覧)
- 2) 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](https://www.mlit.go.jp/toshi/tosiko/toshi_tosiko_tk_000034.html) (2023 年 2 月 6 日閲覧)
- 3) 胡内健一, 長谷川宗昭, 太田恒平: 携帯 GPS 情報を用いた歩行者動線分析, 土木計画学研究発表会・講演集, Vol.52, No.221, pp.1532-1536, 2015
- 4) 守脇幸佑, 村松大吾, 八木康史: 人物属性を考慮した歩容認証, 情報処理学研究報告, 2018 号, CVIM-212, No.22, 1-7, 2018.
- 5) K. Sippakom, 葉健人, 八木優治, 青木保親, 猪井博登, 土井健司: 限定設置条件下での Wi-Fi パケットセンサに基づく来街者の回遊行動の把握, 土木計画学研究発表会・講演集, Vol.59, No.54, pp.1-9, 2019.
- 6) 廣川和希, 笹圭樹, 和泉範之, 絹田裕一, 牧村和彦, 西田純二: Wi-Fi パケットセンサーを用いた人の行動実態の把握～観光都市・飛騨高山での活用に向けて～, 土木計画学研究発表会・講演集, Vol.54, No.162, pp.1180-1185, 2016.
- 7) 新階寛恭, 河井裕紀, 絹田裕一, 和泉範之, 上野優太, 宮内弘太: 携帯基地局データや Wi-Fi パケットセンサーデータ等の統合による多様なスケールでの連続的な人の流動把握手法に関する研究, 土木計画学研究発表会・講演集, Vol.66, 2022.
- 8) Wang, X.: Intelligent multi-camera video surveillance: A review, *Pattern Recognition Letters*, vol. 34, no. 1, pp. 3-19, 2013.
- 9) Gheissari, N. Sebastian, T. and Hartley, R.: Person Re-Identification using spatiotemporal appearance, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1528-1535, 2006.
- 10) Liu, M. Zhao, J. Zhou, Y. Zhu, H. Yao, R. and Chen, Y.: Survey for person re-identification based on coarse-to-fine feature learning, *Multimedia Tools and Applications*, Volume 81, Issue 15, pp. 21939-21973, 2022.
- 11) Zhang, Z. Lan, C. Zeng, W. Jin, X. and Chen, Z.: Relation-aware global attention for person re-identification, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 3183-3192, 2020.
- 12) Zhang, T. Xie, L. Wei, L. Zhuang, Z. Zhang, Y. Li, B. and Tian, Q.: Unrealperson: An adaptive pipeline towards costless person re-identification, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.11506-11516, 2021.
- 13) Epic Games, <https://www.unrealengine.com/ja>
- 14) Su, C. Yang, F. Zhang, S. Tian, Q. Davis, L.S. and Gao, W.:

- Multi-task learning with low rank attribute embedding for person re-identification. IEEE International Conference on Computer Vision (ICCV), pp 3739–3747, 2015.
- 15) Suh, Y. Wang, J. Tang, S. Mei, T. and Lee, K.: Part-Aligned Bilinear Representations for Person Re-identification, ECCV, Computer Vision Foundation, 2018.
  - 16) Szegedy, C. Liu, W. Jia, Y. Sermanet, P. Reed, S. Anguelov, D. Erhan, D. Vanhoucke, D. Rabinovich, A. et al.: Going deeper with convolutions, CVPR, 2015.
  - 17) Cao, Z. Simon, T. Wei, S. and Sheikh, Y.: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, CVPR, pp1302-1310, 2017.
  - 18) Huang, J. Liu, B. and Fu, L.: Joint multi-scale discrimination and region segmentation for person re-id, Pattern Recogn Lett 138, 2020,  
<https://doi.org/10.1016/j.patrec.2020.08.022>
  - 19) Miao, J. Wu, Y. Liu, P. Ding, Y. and Yang Y.: Pose-guided feature alignment for occluded person re-identification, IEEE/CVF International Conference on Computer Vision (ICCV), pp 542–551, 2019.
  - 20) Zheng, W. Gong, S. and Xiang, T.: Person re-identification by probabilistic relative distance comparison, CVPR, 2011.
  - 21) Zheng, W. Li, X. Xiang, T. Liao, S. Lai, J. and Gong, S.: Partial person re-identification, ICCV, 2015.
  - 22) Wang, H. Chen, X. and Liu, C.: Pose-guided part matching network via shrinking and reweighting for occluded person re-identification, Image and Vision Computing, Volume 111, 2021, 104186, ISSN 0262-8856,  
<https://doi.org/10.1016/j.imavis.2021.104186>.
  - 23) Liu, X. Zhao, H. Tian, M. Sheng, L. Shao, J. Yi, S. Yan, J. and Wang, X.: Hydraplus-net: Attentive deep features for pedestrian analysis, IEEE International Conference on Computer Vision (ICCV), pp 350–359, 2017.
  - 24) Gul, S. Malik, M. Khan, G. and Shafait, F.: Multi-view gait recognition system using spatio-temporal features and deep learning, Expert Systems With Applications, Vol. 179, 2021.
  - 25) 松下貴哉：歩行者 OD 自動観測の実装に関する研究, 広島大学工学部卒業論文, 2021.
  - 26) Ildoo Kim, I.: tf-pose-estimation, <https://github.com/ildoonet/tf-pose-estimation>
  - 27) Google Brain: TensorFlow, <https://www.tensorflow.org/>
  - 28) 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.
  - 29) Keras-team: Sequential モデル API, <https://keras.io/ja/models/sequential/>
  - 30) Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, ICLR, 2015.
  - 31) Bewley, A. Ge, Z. Ott, L. Ramos, F. and Upcroft, B.: Simple online and realtime tracking, IEEE, pp. 3464-3468, 2016.
  - 32) Wojke, N. Bewley A. and Paulus D.: Simple online and realtime tracking with a deep association metric, Proc. - Int. Conf. Image Process. ICIP, vol. 2017-Sept, pp. 3645-3649, 2018.
  - 33) Bazarevsky, V. Grishchenko, I. Raveendran, K. Zhu, T. Zhang, F. and Grundmann, M.: BlazePose: On-device Real-time Body Pose tracking, 2020,  
<https://doi.org/10.48550/arXiv.2006.10204>
  - 34) Zhou, K. Yongxin, Y. Cavallaro, A. and Xiang, T.: Omni-scale feature learning for person re-identification, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019.
  - 35) 栗田多喜男：ニューラルネット入門,  
<https://home.hiroshima-u.ac.jp/tkurita/lecture/neuro.pdf> (2023 年 2 月 7 日閲覧)
  - 36) 栗田多喜男：サポートベクターマシン入門,  
<https://home.hiroshima-u.ac.jp/tkurita/lecture/svm.pdf> (2023 年 2 月 7 日閲覧)
  - 37) 波部斉：ランダムフォレスト. 情報処理学会研究報告, Vol.2012-CVIM-182, No.31, 2012.
  - 38) K. Matsumoto, <https://geographica.biz/> (2023 年 2 月 7 日閲覧)
  - 39) DAN 杉本, <https://www.kashmir3d.com/> (2023 年 2 月 7 日閲覧)
  - 40) <https://github.com/albumentations-team/albumentations>
  - 41) McInnes, L. Healy, J. and Melville, J.: Uniform Manifold Approximation and Projection for Dimension Reduction, 2020,  
<https://doi.org/10.48550/arXiv.1802.03426>

(Received March 6, 2023)

(Accepted ????, ????)

## A STUDY ON PEDESTRIAN FEATURE EXTRACTION FROM VIDEO IMAGES FOR PERSON RE-IDENTIFICATION

Kenta YAMAZAKI, Koichi KOIZUMI, Takaya MATSUSHITA, Makoto TSUKAI,  
Masaya HIRANO and Takeshi MIYAJI