ビデオ観測に基づく歩行者 OD 表作成のための人物再同定アルゴリズムの開発

神崎　尭斗[1]・松下　貴哉[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　工学部第 4 類（〒733-8527 東広島市鏡山 1-4-1）
E-mail: b202879@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]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
[4]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
[5]正会員　BRIDGEOVER 株式会社　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，複数地点に設置したカメラで観測した歩行者画像から，地点間を移動する歩行者 OD を取得するために，画像内の人物を個人レベルで特定する Person Re-Identification (ReID) 手法に基づいて，人物識別に有効な画像マッチング方法の検証を行う．昨年の成果に基づいて MediaPipe Pose から得た骨格特徴に基づいて質の悪い画像を削除したうえで，OSNet により外見特徴を抽出することにより人物の特徴のデータセットを作成する．本分析の新規性は，カメラ画角内の人物トラッキング情報を有効活用して同一人物と考えられる一連のフレーム全体を識別単位とすることである．そのために，多対多マッチング手法である最適輸送理論を適用して，人物同定のための分類器を開発する．

Key Words: *Multi class classifier, Image processing, Optimal Transport Theory*

1. はじめに

都市計画マスタープランでは，交通結節点の重要性が再評価され，これらの活性化施策[1)]がとられることが多い．それがハードな施設整備，またはソフト施策のいずれで行われるにせよ，施策効果の検証では，地域の賑わいを定量的に評価できる歩行者 OD 表（Origin to Destination Table）の計測手法が必要である[2)]．

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 技術を歩行者の自動観測に適用できるようになっている．その中でも，カメラ映像から人物の同定を行う手法は，調査前に個人の協力を必要としないという利点がある[3)]．ただしこの手法は，依然として実装に耐える水準の精度が得られないという課題がある．

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[4)]がある．ReID 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である．一般的な ReID は、動画に写る人物の特徴をフレーム毎のベクトル表現として取得して，それらの照合によって画像内の人物を同定する．しかしこの方法には，フレーム間で観測方向が異なると，同定精度が著しく低下する．

本研究では，歩行状況を撮影した各フレームの比較ではなく，連続して撮影した一群のフレーム同士の比較に基づく人物再同定，すなわち多対多マッチングを検討する．簡易な多対多マッチングの方法には，取得した複数フレームから得た特徴量の平均ベクトルに基づく方法や，フレーム間マッチングを多重仮説検定とみなす方法などが考えられる．しかし前者には平均化による情報損失が，後者には連続画像に対する検定の独立性仮定と検定サイズの設定が課題となるなど，手法論の安定性や妥当性に疑問が残る．最適輸送理論は，確率分布同士の距離を比較する方法として提案された．よく知られた同種の手法にカルバック＝ライブラー統計量があるが，この方法は特徴量の分布範囲が比較する分布間で異なる場合（分布の台／support が共通ではない場合）には適用できない．

最適輸送理論は，サンプルを特徴量空間内の点で表すとき，サンプル間の輸送コストを定義することでこの問題の解消を図る手法である．本稿では，最適輸送理論に基づく ReID 手法の開発に関する課題を検討する．

2. ReID 問題と本研究の位置づけ

Person Re-Identification（ReID）は，異なるカメラによって撮影された人物を同定するアルゴリズムを研究する分野である．この手法は Multi-Target Multi-Camera Tracking（MTMCT）のサブタスクとして研究されていたが，2006 年に Gheissari ら[5]が「人物再同定（ReID）」を「重複しない複数のカメラで撮影された人物を照合するための研究テーマ」と定義したことで，独立した研究分野として認知されるようになった．ReID には，訓練データとテストデータに同一人物が含まれる場合（closed set）と，必ずしも同じ人物が含まれない場合（open set）の 2 種類のタスクがあるが，より一般的な ReID は後者である．

ReID モデルでは，第一段階で人物の特徴量抽出を行い，第二段階では，一対の人物フレーム間の特徴量ベクトルの類似度に基づいて同定する．まず訓練データによって ReID モデルを学習して，特徴量抽出器を作成する．テストデータのうち，識別対象の画像群をクエリ，クエリとの類似度を計算する画像群をギャラリーと呼ぶ．ギャラリーの各画像の類似度が得られたら，それを降順に並べたランキングリストを用いて同定精度を検証する．

Open set の場合，識別対象の人物=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．存在すれば人物 ID を出力してタスクが終了するが，存在しない場合は新規人物としてクエリに追加する．歩行者 OD 自動観測は Open set タスクであり，難易度が高い．なお著者のこれまでのレビューの範囲では，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 モデルは存在しておらず，歩行者 OD 自動観測は実装レベルの技術水準には至っていない．

ReID では基本的に外観特徴量を使用するが，歩き方そのものを特徴量とする方法もあり，これは歩容認証と呼ばれる．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[6]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[7]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 では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．しかし GEI に必要な計算量は大きいため，クエリサイズが大きな公共空間での人物同定には不向きである．

ReID の既存研究では，画像内の人物の信頼性の高い外観的特徴（服装の色，人物の形状，または骨格など）に着目して，抽出した特徴量をフレーム間で照合することによって，人物の再同定を行う．しかしこの手順では，画角内の人物トラッキングの成果が，人物同定に活用されない．画角内の人物トラッキングが高精度で可能なことは，ほぼ同一人物と思われる画像群を得ることになる．この先験情報が活用する ReID では，あるカメラ内の画像群と別のカメラ内の画像群に基づく同定，すなわち多対多フレームの人物同定手法が必要である．

山﨑[8]は， ReID タスクに必要な下処理手順を検討した，その結果，人物画像の一部欠損を同定するために，骨格情報に基づく下処理手順が，最も同定精度が高くなることを示した．小泉[9]は，closed set ならびに open set のデータベースを作成して，強化学習による ReID を行った．その結果，ReID について 70～80%程度の精度が得た上で入力画像の処理方法等に関する課題をまとめた．ただしこれらの研究では，多対多マッチングは行っていない．

本研究では，ReID において，複数フレーム間における多対多のマッチングの有効性や，この手法開発に必要な諸問題について検討を行う．昨年度までの研究で，骨格特徴量を用いて前処理した画像の抽出が有効なことが判明している．本課題では，後述するように特徴ベクトルの次元数も検討しつつ，多対多フレームのマッチングができる最適輸送理論[10]を用いる．

3. 人物再同定における特徴量抽出と分類方法

本章では人物再同定のための特徴量抽出と分類方法を概説する．なお以下の特徴量抽出手順は既往研究と同一で新規性はない．最適輸送理論に基づく ReID 手法の開発が，本研究の新規性である．

(1) 画角内の同一人物画像群のトラッキング

2016 年に Bewley が提案した Simple Online and Realtime

Tracking[11]（SORT)は，画角内の物体検出によって捉えた物体を 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．物体追跡またはトラッキングとは，検出された bbox をフレーム間で結びつけることである．SORT は既検出の bbox の位置情報を用いて移動先の bbox を予測し，両者の重なり具合を表す Intersection over Union（IoU）を計算する．その上で結合候補となる物体間の IoU 行列の擬対角成分を最大化するように bbox を結合して，画角内の物体追跡を行う．ただし SORT では，対象物の手前に障害物があるときに人物 ID の消失や振り替わりが起こる．DeepSORT[12] は，外観特徴量と bbox の高さを考慮することによって，画角の出入りを伴う物体のトラッキングが可能になった．

2022 年に，DeepSORT の改良版として StrongSORT[13] が提案された．このアルゴリズムは，特徴量の更新に自己回帰型の慣性項を持たせた点や，カルマンフィルタに基づく非線形な移動に対応できるようになった点，ならびに動き情報を含めたコスト行列の設定を行うようになった点が，主な改良点である．これらの導入により，不規則な動きをする歩行者のトラッキング精度を高めている．本研究では Strong SORT を用いて，それぞれのカメラ画角内から得られる，同一人物画像群を抽出した．

(2)　特徴量抽出手法

StrongSORT から得た同一人物画像群を MediaPipe Pose[14] に入力して，関節座標を抽出する．以下では取得した関節座標を骨格特徴量と呼ぶ．MediaPipe は，Google 社提供のオープンソリューションであり，Mediapipe Pose は，これに含まれる人物姿勢推定す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．取得する関節座標は，33 か所である．これにより，腰の中心を原点として，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の直交座標をメートル単位で取得する．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 Network[15]（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．オムニスケール特徴量は，全身の特徴だけではなく，局所的な特徴も同じく重要という前提の下で，実スケールの違う特徴を統合した特徴量である．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，既存アルゴリズムの強化学習による ReID 性能の改善を図るため，OSNet の ReID 機能は用いずに，オムニスケール特徴量を得るための抽出器として，OSNet を使用する．

提案する分類器は OSNet で得られた特徴量の転移学習を行うニューラルネットと，ここから取得した特徴量ベクトルに人物ラベル分類を適用するアルゴリズムの 2 段階で構成される。提案手法のニューラルネットの構造を図 1 に示す．このニューラルネットを学習すると，訓練データセットに，より適合する特徴量が得られる．バッチサイズは 128，入力は 512 次元，エポック数は 500 である．出力する次元数は 2 次元以上の複数の設定として，その中で最も精度の高い次元数を選択する．

ReID ではニューラルネットの学習時の損失関数に対する研究が多くなされており [16]，距離学習手法を導入したモデルが発展している [17]．損失関数とはモデルの出力と訓練データの差を評価する関数であり，この損失関数の値を最小化するように各層の重みが更新される．本研 TripletLoss 関数 [18),19)] を用いる，通常の学習手法では，異究では最も頻繁に用いられるなる人物ラベル間の距離や同一人物となるフレーム間距離を考慮せずに特徴量空間を学習する．これに対して距離学習手法では，同じ人物ラベルのフレーム間の距離を小さくしつつ，異なる人物ラベルのフレーム間距離を大きくする．以上の手順によって，入力データを人物ラベルごとに分離できるような埋め込み空間への変換方法を学習できるため，人物ラベルの識別性が高まると期待できる．

既往研究にならって，損失関数に Triplet Loss 関数 [20)]を採用する．Triplet Loss 関数では Anchor と呼ばれる参照元画像 x_A と，Positive（正例）と呼ばれる Anchor と同一人物の画像 x_P，ならびに Negative（負例）と呼ばれる Anchor：x_A とは異なる人物の画像 x_N である．このとき，Triplet Loss 関数は以下のように定義される．

$$L_{triplet\ Loss} = [d(Z_A, Z_P) - d(Z_A, Z_N) + m] \qquad (1)$$

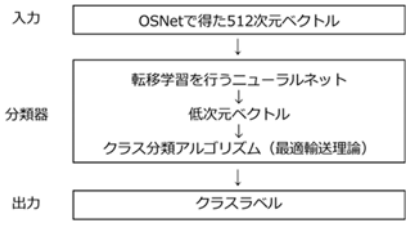


図 1　提案手法の分類器

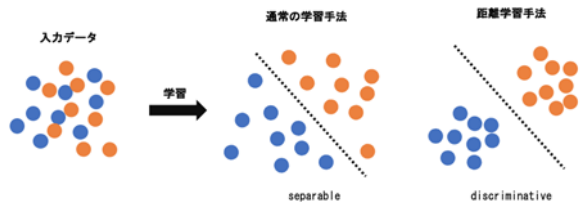


図 2　通常の学習手法と距離学習手法の特徴量

ここで$Z_i = \varphi(x_i)$は, 学習されたx_iの特徴量, $d(\quad,\ )$は, 引数に現れる 1 組のサンプルの距離，m は予め設定するマージンであり，$[c]_+$+は,$\max(0,c)$を表す．式(1)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が学習できる．

(3) 人物同定のための分類手法

本研究では，人物ラベルの分類手法として，同一人物と思われる一連の画像群同士の距離を計測する最適輸送理論[10]を用いる．この問題は，たとえば 1 次元上の離散点x_iにそれぞれ分布する質量$a x_i$の山を，同じ軸上の別の離散点x'_iにそれぞれ分布する質量$b x'_i$の山に移す際の効率的な移動方法を考えることである．ここで山を移動させる方法は無数にあるが，最も手間の小さな移動方法を採用したときの移動コストの総和を二つの分布間の距離とみなすのが，最適輸送の考え方である．この問題は，線形計画問題として定式化できる．

$$\min_{P \in R^{n \times m}} \quad \sum_{i=1}^{n}\sum_{j=1}^{m} C(x_i, y_j) P_{ij} \qquad (2)$$

subject to

$$P_{ij} \geq 0 \quad (\forall i \in [n],\ \forall j \in [m]) \qquad (3)$$

$$\sum_{j=1}^{m} P_{ij} = a_i \quad (\forall i \in [n]) \qquad (4)$$

$$\sum_{j=1}^{n} P_{ij} = b_i \quad (\forall i \in [m]) \qquad (5)$$

上式で$C(x_i, y_j)$はコスト関数と呼ばれる外生変数であり，空間 X 上の各対(x_i, y_j)において，点x_iから点y_jに一単位の質量を輸送するのにかかるコストをあらわす．たとえば 2 次元空間では $C(x,y) = \|x-y\|_2$などが考えられる。またP_{ij}は点x_iから点y_jに輸送される質量である．

この最適化問題の最適解の目的関数の値を最適輸送コストと呼び, $\mathrm{OT}(\alpha, \beta, C)$と表記する. 式(2)が示すように, 本問は線形計画法のソルバーによって求解できる．

人物同定手法としての最適輸送理論と距離学習の関係について式(1)と(2)を比較すると，最適輸送理論では所与とされるコスト関数が，機械学習の対象となっている特徴量空間の計量, すなわち距離に対応することが分かる．よって同理論を活用した多対多分類手法アルゴリズムの開発方針は，1) 図 1 に示すようにコスト関数自体を最適輸送理論に置換して転移学習を行う距離空間の一括学習手法と，2)転移学習したニューラルネットワークから重要なオムニスケール特徴量を抽出して，その重要度を反映させるように定義したコスト関数に基づいて最適輸送距離を定義して多対多分類を実施する二手法が考えられる．なお最適輸送理論が出力するP_{ij}は各画像のマッチング割合を表わす．したがって距離計量空間が ReID 精度を高めるように定義されていれば，画像群の中でも撮影角度が類似した画像同士がマッチングされる等の対応が得られると考えられる．

4.　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た実験を行った．歩行者を撮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 名の被験者の協力の下で，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 の間で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，本人の写っている箇所を削除できる．これは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手続きであり，画像を撮影してアルゴリズム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た．

歩行経路の確認のため，被験者に「ジオグラフィカ|登山用 GPS」という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．

本研究では，骨格特徴量の visibility によって下処理をした外観特徴量を抽出する．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 に変更した，実験 1,2 で被験者が指定した地点を回り終わるまでの時間は，それぞれ 10:00~10:15 の 15 分間および 10:20~10:31 の 11 分間だった．実験 3 では巡回順序を指定せず，10:45~11:00 の 15 分間で任意の地点



図 3　学内のカメラ設置場所と画角

を巡回させた.

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．この信頼度とは，DeepSORT が人物を検出する際の，人物物体検知の信頼度を指す．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，その他の一般通行者を分けるアノテーション作業を人手で行って，同一人物画像群を作成した．

機械学習モデルの訓練に入力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て，訓練に使用できる画像枚数を人為的に増やす手法である．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[21] を使用した．画像をぼかす Blur，ランダムに明るさを変える Random_Brightness，ランダムにコントラストを変える Random_Contrast 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），Random_Brightness は（-0.2，0.2），Random_Contrast は（-0.2，0.2）である．これらのによって取得した画像　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 に入力する画像のサイズに制限はないが，対象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．骨格特徴量は，MediaPipe Pose により抽出した．こちらも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．なお MediaPipe Pose では人物の検出に失敗することがあり，その場合は関節座標が抽出されない．よって一連の処理によって出力された画像枚数を比較すると，外観特徴量から取得できる画像枚数よりも，骨格特徴量から取得できる画像枚数は少なくなる.

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) 最適輸送理論を用いた分類器の精度

検証結果がまだ出ていないので当日の発表とする

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（2023 年 2 月 6 日閲覧）
3) 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.
4) Wang, X.: Intelligent multi-camera video surveillance: A review, Pattern Recognition Letters, vol. 34, no. 1, pp. 3-19, 2013.
5) 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.
6) 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.
7) 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.
8) 山﨑健太，小泉鴻一，松下貴哉，塚井誠人，平野雅也，宮地岳：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，第 68 回土木計画学研究発表会，2023.
9) 小泉鴻一：深層学習に基づく複数カメラ間の人物再同定手法の開発，広島大学修士論文, 2022.
10) 佐藤竜馬 : 最適輸送の理論とアルゴリズム，講談社，2023.
11) Bewley, A. Ge, Z. Ott, L. Ramos, F. and Upcroft, B.: Simple online and realtime tracking, IEEE, pp. 3464-3468, 2016.
12) 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-Septe, pp. 3645-3649, 2018.
13) Du. Y, Z. Zhao, Y Song, Y. Zhao, Y. Su, T. Gong, M and Hongying : StrongSORT: Make DeepSORT Great Again, Computer Vision and Pattern Recognition(CVPR), 2022. https://doi.org/10.48550/arXiv.2202.13514
14) 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
15) 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.
16) Yan, C., Pang, G., and Bai, X. : Beyond Triplet Loss: Person Re-identification with Fine-grained Difference-aware Pairwise Loss, IEEE Transactions on Multimedia, Vol. 242022, pp. 1665–1677, 2022.
17) Zou, G., Fu, G., Peng, X., Liu, Y., Gao, M., and Liu, Z. : Person re-identification based on metric learning: A survey, Multimedia Tools Appl., Vol. 80, No. 17, pp. 26855-26888, 2021.
18) Ding, S., Lin, L., Wang, G., and Chao, H. : Deep feature learning with relative distance comparison for person reidentification. Pattern Recognition, 48(10), pp. 2993–3003, 2015.
19) Su, C., Zhang, S., Xing, J., Gao, W., and Tian, Q. : Deep attributes driven multi-camera person re-identification, European Conference on Computer Vision, pp. 475–491, 2016.
20) Hoffer, E., and Ailon, N. : Deep metric learning using triplet network, International Workshop on Similarity-Based Pattern Recognition. Springer, Cham, 2015.
21) Su, C., Zhang, S., Xing, J., Gao, W., and Tian, Q. : Deep attributes driven multi-camera person re-identification, European Conference on Computer Vision, pp. 475–491, 2016.

(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

8lines for abstract