Vol.2021-MBL-101 No.29
Vol.2021-ITS-87 No.29
2021/12/8

スマートグラスの可視領域情報を用いた不連続動画上の人物追跡

高橋 直也[1]　天野 辰哉[1]　山口 弘純[1]

概要：
本研究では，スマートグラスが備える RGB カメラにより撮影される，可変視点での動画像に出現する人物の，動画像内位置の追跡手法を提案する．一般に，既存の固定視点の動画像上の人物追跡手法では，動画像フレーム上の人物検出とカルマンフィルタなどによる移動体追跡の組み合わせにより追跡を実現する．しかし，ウェアラブルデバイスであるスマートグラスは頭部の移動や装着者の移動により頻繁に位置や方向が変化する．それに伴い．動画像内の人物の位置が連続フレーム間でも大きく変化したり，画角端から人物が消失，あるいは再登場するような状況が多く発生する．このような場合，安定的移動を想定した既存の固定視点の人物追跡を適用できない課題がある．これに対し本研究では，固定視点の動画像上での追跡手法 DeepSORT に対し，人物の外見特徴の深層距離学習による人物再識別技術を適用することで，DeepSORT が追跡に失敗することで分断された同一人物の軌跡を接合する．さらに，スマートグラスの可視領域（視野）情報から得られる周辺人物の推定位置情報を用い，空間内位置に関する不整合が生じる人物再識別を抑制することで，視界の移動に対して堅牢な人物追跡を実現する．大学キャンパスの屋外環境において移動カメラにより周辺歩行者を撮影し，提案手法による人物追跡精度を評価した結果，DeepSORT と比較して連続フレームにおいて同一人物であるにも関わらず異なると識別される回数（ID-Switch）を 61% 削減できた．

キーワード：Person Re-ID, 人物トラッキング, 深層距離学習, スマートグラス

1. はじめに

近年，Augmented Reality（AR）グラスやスマートグラスが備える RGB カメラや深度カメラを用いて現実空間の人やモノを認識し，それらに関するサイバー情報を現実空間映像やグラスに重畳して表示する様々なアプリケーションが提案されている．例えばスポーツ分野では，カメラによって試合やトレーニング中の選手を検出・識別し，スマートグラスを装着する観客やコーチに対し，それらのデバイスを通して選手情報を提供するシステムも提案されている [1]．このようなスマートグラス用アプリケーションの実現のためには，カメラが捉えた映像中の人の追跡が不可欠である．映像中の物体追跡手法は数多く提案されており，例えば DeepSORT[2] は，物体検出手法 Yolo を用いて画像内の対象オブジェクトを検出し，2 次元画像フレーム内での移動を想定したカルマンフィルタを用いて連続フレーム間で対象物体の追跡を実現する．しかし DeepSORT を含め，既存の映像中の人物追跡手法は定点観測映像を想定しているものが多く，カメラの位置や向きが変化する状況では，数フレーム以上に渡る遮蔽（オクルージョン）や追跡対象の画像フレーム内，あるいはフレーム内外での想定以上の移動が頻繁に発生し，安定的な移動を想定しているカルマンフィルタでは追跡精度が大幅に低下するといった課題がある．特に本研究が想定するスマートグラスに内蔵されたカメラからの映像は，装着者の頭部の移動や歩行などにより視野方向や撮影地点が頻繁に変化するため，連続フレーム間での物体検知と移動予測に依存する既存手法の適用は適切ではない．

一方で，カメラの移動量が少なく，視点や視野が比較的安定している状況においては，数フレーム以内では大きなフレーム内移動が発生しないことを前提にフレーム内での人物群を追跡する DeepSORT など，単一の定点観測カメラ映像からの複数人物の追跡手法（Multi-Object Tracking, MOT）が有効に動作することが多い．したがって，カメラ視野の大きな変化により MOT が人物追跡に失敗した場合の処理を適切に行うことが，そういった視点移動が発生する映像における人物追跡の精度向上に大きく寄与すると

[1] 大阪大学情報科学研究科

考えられる．

本研究では，定点観測映像の既存の複数人物追跡（MOT）手法に対し，映像内で新たに検出された人物が過去に追跡されていた人物かを判定し，そうであればどの人物かを識別する人物再識別（Re-ID）を適切に適用することで，スマートグラスが捉える可変視点の動画像に出現する人物を継続的に追跡する手法を提案する．具体的には，MOT（本研究では DeepSORT を用いる）が追跡する各人物に対し，ある人物を見失ったタイミングでその人物のそれまでの軌跡をセグメント化する．得られた軌跡セグメント群に対し，同一人物であるにも関わらず複数の軌跡セグメントに分断されている状況を検出し，人物再識別を用いて分断された軌跡セグメントを接合する．また，別人物であるにも関わらず，MOT で 1 つの軌跡セグメントと判定される場合も多い．こういったいわゆる人物のすり替わり（ID-Switch）に対しては，ID-Switch の前後で軌跡セグメント内の画素値ヒストグラムが大きく変化することを利用することで検知し，当該軌跡セグメントを正しく分割する．また，一度検出した人物は，最後に映像から消失した位置と時刻からその存在位置を推定し続けるとともに，スマートグラスの視線方向および位置を元に，現在の視野内にその人物が含まれるかを推定することで，明らかに存在し得ない人物を再識別対象とすることを抑制する．これにより人物再識別精度の向上を図る．

人物再識別には，深層距離学習に基づく手法 ArcFace[3] を利用する．深層距離学習ではデータの特徴量空間における距離を学習することができるため同一人物の画像同士の距離は小さく，異なる人物画像同士の距離は大きくなるように学習しておくことにより，検出された人物と過去に出現した人物同士の特徴距離を計算する．

大学キャンパスの屋外環境において，歩行者が通行する様子を移動するカメラで 10 分間撮影し，得られた動画像に対して提案手法を適用した．その結果，DeepSORT と比較して連続フレームにおいて同一人物であるにも関わらず異なると識別される回数（ID-Switch）を 61% 削減できた．また 10 分間において DeepSORT が追跡した人物数は実際の 8 人に対して 358 人であった一方，提案手法は 21 人であり，同一人物判定精度が大幅に向上した．

2. 関連研究

2.1 Multi-Object Tracking

単一カメラ内で人物や車両を追跡する手法は Multi-Object Tracking (MOT) として以前より多く研究されている．MOT 手法は通常, 検出ベーストラッキング (Detection-Based Tracking, DBT) および検出フリートラッキング (Detection-Free Tracking DFT) に分類される．DBT では物体を事前学習した検出器により検出し，それらを繋ぎ合わせた軌跡を導出する [4] のに対し，DFT では，初期フレームにおいて手動によるオブジェクト指定を行ったあと，その移動を追跡する [5]． オブジェクトが消滅，出現を繰り返すより一般的な環境に対応可能である点と，近年の物体認識の高度化により，現在では DBT がより一般的となっている．MOT では大規模なベンチマークとデータセットは MOT Challange として公開されており [6]，任意のオブジェクトのトラッキングや，混雑時におけるトラッキングなど様々なシーンにおけるトラッキング手法の比較が可能となっている．また，Multi-Object Tracking and Segmentation [7] では，トラッキングとセグメンテーションを同時に行う方法を提案しており，KITTI データセットを用いた評価を行っている．MOT に関しては最新のサーベイ論文 [8] を参照されたい．

2.2 オブジェクト再識別（Re-ID）

異なる地点に設置されたカメラによる複数の定点観測映像において，複数オブジェクトが出現する中で各オブジェクトをトラッキングする問題（Multi-Target Multi-Camera Tracking）手法が研究されてきている．MTMCT 問題はカメラの設置位置により光量や人物までの距離，撮影方向が大きく異なる中で同一人物を発見できる必要があるため，その高精度化は非常に挑戦的な問題である．

MTMCT は R-CNN や SSD といった深層学習ベースの高精度なオブジェクト検出器や OpenPose[9] などの姿勢検出器の登場により，それらを用いた高精度化に向けて一層注目を集めている研究分野である．MTMCT 手法の多くは，カメラ同士の FOV（Field of View）の重畳を仮定できないため，単一カメラ内ではオブジェクト追跡（MOT）を用い，複数カメラ間ではオブジェクト再識別 (Re-idenfication, Re-ID）をもとに人物同定を行う．Re-ID 問題は与えられたターゲット（人物）画像と画像集合に対し，その類似度でランク付けを行う問題であり，MTMCT はそれを用い，2 つの画像が与えられたときにそれが同じ人物か否かを判定する 2 値分類器を構成する問題である [10]．[10] では MTMCT 問題と Re-ID 問題に共通する特徴量を発見し，損失関数として動的重み付き Triplet Loss を用いる手法を提案している．

車両の MTMCT 手法に関しては複数の路側監視カメラを用いた CityFlow [11] などが知られている．CityFlow では, 最大 2.5km 離れた 10 の交差点に設置された 40 台のカメラから得られた 3 時間以上の交通 HD 映像に 20 万以上のラベル付けを行ったデータを提供しており，車両再識別のためのベンチマークを提供している．車両検出，トラッキング，車両再識別において，様々な追跡手法やそれらの組み合わせの性能評価を行っており，2019 AI City Challenge (https://www.aicitychallenge.org/) に向けたサーバーを提供した実績がある．

なお, 人物再識別用データセットとして, Market-1501[12]

や Motion Analysis and Re-identification Set (MARS) [13] が知られている．Market-1501 は Deformable Part Model (DPM) による人物検出器を用いた 3 万 2 千以上のバウンディングボックスがアノテーションされており，50 万以上の不正解データも含まれている．また MARS はビデオデータを対象としており，Market-1501 同様，DPM を用いて，1,261 の人物による約 2 万の Tracklet 特徴量（追跡点の軌跡）を提供している．

2.3 提案手法の位置づけ

前述したような MOT または MTMCT では，固定設置された単一または複数カメラを使用することが前提となっている．しかし，スマートグラスを通したトラッキングシステムのように，カメラ視点が固定されておらずシーン切り替えが頻繁に発生する状況を想定したトラッキング手法ではない．これに対して提案手法では，連続フレームを想定した MOT 向けトラッキング手法を活用し，大量の軌跡セグメントを形成する．その軌跡セグメントを，人物再認識と移動予測によってつなぎ合わせ，トラッキングを実現する．トラッキングの実現のためにスマートグラス特有の空間情報を導入している点で従来手法とは異なる．

3. 手法概要

提案手法の概要を図 1 に示す．MOT[2] により生成した軌跡セグメント群を，ArcFace による人物再識別モデルとスマートグラスを利用した位置予測と組み合わせた人物追跡手法を提案する．

まずスマートグラス装着者が取得する映像データに対し，連続フレームに対するトラッキング手法である MOT[2] により連続してフレームに映り込む人物のフレーム間同定（フレーム間トラッキング）を行う．システムの時刻を t で表し，実行開始時は $t = 0$，i 枚目の画像フレームを得た時刻を $t = i$ とする．また，検出した人物の画像セグメントを保持する画像セグメントデータベース（以下，単にデータベースとよぶ）の時刻 i における処理後の内容を DB_i で表すとし，$DB_0 = \{\}$ とする．

以下，時刻 k に得られた画像フレーム内で検出された人物の画像セグメント集合を P_k，$p \in P_k$ に対し，MOT が出力する ID を $ID_{mot}(p)$, 提案システムが最終的に決定した ID を $ID_{final}(p)$ とする．また，一般的なマップ関数 $map(f, L) = \{\forall x \in L | f(x)\}$ を用いる．

スマートグラス装着者の視点は常に変化するため，人物が視界からフレームアウトしたり，図 2 で示すような人物同士の重なり（オクルージョン）が発生したりした場合，MOT は類似オブジェクトを発見できない可能性が高い．また，スマートグラスによって各軌跡セグメント群の位置を特定することが出来る．これらを利用し，$map(ID_{mot}, P_k) \neq map(ID_{final}, P_{k-1})$ であれば，シーンの切り替わりにより新しい ID が与えられたとみなす．各画像セグメント $p \in P_k$ とデータベース DB_k に対し，ArcFace による人物再識別器（$af : S \times 2^S \rightarrow S \cup \epsilon$) を適用する．ここで，$af$ はある画像 s および画像集合 S が与えられた場合，s に最も類似度が高く，かつ一定の閾値以上の類似度がある S に含まれる画像もしくは不一致シンボル（ϵ）を返す関数であり，提案手法はこれを深層距離学習 ArcFace を事前に（あるデータセットで）訓練した関数を用いる．$p' = af(p, DB_{k-1})$ において，$p' \neq \epsilon$ であれば，$ID_{final}(p') = ID_{final}(p)$，そうでなければ $ID_{final}(p') = ID_{mot}(p)$（MOT による新しい ID）とし，$DB_k = DB_{k-1} \cup P_k$ とする．

ここで，トラッキングしたい人物が事前に把握できるのであれば，人物検出モデルおよび分類モデルを既存の深層学習等で事前に構築し，各映像フレームごとに人物を識別することも可能である．しかし本研究で想定する利用シーンでは，事前に特定の人物を追跡するためのデータを取得することは困難であることが多い．そこで提案手法では，与えられた 2 つの画像（検出した選手の画像とデータベースにある選手の画像）が同一人物であるかどうかを判定するように学習させた Verification モデルを利用してデータベースとの照合を行う．

本研究では，深層距離学習の一種である ArcFace を利用し，Verification モデルを構築する．深層距離学習では，データの特徴量空間における距離関数を学習することができる．同一人物の画像同士の距離は小さく，異なる人物画像同士の距離は大きくなるよう学習することで，検出された人物と過去に追跡されていた人物同士の特徴距離を計算できる．これは通常のクラス分類モデルに ArcFace 独自のレイヤを追加することで構築可能である．

また，新たに検出した人物はスマートグラスにより位置が特定できている．そこで移動予測モデルを用いて図 1 のように検出エリア外の人物の移動を予測する（Virtual space の青やオレンジの範囲）．位置予測をしない場合，新たに人が検出されたとき全ての軌跡セグメントと人物再認識を行う必要がある．その一方で，位置予測ができていれば図 1 の Prediction で比較すべき軌跡セグメントは限定される．この手法の詳細に関しては 5 章にて述べる．

3.1 軌跡セグメント群の分断

MOT によってトラッキングが連続して成功している間の画像を 1 つの軌跡セグメントとして生成した．しかし，オクルージョンや視野の変動によって IDSwitch が発生し，1 つの軌跡セグメント内に複数の ID が混入する．例えば図 2 はオクルージョン発生時に MOT で切り出された一つの軌跡セグメント群であるが，1 フレームごとに少しずつ検出対象が替わっていることがみてとれる．そこで作成した各軌跡セグメントに 1 人の追跡対象となるよう，画像の

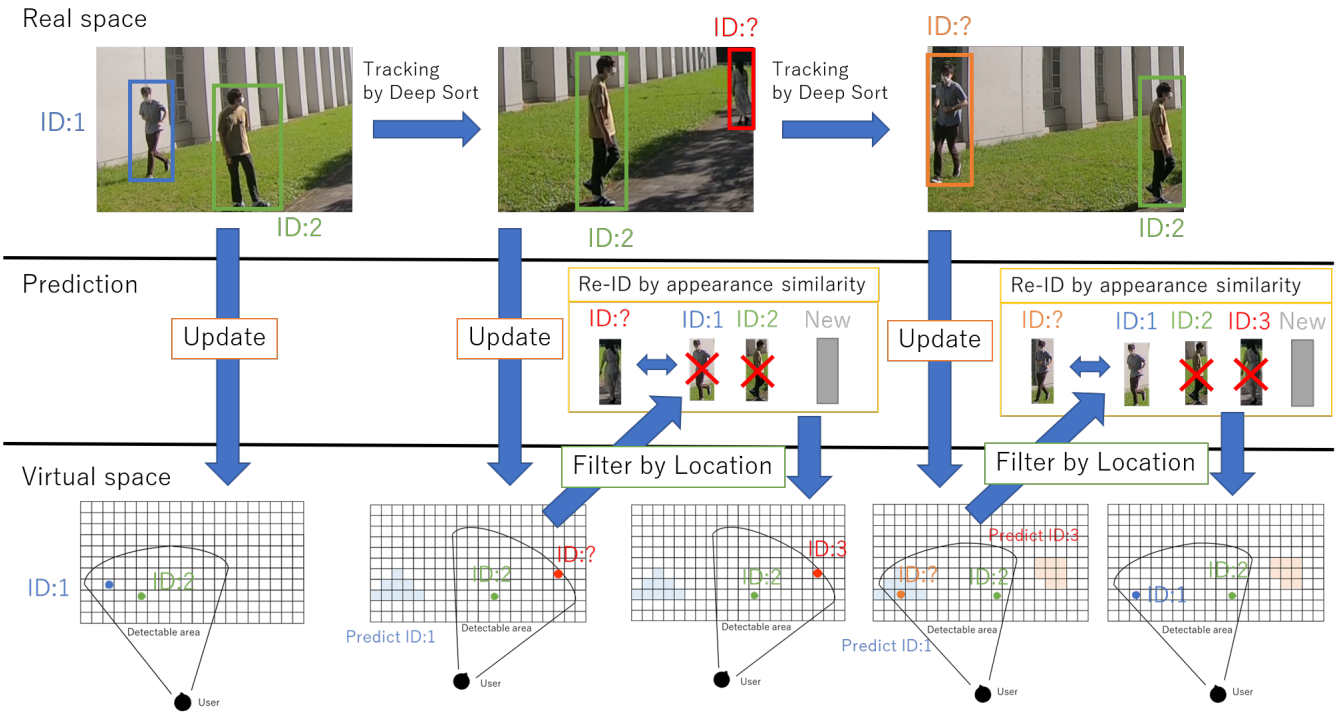


図 1: 手法概要

図 2: オクルージョンの例

ヒストグラムを比較することによる軌跡セグメント群の分割を行う．MOT のトラッキング特性上，時間的に近接するフレーム間の画像の差異は微小であるが，離れたフレーム間の画像の差異はその間に入れ替わりが発生していれば大きくなる．そこで，同一軌跡セグメント群でフレーム間距離が 5 である 2 画像のヒストグラムを比較し，一定以上の変化があれば入れ替わり発生とみなし，特定したタイミングの前後で軌跡セグメント分割を行う．これにより純度の高い軌跡セグメント群を生成した．純度が高い軌跡セグメント群を生成するため，不要な分割（同一人物の軌跡セグメントの分割）も発生するが，4 章，5 章で述べる人物再識別と移動予測を利用することで正しく再接合されることが期待される．

4. 人物再識別手法

Additive Angular Margin Loss (ArcFace) 関数は顔認識において高い分類性能を発揮することが示された損失関数である [3]．距離学習に用いられる損失関数であり，類似度を角度で表現する．提案手法では図 3 に示すように，ResNet50 と ArcFace を組み合わせた深層距離学習で人物再識別を行うための Verfication モデルを構築する．

分類問題でよく利用される損失関数であるソフトマックス関数は次式で表される．

$$L_1 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}} \quad (1)$$

ここで $x_i \in R^d$ は i 番目の層から出力される特徴量を示し，図 3 の ResNet50 からの出力に相当する．本稿では ResNet50 から得られる特徴量の次元を 2048 とした．また，$W_j \in R^d$ は $W \in R^{d\times n}$ の j 列目の重み，$b_j \in R^n$ はバイアス項を示す．N はバッチサイズ，n はクラス数を示す．しかし，ソフトマックス関数はクラス内分散を小さくしてクラス間分散を多様化することには適していない．そこで x_i と行列 W を列ごとに L2 正規化し，それぞれのベクトルの積を取ると次のように表せる．

$$L_2 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos\theta_{y_i}}}{e^{s\cos\theta_{y_i}} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}} \quad (2)$$

この式に $(\theta_{y_i} + m)(m > 0)$ のペナルティを加えると次のように表せる．

$$L_3 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos(\theta_j+m)}} \quad (3)$$

ペナルティを与えることで正解クラスに対する logit は小さくなるが，これによってクラス内分散が小さく，クラス間分散が大きくなるように学習を進めることができる [3].

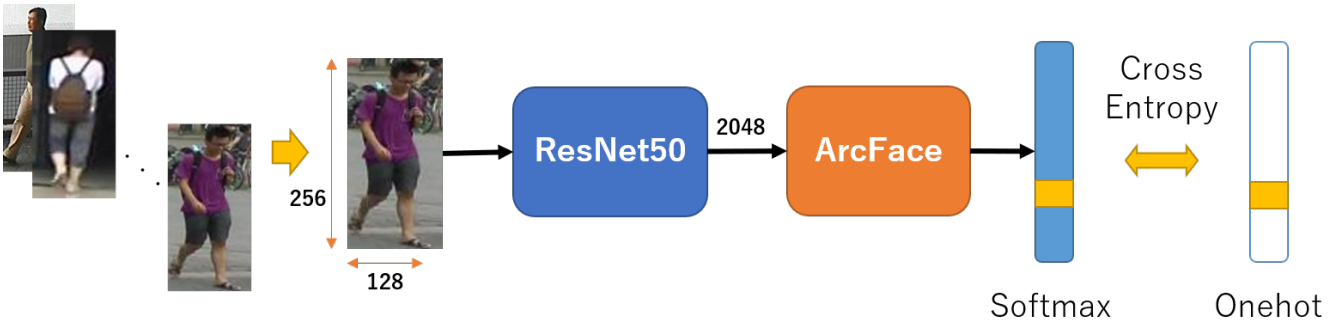


図 3: Verification モデル概要

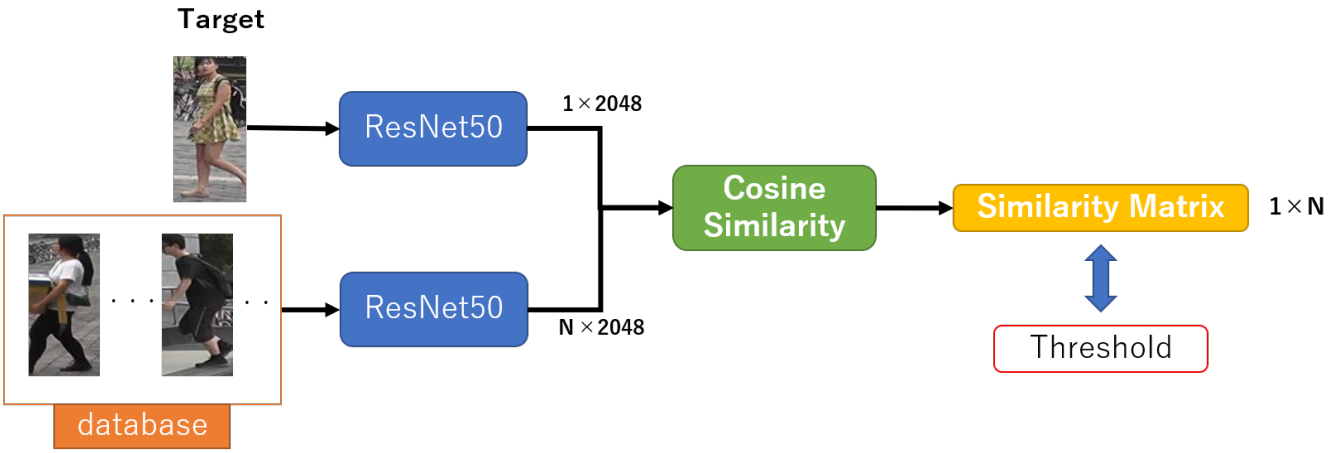


図 4: 人物再識別方法概要

提案手法における人物の再識別方法の概要を図 4 に示す．まず ArcFace を用いて事前学習した ResNet50 に対し，時刻 k で取得した再識別対象の人物画像セグメント p およびデータベース（DB_k）内の全人物画像セグメント（その総数を N とする）を入力とし，特徴量を得る．次に DB_k の N の各特徴量に対し，p の特徴量とのコサイン類似度を算出する．そのうち最大のコサイン類似度が予め定めた閾値を超えた場合，それに対応する画像セグメントと同一人物と判定しトラッキングを継続する．そうでなければ，新たに検出した人物とし新しい ID を付与する．

5. 移動予測に基づく人物再識別

時刻 t における，スマートグラスによって特定された $ID = A$ の軌跡セグメント P_A^t の位置を $l_{AR}(P_A^t)$，移動予測モデルによって推定した位置を $l_{estimate}(P_A^t)$ とする．また，図 5 で示すように軌跡セグメントの中で最後に連続してトラッキングできた時間を Δt，観測できていない時間を $\Delta t_{disappear}$ とする．

人が動く方向は座標軸に対して水平もしくは斜めの 2 つのパターンに分類され，それぞれについて図 6 に示すような前進（斜めを含む 3 方向），後退，停止の 5 つの移動パターンを考える．ここで，人が動くと予想される方向は図 7(a) に示すように，$\{l_{AR}(P_A^t) - l_{AR}(P_A^{t-\Delta t})\}/|\{l_{AR}(P_A^t) - l_{AR}(P_A^{t-\Delta t}\}|$ で推定できる．人の動く速さを $1m/s$ と仮定し，1 秒ごとに 5 つの移動パターンから 1 つ選択し移動すると考えると，図 7(b) のように軌跡セグメントのセルごとの存在領域 $R_{P_k}^{t'}$ を算出できる．図 7(c) のように各軌跡セグメントごとに存在領域を求めることで比較すべき軌跡セグメント群 T_{P_k} は，$T_{P_k} = \{P_k | l_{AR}(P_?^{t'}) \in R_{P_k}^{t'}\}$ と限定できる．

時刻 t における，スマートグラスによって特定された $ID =?$ の軌跡セグメントを移動予測と人物再識別の組み合わせて ID を特定する．前節で求めた比較すべき軌跡セグメント群 T_{P_k} の $l_{AR}(P_?^t)$ での存在確率は $Probability_{P_k^t}$ となる．人物再識別モデルによって求められた $P_?$ と P_k の類似度 $Similarity(P_?, P_k)$ とすると移動予測と人物再識別によって求められる $Score(P_?, P_k)$ は以下のように表せる．

$$Score(P_?, P_k) = \alpha \times Similarity(P_?, P_k) + (1-\alpha) \times Probability_{P_k^t} \quad (4)$$

ここで求まった最も高かった $Score(P_?, P_k)$ が閾値を超えたとき，軌跡セグメント P_k につなぎ合わせる．閾値を下回った場合は，新たな ID を付与する．

6. 実験と評価

6.1 実験環境とデータセット

本研究では，提案手法の性能評価のためスマートグラスでの利用を想定して撮影した 10 分程度歩行者の歩行者映像を用いた．映像内では 8 人の人物が約 15m × 40m の範

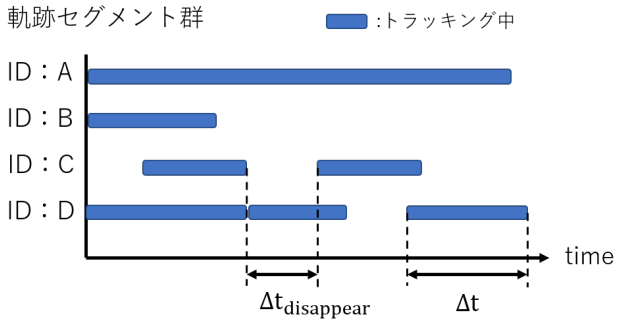
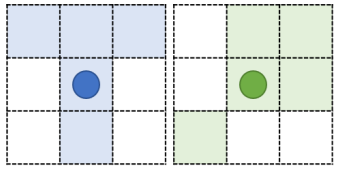


図 5: 軌跡セグメント群の概要



(a) 座標軸に並行　(b) 斜め

図 6: 移動パターン

囲をランダムに歩行している．撮影者もランダムに歩行し，その中で視点を左右に動かしている．このデータに対して MOT を適用し，人物が検出される様子を図 8 に示す．

6.2 評価指標

移動推定モデルの推定精度を評価するため，False Reduction Rate (FRR)，True Reduction Rate (TRR) を用いた．FRR は新たに人を検出したとき，移動推定モデルが出力した比較すべき軌跡セグメント群 T_{P_k} の中に正しい軌跡セグメントが含まれない割合を表す．FRR が低いほど誤って正解軌跡セグメントを削減していないことを表す．TRR は，FRR と同様に新たに人を検出したとき，移動推定モデルが比較すべき軌跡セグメント群 T_{P_k} を正しく削減できた割合を表す．TRR が高いほど削減できた軌跡セグメントが多い事を表す．

また，人物再認識モデルと移動推定モデルを組み合わせた提案手法の評価を行うための評価指標として軌跡セグメント数と IDSwitch を用いた．軌跡セグメント数は，使用した手法が推定した登場した人物の数を表すため，本実験環境では 8 人になることが望ましい．しかし，軌跡セグメント数を減らすと誤って IDSwitch が発生する確率が増加するためトレードオフの関係にある．

6.3 評価結果と考察

移動推定モデルの評価結果は，FRR は 7.41%，TRR は 42.5% となった．TRR の結果から移動推定モデルを導入することにより比較対象を大幅に削減しすることが出来る．FRR の結果から，削減してはいけない軌跡セグメントを削除してしまう可能性がある．この削減によって正しい軌跡セグメントとのつなぎ合わせが不可能となってしまうため，IDSwitch が増えることが考えられる．FRR が高くなってしまった原因として，移動推定モデルの単純さが考えられる．今回の検証では，移動パターンは 6 に示した 2 つのパターンに限定している．また，全ての軌跡セグメントが同じ速さで移動すると仮定したが実際とは異なる．これらの理由から実際の人の動きと移動予測との乖離度が大きくなってしまったと考えられる．

表 1: 提案手法の評価結果

手法	軌跡セグメント数	IDSwitch 数
DeepSort	358	433
提案手法	21	170

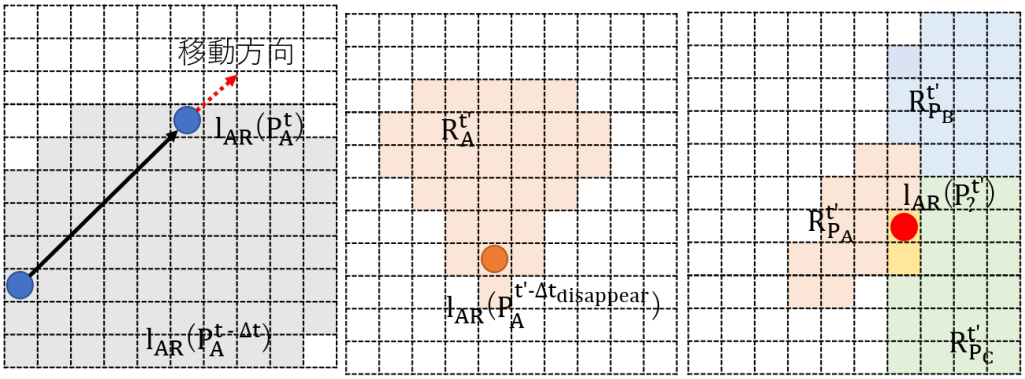
人物再認識モデルと移動推定モデルを組み合わせた提案手法を DeepSORT と比較した評価結果を表 1 に示す．実際に観測された人数 8 人に対して DeepSORT ではデータセットに対して DeepSort を適用した結果，軌跡セグメントの数は 358 個，IDSwitch は 433 回となった．DeepSORT は定点観測映像を想定しているものであるため，今回のようにカメラ自体が移動する映像に対しては，オクルージョンや画角の変動によって継続して正しく追跡できない．そのため，トラッキングが途切れるたびに新たな ID を付与し，軌跡セグメントの数が登場人物数に対して多くなった．また，オクルージョンが検出できないため，1 つの軌跡セグメントの中で IDSwitch が頻繁に発生する．

それに対して提案手法では，軌跡セグメントの数が 21，IDSwitch が 170 であり，軌跡セグメントの数は 94%，IDSwitch を 61% 削減することができた．IDSwitch が 170 であることから軌跡セグメントの数は減らせたが，1 つの軌跡セグメントの中に複数の ID が含まれていることが分かる．原因の 1 つとして Re-ID 精度の低さが考えられる．一度 Re-ID に失敗すると，異なる人物の画像が同一の人物であると画像セグメントデータベースに記録されるため，それ以降の Re-ID 処理の精度が連鎖的に低下する．

また，もう一つの原因として画角外へ消えた人物の位置推定精度の低さが考えられる．位置推定精度の低さは軌跡セグメントのフィルタリングに失敗につながり，全体の追跡精度を低下させる．

7. まとめと今後の課題

本研究ではスマートグラスで利用できる，人物追跡アプリケーションを目指し，既存の MOT に加えて人物再認識と移動推定モデルを組み合わせた新たな手法を提案した．既存のマルチオブジェクトトラッキングでは，連続して出現する連続フレームのトラッキングは出来るが，オクルージョンが起きると追跡対象を見失ったり，あるいは誤ってトラッキングしたりする．一方で提案手法では，オクルージョンや画角の変動が起きても軌跡セグメントを細かく分



(a) 移動方向予測　(b) 存在領域　(c) 各軌跡セグメントの存在領域

図 7: 位置予測モデル概要

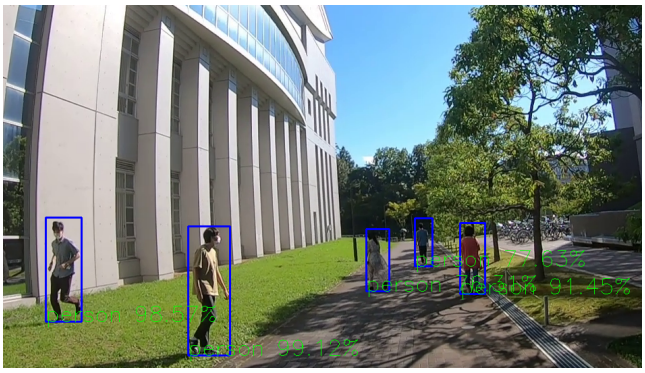


図 8: MOT による人物検出の様子

断することで各軌跡セグメント内で IDSwitch は発生しない．また，提案手法ではスマートグラスならではの位置情報を用い，人物再認識と移動推定を組み合わせて実装することで分断した軌跡セグメントを正しくつなぎ合わせる．提案手法を導入することにより DeepSORT と比較して軌跡セグメントの数を 94% 削減，IDSwitch を 61% 削減できた．

今後の課題としては，移動予測および人物再認識の精度向上に取り組む予定である．

謝辞

本研究成果は国立研究開発法人情報通信研究機構（NICT）の委託研究「ウイルス等感染症対策に資する情報通信技術の研究開発（課題番号 222）」により得られたものです．

参考文献

[1] Pooya Soltani and Antoine H.P. Morice. Augmented reality tools for sports education and training. *Computers Education*, Vol. 155, p. 103923, 2020.

[2] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649. IEEE, 2017.

[3] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699, 2019.

[4] Biswajit Bose, Xiaogang Wang, and Eric Grimson. Multi-class object tracking algorithm that handles fragmentation and grouping. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[5] L. Zhang and L. van der Maaten. Preserving structure in model-free tracking. *IEEE Transactions on Pattern Analysis Machine Intelligence*, No. 04, pp. 756–769, apr 2014.

[6] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixé. Mot20: A benchmark for multi object tracking in crowded scenes. *arXiv:2003.09003[cs]*, March 2020. arXiv: 2003.09003.

[7] Paul Voigtlaender, Michael Krause, Aljosa Osep, Jonathon Luiten, Berin Balachandar Gnana Sekar, Andreas Geiger, and Bastian Leibe. Mots: Multi-object tracking and segmentation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7934–7943, 2019.

[8] Wenhan Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, and Tae-Kyun Kim. Multiple object tracking: A literature review. *Artificial Intelligence*, Vol. 293, p. 103448, 2021.

[9] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1302–1310, 2017.

[10] Ergys Ristani and Carlo Tomasi. Features for multi-target multi-camera tracking and re-identification. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6036–6046, 2018.

[11] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stan Birchfield, Shuo Wang, Ratnesh Kumar, David Anastasiu, and Jenq-Neng Hwang. Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8789–8798, 2019.

[12] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp.

1116–1124, 2015.

[13] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. Mars: A video benchmark for large-scale person re-identification. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pp. 868–884, Cham, 2016. Springer International Publishing.