広角カメラの画像歪みを考慮した物流倉庫内マルチカメラトラッキング手法

森 裕輝[1,a)]　加納 一馬[1]　浅井 悠佑[1]　片山 晋[1]　浦野 健太[1]　米澤 拓郎[1]　河口 信夫[1,2]

概要：電子商取引（EC：Electronic Commerce）の普及に伴い，世界的に物流市場の拡大が続いている．これに伴い，物流倉庫内の業務量が増加しており，業務の効率化が重要な課題となっている．現在様々な試みが行われているが，その中でも，デジタルツインを用いたアプローチが注目されている．このアプローチを実現するためには，物流倉庫内の人や物の位置情報の正確な取得が不可欠である．しかし，単一のカメラでは視野の制限や死角の問題が生じるため，複数のカメラを用いたセンシングが必要となる．本研究では，物流倉庫の天井から床面を見下ろす形で設置した 19 台の広角カメラを活用し，マルチカメラによる作業員の追跡手法を検討した．カメラ座標と実際の倉庫内の位置との対応関係を理解するために，床面基準の位置合わせを行った．しかし，広角カメラの特性上，画面端では歪みが大きくなり，特に高さ方向の歪みの影響が顕著となる．そこで，各カメラで得られた作業員の検出情報を足元基準で整合させ，広角映像の歪みや位置合わせのズレの影響を抑制し，20%以上のトラッキング精度向上を確認した．さらに，外観特徴の統合手法についても比較検討し，提案手法の有効性を確認した．

キーワード：画像処理，物体検出，マルチカメラトラッキング，トラッキング

1. はじめに

近年，電子商取引（EC: Electronic Commerce）の普及により物流市場が拡大し，倉庫面積の増加が続いている．その結果，倉庫内業務の負担が増大し，業務の効率化が重要な課題となっている．この課題に対し，ロボットの最適経路探索を用いた効率化 [1] や倉庫レイアウト最適化 [2] など，様々な効率化手法が提案されており，その中でもデジタルツインを用いたアプローチが注目されている [3]．デジタルツインとは，IoT 機器などを用いて物理空間から取得した情報を基に，仮想空間上で人やモノの動きを再現・シミュレーションする技術である [4]．これにより，現場への影響を抑えつつ，効率化に向けた実験を低コストで実施でき，課題に対する有望な解決策が期待できる．しかし，その構築には物理空間の高精度なデジタル化が必要であり，物理空間のセンシングが不可欠である．特に，人や物の位置情報の正確な取得は重要な要素である．

位置情報の取得にはビーコンを用いる手法もあり，我々も作業員に装着して取り組んだが [5]，これは荷物や道具への応用が困難である．一方，カメラによるセンシングは，作業員や荷物に追加の機器を装着する必要がなく，比較的容易に様々な対象に応用できる．しかし，単一カメラでは視野の制限や死角が生じやすく，倉庫全体の網羅的な把握は困難である．したがって，複数カメラを用いたマルチカメラシステムが必要となる．

我々は，物流倉庫内に，図 1 に示す 1 階を含む，1〜5 階にわたって 80 台以上の定点カメラで構成される大規模カメラ基盤を構築した．本研究では，そのうち 1 階の天井に床面を見下ろす形で設置した 19 台の広角カメラを用いて，マルチカメラによる作業員の追跡手法を検討した．具体的には，各カメラで検出した作業員の位置を，カメラ固有の座標系から統一されたグローバル座標系に変換し，各カメラで得られたトラッキング結果を統合して，倉庫全体での高精度な追跡を実現した．特に本研究では，従来広く用いられてきた検出バウンディングボックス（bbox）の中心座標ではなく，作業員の足元位置を使用する手法を導入した．この工夫により，広角カメラの映像歪み，特に画面端で顕著となる高さ方向の歪みや，カメラ間の位置合わせにおけるズレの影響を抑制し，精度の高い位置整合を実現した．その結果，広範なエリアにおいてより高精度な人物追跡が可能になった．さらに，外観特徴情報の活用で，トラッキ

[1] 名古屋大学大学院 工学研究科
Graduate School of Engineering, Nagoya University
[2] 名古屋大学 未来社会創造機構
Institutes of Innovation for Future Society, Nagoya University
[a)] ymori@ucl.nuee.nagoya-u.ac.jp

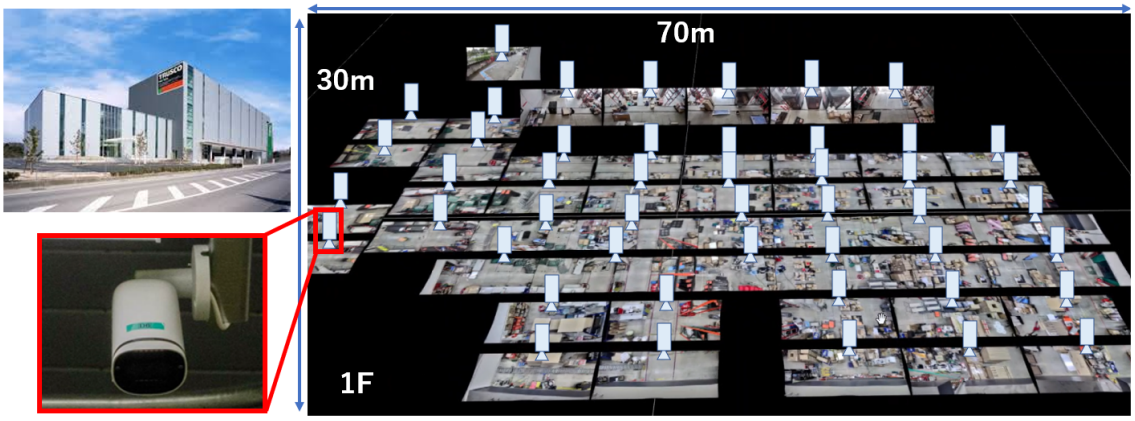

図 1: 研究対象の倉庫の大規模カメラ基盤

ング精度のさらなる向上が期待できる．ただし，広角カメラでは画面上の位置によって外観が大きく変化し，倉庫内の物体の影響で部分的な外観しか得られない場合も多い．そこで，外観特徴の利用方法（単純平均および位置と移動方向考慮型）を比較検証し，複数カメラ間での作業員の識別精度向上への有効的な手法を示した．

本研究の貢献は以下のようにまとめられる．

- 物流倉庫に設置した 19 台の広角カメラを用いて作業員の検出から倉庫全体でのトラッキングまでを一貫して実現する手法を提案し，HOTA 51.019，IDF1 54.655，MOTA 79.664 の評価スコアを達成した点．
- 検出 bbox 中心座標と足元座標の両手法について比較検証を実施し，足元座標の採用により広角映像の歪みや複数カメラの位置合わせのズレの影響を抑え，高精度なトラッキングが可能なことを明らかにした点．
- 外観特徴の利用方法として単純平均および位置移動方向考慮型の両手法を比較検証し，各手法の有効性と限界を示し，適用場面に応じた最適な特徴利用手法の選択指針を提供した点．

2. 関連研究

2.1 シングルカメラトラッキング

近年，物体検出の高性能化に伴い，単一カメラでの多物体追跡（Multi-Object Tracking; MOT）精度も向上し，多数の改良手法が提案されている．その中でも，カルマンフィルタを用いた様々な MOT 手法が提案されている [6], [7], [8]．特に，ByteTrack[6] は従来使用しなかった信頼度の低い検出結果も用いて，高精度な追跡を実現した．また，検出と追跡を統合的に学習する手法も現れている．FairMOT[9] は One-shot トラッカーの代表例で，単一のニューラルネットワーク内に物体検出と人物再識別（ReID）の 2 つのブランチを持たせ，検出精度と ReID 精度のバランスを最適化し最高性能を達成した．さらに TransTrack[10] に代表される Transformer ベースの手法では，前フレームの物体特徴をクエリとして現フレームの検出にエンコードし，注意機構によって新規検出と過去物体の対応付けを同時に行う枠組みが提案されている．このようにシングルカメラ MOT では，従来の単純なモーション予測+対応付けから，外観特徴の統合やエンドツーエンド学習，Transformer による高度な対応付けまで，多様な手法が発展してきている．

2.2 ReID と外観特徴の活用

ReID は，防犯カメラネットワークなどで写った人物画像から同一人物を特定する技術であり，特にマルチカメラ追跡において重要な役割を果たす．近年のディープラーニングの発展により，個人を特徴付ける高次元な外観特徴ベクトルを抽出する学習モデルが多数提案されてきた．Omni-Scale Network (OSNet)[11] はその代表例で，様々なスケールの特徴を捉える軽量な CNN 構造を導入し，小型モデルながら複数の ReID ベンチマークで最先端の識別精度を達成した．また，He ら [12] は，Transformer をベースとした先駆的手法であり，多くのベンチマークで CNN ベースの手法を上回った．Luo ら [13] は，バッチ正規化の調整やラベル平滑化に加え，類似・非類似な人物画像を比較して識別力を高める損失関数の工夫など，既存技術の巧みな組み合わせを体系化した Bag of Tricks（BoT）を提案した．このような高性能 ReID モデルの出現により，外観特徴の類似性に基づく高精度な人物対応付けが可能になってきている．実際，ReIDTrack[14] のように外観情報に依存した追跡手法も登場し，外観特徴の識別能力向上が追跡精度を押し上げる傾向が確認されている．

2.3 マルチカメラトラッキング

マルチカメラ多物体トラッキング（Multi-Camera Multi-Object Tracking; MCMOT）は，複数カメラを用いて広範囲な追跡を実現する技術である．MCMOT はカメラ配置により，視野が重複する場合としない場合に大別される [15]．

視野が重複しない場合には，ReID が重要な役割を担う．He ら [16] は視覚特徴と時空間情報を組み合わせ，複数カメラ間での車両追跡を実現した．また，Bipin ら [17] は，

図 2: カメラ配置の様子

人物検出，トラッキング，ReID を組み合わせるともに，エッジデバイスを用いてリアルタイム性を重視した手法を提案した．一方，視野が重複するカメラ配置においては，多くの研究で ReID と位置情報と組み合わせて精度向上を図っている [18], [19]．Yoshida ら [18] は，グローバル座標系の位置情報と，姿勢推定に基づく代表画像選択や平均連結法によるクラスタリング再識別などを組み合わせ，MCMOT の精度を競う AI City Challenge 2024 で優勝した．Xie ら [19] は，幾何学的一貫性（2D 空間親和性，3D エピポーラ親和性，ホモグラフィ親和性）と状態認識型の ReID 補正の組み合わせにより，遮蔽時の ID スイッチを効果的に防ぎ，リアルタイム追跡で高精度を実現した．

しかし，既存の MCMOT 手法の多くは，特定のカメラ設置条件において生じる，広角レンズの映像歪みやカメラ間の位置合わせの不整合といった課題への対応が十分に検討されていない．本研究では，このような構成に特有のトラッキング課題に着目し，位置情報と外観特徴の活用方法に関する比較検証を通じて，実環境における実用性の高い対策を提示する．

3. 対象環境

3.1 倉庫環境について

本研究は，愛知県にある物流倉庫を対象に検証を行った．本倉庫は，図 1 のように，80 台以上の定点カメラで構成される大規模カメラ基盤を構築している．そのうち，本研究では，床面を真上から撮影する H.View 製 HV-800G2A5 カメラ 19 台を用いた．使用動画は FullHD，5fps である．

3.2 カメラ配置と複数カメラ映像統合

設置上の制約から，視野角 110 度の広角カメラを倉庫の天井に不均一に設置しているため，各カメラ画像に対して歪み補正とカメラ位置の登録が必要となる．そこで，速度と性能のバランスが良好な Double Sphere Model[20] を用いて歪み補正を行った．しかし，天井に固定した後にカメラ固有の歪みの問題が顕在化し，キャリブレーションにおける課題が残ってしまった．現時点ではその修正は不十分であり，今後の課題である．

歪み補正後の画像は物体認識やカメラ間の位置合わせに利用される．カメラ位置の登録には，Leica 社製 BLK2GO[21] で取得したカラーマッピング済みの床面点群と各カメラで撮影された画像の床面の特徴マッチングにより実施する．具体的には，SuperPoint[22] と LightGlue[23] を用いて特徴点の対応付けし，各カメラの相対的な位置・姿勢を推定した．この画像間対応に基づき，後述する検出結果のグローバル座標変換に必要な変換行列（プロジェクション行列）を導出した．図 2 は，歪み補正後の画像に対して位置合わせを行った結果である．ただし，このプロセスにより得られたカメラ間の位置合わせは，床面領域において比較的高精度である一方，ズレ（以下，位置合わせのズレと呼ぶ）が残ってしまい，その修正も今後の課題である．

4. 提案手法

本研究のマルチカメラトラッキング手法のフレームワークを図 3 に示す．本手法は，大きく分けて，「各カメラでの作業員検出とトラッキング」「外観特徴の活用」「検出座標のグローバル座標への変換」「移動軌跡比較と重複処理」「カルマンフィルタを用いた結合」の 5 つのプロセスから構成される．「外観特徴の活用」の節では，外観特徴の抽出方法と「移動軌跡比較と重複処理」「カルマンフィルタを用いた結合」の節での外観特徴の活用方法について記載する．

4.1 各カメラでの作業員検出とトラッキング

各カメラ映像中の作業員を検出して追跡するために，YOLOv8（You Only Look Once）[24] と ByteTrack を組み合わせる．はじめに，各カメラ映像に対して，YOLOv8x モデルで各フレームの作業員を検出する．このとき得られた結果は，以降「検出結果」と呼ぶ．そして，ByteTrack を用いて，フレーム間での作業員の対応付けを行う．図 4 にトラッキングの様子を示す．このようにして，連続したフレーム上で同一人物と判断された検出結果のまとまりを，トラックと呼ぶ．また，各トラックには一意のトラッキング ID が割り当てられる．

4.2 外観特徴の活用（比較手法）

4.2.1 外観特徴抽出

カメラ間の移動においては，位置情報だけでなく外観特徴の活用で，トラッキング精度の向上が期待できる．そこで本研究では，OSNet_x1_0 を用いて，4.1 節で検出された bbox 内の画像から，作業員の外観特徴を抽出する．

4.2.2 外観特徴利用法

トラッキングにおいて外観特徴の利用は精度を高めるための有用な情報になり得る．しかし，広角カメラの利用により，図 5 のように，画面内の位置で見た目が大きく異なる．また，物流倉庫という多種多様な物にあふれた環境に

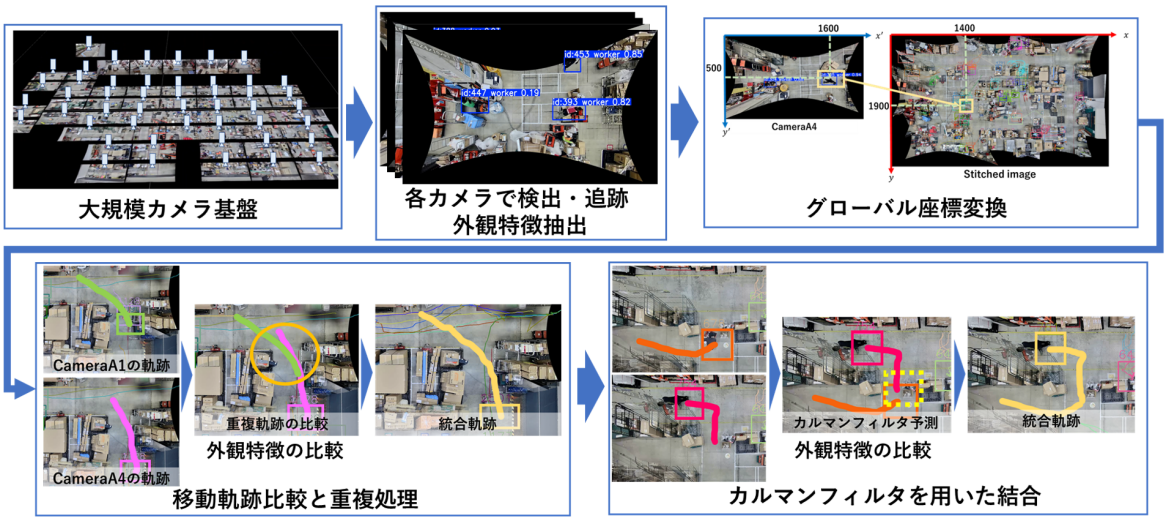

図 3: 提案手法のフレームワーク

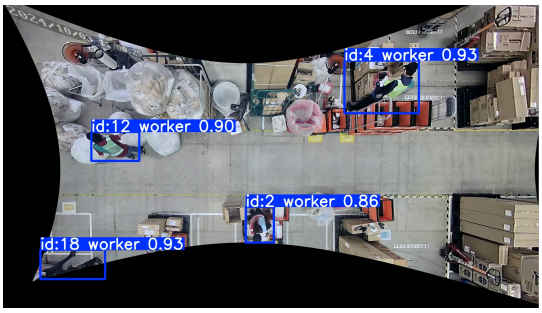

図 4: 単一カメラでのトラッキングの様子

図 5: 同一人物の見た目の変化

図 6: 画面上の位置と移動方向が近い検出結果の例

より，身体の一部のみが見える場合や道具を持つ状況など多様な状態が起こる．このような外観特徴のばらつきは，外観類似度によるトラッキングにおいて誤判定の原因となる．したがって，本研究では次の 2 種類の外観特徴の利用法を比較し，有効な手法を検討する．

4.2.2.1 単純平均

同一トラック内の複数の検出結果から得られる特徴量を単純平均により統合し，各トラックの代表値として比較する．各検出の特徴量を平均すれば，一時的な見た目の変化や身体の一部のみが映る状況などの影響を抑えられる．

4.2.2.2 位置と移動方向考慮型

信頼性の高い外観特徴類似度評価の実現するため，図 6 のように，画面上の位置と移動方向が近い検出結果に限定して外観特徴の比較を行う．

以下に，各検出結果の外観特徴比較手順を示す．また，比較を行う検出結果の組を検出ペアと表現する．

(1) 各検出結果のカメラ座標系における bbox 中心位置を用いて，検出間距離 d を算出し，距離が近い検出ペアほど高い重みを持つように，位置重み w_{pos} を計算する．ここで，σ_p は重み付けのパラメータである．

$$w_{pos} = \exp\left(-\frac{d^2}{\sigma_p^2}\right)$$

(2) 検出ペアに対し，それぞれの移動方向ベクトル $\boldsymbol{v}_1$，$\boldsymbol{v}_2$ 間の方向の一致度を評価するため，コサイン類似度 $\cos_{vel}$ を用いる．移動方向が一致するほど重みが高くなるように，移動方向の重み w_{vel} を算出する．

$$w_{vel} = \frac{1 + \cos_{vel}}{2}$$

(3) 位置と移動方向の両方が類似している検出ペアを優先するため，位置重み w_{pos} と移動方向重み w_{vel} の積により，総合重み w を求める．

$$w = w_{pos} \times w_{vel}$$

(4) 比較を行うトラック内の検出ペアが M 件以上ある場合は，重み w の上位 M 件について，外観特徴の類似度を算出し，その上位 75% の検出ペアの平均値を外観類似度として評価する．これは，外れ値の影響を抑えつつ，有効な情報を活用するためである．一方で，検出ペアが M 件未満の場合は，位置と移動方向が近い検出ペアを比較するという意図が十分に機能しない可能性や外れ値の影響を強く受ける可能性があるため，外観類似度は単純平均の値を用いる．

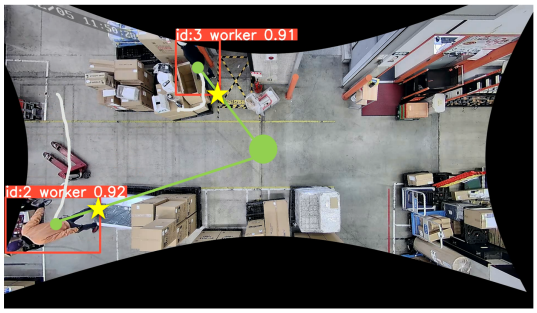

図 7: 足元座標の算出

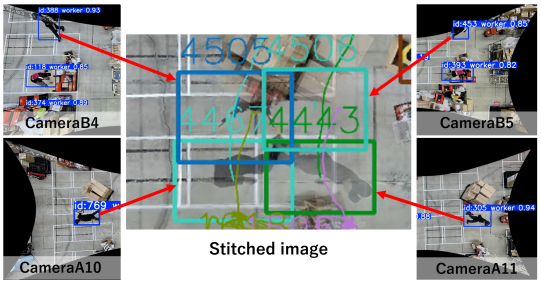

図 8: 重複検出の様子

4.3 検出座標のグローバル座標への変換

各カメラでの検出結果をグローバル座標に変換し，倉庫全体でのトラッキングに繋げる．そこで，3.2 節で作成したプロジェクション行列を用い，各カメラでの検出結果をグローバル座標系に変換する．ただし，グローバル座標への変換行列は 3.2 節で記載の通り床面の特徴マッチングにより生成している．したがって，カメラ間の位置合わせの精度は床面に近い領域ほど高く保たれやすく，一方で物体の上部など高さ方向の情報は，広角カメラの映像歪みや位置合わせのズレの影響を受けやすい．そのため，本研究では，以下の 2 通りの方法で検出結果をカメラ座標系からグローバル座標系へと変換し，精度を比較する．1 つ目は，検出した bbox の中央座標を利用する方法である．2 つ目は，図 7 に示すように，検出した bbox の中心とカメラ中心を結ぶ線分と，bbox 各辺との交点を求めて，足元座標として利用する方法である．本研究では，天井から床面を見下ろす形で設置したカメラを用いているため，人物は頭部よりも足元のほうが，画面中心近くに映る．したがって，検出された bbox の中心とカメラ中心を結ぶ線分と，bbox の辺との交点は，床面に近い位置，すなわち足元に最も近い点とみなせる．このようにして算出された足元座標を活用して，広角カメラの映像歪みや位置合わせのズレの影響を低減し，より正確なトラッキングの実現を図る．

4.4 移動軌跡比較と重複処理

3.2 節の通り，画角が重複する関係上，図 8 のように，複数カメラで同一作業員を検出・追跡する場合がある．そこで，以下の要素を組み合わせて，各カメラでの移動軌跡を比較し，重複・分断されたトラックの統合を行う：

- **位置的評価:** 重複フレームにおける各トラックの検出座標間の距離の平均や，各対応検出間の最大距離を算出し，距離が近い軌跡同士を統合候補とする．また，統合候補は同一カメラ内の検出ではなく，カメラ ID が異なる場合に限定する．
- **移動方向の一貫性:** 比較するトラックの重複フレーム部分の開始と終点位置の変位方向を，コサイン類似度で評価し，類似度が高い場合のみ統合候補とする．
- **外観的評価:** 4.2.2 項の外観特徴の類似性を評価し，類似性が高い場合に統合候補とする．

上記すべての要素を満たすトラックの組み合わせに対して，同一のトラッキング ID を付与する．そして，各トラックの統合後は，各フレーム内で同一 ID が複数存在する場合，検出したカメラ座標が中央座標に近い方を残し，映像歪みの影響を最小限に抑える．

4.5 カルマンフィルタを用いた結合

4.4 節では，重複フレームにおける距離，移動方向と外観特徴の類似性を基に，同一作業員のトラックを統合する．しかし，検出結果の重複がないフレーム間では直接比較が困難となり，トラッキングが断片化する．そこで，各トラックにカルマンフィルタを適用し，予測された位置，移動方向と外観特徴を用いて，検出フレーム間のギャップがある場合でも信頼性の高い統合を実現する．

具体的には，トラックの終了時点からカルマンフィルタで次の位置を予測し，その移動方向と他トラックの開始時の移動方向のコサイン類似度を求める．また，4.2.2 項の手法により，トラック間の外観特徴の類似度を算出する．外観類似度が高い場合には，統合の許容距離を拡大し，移動方向の閾値も緩和する．一方で，類似度が低い場合は，許容距離を縮小する．最後に，カルマンフィルタの予測と次のトラックの開始位置の距離が閾値以下かつ移動方向の類似度が一定値以上であれば，同一作業員として統合する．

5. 評価実験

5.1 実験に使用したデータ

19 台のカメラで撮影された 5fps，FullHD の 30 分の動画を用いた．各カメラの配置は 3.2 節の通りである．

5.2 評価実験準備

5.2.1 物体検出モデルの学習

本実験では，物体検出に YOLOv8x モデルを用いた．学習には，5.1 節とは異なる時間の映像に対して，Kano ら [25] のアノテーション手法と手動アノテーション（人手で対象オブジェクトを矩形で囲い，適切なラベルを付与する）手法で生成したデータセットを用いた．表 1 に，学習（Train）と検証（Validation）に用いたデータ数を示す．また，モデルの入力サイズは 640 × 640 である．

表 1: 物体検出モデルの学習に使用したデータセットの統計

データセット	画像数	データ件数
Train	11,459	29,473
Validation	1,373	3,629

5.2.2 外観特徴抽出モデルの学習

作業員の外観特徴抽出のために OSNet_x1_0 モデルを用い，我々が構築したデータセットを使用してファインチューニングを行った．このデータセットは，5.1 節とは異なる時間帯に撮影された 10 分間の映像から作成した．まず，YOLO を用いて各フレームから作業員を検出し，検出された bbox で人物画像をトリミングした．次に，提案手法に基づいてトラッキングを実施した．トラッキング結果にはトラッキングミスが含まれるため，それらをすべて手動で修正し，複数のカメラに映る同一人物の対応関係が正確なデータを作成した．また，連続フレームからは似た画像が多く得られるため，データの多様性を確保する目的で，20 フレームごとにサンプリングして学習用画像とした．学習データセットは 19 台のカメラで撮影された計 40 人の 3029 枚の画像で構成され，画像は 256 × 128 にリサイズし，トレーニングはバッチサイズ 64，最大 250 エポックの設定で実施した．

5.2.3 評価指標と正解データ

トラッキング結果の評価には，3 つの指標を用いた．

- **Higher Order Tracking Accuracy (HOTA)**[26]：検出精度（検出の正確さ）と関連付け精度（ID の正しさ）をバランスよく評価する指標であり，トラッキング手法全体の性能を包括的に示すことができる．
- **ID F1-score (IDF1)**[27]：同一物体の ID がどれだけ正しく保たれているかを評価する指標であり，ID の一貫性に特化している．
- **Multiple Object Tracking Accuracy (MOTA)**[28]：誤検出（False Positives），検出漏れ（False Negatives），トラッキングの誤り（ID スイッチ）を統合的に考慮した評価指標であり，誤り率に対する感度が高い．

これらの指標は，トラッキングの評価で広く用いられており，すべて 0-1 で評価され，1 に近いほど優秀である．本論文では，これらの値を 100 倍し，0〜100 のスケールで記載している．評価指標の計算には，評価ツールである TrackEval[29] を使用した．

また，正解データは，提案手法に基づいてトラッキングを行った結果に対し，ID スイッチや検出漏れ，誤検出があった箇所を全て手動で修正・補完して作成した．

5.3 評価実験

5.3.1 比較パターン

座標種（足元座標または検出中心座標）と外観特徴の使用方法（非利用／単純平均／位置と移動方向考慮型）の組み合わせにより，各要素がトラッキング精度向上にどの程度寄与するかを比較検証した．

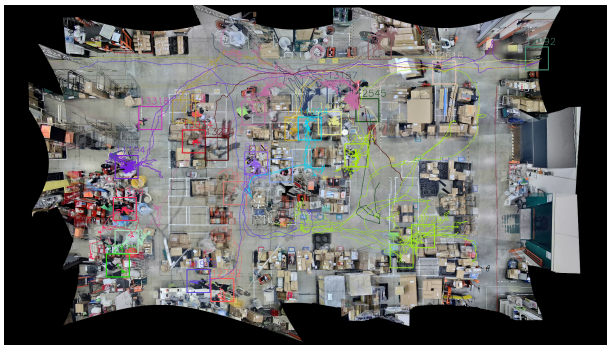

図 9: トラッキング結果の様子

5.3.2 実験手法

提案手法の各構成要素がトラッキング精度に及ぼす影響を検証するため，提案手法に基づく比較実験を実施した．まず，5.1 節の各カメラ映像に対して物体検出とトラッキングを行う．物体検出には，5.2.1 項の通りファインチューニングした YOLOv8x モデルを用いた．次に，5.2.2 項に記す OSNet_x1_0 モデルを用いて，外観特徴量を抽出する．また，4.3 節の通り，検出 bbox の中心座標または足元の座標を，プロジェクション行列を用いてグローバル座標系に変換する．そして，グローバル座標に変換した重複または断片化した軌跡に対して，4.4 節の処理を行う．以下の条件のもと，2 つのトラックのグローバル座標における距離の平均が 130 ピクセル以内の場合にトラッキング ID を統合した．また，これらの閾値は，複数の候補値を比較し，トラッキング指標が最も良好なスコアを示した閾値構成を採用した．

- いずれかのフレームにおいても 2 物体間の距離が 300 ピクセル以下である．
- 移動方向のコサイン類似度が 0.8 以上である．
- 4.2.2 項の外観特徴のコサイン類似度が 0.85 以上である．

そして，最後に 4.5 節の統合処理を適用し，以下の条件を満たす場合に統合を行った．

- トラック間のフレームギャップが 10 フレーム以内である．
- 一方のトラックの予測移動方向ベクトルと，他方トラックの開始時の移動方向ベクトルとのコサイン類似度が 0.8 以上である．
- 予測位置と次トラックの開始位置の距離が許容距離（130 ピクセル）以内である．ただし，4.2.2 項の外観類似度が 0.85 以上の場合は許容距離を 2 倍に拡大し，低い場合（〜0.5）は 0.5 倍に縮小する．

また，その他のパラメータは，$\sigma_p = 500$，$M = 8$ とした．

表 2: 実験条件ごとのトラッキング評価結果

条件	HOTA	IDF1	MOTA
(1) bbox 中心座標のみ	39.6	39.7	65.7
(2) bbox 中心座標＋外観特徴（単純平均）	48.5	49.7	77.0
(3) bbox 中心座標＋外観特徴（位置と移動方向考慮型）	46.9	47.0	75.4
(4) 足元座標のみ	49.1	50.2	78.9
(5) 足元座標＋外観特徴（単純平均）	**51.0**	**54.7**	**79.7**
(6) 足元座標＋外観特徴（位置と移動方向考慮型）	50.8	54.5	79.2

5.4 結果と考察

表 2 に，計 6 条件でのトラッキング結果を示す．また，図 9 に最も高精度な条件でのトラッキングの様子を示す．

まず，座標系の違いによる影響について述べる．外観特徴を利用しない条件では，bbox 中心座標での HOTA，IDF1，MOTA は 39.6，39.7，65.7 に対し，足元座標では 49.1，50.2，78.9 と，いずれも大幅に向上した．これは，足元座標の利用により，広角カメラの映像歪みや位置合わせのズレの影響を低減できたためと考えられる．

次に，外観特徴の利用効果を比較すると，両座標系において外観特徴の導入により，識別・対応付け精度が向上した．単純平均による利用では，足元座標での HOTA，IDF1，MOTA が 51.0，54.7，79.7，bbox 中心座標では 48.5，49.7，77.0 となった．一方，位置と移動方向を考慮した場合，足元座標では 50.8，54.5，79.2，bbox 中心座標では 46.9，47.0，75.4 であった．いずれの手法でも，IDF1 を中心に精度が向上し，特に足元座標と組み合わせた場合，最も高い精度が得られた．また，単純平均を用いた方が，位置や移動方向を考慮した場合よりもわずかに高い評価を示した．このことから，一時的な見た目の変化や身体の一部しか映っていない状況の影響を抑えることが重要だと考えられる．

以上の結果から，足元座標の利用がトラッキングの局所化と対応付け精度の向上に大きく寄与することが明らかとなった．さらに，外観特徴の活用は対象識別の安定性を高め，トラッキング精度の向上が確認された．特に，足元座標と単純平均による外観特徴利用の組み合わせが最も高精度であるため，多視点映像での正確なトラッキングには，空間情報と外観情報の適切な統合が重要だと考えられる．

6. まとめと今後の展望

本研究では，広角カメラを用いたマルチカメラトラッキング精度の向上を目的とした手法を提案した．特に，検出 bbox の中心座標と足元座標の 2 種類を用いて比較検証を行い，広角カメラの映像歪みや位置合わせのズレに対して，足元座標の有効性を確認した．特に，外観特徴を使用しない条件下においては，HOTA 24%, IDF1 26%，MOTA 20%の向上が確認できた．さらに，外観特徴の利用方法として，OSNet により抽出した特徴を単純平均する方法と，位置と移動方向を考慮する方法を評価し，各手法の特性と精度への寄与を明らかにした．提案手法の有効性は，物流倉庫内に設置した 19 台の広角カメラを用いて実環境下で検証し，空間情報と外観情報の適切な統合がトラッキング精度向上に有効であると示した．

一方で，トラッキングが途切れてしまう最も大きな要因は，位置合わせのズレの影響である．本研究では，作業員の足元座標を用いることで，位置合わせのズレの影響を軽減し，より安定した位置変換を実現した．しかし，足元座標を用いても，位置合わせのズレの影響は完全には解消されず，各カメラのトラッキング結果の統合における位置的整合性を損ねる要因となり，結果として同一人物の軌跡であっても別 ID として処理されてしまうケースがあった．また，作業員の下半身が物に隠れ，上半身しか検出できないケースなども存在し，足元の座標を取得できていない状況も存在した．

今後は，位置合わせのズレをさらに低減するために，時系列情報や動きの一貫性を活用した統合手法の改良を進める．また，足元が検出できない状況への対応として，上半身など他の部位から足元位置を推定する補完的な手法の導入も検討する．さらに，ビーコンの情報 [5] を融合し，トラッキングの頑健性を向上させる．

これらの方向性を追求し，現実の物流倉庫における作業員トラッキングの精度と信頼性をさらに向上させ，効率的な業務運用への貢献を目指す．

謝辞 本研究の一部は，国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の委託業務（JPNP23003），科研費挑戦的研究（開拓）22K18422，トラスコ中山株式会社に支援いただいている．

参考文献

[1] Ping Li and Jiachen Zhao. Optimal path allocation of robot based on modern logistics warehouse. In *Proceedings of the 2022 5th International Conference on E-Business, Information Management and Computer Science*, pp. 378–383, 2022.

[2] Xiulian Hu and Yi-Fei Chuang. E-commerce warehouse layout optimization: systematic layout planning using a genetic algorithm. *Electronic Commerce Research*, Vol. 23, No. 1, pp. 97–114, 2023.

[3] Praveen Kumar Reddy Maddikunta, Quoc-Viet Pham, B Prabadevi, Natarajan Deepa, Kapal Dev, Thippa Reddy Gadekallu, Rukhsana Ruby, and Mad-

husanka Liyanage. Industry 5.0: A survey on enabling technologies and potential applications. *Journal of Industrial Information Integration*, Vol. 26, p. 100257, 2022.

[4] Barbara Rita Barricelli, Elena Casiraghi, and Daniela Fogli. A survey on digital twin: Definitions, characteristics, applications, and design implications. *IEEE Access*, Vol. 7, pp. 167653–167671, 2019.

[5] Kazuma Kano, Takuto Yoshida, Nozomi Hayashida, Yusuke Asai, Hitoshi Matsuyama, Shin Katayama, Kenta Urano, Takuro Yonezawa, and Nobuo Kawaguchi. Smartphone localization with solar-powered ble beacons in warehouse. In *International Conference on Human-Computer Interaction*, pp. 291–310. Springer, 2022.

[6] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. In *Proceedings of the European Conference on Computer Vision*, pp. 1–21. Springer, 2022.

[7] Yunhao Du, Zhicheng Zhao, Yang Song, Yanyun Zhao, Fei Su, Tao Gong, and Hongying Meng. Strongsort: Make deepsort great again. *IEEE Transactions on Multimedia*, Vol. 25, pp. 8725–8737, 2023.

[8] Yu-Hsiang Wang, Jun-Wei Hsieh, Ping-Yang Chen, Ming-Ching Chang, Hung-Hin So, and Xin Li. Smiletrack: Similarity learning for occlusion-aware multiple object tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38, pp. 5740–5748, 2024.

[9] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu. Fairmot: On the fairness of detection and re-identification in multiple object tracking. *International journal of computer vision*, Vol. 129, pp. 3069–3087, 2021.

[10] Peize Sun, Jinkun Cao, Yi Jiang, Rufeng Zhang, Enze Xie, Zehuan Yuan, Changhu Wang, and Ping Luo. Transtrack: Multiple object tracking with transformer. *arXiv preprint arXiv:2012.15460*, 2020.

[11] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3702–3712, 2019.

[12] Shuting He, Hao Luo, Pichao Wang, Fan Wang, Hao Li, and Wei Jiang. Transreid: Transformer-based object re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15013–15022, 2021.

[13] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 0–0, 2019.

[14] Andreas Specker and Jürgen Beyerer. Reidtrack: Reid-only multi-target multi-camera tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5442–5452, June 2023.

[15] Temitope Ibrahim Amosa, Patrick Sebastian, Lila Iznita Izhar, Oladimeji Ibrahim, Lukman Shehu Ayinla, Abdulrahman Abdullah Bahashwan, Abubakar Bala, and Yau Alhaji Samaila. Multi-camera multi-object tracking: A review of current trends and future advances. *Neurocomputing*, Vol. 552, p. 126558, 2023.

[16] Zhiqun He, Yu Lei, Shuai Bai, and Wei Wu. Multi-camera vehicle tracking with powerful visual features and spatial-temporal cue. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vol. 1, p. 1, 2019.

[17] Bipin Gaikwad and Abhijit Karmakar. Smart surveillance system for real-time multi-person multi-camera tracking at the edge. *Journal of real-time image processing*, Vol. 18, No. 6, pp. 1993–2007, 2021.

[18] Ryuto Yoshida, Junichi Okubo, Junichiro Fujii, Masazumi Amakata, and Takayoshi Yamashita. Overlap suppression clustering for offline multi-camera people tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7153–7162, 2024.

[19] Zhenyu Xie, Zelin Ni, Wenjie Yang, Yuang Zhang, Yihang Chen, Yang Zhang, and Xiao Ma. A robust online multi-camera people tracking system with geometric consistency and state-aware re-id correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7007–7016, 2024.

[20] Vladyslav Usenko, Nikolaus Demmel, and Daniel Cremers. The double sphere camera model. In *International Conference on 3D Vision, 3DV 2018*, pp. 552–560. Institute of Electrical and Electronics Engineers Inc., 2018.

[21] BLK2Go. Leica geosystem. accessed 2025/3/10. `https://leica-geosystems.com/products/laserscanners/autonomous-reality-capture/leica-blk2go-handheld-imaginglaser-scanner`.

[22] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 224–236, 2018.

[23] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17627–17638, 2023.

[24] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.

[25] Kazuma Kano, Yuki Mori, Keisuke Higashiura, Tahera Hossain, Shin Katayama, Kenta Urano, Takuro Yonezawa, and Nobuo Kawaguchi. Composite image generation using labeled segments for pattern-rich dataset without unannotated target. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 507–512, 2024.

[26] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixé, and Bastian Leibe. Hota: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision*, pp. 1–31, 2020.

[27] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European conference on computer vision*, pp. 17–35. Springer, 2016.

[28] Keni Bernardin and Rainer Stiefelhagen. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, Vol. 2008, pp. 1–10, 2008.

[29] Arne Hoffhues Jonathon Luiten. Trackeval. `https://github.com/JonathonLuiten/TrackEval`, 2020.