

スポーツ選手を対象とする姿勢特徴を考慮した複数人物追跡

小林 万葉^{1,a)} 小池 和輝¹ 植田 諒大¹ 滝本 裕則^{2,b)}

概要：複数物体追跡 (Multi-Object Tracking, MOT) は、コンピュータビジョン分野における重要なタスクの 1 つであり、ビデオシーケンスに登場する複数の物体を一意に識別し、その動きを正確に追跡することを目的とする。近年、スポーツ領域においても運動解析や戦術分析への応用を背景として、試合映像中の選手を対象とした MOT の研究が注目されている。しかし、スポーツシーンでは、選手が加減速や急な方向転換を伴う不規則な運動を示す上に、チーム競技では共通したユニフォームを着用することで、選手間でその外観が類似しやすくなる。そのため、位置情報や外観情報に基づく従来の関連付けでは物体の誤割当てが生じやすく、異なる物体同士を時間的に結び付けてしまうことで追跡精度の低下を招く。本研究では、スポーツ選手を対象とした MOT の高精度化を目的とし、姿勢情報を考慮した追跡手法を提案する。具体的には、試合中に選手が多様な姿勢をとる点に着目し、深層学習に基づく姿勢推定モデルの中間層から抽出される特徴量の類似度を、物体検出結果と過去の Tracklet の関連付けに導入した。これにより、複雑な運動や類似した外観を持つ物体が登場するシナリオでも、安定した物体の関連付けが期待できる。評価実験では、アイスホッケー選手を追跡対象とする VIP-HTD データセットを用い、MOTA や IDF1 をはじめとする主要評価指標により提案手法の有効性を検証した。

1. はじめに

複数物体追跡 (Multi-Object Tracking, MOT) は、コンピュータビジョン分野における重要なタスクの 1 つであり、ビデオシーケンスに登場する複数の物体を一意に識別し、その動きを正確に追跡することを目的とする。近年、スポーツ領域においても運動解析や戦術分析への応用を背景とし、試合映像中の選手を対象とした MOT の研究が注目されている。

本研究では、スポーツ選手を対象とした MOT の高精度化を目的とし、姿勢情報を考慮した追跡手法を提案する。具体的には、これにより、

2. 関連研究

MOT を解く代表的なパラダイムは “Tracking-by-Detection” である。このパラダイムでは、まず各フレームに登場する物体を検出し、その後、既存のトラックと検出

された物体を何かしらの指標に基づいて結びつける。これにより、異なるフレーム間で同じ物体には同じ識別子 (ID) が割り当てられ、時間的な物体追跡が実現する。ここで、既存トラックと検出結果の間に同一物体を結びつける操作を関連付け (Association) と呼ぶ。Association は、トラックの集合を $\hat{\mathcal{X}} = \{\hat{\mathbf{x}}_i\}_{i=1}^M$ 、検出結果の集合を $\hat{\mathcal{Y}} = \{\hat{\mathbf{y}}_j\}_{j=1}^N$ とすると、両集合の要素を重複なく 1 対 1 で対応付け、コスト $\mathcal{L}_{\text{match}}(\cdot)$ の総和を最小化するインデックス割当 \hat{P} を求める問題に帰着する。これは、式 1 のように線形割当問題として定式化され、ハンガリアン法 [1] などのアルゴリズムによって解かれる。

$$\hat{P} = \underset{P \in \mathcal{P}_{M \times N}}{\operatorname{argmin}} \sum_{i=1}^M \sum_{j=1}^N \mathcal{L}_{\text{match}}(\hat{\mathbf{x}}_i, \hat{\mathbf{y}}_j) \quad (1)$$

SORT[2] は Tracking-by-Detection に基づく代表的な追跡手法である。SORT では、物体の位置とその速度によって状態変数を記述し、Kalman Filter[3] を用いて状態の予測・更新を行う。この際、Association のコストには、IoU(Intersection over Union) が用いられる。IoU は 2 つの図形の重なり度合いを評価する指標であり、SORT では前時刻から状態予測されたトラックと現在の検出結果の Bounding-Box 間で IoU を計算し、物体の位置的類似性に基づいた Association を行う。SORT は、非常にシンプルな枠組みで物体追跡を実現する一方で、複雑に動く物体の追跡を苦手とする。これは、Kalman Filter が前提とする

¹ 岡山県立大学大学院 情報系工学研究科 システム工学専攻
Okayama Prefectural University Graduate School of Systems Engineering Course of Advanced Systems Engineering

² 岡山県立大学 情報工学部 情報通信工学科
Okayama Prefectural University Faculty of Computer Science and System Engineering Department of Communication Engineering

a) sk625018@c.oka-pu.ac.jp

b) takimoto@c.oka-pu.ac.jp

状態空間が、状態の遷移に線形性を仮定することに起因する。物体が不規則な動きを示したり、カメラモーションが加わると、物体の見かけの動きが複雑化し、線形運動を仮定した状態予測は、予測トラックと検出結果の位置的類似性を低下させる。その結果、IoUに基づくAssociationではID Switch（フレーム間で同一物体のIDが切り替わる現象）が発生しやすくなる。

これに対して、複雑な物体の動きに対応するために、パーティクルフィルタやTransformerに基づいたより高精度な予測器を用いた追跡手法が提案されているが、いずれも学習に要するコストや、推論自体に時間がかかるてしまうことから実用性に課題を抱えている。

DeepSORT[4]は、SORTの課題を克服することを目的に設計された追跡手法である。DeepSORTでは、AssociationのコストにCNN(Convolutional Neural Network)から抽出された外観特徴量のコサイン類似度を導入している。これにより、物体の複雑な運動や、Occlusion（物体の一部または全体が遮蔽される現象）が生じても、外観特徴に基づいたAssociationによって、より頑健な追跡を実現している。しかし、共通のユニフォームを着用するスポーツシーンや、野生動物を追跡対象とするシナリオでは、各物体の外観が類似しやすいことから、外観類似性に基づいたAssociationはむしろ追跡精度低下の要因となることが報告されている[5], [6]。

近年、外観情報を用いずに運動情報を強化することで追跡精度を向上させる手法も提案されている。例えばOC-SORT[7]は、。また、PoseTrack

本研究では、

3. 提案手法

本研究では、を提案する。具体手には、。さらに、することで必要としない。

狙い。なぜいいのか？

OC-SORTの上に設計することを述べる。

全体図を示す。

3.1 OC-SORT

あいうえお。

3.2 姿勢特徴量の抽出

ある時刻において、検出結果の集合 $\hat{\mathcal{Y}} = \{\hat{\mathbf{y}}_j\}_{j=1}^N$ に対して得られる姿勢特徴量の集合を $\mathcal{F} = \{\mathbf{f}_j\}_{j=1}^N$ とすると、検出結果 $\hat{\mathbf{y}}_j$ の姿勢特徴量 \mathbf{f}_j は、次のように定義される。

$$\mathbf{f}_j = \text{GAP}(F_j) \in \mathbb{R}^C \quad (2)$$

ここで、 $F_j \in \mathbb{R}^{C \times H \times W}$ は、 $\hat{\mathbf{y}}_j$ の Bounding-Box で切り取られるフレーム領域を Top-Down 方式の姿勢推定器に入力した際に、その中間層が出力する特徴マップである。

この特徴マップ F_j は、対象物体 $\hat{\mathbf{y}}_j$ の局所的な間接配置や身体構造を捉えており、外観が類似する物体が多く登場するシナリオでは、比較的頑健な物体表現が得られることが期待される。また、 $\mathbf{f}_j \in \mathbb{R}^C$ は、 F_j に GAP(Global Average Pooling) を適用し、 C 次元のベクトルに変換することで得られる。

3.3 姿勢特徴を考慮した Association

ある時刻において、トラックの集合 $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^M$ が保持する姿勢情報の集合を $\mathcal{G} = \{\mathbf{g}_i, c_i\}_{i=1}^M$ とする。ここで、 $\mathbf{g}_i \in \mathbb{R}^C$ はトラック \mathbf{x}_i が持つ姿勢特徴量であり、 $c_i \in [0, 1]$ はその姿勢信頼度を表す。集合 \mathcal{G} は、Association の結果に応じて時間的に更新される（後述）。

本研究では、姿勢特徴を考慮した Association を行うにあたり、トラックと検出結果の姿勢類似度を式 3 で定まる姿勢コスト関数 $C_{\text{pose}}(\cdot)$ を用いて求める。

$$C_{\text{pose}}(\mathbf{x}_i, \hat{\mathbf{y}}_j) = \frac{\mathbf{g}_i^T \mathbf{f}_j}{\|\mathbf{g}_i\|_2 \|\mathbf{f}_j\|_2} \in [0, 1] \quad (3)$$

式 3 は \mathbf{g}_i と \mathbf{f}_j のコサイン類似度を表し、その値が 1 に近づくほどトラック \mathbf{x}_i と検出結果 $\hat{\mathbf{y}}_j$ の姿勢類似性が高いことを意味する。最終的なコスト関数 $\mathcal{L}_{\text{match}}(\cdot)$ は、OC-SORT の既存コスト関数 $C_{\text{ocsort}}(\cdot)$ に姿勢コスト関数 $C_{\text{pose}}(\cdot)$ を加えて式 4 で与える。

$$\mathcal{L}_{\text{match}}(\hat{\mathbf{x}}_i, \hat{\mathbf{y}}_j) = C_{\text{ocsort}}(\hat{\mathbf{x}}_i, \hat{\mathbf{y}}_j) + \lambda_p c_i C_{\text{pose}}(\hat{\mathbf{x}}_i, \hat{\mathbf{y}}_j) \quad (4)$$

ここで、 $C_{\text{ocsort}}(\cdot)$ は IoU と速度方向の類似性に基づくコスト関数であり、 λ_p は姿勢項の寄与を制御する重みである。また、 c_i を掛けることで、トラック \mathbf{x}_i が信頼度の低い姿勢特徴量 \mathbf{g}_i を持つ場合には、 \mathbf{g}_i から算出される姿勢項の影響を抑える狙いがある。実装上は、 $\mathcal{L}_{\text{match}}(\cdot) \leftarrow -\mathcal{L}_{\text{match}}(\cdot)$ として、ハンガリアン法によって式 1 で表される総コストの最小化問題を解く。

3.4 姿勢情報の更新

ある時刻において、式 4 のコスト関数 $\mathcal{L}_{\text{match}}(\cdot)$ を用いた Association の結果、トラック \mathbf{x}_i と検出結果 $\hat{\mathbf{y}}_j$ が対応付けられたとする。このとき、 \mathbf{x}_i が保持する姿勢情報 $\{\mathbf{g}_i, c_i\}$ は次のように更新する。

$$\mathbf{g}_i \leftarrow \mathbf{f}_j, \quad c_i \leftarrow 1.0 \quad (5)$$

これは、最新の検出結果が持つ姿勢特徴量を対応付けられたトラックが引継ぎ、信頼度を最大値に戻す操作である。一方、どの検出結果にも対応付けられなかったトラック \mathbf{x}_k については、保持している姿勢特徴量 \mathbf{g}_k を維持しつつ、姿勢信頼度 c_k のみを時間減衰させる。

$$\mathbf{g}_k \leftarrow \mathbf{g}_k, \quad c_k \leftarrow \mathbb{1}_{\{\gamma c_k > \tau\}} \gamma c_k \quad (6)$$

ここで、 $\gamma \in (0, 1]$ は減衰率、 τ は姿勢信頼度の閾値であり、 $\mathbb{1}_{\{\gamma c_k > \tau\}}$ は減衰値 γc_k が閾値 τ 以上の場合に 1、それ以外の場合に 0 を出力する指示関数である。式 6 による信頼度の減衰挙動を図??に示す。式 5,6 に基づいて集合 \mathcal{G} を更新することで、時間的に物体の姿勢が変化する中で、長期間対応付けられていないトラックの姿勢情報が Association に与える影響を抑制し、追跡精度の低下を防ぐ狙いがある。

4. 評価実験

4.1 評価データセット

本研究では、評価データセットとして、アイスホッケー選手を追跡対象とする VIP-HTD データセット?を用いた。

4.2 評価指標

本研究では、評価指標として、MOTA(Multi-Object Tracking Accuracy)を用いた。

$$MOTA = 1 - \frac{|FN| + |FP| + |IDs|}{|GT|} \quad (7)$$

ここで、 $|FN|$ は、

また、 $IDF1$ は、 IDP と IDR の調和平均として式 8 で定義される。

$$IDF1 = \frac{2 \times IDP \times IDR}{IDP + IDR} \quad (8)$$

ここで、

4.3 評価結果および考察

ここに評価結果の内容を書く。

5. おわりに

本研究では、

参考文献

- [1] Kuhn, H. W.: The Hungarian method for the assignment problem, *Naval Research Logistics (NRL)*, Vol. 52 (1955).
- [2] Bewley, A., Ge, Z., Ott, L., Ramos, F. and Upcroft, B.: Simple online and realtime tracking, *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3464–3468 (online), DOI: 10.1109/ICIP.2016.7533003 (2016).
- [3] Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering*, Vol. 82, No. 1, pp. 35–45 (1960).
- [4] Wojke, N., Bewley, A. and Paulus, D.: Simple online and realtime tracking with a deep association metric, *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649 (2017).
- [5] Sun, P., Cao, J., Jiang, Y., Yuan, Z., Bai, S., Kitani, K. and Luo, P.: DanceTrack: Multi-Object Tracking in Uniform Appearance and Diverse Motion, *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20961–20970 (2022).
- [6] Zhang, L., Gao, J., Xiao, Z. and Fan, H.: AnimalTrack: A

- Benchmark for Multi-Animal Tracking in the Wild (2022).
[7] Cao, J., Pang, J., Weng, X., Khirodkar, R. and Kitani, K.: Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking, *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9686–9696 (2023).

付 錄

A.1 付録 1 節の表題

ここに付録の内容を書く。