

## 複数物体追跡における ID スイッチの抑制のための Motion SORT の提案

### 1. はじめに

複数物体追跡 (Multi Object Tracking, MOT) は動画シーケンス中の複数の物体を識別・追跡するコンピュータビジョンにおけるタスクの一つである。近年の深層学習に基づく Tracking-by-detection の代表的な手法として Deep SORT(Simple Online Realtime Tracker)[1] が挙げられる。Deep SORT をはじめとした MOT モデルにおける一般的な課題は、ID スイッチと呼ばれる同一物体の ID が誤った ID 割り当てによって切り替わってしまうことである。ID スイッチは、特に密集度の高いシーンや急激な物体の動きを含む動画シーケンスにおいて発生する。密集度が高くなった場合、外観のオクルージョンが発生するため外観特徴による物体の識別が困難となる。また、位置特徴に関しても物体の相互作用によって移動の不確実性が増加し、移動位置の正確な予測が困難となる。

本研究では、人が密集した環境下での MOT における ID スイッチ発生抑制を目的とし、Deep SORT を拡張した Motion SORT を提案した。Motion SORT は2つの主要な提案手法からなる。1つ目に物体の移動量の情報を考慮したマッチングモジュールである Motion matching を提案した。次に、複数物体の相互作用を考慮した軌道予測モデルである Social-LSTM(Long Short-Term Memory)[2] を採用した社会性を考慮した予測モジュールを提案した。

### 2. 提案手法 (Motion SORT)

1つ目の提案である Motion matching の目的は、適切にマッチングされなかった割り当て猶予中のトラックを追跡中のトラックと結びつけることで、本来は同一であるトラックを単一のトラックとして修復することにより ID スイッチの発生を抑制することである。猶予中のトラックとは、新たに検出された物体を新規に追跡する際に、最初の3フレームの間だけ暫定的なトラックとして追加されたトラックのことである。

2つ目の提案は、複数人物が持つ移動の社会性を考慮した非線形な位置予測である。人の社会性を考慮した位置予測を行うため、位置予測モジュールとして Social-LSTM を採用した。Social-LSTM は人の社会性に基づく相互作用を考慮した軌道予測モデルである。Social-LSTM は各 LSTM に物体を対応させ、入力として過去の軌道を与え、出力として予測位置を出力させる。LSTM は予測の過程で近傍の物体に対応する LSTM と情報共有を行うことで社会性を考慮した予測を行うことが可能となる。

提案手法では、上記の2つの主要な提案モジュールの他に2つの変更を行った。まず、BB サイズの急激な変化の抑制機構を加えることである。加えて、外観特徴マッチングにおいて割り当ての重みづけに適用される閾値を文献 [1] で用いられていたものより高く設定することである。図1に示すように、Motion matching モジュールは外観特徴マッチングモジュールの下流に存在する。外観特徴マッチングにおいて信頼性が

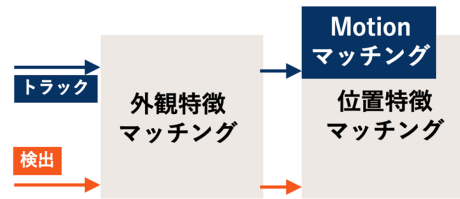


図1 マッチング方式の概要

表1 ベースモデルとの比較実験の結果

	IDSW	MOTA
ベースモデル	172	30.5
提案手法	164	30.6

低いマッチングが行われることを避け後段の Motion matching で信頼性の高いマッチングを行うことが可能となる。

### 3. 評価実験

実験には、MOT の標準的なベンチマーク用のデータセットである MOT16 Challenges データセット [3] を用いる。評価指標として、MOTA (Multiple Object Tracking Accuracy) と IDSW (ID SWitch) を使用する。MOTA は総合的な追跡精度を測る指標であり、IDSW はベンチマーク全体における ID スイッチの回数の合計値である。この実験では、提案手法のベースモデルである Deep SORT[1] との比較を行った。

ベースモデルとの比較結果を表1に示す。結果より、両方のメトリクスに関して提案手法はベースモデルよりも高精度に追跡できていることを確認した。

### 4. おわりに

本研究では、Deep SORT[1] をベースとして、位置特徴マッチングを改良した Motion SORT を提案した。本研究の提案手法は、MOT16 Challenges データセット [3] において、ベースモデルの Deep SORT に対して MOTA と ID スイッチの削減の観点から性能が向上したことを示した。

### 参考文献

- [1] N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric," Proc. of ICIP, 2017, doi:10.1109/ICIP.2017.8296962.
- [2] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," Proc. of CVPR, 2016, doi:10.1109/CVPR.2016.110.
- [3] A. Milan, L. Leal-Taixé, I. D. Reid, S. Roth and K. Schindler, "MOT16: A Benchmark for Multi-Object Tracking," Arxiv, 2016, <https://doi.org/10.48550/arXiv.1603.00831>.