I-013

Attention 機構を用いた道路交通流の複数視点観測
Multi-viewpoint observation of road traffic using Attention mechanism

藤原 友[†]　　全 炳東[†‡]
Tomo Fujiwara[†]　　Heitoh Zen[†‡]

1. はじめに

複数物体追跡（Multiple Object Tracking）では、物体同士が重なることにより、追跡が途絶すること（オクルージョン）が問題となっている。本研究では、特に隠蔽がよく発生する交差点において、カメラ 2 台を用いてカメラ間の連携を取ることにより、追跡が途絶しないアルゴリズムを提案する。具体的には、カメラ 2 台を交差点の対向方向に設置し、それぞれのカメラで追跡結果をリアルタイムで取得する。その最中、2 つのカメラで同時刻に映る同一車両を対応付けすることにより、片方のカメラにおいて追跡が途絶しても、もう片方のカメラで追跡を継続することで全体として途切れない追跡を目指している。

2. 提案手法

本研究では、図 1 のように 2 台のカメラを交差点の対向方向に設置し、それぞれのカメラで複数物体追跡を行う（以下、カメラ内対応付け）。また、Triplet Network[1]を用いたアルゴリズムにより、カメラ間での対応付けを行う（以下、カメラ間対応付け）。

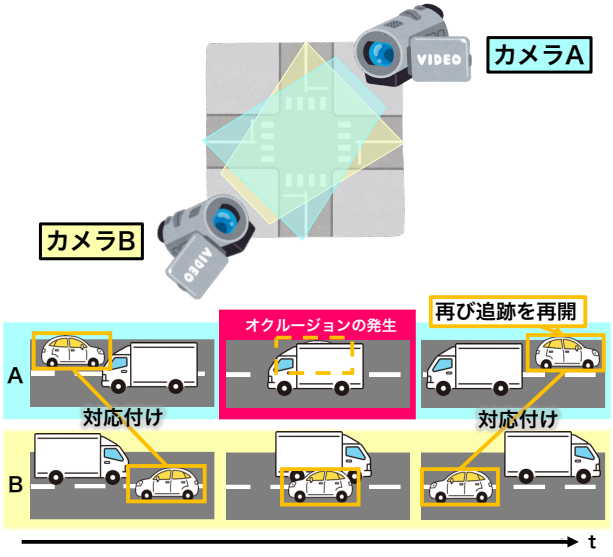


図 1　提案手法概要

カメラ内対応付けには ByteTrack[2]を用いる。それぞれのカメラにおいてこの ByteTrack を用いて追跡結果を得る。得られた追跡結果をカメラ間対応付けに用いる。

カメラ間対応付けでは、それぞれの追跡結果をもとに Triplet Network を用いて同一車両同士を対応付けする。Triplet Network は、ByteTrack で用いる物体検出器 YOLOX[3]により得られた車両の Bounding Box を入力し、写っている車両の特徴ベクトルを出力する。車両ごとに得られた特徴ベクトル同士でユークリッド距離を計算し、特徴ベクトルの距離が近い車両同士を対応付けする。

Triplet Network のモデル構造の概要は図 2 の通りである。また、図 2 に示す body、attention branch、head の各部分の入出力データの形状は表 1 に示している。本研究における実装では、通常の CNN の構造に加えて、attention branch[4]を導入している。車両画像を入力する際、背景情報を除去し純粋な特徴ベクトルを得るためである。

推論時は、body と呼ぶ CNN で入力画像を特徴マップに変換し、attention branch で body から得られた特徴マップから attention map を計算する。その後、body から得た特徴マップに attention map を乗算し、head と呼ぶ CNN で 1024 次元の特徴ベクトルへ変換する。

学習時は、Triplet Loss[5]を用い、NVIDIA AI City Challenge[6]の車両の種類が 184、画像総数 43794 のデータセットを用いる。各車両の様々な方向から撮影された画像を用いるため、本研究のような異なる方向から映る車両の画像であっても適切に対応付けすることができるような学習が可能である。

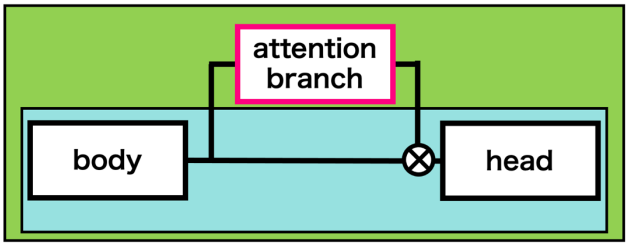


図 2　Triplet Network 概要

表 1 Triplet Network を順伝播するデータ形状

name	Input Size	Output Size
body	[3, 256, 256]	[512, 64, 64]
attention branch	[512, 64, 64]	[1, 64, 64]
head	[512, 64, 64]	[1024, 1, 1]

3. 実験

表 2 に示す通りの撮影環境で動画を撮影し、アノテーションを施した。この動画を用いて、精度評価を行った。

表 2　動画の撮影環境

場所	国道16号線穴川交差点（千葉市稲毛区）
カメラの位置設定	信号機と同じ高さ
環境	晴れの昼
動画の長さ	9400フレーム（5分13秒）

† 千葉大学大学院 融合理工学府 数学情報科学専攻
Chiba University, Graduate School of Science and Engineering
‡ 千葉大学 情報戦略機構
Chiba University, Digital Transformation Enhancement Council

3.1 実験結果

表 2 の動画にて本アルゴリズムを適用し、精度を求めた。表 3 は、図 2 の attention branch を用いた場合と attention branch を用いなかった場合で、MOTA[7]と IDF1[8]を比較した結果である。branch を用いなかった場合と比較して、若干の精度向上が見られ attention の効果が確認できた。

図 3 は、ID15 番の車において、長いオクルージョンが発生する前後でも、ID が切り替わらず適切に追跡が継続できている様子を示している。

表 3　MOTA と IDF1 の結果

カメラ名	branch有無	MOTA	IDF1
カメラA	branch無	78.7	79.5
	branch有	**78.8**	**80.5**
カメラB	branch無	75.8	82.7
	branch有	75.8	**83.5**

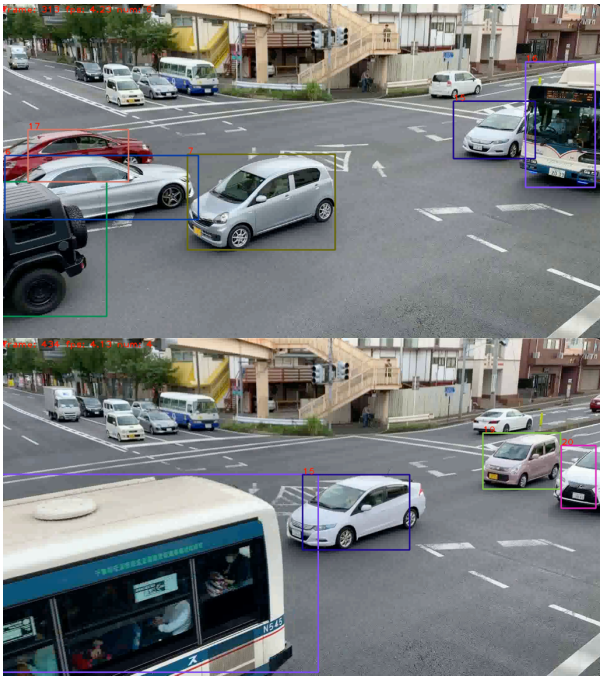
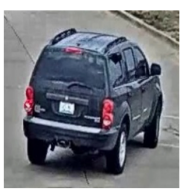
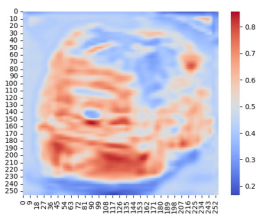


図 3 ID15（上段）オクルージョン前（下段）オクルージョン後

Triplet Network の attention map を入力画像と同じサイズに拡大し可視化した例が下の図 4 である。attention branch は、attention map を乗算することにより精度が向上するように学習されたネットワークである。背景部分がマスクされ、車両部分を重視した特徴ベクトルの算出が期待できることが確認できる。

図 4 attention map の可視化





4. 考察

図 3 のように片方のカメラで完全に車両が隠れても、もう一方のカメラの情報との対応付けができれば、オクルージョン発生時にも追跡を継続できる。MOTA や IDF1 の精度が十分かどうかは応用に依るので、さらに調査が必要である。

さらに精度を改善するための検討事項について述べる。追跡に失敗する原因として、両方のカメラで同時にオクルージョンが発生する事象が挙げられる。このような事象に対応するためには、カメラ内対応付けの改良が必要である。どちらか片方のカメラ内対応付けに成功する時間帯が増えれば、提案した方法で追跡は継続できる。

そこでカメラ内対応付けに Bounding Box の特徴ベクトルだけを用いるのではなく動きベクトルを併用することが考えられる。車両が隠れている間に大きく進路を変えない場合は、隠れる直前までの動きから隠ぺい中やその後の位置を予測できる。この考えに基づく手法は別途開発している。

5. おわりに

本研究では、2 台のカメラを用いてオクルージョンに対応する複数物体追跡手法を提案した。実験では片方のカメラでオクルージョンが発生した場合でも、追跡を継続できることを確認した。さらに精度を向上させるためには、動き情報などを利用したカメラ内対応付けの改良が考えられる。

参考文献

[1] Hoffer Elad, Ailon Nir, "Deep metric learning using Triplet network", ICLR Workshop (2015)

[2] Zhang Yifu, Sun Peize, Jiang Yi, Yu Dongdong, Weng Fucheng, Yuan Zehuan, Luo Ping, Liu Wenyu, Wang Xinggang, "ByteTrack: Multi-Object Tracking by Associating Every Detection Box", ECCV (2022)

[3] Ge, Zheng; Liu, Songtao; Wang, Feng; Li, Zeming; Sun, Jian, "YOLOX: Exceeding YOLO Series", Workshop on Autonomous Driving at CVPR, (2021)

[4] Fukui Hiroshi, Hirakawa Tsubasa, Yamashita Takayoshi, Fujiyoshi Hironobu, "Attention Branch Network: Learning of Attention Mechanism for Visual Explanation" In CVPR, (2019)

[5] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu. "Learning Fine-grained Image Similarity with Deep Ranking." In CVPR, (2014)

[6] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stan Birch- field, 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 CVPR, (2019)

[7] Keni Bernardin and Rainer Stiefelhagen. Evaluating multiple object tracking performance: The CLEAR MOT metrics. Eurasip Journal on Image and Video Processing, Vol. 2008, , 2008.

[8] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. Lecture Notes in Computer Science (including subseries Lecture Notes in Arti- ficial Intelligence and Lecture Notes in Bioinformatics), Vol. 9914 LNCS, No. c, pp. 17–35, 2016.