# イベントカメラを用いた
# 悪環境に対して頑健かつ相対速度に依存しない物体検出手法

○田倉 竜也（東京大学），井倉 幹大（Italian Institute of Technology，東京大学）
安 琪（東京大学），山下 淳（東京大学）

# Robust Event-based Object Detection
# Invariant to Adverse Environments and Relative Velocity

○ Tatsuya TAKURA（The University of Tokyo）
Mikihiro IKURA（Italian Institute of Technology, The University of Tokyo）
Qi AN（The University of Tokyo）
Atsushi YAMASHITA（The University of Tokyo）

Abstract: Object detection is an important task for autonomous mobile robots to recognize the environments and to plan their paths. Unlike RGB cameras, event cameras can be used in high dynamic range but cannot capture relatively stationary objects. In this paper, we propose an algorithm to detect both dynamic and relatively stationary objects by estimating the velocities of bounding boxes and retaining the objects considered stationary. An experiment indicates that the proposed method can detect not only dynamic but also relatively stationary object not detected by the conventional methods.

## 1. 序論

自動運転技術において，周囲に存在する他の車両や歩行者，信号などを認識する必要がある．そのためには，画像内に映る複数の物体の位置，およびそれらが何であるかを推定する技術が求められる．このような技術は物体検出（Object Detection）と呼ばれ，自動運転技術[1] をはじめとして，多くの応用先が存在する非常に重要な技術となっている．

物体検出技術を実環境に適用するにあたって，周囲の物体との相対速度が大きい状況や明暗差の激しい環境での利用が求められる．RGB カメラは絞りや感度，シャッタースピードなどが画素全体で統一されているため，ハイビームで照らされている道路の横を人が歩いているような，同じ画像内で明暗差が激しい環境ではダイナミックレンジが足りず，黒潰れや白飛びが発生する．したがって，RGB カメラを用いた物体検出手法は明暗差が激しい環境での運用が難しい．

一方でイベントカメラは，各画素が独立に，輝度の対数値の変化が一定量に達したタイミングで信号（イベントと呼ばれる）を出力するセンサであり，小さい光量では小さい変化を，大きい光量では大きい変化を，それぞれ同時に捉えることができる．その結果としてハイダイナミックレンジを実現しており，明暗差が激しい環境であっても物体を検出することができる．また，高い時間分解能を持つため，周囲の物体との相対速度が大きい状況での利用も可能である．**Fig. 1** はイベントカメラの出力を可視化したものである．一定時間内に生じたイベントに対し，対応する画素を順に，輝度の変化が正ならば白く，負ならば黒く塗り潰した．

しかし，イベントを生成するために必要となる輝度の対数値の変化量の閾値は事前に決めるため，速度の大きさによって発生するイベント数が異なる．そのため，カメラとの相対速度が非常に小さい物体は，発生するイベント数が非常に少なくなる．（以後，消失物体，Vanished Object と呼ぶ）．**Fig. 2** 中の黄色の枠中には車両が存在しているが，カメラとの相対速度が非常に小さいためにイベントが殆ど生じていない．そのため，センサ出力が速度依存となるイベントカメラを用いた従来の物体検出手法[2][3] では低相対速度の物体の検出を行うことが難しい．時間方向の依存関係を考慮することで，消失物体の検出が可能となる手法が研究されている[4] が，消失物体の前後を別の物体が通過したのちに，検出が途切れるという課題が存在する．このことは，信号待ちの最中などに，停止している自身の車両と他の物体の間を車両や自転車などが通過するような状況で問題となる．停止している前方の車両や自転車の検出が途切れることとなり，結果として加減速の判断に影響が生じ，加速後急停止のような危険な制御を行うなどの恐れがある．

以上を踏まえ，本研究では，消失物体を継続的に検出可能な，イベントカメラを用いた物体検出手法の構築を目的とする．

## 2. 関連研究

### 2.1 RGB カメラを用いた物体検出手法

RGB カメラを用いた物体検出については，以前から多くの研究がされてきたが，深層学習の発展により，深層ニューラルネットワーク（DNN），特に CNN によって画像から優れた特徴量を得ることができるようになり，大幅な発展を遂げた．深層学習の流行以降はじめに主流となったのが，物体が存在する領域を推論する領域提案（Region Proposal）を行い，その後に領域中の物体が何であるかを推論する物体認識を行


SY0009/24/0000 - 0822 

**Fig. 1** Visualized events

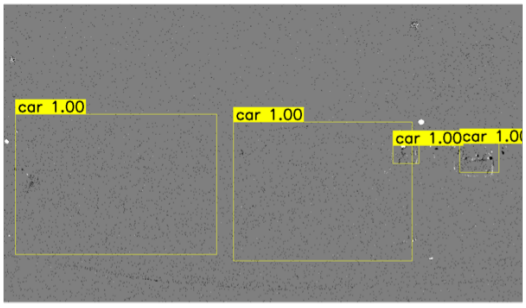


**Fig. 2** “Vanished” objects with their labels

う 2 ステージ型の手法である．この例としては Faster R-CNN[5] が挙げられる．2 ステージ型の手法は 2 つの処理が直列に行われるため，推論に時間がかかるという課題があった．続いて登場したのが，領域提案と物体認識を同時に行うことで推論の高速化を図った，1 ステージ型の手法である．代表的な例としては You Only Look Once（YOLO）[6] が挙げられる．YOLO では画像全体を複数のマス目に分割し，それぞれのマス目に対してバウンディングボックスの中心と形状，信頼度および各クラスの確率を同時に推論する．YOLO には多くの後継モデル[7][8] が存在しており，様々な手法を取り入れて精度や推論速度の向上がされてきた．また，自然言語処理分野で登場した Transformer を利用した手法[9] も提案されている．

### 2.2 イベントカメラを用いた物体検出手法

イベントカメラに対する物体検出手法として，DNN ベースの手法[2]，スパイキングニューラルネットワーク（SNN）ベースの手法[10]，グラフニューラルネットワーク（GNN）ベースの手法[3] が提案されている．しかし，SNN ベースの手法はモデルの最適化が困難である．また，DNN，GNN ベースの手法は限られた時間の情報のみを利用するために長時間の依存関係を考慮することが困難であるため，特に相対速度が小さい物体に対する検出精度が悪いという課題がある．これらに対し，時空間方向の依存関係を保持するメモリ機構を組み込むことで検出精度を向上した手法が複数提案されており，ConvLSTM を用いた RED[11] や RVT[4]，Attention ベースのメモリ機構を用いた HMNet[12] などが挙げられる．このうち RVT は，相対速度が非常に小さく，消失物体を検出することが可能であるが，消失物体に他の物体が一時的に重なると，その後は検出が途切れるという課題が存在する．

## 3. 提案手法

### 3.1 検出結果の表現

物体検出において，検出結果は Fig. 2 中の黄色の枠のようなバウンディングボックス（Bounding Box，BB）として表現される．この BB に対して，中心座標 $(x, y)$，縦横のサイズ $(w, h)$ を状態変数とする．

### 3.2 複数物体追跡手法

検出した物体に対し，複数フレーム間で追跡して固有の ID を振るタスクを複数物体追跡（Multi-Object Tracking, MOT）と呼ぶ．MOT 手法は大きく分けて，状態（追跡中の物体の状態変数）の推定，状態と観測（物体検出手法の出力）の対応づけ，状態の更新の 3 つの段階からなる．状態の推定では，1 つ前の時刻の状態から現在の時刻の状態を推定する．状態と観測の対応づけでは，BB の重なり具合（Intersection over Union, IOU）や BB 内のテクスチャの類似度などを基準として両者の紐付けを行う．状態の更新では，対応づけられた観測の情報を用いて状態を更新する．

### 3.3 提案手法の概要

本手法は，MOT と消失物体の管理によって消失物体の判定，保持を行う．ある時刻において，ある物体が検出器で検出されなかったとき，その物体の追跡が途切れることとなる．ここで，MOT によって BB の相対速度を推定し，その情報を用いてその物体が消失物体であるかの判定を行う．消失物体である場合は最後の追跡結果を保持，消失物体でない場合は追跡を終了する．消失物体として追跡中の物体と検出器の出力の和集合が本手法の最終的な出力となる．

具体的な流れを **Fig. 3** に示す．物体検出器の出力として物体 A，B，追跡中の物体として物体 1，2，3，消失物体として追跡中の物体として物体 1’，2’ がそれぞれ存在しているとする．ここで，追跡中の物体のうち，物体 2 が消失物体である．

追跡中の物体のうち対応する物体検出器の出力（物体 A）が存在するもの（物体 1）は，引き続き追跡中の物体として扱う（Fig. 3 の Tracked Objects）．追跡中の物体のうち対応する物体検出器の出力が存在しないもの，すなわち検出が途切れた物体（Fig. 3 の物体 2，3）のうち，消失物体であると判定されたもの（物体 2）は，消失物体として追跡を続け（Fig. 3 の Vanished Objects），消失物体ではなく遮蔽により検出が途切れたと判断したもの（物体 3）は追跡を中断する（Fig. 3 の Removed Objects）．消失物体として追跡中の物体のうち，その位置から再度対応する物体検出器の出力（物体 B）が検出されたもの（物体 2’）は，消失物体としての扱いを終え，通常の追跡中の物体（Tracked Objects）として扱う．

### 3.4 消失物体の判定

MOT によって複数フレーム間で BB の対応づけが行えるため，BB の状態変数の時間微分，すなわち BB の画像内での移動速度 $\dot{x}, \dot{y}$，変形速度 $\dot{w}, \dot{h}$ の推定が可能となる．本手法では，画像の縦横のサイズ $(W, H)$，閾値 $v_{th}$ に対し，

$$\max\left(\frac{|\dot{x}|}{W}, \frac{|\dot{y}|}{H}, \frac{|\dot{w}|}{W}, \frac{|\dot{h}|}{H}\right) \leq v_{th}$$

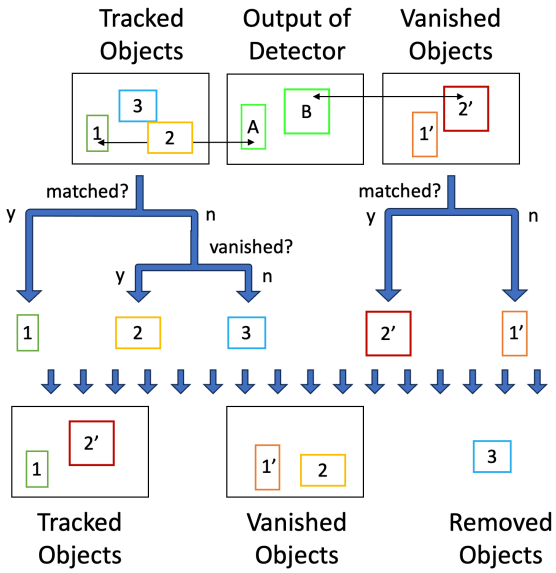


**Fig. 3** Architecture of the proposed method

が成立し，かつ $N_{th}$ フレーム以上追跡が続いている場合に消失物体であると判断する．

## 4. 実験

提案手法により検出漏れが減るかどうかの定量的検証，および消失物体の継続的な検出が可能であるかどうかの定性的検証を行った．

### 4.1 実験方法

物体検出器として RVT[4] を，検出結果の紐付けを行うための MOT 手法として ByteTrack[13] をそれぞれ採用した．検証用のデータとして，1Mpx データセット[11] を用いた．1Mpx は走行中の自動車に搭載されたイベントカメラで撮影された映像で構成されたデータセットである．カメラの解像度は 1280 × 720 であり，60Hz の周期で Car, Pedestrian, Bicycle の 3 つのクラスに対してバウンディングボックスのラベリングが行われている．イベントカメラと RGB カメラの両方を車に載せて運転中の車外の光景を同時に撮影し，RGB カメラの情報を利用してイベントカメラの出力にラベリングを施しているため，信号が極端に少ないような低相対速度の物体にもラベリングが行われていることが特徴である．本実験においては，10Hz で検出を行い，対応する正解データとの比較によって評価を行った．定量的な評価指標として，Recall（= TP / (TP+FN)）を採用した．Recall が大きいほど TP に対する FN の割合が少ない，すなわち検出漏れが少ないことを意味する．

### 4.2 実験結果

定量的な実験結果を **Table 1** に示す．IoU=0.5（0.75）は正解ラベルとの IoU が 0.5（0.75）以上であるときに TP と判定することを意味する．また，class 0，1，2 はそれぞれ Car, Pedestrian, Bicycle を表す．すべてのケースにおいて，従来手法と比較して提案手法の方が Recall が大きいことが確認できる．

また，消失物体に着目して定性的な評価を行うため，検出結果を可視化した．その例を **Fig. 4**，**Fig. 5** に示す．Fig. 4 は正しく消失物体を検出できた例であり，それぞれ正解ラベル（GT）（上）, RVT 単体（中），提案手法（下）による検出結果となっている．Fig. 4 下の左

**Table 1** Recall of each method

| IoU | class | RVT | Proposed Method |
|---|---|---|---|
| 0.5 | 0 | 0.8176 | 0.8201 |
| | 1 | 0.8368 | 0.8393 |
| | 2 | 0.9093 | 0.9110 |
| 0.75 | 0 | 0.4887 | 0.4894 |
| | 1 | 0.6020 | 0.6025 |
| | 2 | 0.7529 | 0.7541 |

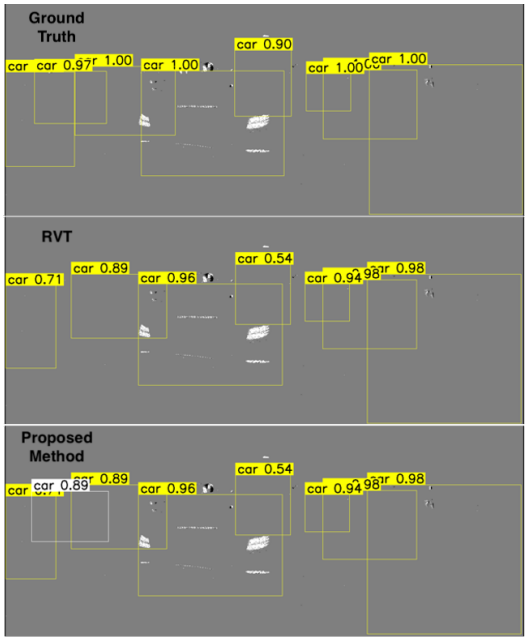


**Fig. 4** Detected objects and Ground Truth

下の白い BB が消失物体を表している．BB 内の領域でイベントが発生していないこと，対応する BB が GT では存在するが RVT 単体では存在しない（=検出されなかった）ことが確認できる．消失物体はその位置から同じ物体が検出されるまで 120 フレームの間存在した．Fig. 5 は実際には物体が存在しないにも関わらず，消失物体として誤って検出したものである．左下の白い BB がその例であり，最終フレームまで白い BB が消えることはなかった．この理由として，消失物体であると判定された物体が動き出した際に，その物体が他の物体の後方に存在したために検出されず，結果として消失物体と判定された BB が残り続けてしまったためであると考えられる．

## 5. 結論

本研究では，消失物体を継続的に検出可能な，イベントカメラを用いた物体検出手法を構築した．また，

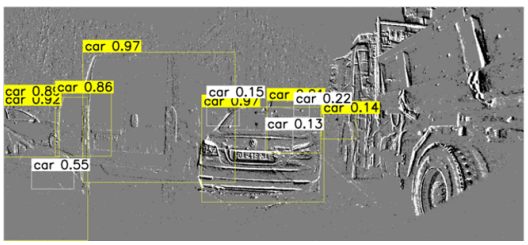

**Fig. 5** Vanished objects remained as False Positive

提案した手法によって実際にそのような物体を検出することが可能であることを確認した．今後の展望としては，Fig. 5 のような消失物体の誤検出を可能な限り防ぐことが挙げられる．このようなことが頻発すると，行動計画の際にスタックしてしまうようなことが考えられるためである．

## 謝辞

本研究の一部はソニーセミコンダクタソリューションズ株式会社から支援を受けて実施したものである．ここに謝意を表する．

## 参考文献

[1] H. Y. Yatbaz, M. Dianati, K. Koufos, and R. Woodman: “Run-Time Introspection of 2D Object Detection in Automated Driving Systems Using Learning Representations.” *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 6, pp. 5033–5046, (2024).

[2] Z. Jiang, P. Xia, K. Huang, W. Stechele, G. Chen, Z. Bing, and A. Knoll: “Mixed Frame-/Event-Driven Fast Pedestrian Detection.” *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA)*, pp. 8332–8338, (2019).

[3] S. Schaefer, D. Gehrig, and D. Scaramuzza: “AEGNN: Asynchronous Event-based Graph Neural Networks.” *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12361–12371, (2022).

[4] M. Gehrig and D. Scaramuzza: “Recurrent Vision Transformers for Object Detection with Event Cameras.” *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13884–13893, (2023).

[5] S. Ren, K. He, R. Girshick, and J. Sun: “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, (2017).

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi: “You Only Look Once: Unified, Real-Time Object Detection.” *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, (2016).

[7] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun: “YOLOX: Exceeding YOLO Series in 2021.” *arXiv preprint arXiv:2107.08430*, (2021).

[8] R. Varghese and S. M.: “YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness.” *Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, pp. 1–6, (2024).

[9] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko: “End-to-End Object Detection with Transformers.” *Proceedings of the 16th European Conference on Computer Vision (ECCV)*, pp. 213–229, (2020).

[10] L. Cordone, B. Miramond, and P. Thierion: “Object Detection with Spiking Neural Networks on Automotive Event Data.” *Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, (2022).

[11] E. Perot, P. de Tournemire, D. Nitti, J. Masci, and A. Sironi: “Learning to Detect Objects with a 1 Megapixel Event Camera.” *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)*, pp. 16639–16652, (2020).

[12] R. Hamaguchi, Y. Furukawa, M. Onishi, and K. Sakurada: “Hierarchical Neural Memory Network for Low Latency Event Processing.” *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22867–22876, (2023).

[13] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang: “ByteTrack: Multi-object Tracking by Associating Every Detection Box.” *Proceedings of the 17th European Conference on Computer Vision (ECCV)*, pp. 1–21, (2022).