Причинная кросс-видовая память для онлайн-отслеживания трехмерных объектов

Ц. Ли*, Т. Е. Мамонова‡

В DeepFusionMOT успешный Stage-4 обновляет счетчики и жизненный цикл, но не сохраненную двумерную память, которая затем используется при недоступной проекции или пропуске лидара. Предложен причинный контур: P3DR использует проекцию прогнозируемого трехмерного параллелепипеда как свидетельство для Stage-4, а C2MC записывает принятое двумерное состояние в память для последующих кадров. Относительно DeepFusionMOT на KITTI Tracking Val11 HOTA повышается с 79,876% до 80,385%, sAMOTA — с 93,01% до 95,42%, FP сокращается с 294 до 110. В контроле при отключенном RRC двумерная ветвь получает штатный или спроецированный прямоугольник того же трехмерного детектора, и выходные файлы последовательностей совпадают с базовой моделью.

Ключевые слова: трехмерное многообъектное отслеживание, камера и лидар, кросс-видовая память, онлайн-ассоциация.

1. Введение

Трехмерное многообъектное отслеживание необходимо для автономного вождения. DeepFusionMOT объединяет детекции камеры и лидара на четырех этапах ассоциации [1]; CrossTracker использует двустороннюю кросс-коррекцию [2]; 3DMOTFormer применяет графовый Transformer к связям траекторий и детекций [3].

В четырехуровневой схеме после восстановления camera-only и трехмерной траекторий обновляются счетчики и жизненный цикл, но двумерная память трехмерной траектории остается устаревшей. При пропуске

* *Ли Цзюмин* — аспирант, Томский политехнический университет, Россия, Томск, e-mail: czyumin1@tpu.ru.

Li Jiuming — Postgraduate Student, Tomsk Polytechnic University, Tomsk, Russia.

‡ *Мамонова Татьяна Егоровна* — канд. техн. наук, доцент, доцент отделения интеллектуальных систем, и. о. заведующего кафедрой — руководителя ОИС на правах кафедры ИШИнЭС, Томский политехнический университет, Россия, Томск, e-mail: stepte@tpu.ru.

Mamonova Tatiana Egorovna — Cand. Tech. Sci., Assoc. Prof., Intelligent Systems Division; Acting Head of Department, School of IT and Intelligent Systems, Tomsk Polytechnic University, Tomsk, Russia.

Опубликовано на условиях лицензии CC-BY 4.0
Published under CC-BY 4.0

лидара проекция текущего трехмерного прогноза служит геометрическим ориентиром, а двумерная траектория — свидетельством идентичности. После успешной ассоциации это наблюдение должно записываться в память; при одноисточниковых наблюдениях такого разрыва нет.

Предлагаются проекционная трехмерная реидентификация (P3DR) и причинная передача двумерной памяти (C2MC). P3DR дополняет Stage-4 проекцией текущего прогноза, а C2MC синхронизирует состояние без изменения трехмерного положения и без будущих кадров.

2. Метод

Состояние трехмерной траектории $\mathbf{X}_i^t = [x_i^t, y_i^t, z_i^t, \theta_i^t, l_i^t, w_i^t, h_i^t]$ прогнозируется фильтром Калмана. Восемь углов прогнозируемого параллелепипеда переводятся в систему camera_02, проецируются матрицей P_2, после чего вычисляется охватывающий прямоугольник $b_{\mathrm{proj},i}^t$. Проекция допустима, если глубина всех восьми углов не меньше 0,1 м. Для двумерной траектории j после этапов 1–3 используется ее текущий фильтрованный прямоугольник $u_j^t = \mathrm{Box2D}(S_j^t)$; исходная детекция в формуле (1) не применяется.

$$s_{ij}^t = (1-\alpha)\,\mathrm{IoU}(u_j^t, m_i^{t-1}) + \alpha\,\mathrm{IoU}(u_j^t, b_{\mathrm{proj},i}^t), \tag{1}$$

где m_i^{t-1} — сохраненный фильтрованный прямоугольник двумерной траектории. При допустимой проекции $\alpha = 1{,}0$, и основным свидетельством служит $b_{\mathrm{proj},i}^t$. При недопустимой проекции $\alpha = 0$, поэтому Stage-4 использует сохраненную память m_i^{t-1}. Пары выбираются взаимно-однозначным назначением при пороге $\tau_{2\mathrm{D}} = 0{,}5$.

Для каждой принятой пары (i, j) память обновляется как

$$m_i^t = u_j^t. \tag{2}$$

Обновленная память используется в последующих Stage-4 при недопустимой проекции и при временном отсутствии объединенного или LiDAR-наблюдения; новое объединенное наблюдение повторно инициализирует ее. Метод не изменяет трехмерный центр, размеры или ориентацию. Здесь N_t — число несопоставленных трехмерных траекторий, K_t — число доступных двумерных траекторий на входе Stage-4. Дополнительная стоимость составляет $O(N_t)$ на проекцию, $O(N_t K_t)$ на IoU и $O(\max(N_t, K_t)^3)$ на назначение; обучаемые параметры отсутствуют.

3. Результаты

Эксперименты выполнены для класса Car на KITTI Tracking Val11 по последовательностям 0001, 0006, 0008, 0010, 0012, 0013, 0014, 0015, 0016,

0018 и 0019. Используются PointRCNN и RRC [1]; все методы работают онлайн. В TrackEval Kitti2DBox оценивается текущая фильтрованная траектория u_j^t, а $b_{\text{proj},i}^t$ служит только свидетельством Stage-4. Оценка включает HOTA, IDSW, sAMOTA, AMOTA и FP [4].

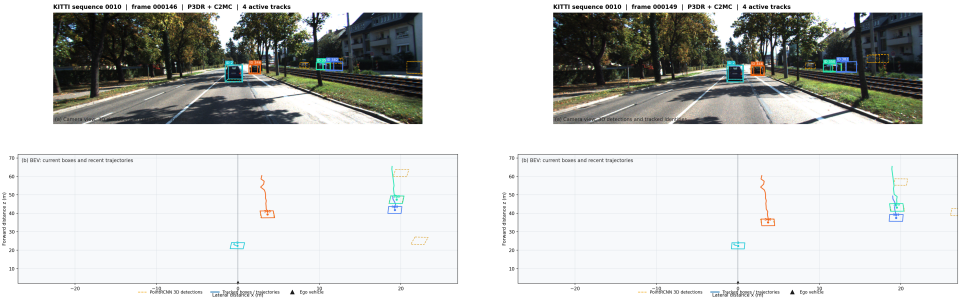


Рис. 1: Качественные результаты для последовательности 0010 (кадры 146 и 149): проекции 3D-обнаружений, отслеживаемые 3D-траектории и BEV-история.

Относительно DeepFusionMOT полная модель повышает HOTA с 79,876 до 80,385, sAMOTA — с 93,01 до 95,42 и AMOTA — с 45,43 до 47,73, а FP сокращает с 294 до 110. P3DR отдельно уменьшает IDSW с 20 до 18, тогда как C2MC повышает HOTA, но временно увеличивает IDSW до 26. Совместная конфигурация дает лучшие HOTA и sAMOTA: P3DR стабилизирует текущую повторную ассоциацию, а C2MC сохраняет обновленное состояние для следующих кадров.

В 9 из 11 последовательностей HOTA возрастает; приросты в 0014 и 0010 равны 1,447 и 1,271, а снижение в 0016 — 0,105 процентного пункта.

Таблица 1: Абляция на KITTI Tracking Val11

Метод	HOTA	sAMOTA	AMOTA	FP
DeepFusionMOT	79,876	93,01	45,43	294
P3DR	79,913	95,37	47,71	115
C2MC	80,307	93,02	45,43	128
P3DR + C2MC	**80,385**	**95,42**	**47,73**	**110**

4. Заключение

Предложен причинный контур кросс-видовой памяти для онлайн-трехмерного отслеживания. P3DR использует текущую проекцию трехмерного прогноза как геометрическое свидетельство, а C2MC сохраняет согласованное двумерное состояние после успешной повторной ассоциации.

Эксперименты на валидационном разбиении KITTI Tracking показывают, что совместное использование компонентов улучшает двумерную и трехмерную согласованность траекторий. Основные ограничения связаны с одним классом объектов, фиксированным набором валидационных последовательностей и отсутствием оценки на официальном тестовом наборе.

Список литературы

[1] X. Wang, C. Fu, Z. Li, Y. Lai, J. He, "DeepFusionMOT: A 3D Multi-Object Tracking Framework Based on Camera–LiDAR Fusion with Deep Association", *IEEE Robotics and Automation Letters*, **7**:3 (2022), 8260–8267. DOI: 10.1109/LRA.2022.3187264.

[2] L. Gu, X. Yan, W. Wang, H. Chen, D. Zhu, L. Nan, M. Wei, "CrossTracker: Robust Multi-modal 3D Multi-Object Tracking via Cross Correction", 2024, arXiv: 2411.18850.

[3] S. Ding, E. Rehder, L. Schneider, M. Cordts, J. Gall, "3DMOTFormer: Graph Transformer for Online 3D Multi-Object Tracking", *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, 9750–9760. DOI: 10.1109/iccv51070.2023.00897, arXiv: 2308.06635.

[4] Z. Yan, S. Feng, X. Li, Y. Zhou, C. Xia, S. Li, "S3MOT: Monocular 3D Object Tracking with Selective State Space Model", 2025, arXiv: 2504.18068.

Статья поступила 9 ноября 2026 г.

Causal Cross-View Memory for Online 3D Multi-Object Tracking

J. Li, T. E. Mamonova

In DeepFusionMOT, Stage-4 updates counters but not the stored 2D memory, which is reused when projection is unavailable or LiDAR misses occur. P3DR projects the predicted 3D box as evidence; C2MC writes the accepted 2D state for later frames. On KITTI Tracking Val11, relative to DeepFusionMOT, HOTA rises from 79.876% to 80.385%, sAMOTA from 93.01% to 95.42%, and FP falls from 294 to 110. In a same-source control (RRC disabled; native or projected boxes from one 3D detector), sequence outputs match the baseline.

Keywords: 3D multi-object tracking, camera–LiDAR fusion, cross-view memory, online association.

Received on November 9, 2026