

УДК 004.932

## КЛАССИФИКАЦИЯ МЕТОДОВ ПРЕОДОЛЕНИЯ ОККЛЮЗИЙ В МУЛЬТИОБЪЕКТНОМ СОПРОВОЖДЕНИИ

**Никашов Илья Алексеевич,**

аспирант кафедры ВТиСУ Владимирского государственного университета имени А.Г. и Н.Г. Столетовых, г. Владимир, Россия, email: inikashov@vlsu.ru

### Аннотация

Рассмотрена проблема устойчивости мультиобъектного сопровождения в условиях взаимных перекрытий объектов. Окклюзии приводят к фрагментации траекторий, переключению идентификаторов и потере объектов, что существенно снижает качество работы алгоритмов сопровождения. Существующие обзоры систематизируют методы преодоления окклюзий по архитектурным признакам или этапам алгоритма, что затрудняет их сравнительную оценку применительно к конкретным условиям наблюдения. Предложена классификация, основанная на трёх признаках: типе используемой информации, горизонте принятия решений и способе представления неопределённости состояния объекта. По сочетанию этих признаков выделены пять групп методов: кинематическое рекуррентное оценивание, ассоциация по визуальному описанию, глобальная оптимизация во временном окне, вероятностное прогнозирование траекторий и учёт пространственного контекста. Обоснованность классификации подтверждена сопоставлением опубликованных результатов шести алгоритмов сопровождения объектов на наборе данных MOT17. Показано, что методы, комбинирующие несколько механизмов, обеспечивают наиболее стабильные значения метрик IDF1 и HOTA, тогда как наибольшее значение IDF1 достигнуто алгоритмом из группы ассоциации по визуальному описанию. Для каждой группы описаны предпосылки применимости, типичные источники ошибок и вычислительные ограничения, что позволяет обоснованно выбирать механизм в зависимости от характеристик целевой сцены.

**Ключевые слова:** сопровождение, окклюзия, классификация, ассоциация, дескрипторы, прогнозирование, контекст, MOT17, IDF1, HOTA

## CLASSIFICATION OF OCCLUSION HANDLING METHODS IN MULTI- OBJECT TRACKING

**Nikashov Ilya Alekseevich,**

postgraduate student of the Department of Computer Engineering and Control Systems, Vladimir State University named after A.G. and N.G. Stoletovs, Vladimir, Russia, email: inikashov@vlsu.ru

ABSTRACT

The problem of multi-object tracking robustness under mutual object occlusions is considered. Occlusions cause trajectory fragmentation, identity switches, and object loss, significantly degrading tracking algorithm performance. Existing surveys systematize occlusion handling methods by architectural features or algorithm stages, which complicates their comparative evaluation for specific observation conditions. A classification based on three criteria is proposed: the type of information used, the decision horizon, and the representation of object state uncertainty. Five method groups are identified by the combination of these criteria: kinematic recursive estimation, appearance-based association, global optimization over a temporal window, probabilistic trajectory forecasting, and spatial context modeling. The validity of the classification is confirmed by comparing published results of six tracking algorithms on the MOT17 dataset. It is shown that methods combining multiple mechanisms provide the most stable IDF1 and HOTA metric values, while the highest IDF1 value is achieved by an algorithm from the appearance-based association group. For each group, applicability conditions, typical error sources, and computational constraints are described, enabling informed selection of a mechanism based on target scene characteristics.

---

**Keywords:** tracking, occlusion, classification, association, descriptors, forecasting, context, MOT17, IDF1, HOTA

---

#### Актуальность

Окклюзии выступают одним из доминирующих факторов снижения качества мультиобъектного сопровождения (МОС). На стандартных наборах данных MOT17 [1] и MOT20 [2] значительная доля фрагментаций траекторий, переключений идентификаторов и потерь объектов обусловлена периодическим перекрытием целевых объектов другими элементами сцены. В периоды потери наблюдаемости нарушается непрерывность детекций и возрастает неопределённость пространственного положения объекта, что существенно усложняет ассоциацию новых измерений с ранее сформированными траекториями. Тематика окклюзий исследуется в обзорных работах по МОС, однако принципы систематизации методов в них существенно различаются. В обзоре [3] алгоритмы систематизированы по архитектурному признаку, но детальный анализ того, какой именно компонент обеспечивает робастность к окклюзиям и при каких условиях, не проводится. Специализированный обзор [4], посвящённый сопровождению пешеходов в условиях перекрытий, предлагает развёрнутую классификацию и вводит метрику ОНР для оценки устойчивости алгоритмов. Тем не менее предложенная систематизация ориентирована на конкретные детали программной реализации, что затрудняет сравнительный анализ применимости методов в различных сценариях. Несмотря на наличие эффективных решений для прогнозирования траекторий [5], мультикамерного сопровождения [6] и глубокой интеграции признаков [7], в современной литературе отсутствует системная классификация методов обработки окклюзий, учитывающая параметры целевой сцены. Научная новизна исследования заключается в разработке архитектурно-независимой типологии, позволяющей впервые формализовать критерии выбора механизмов восстановления сопровождения после перекрытия на основе анализа динамических и геометрических характеристик среды.

#### Цель исследования

Целью настоящей работы является разработка единой классификации методов преодоления окклюзий в МОС. Данная классификация связывает условия корректного применения и типичные источники ошибок каждой группы с характеристиками сцены,

независимо от программной и архитектурной реализации алгоритмов. Практическая обоснованность предложенной типологии подтверждается путём сопоставления опубликованных показателей качества представительных алгоритмов из каждой выделенной группы на наборе данных MOT17. Для достижения поставленной цели вводится классификация по трём ортогональным признакам, комбинация которых определяет принадлежность метода к одной из пяти основных групп.

#### Материалы и методы исследования

Материал исследования составили публикации 2016–2025 годов, отобранные по базам данных IEEE Xplore, Semantic Scholar и arXiv. В выборку включались работы, в которых обработка перекрытий является основным вкладом или определяющим механизмом ассоциации при утрате наблюдаемости объекта. Работы, ограниченные совершенствованием детектора без специализированной стратегии обработки окклюзий, из рассмотрения исключались.

Классификация строится по трём ортогональным признакам. Первый признак, тип используемой информации, принимает значения кинематических характеристик траектории, визуальных дескрипторов объекта и геометрических параметров сцены. Вторым признаком является горизонт принятия решений, различает однокадровую ассоциацию, обработку окна фиксированной длительности и многошаговое вероятностное прогнозирование. Третьим признаком является представление неопределённости состояния, выделяет четыре варианта, а именно ковариационную параметризацию фильтра Калмана [8], метрику в пространстве дескрипторов, детерминированную оптимизацию по интервалу наблюдения и многомодальное распределение по возможным продолжениям траектории.

Применимость предложенной типологии оценивалась по опубликованным показателям качества на MOT17 для представительных методов из каждой выделенной группы. Для обеспечения методологической корректности сопоставления рассматривались результаты алгоритмов, полученные в рамках единого протокола. Поскольку метрики MOS по-разному агрегируют составляющие ошибки [9], устойчивость алгоритмов к окклюзиям оценивалась исключительно по метрикам IDF1 и NOTA.

#### Результаты и их обсуждение

Сочетание трёх классификационных признаков определяет пять основных групп методов, первую из которых образуют методы кинематического рекуррентного оценивания, опирающиеся исключительно на координаты и скорости объекта с однокадровым горизонтом принятия решений. Неопределённость параметризуется ковариационной матрицей фильтра Калмана, а в период окклюзии оценка формируется на основе прогноза без обновления по измерению [10]. Преимуществом группы является постоянная вычислительная трудоёмкость и отсутствие требований к дополнительным данным, однако с каждым шагом без наблюдений неопределённость координат нарастает, и при продолжительной потере наблюдаемости вероятность ошибочной ассоциации траекторий остаётся высокой.

Вторую группу образуют методы ассоциации по визуальному описанию. Для каждого трека поддерживается признаковое представление, с которым сопоставляются новые детекции методами метрического обучения [11]. Неопределённость выражается расстоянием в пространстве признаков, а не ковариационной матрицей. Ключевым ограничением является искажение признакового описания при частичных перекрытиях, когда фрагменты заслоняющего объекта включаются в формируемое представление и снижают дискриминативность, особенно в сценах с высокой плотностью и визуальной однородностью объектов.

Третья группа объединяет методы совместной оптимизации на множестве кадров. В классическом варианте последовательность детекций представляется ориентированным ациклическим графом, а восстановление траекторий сводится к поиску непересекающихся путей минимальной стоимости. Альтернативный вариант реализуется трансформерными архитектурами, в которых ассоциация обеспечивается механизмом внимания по совокупности запросов треков и текущих детекций [12]. Общим свойством группы является выход за рамки покадровой ассоциации, однако вычислительные затраты возрастают с увеличением интервала наблюдения, что ограничивает область применения задачами постобработки или сценариями с коротким окном наблюдения.

Четвёртая группа включает методы вероятностного прогнозирования траекторий, описывающие будущее состояние объекта условным распределением по возможным продолжениям, а не точечной оценкой. Многомодальность распределения повышает вероятность корректного восстановления ассоциации после длительной потери наблюдаемости. В работе [13] показано, что даже компактное множество разнообразных прогнозов движения существенно сокращает пространство поиска ассоциаций и снижает число переключений идентификаторов. Отличительной особенностью группы является то, что её методы функционируют не как самостоятельные алгоритмы сопровождения, а как модули постобработки, применяемые поверх готовых результатов трекеров из групп 1–3. По этой причине они не представлены отдельными строками в стандартных таблицах MOT17 и не включены в таблицу 1, однако способны повышать качество любого базового алгоритма.

Пятая группа объединяет методы учёта пространственного контекста, дополняющие модель состояния геометрическими параметрами сцены. Использование информации о глубине позволяет различать объекты, совпадающие в проекции, но разделённые по расстоянию до камеры. В работе [14] предложен метод оценки псевдоглубины непосредственно из монокулярных изображений и каскадное сопоставление по глубинным слоям, что свидетельствует о возможности учёта пространственного контекста без специализированного оборудования. Результаты сопоставления алгоритмов из каждой выделенной группы приведены в таблице 1. Все значения получены на тестовом наборе MOT17 по приватному протоколу с единым детектором YOLOX. Следует учитывать, что SUSHI является оффлайн-методом с обработкой всей видеопоследовательности, тогда как остальные алгоритмы работают в режиме реального времени.

Таблица 1 – Сравнение алгоритмов по группам классификации на наборе данных MOT17

Группа	Алгоритм	НОТА	МОТА	IDF1
1 – Кинематика	ByteTrack	63,1	80,3	77,3
	OC-SORT	63,2	78,0	77,5
2 – Визуальное описание	SUSHI	76,1	82,0	83,1
3 – Совместная оптимизация	MOTRv2	60,3	76,2	72,2
1+2 – Комбинированные	BoT-SORT	65,0	80,5	80,2
	Deep OC-SORT	64,9	80,6	79,4

Сопоставление выявляет ряд закономерностей, согласующихся с предсказаниями классификации. Алгоритм ByteTrack (группа 1) показал значение IDF1 77,3 при полном отсутствии признаков внешнего вида, что обусловлено стратегией двухпроходной ассоциации детекций с низким порогом уверенности [15]. Модифицированный кинематический подход OC-SORT (группа 1), опирающийся на виртуальные траектории для коррекции ошибок фильтра Калмана, демонстрирует сопоставимый результат (IDF1 77,5), что указывает на достаточность усовершенствованной кинематической модели при качественном детекторе. Алгоритм MOTRv2 (группа 3, IDF1 72,2) уступает группе 1, что согласуется с предсказанным ограничением группы. Трансформерный декодер MOTRv2 выполняет совместную оптимизацию ассоциации, однако в реальном масштабе обработки его рассуждение ограничено фиксированным числом запросов и кадров, что нивелирует преимущество перед однокадровыми методами с качественным детектором. Наибольшее значение IDF1 (83,1) получено алгоритмом SUSHI (группа 2), что демонстрирует высокую эффективность иерархической кластеризации признаков. Комбинированные алгоритмы BoT-SORT [16] и Deep OC-SORT [17] (группа 1+2) показали значения IDF1 в диапазоне 79,4–80,2, что свидетельствует об отсутствии универсального механизма преодоления окклюзий, поскольку объединение источников информации снижает дисперсию результатов относительно однородных групп. Вместе с тем превосходство SUSHI над комбинированными методами указывает на то, что в условиях набора данных MOT17 полноценного признакового описания оказывается достаточно без явной кинематической составляющей. Однако для сцен с визуально однородными объектами данный вывод требует дополнительной проверки на специализированных наборах данных.

#### Выводы

Сопоставление опубликованных результатов алгоритмов из каждой выделенной группы на наборе данных MOT17 подтвердило практическую обоснованность предложенной классификации. Кинематические методы при качественной двухпроходной ассоциации показывают значения IDF1, сопоставимые с многокомпонентными подходами, тогда как методы визуального описания обеспечивают наибольшее значение метрики среди отдельных алгоритмов, что указывает на их существенную зависимость от качества признакового представления. Методы совместной оптимизации на множестве кадров в режиме реального времени уступают более простым подходам из-за ограниченного окна наблюдения. Комбинирование механизмов из нескольких групп снижает дисперсию результатов и обеспечивает наиболее стабильные значения метрик.

Полученные результаты позволяют сформулировать практические рекомендации по выбору механизма в зависимости от характеристик целевой сцены. В сценах с визуально однородными объектами предпочтительны кинематические методы, в том числе с механизмами коррекции фильтра Калмана, позволяющие повысить надёжность сопровождения без привлечения признаков внешнего вида. При высоком разнообразии внешнего вида объектов и надёжном детекторе наибольшую устойчивость обеспечивают методы визуального описания или их комбинация с кинематикой. Методы вероятностного прогнозирования и пространственного контекста наиболее перспективны для специализированных сценариев со сложной нелинейной динамикой движения или в условиях, когда доступна пространственная информация о сцене.

Предложенная классификация служит объективной основой для систематического анализа основных источников ошибок сопровождения в конкретном приложении, определяя тем самым выбор оптимального алгоритма. Дальнейшие исследования целесообразно направить на валидацию предложенной типологии на более сложных наборах данных, например DanceTrack и MOT20, а также на разработку методики

количественной оценки степени окклюзии в видеопоследовательности для динамического переключения между группами методов.

### Список литературы:

1. MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking / P. Dendorfer, A. Ošep, A. Milan [et al.] // International Journal of Computer Vision. – 2021. – Vol. 129, № 4. – P. 845–881. – DOI: 10.1007/s11263-020-01393-0;
2. MOT20: A Benchmark for Multi Object Tracking in Crowded Scenes / P. Dendorfer, H. Rezatofighi, A. Milan [et al.]. – 2020. – URL: <https://arxiv.org/abs/2003.09003> (дата обращения: 10.03.2026);
3. Multiple object tracking: A literature review / W. Luo, J. Xing, A. Milan [et al.] // Artificial Intelligence. – 2021. – Vol. 293. – P. 103448. – DOI: 10.1016/j.artint.2020.103448;
4. Multiple Pedestrian Tracking Under Occlusion: A Survey and Outlook / Z. Sun, G. Wei, W. Fu [et al.] // IEEE Transactions on Circuits and Systems for Video Technology. – 2025. – Vol. 35, № 2. – P. 1009–1027. – DOI: 10.1109/TCSVT.2024.3481425;
5. Прогноз состояния объекта на основе применения фильтра Калмана и глубоких нейронных сетей / А. Ю. Пучков, М. И. Дли, Е. И. Лобанева, М. А. Василькова // Программные продукты и системы. – 2019. – Т. 22. – С. 368–376. – DOI: 10.15827/0236-235X.127.368-376;
6. SparseTrack: Multi-Object Tracking by Performing Scene Decomposition based on Pseudo-Depth / Z. Liu, X. Wang, C. Wang [et al.]. – 2023. – URL: <https://arxiv.org/abs/2306.05238> (дата обращения: 10.03.2026);
7. Багиров, М. Б. Многокамерное сопровождение объектов с учетом глобальных и локальных признаков / М. Б. Багиров, Э. С. Соколова // Труды НГТУ им. П. Е. Алексеева. – Нижний Новгород, 2025. – Т. 150, № 3. – С. 7–14. – DOI: 10.46960/1816-210X\_2025\_3\_7;
8. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems / R. E. Kalman // Journal of Basic Engineering. – 1960. – Vol. 82, № 1. – P. 35–45. – DOI: 10.1115/1.3662552;
9. HO TA: A Higher Order Metric for Evaluating Multi-object Tracking / J. Luiten, A. Ošep, P. Dendorfer [et al.] // International Journal of Computer Vision. – 2021. – Vol. 129, № 2. – P. 548–578. – DOI: 10.1007/s11263-020-01375-2;
10. Simple Online and Realtime Tracking / A. Bewley, Z. Ge, L. Ott [et al.] // 2016 IEEE International Conference on Image Processing (ICIP). – 2016. – P. 3464–3468. – URL: <http://arxiv.org/abs/1602.00763> (дата обращения: 10.03.2026);
11. Cetintas, O. Unifying Short and Long-Term Tracking with Graph Hierarchies / O. Cetintas, G. Brasó, L. Leal-Taixé // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2023. – P. 22877–22887;
12. Zhang, Y. MOTRv2: Bootstrapping End-to-End Multi-Object Tracking by Pretrained Object Detectors / Y. Zhang, T. Wang, X. Zhang // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2023. – P. 22056–22065;
13. Quo Vadis: Is trajectory forecasting the key towards long-term multi-object tracking? / P. Dendorfer, V. Yugay, A. Ošep, L. Leal-Taixé // Proceedings of the Conference on Neural

- Information Processing Systems (NeurIPS). – 2022. – Vol. 35. – P. 20054–20067. – DOI: 10.48550/arXiv.2210.07681;
14. PD-SORT: Occlusion-Robust Multi-Object Tracking Using Pseudo-Depth Cues / Y. Wang, D. Zhang, R. Li [et al.] // IEEE Transactions on Consumer Electronics. – 2025. – Vol. 71. – P. 165–177;
15. ByteTrack: Multi-Object Tracking by Associating Every Detection Box / Y. Zhang, P. Sun, Y. Jiang [et al.] // Computer Vision – ECCV 2022. – 2022. – URL: <https://arxiv.org/abs/2110.06864> (дата обращения: 10.03.2026);
16. Aharon, N. BoT-SORT: Robust Associations Multi-Pedestrian Tracking / N. Aharon, R. Orfaig, B.-Z. Bobrovsky. – 2022. – URL: <https://arxiv.org/abs/2206.14651> (дата обращения: 10.03.2026);
17. Deep OC-SORT: Multi-Pedestrian Tracking by Adaptive Re-Identification / G. Maggiolino, A. Ahmad, J. Cao, K. Kitani // 2023 IEEE International Conference on Image Processing (ICIP). – 2023. – P. 3025–3029. – URL: <https://doi.org/10.1109/ICIP49359.2023.10222576> (дата обращения: 10.03.2026).

#### References:

1. MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking / P. Dendorfer, A. Ošep, A. Milan [et al.] // International Journal of Computer Vision. – 2021. – Vol. 129, № 4. – P. 845–881. – DOI: 10.1007/s11263-020-01393-0;
2. MOT20: A Benchmark for Multi Object Tracking in Crowded Scenes / P. Dendorfer, H. Rezatofighi, A. Milan [et al.]. – 2020. – URL: <https://arxiv.org/abs/2003.09003> (accessed: 10.03.2026);
3. Multiple object tracking: A literature review / W. Luo, J. Xing, A. Milan [et al.] // Artificial Intelligence. – 2021. – Vol. 293. – P. 103448. – DOI: 10.1016/j.artint.2020.103448;
4. Multiple Pedestrian Tracking Under Occlusion: A Survey and Outlook / Z. Sun, G. Wei, W. Fu [et al.] // IEEE Transactions on Circuits and Systems for Video Technology. – 2025. – Vol. 35, № 2. – P. 1009–1027. – DOI: 10.1109/TCSVT.2024.3481425;
5. Object State Prediction Using Kalman Filter and Deep Neural Networks / A. Yu. Puchkov, M. I. Dli, E. I. Lobaneva, M. A. Vasil'kova // Software Products and Systems. – 2019. – Vol. 22. – P. 368–376. – DOI: 10.15827/0236-235X.127.368-376;
6. SparseTrack: Multi-Object Tracking by Performing Scene Decomposition based on Pseudo-Depth / Z. Liu, X. Wang, C. Wang [et al.]. – 2023. – URL: <https://arxiv.org/abs/2306.05238> (accessed: 10.03.2026);
7. Bagirov, M. B. Multi-Camera Object Tracking Using Global and Local Features / M. B. Bagirov, E. S. Sokolova // Proceedings of NNSTU n.a. R. E. Alekseev. – Nizhny Novgorod, 2025. – Vol. 150, № 3. – P. 7–14. – DOI: 10.46960/1816-210X\_2025\_3\_7;
8. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems / R. E. Kalman // Journal of Basic Engineering. – 1960. – Vol. 82, № 1. – P. 35–45. – DOI: 10.1115/1.3662552;
9. HOTA: A Higher Order Metric for Evaluating Multi-object Tracking / J. Luiten, A. Ošep, P. Dendorfer [et al.] // International Journal of Computer Vision. – 2021. – Vol. 129, № 2. – P. 548–578. – DOI: 10.1007/s11263-020-01375-2;

10. Simple Online and Realtime Tracking / A. Bewley, Z. Ge, L. Ott [et al.] // 2016 IEEE International Conference on Image Processing (ICIP). – 2016. – P. 3464–3468. – URL: <http://arxiv.org/abs/1602.00763> (accessed: 10.03.2026);
11. Cetintas, O. Unifying Short and Long-Term Tracking with Graph Hierarchies / O. Cetintas, G. Brasó, L. Leal-Taixé // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2023. – P. 22877–22887 (accessed: 10.03.2026);
12. Zhang, Y. MOTRv2: Bootstrapping End-to-End Multi-Object Tracking by Pretrained Object Detectors / Y. Zhang, T. Wang, X. Zhang // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2023. – P. 22056–22065;
13. Quo Vadis: Is trajectory forecasting the key towards long-term multi-object tracking? / P. Dendorfer, V. Yugay, A. Osep, L. Leal-Taixe // Proceedings of the Conference on Neural Information Processing Systems (NeurIPS). – 2022. – Vol. 35. – P. 20054–20067. – DOI: 10.48550/arXiv.2210.07681;
14. PD-SORT: Occlusion-Robust Multi-Object Tracking Using Pseudo-Depth Cues / Y. Wang, D. Zhang, R. Li [et al.] // IEEE Transactions on Consumer Electronics. – 2025. – Vol. 71. – P. 165–177;
15. ByteTrack: Multi-Object Tracking by Associating Every Detection Box / Y. Zhang, P. Sun, Y. Jiang [et al.] // Computer Vision – ECCV 2022. – 2022. – URL: <https://arxiv.org/abs/2110.06864> (accessed: 10.03.2026);
16. Aharon, N. BoT-SORT: Robust Associations Multi-Pedestrian Tracking / N. Aharon, R. Orfaig, B.-Z. Bobrovsky. – 2022. – URL: <https://arxiv.org/abs/2206.14651> (accessed: 10.03.2026);
17. Deep OC-SORT: Multi-Pedestrian Tracking by Adaptive Re-Identification / G. Maggolino, A. Ahmad, J. Cao, K. Kitani // 2023 IEEE International Conference on Image Processing (ICIP). – 2023. – P. 3025–3029. – URL: <https://doi.org/10.1109/ICIP49359.2023.10222576>.