

Никитин Макар Дмитриевич

магистр

Научный руководитель

Никонов Вячеслав Викторович

доцент

ФГБОУ ВО «МИРЭА – Российский технологический университет»

г. Москва

**ОБ ОДНОМ ПОДХОДЕ К ИНТЕЛЛЕКТУАЛЬНОМУ АНАЛИЗУ
ПРОСТРАНСТВЕННО-ВРЕМЕННЫХ ВЗАИМОДЕЙСТВИЙ
ДЛЯ АВТОМАТИЗАЦИИ РЕТРОСПЕКТИВНОГО ПОИСКА
ИНЦИДЕНТОВ В ВИДЕОАРХИВАХ БОЛЬШОЙ ДЛИТЕЛЬНОСТИ**

***Аннотация:** в работе рассматривается задача автоматического обнаружения инцидентов взаимодействия с заданным объектом. Проведенный анализ существующих решений видеоаналитики показал их ограничения в контексте рассматриваемой задачи. Предложен метод анализирования видео для обнаружения инцидентов, основанный на пространственно-временном анализе взаимодействий, включающий формализацию понятия контакта. Разработана методика оценки, которая содержит набор сценариев и возможности выбора оценок качества. Получены результаты, доказывающие повышение точности и частоты обнаружения инцидентов, по сравнению со стандартными подходами.*

***Ключевые слова:** интеллектуальное видеонаблюдение, машинное обучение, детекция объектов, трекинг объектов, ретроспективный поиск, коэффициент сжатия внимания, безопасность статических объектов, автоматизация мониторинга.*

В современных городах системы видеонаблюдения являются частью инфраструктуры и являются инструментом для поиска инцидентов на видео для выявления преступлений различного рода. Рост количества и качества камер приводит

к большому количеству данных, которые необходимо анализировать. Традиционный подход, основанный на участии операторов, оказывается недостаточно эффективным вследствие ограниченности человеческого внимания и высокой нагрузки.

Появление методов машинного зрения и глубокого обучения позволило автоматизировать процессы анализа видео. Развитие алгоритмов детекции обеспечивают высокую точность распознавания людей, машин и прочих движущихся объектов. Однако подавляющее большинство существующих решений направлены на работу с заранее прописанными сценариями и не обеспечивают гибкость в анализе и выявлении взаимодействия объектов.

На текущий момент существующие решения работы с видеопотоком предлагают стандартный набор инструментов видеоаналитики. Однако применительно к задаче ретроспективного поиска инцидентов данные решения демонстрируют низкую операционную эффективность.

Таблица 1

Основные недостатки существующих решений

Технология/ Система	Основной механизм работы	Ключевые недостатки в контексте поиска инцидентов	Экспертная оценка трудозатрат оператора
Программный детектор движения (Motion Detection)	Анализ изменения интенсивности групп пикселей в кадре	Высокая чувствительность к «шумам»: изменение освещения, тени, осадки, движение растительности	Критически высокая. Оператор вынужден просматривать до 90% ложных срабатываний
Объектная аналитика (Human/Vehicle Detection)	Классификация объектов с помощью нейросетей (YOLO и аналоги)	Не фиксирует факт контакта. На оживленных участках (парковки, тротуары) генерирует избыточное количество меток «прохождения мимо»	Высокая. Фильтрация по типу объекта лишь незначительно сокращает объем выборки для анализа
Зональная аналитика (Intrusion Detection)	Фиксация пересечения виртуальной линии или входа в заданную область	Не учитывает перспективу и наложение объектов (occlusion). Ложно срабатывает, если человек проходит «перед» машиной, не касясь её	Средняя. Требуется предварительная настройка зон, что невозможно сделать ретроспективно для случайного объекта

Системы ускоренного просмотра (Time-lapse / Video Synopsis)	Наложение движущихся объектов из разных временных отрезков на один фон	Высокая когнитивная нагрузка. При большой плотности событий в кадре возникает «каша», в которой легко пропустить момент микро-повреждения (царапины)	Средняя. Эффективно для поиска краж, но малоприспособлено для поиска едва заметных контактов
---	--	--	--

Представленные данные подтверждают, что современные коммерческие решения ориентированы на мониторинг событий, а не на анализ взаимодействий. В сценарии, когда автомобиль стоит на парковке 48 часов, ни одна из систем не позволяет автоматически выделить те 2–3 секунды, когда произошел физический контакт, если рядом с объектом за это время прошло, например, 200 человек.

Основная задача заключается в отсутствии возможности оптимизировать работу оператора. Затрачиваемое время на поиск инцидента в большинстве случаев прямо пропорционален длительности видеофрагмента.

Главная метрика, которую необходимо оптимизировать – это коэффициент сжатия внимания. Если оператор тратит T_{man} времени на поиск в архиве длительностью T_{arc} , то эффективность системы η выражается как:

$$\eta = \frac{T_{\text{man}}}{T_{\text{arc}}}$$

В существующих системах η редко превышает 5–10. Наша цель – поднять этот показатель до 50–100 за счет автоматического исключения «пустых» взаимодействий.

Предложенный подход базируется на гибридной модели, сочетающей детектор объектов архитектуры YOLOv8 (для локализации), трекер ByteTrack (для анализа траекторий) и оригинальный модуль Spatial-Temporal Interaction (STI) для верификации контакта.

Процесс поиска инцидента разделен на четыре последовательных этапа.

1. Статическая маскировка ROI Selection: пользователь или система определяет целевой объект O (например, припаркованный автомобиль). Создается бинарная маска M_O .

2. Детекция и фильтрация агентов: нейросеть идентифицирует потенциальных «объектов взаимодействия» А (люди, другие ТС).

3. Анализ пересечения и близости: вычисляется геометрическое взаимодействие между боксами агентов B_a и маской объекта M_o .

4. Валидация инцидента: проверка временной задержки и изменения вектора движения в зоне контакта.

Для исключения ложных срабатываний, когда объект просто проходит «на фоне» или «перед» целью, вводится комбинированный критерий инцидента.

1. Коэффициент пространственного перекрытия K_{ovl} : вместо стандартного IoU, который плохо работает с объектами разных порядков (человек и машина), мы используем отношение площади пересечения к площади меньшего объекта:

$$K_{(ovl)(t)} = \frac{Area(B_a(t) \cap M_o)}{Area(B_a(t))},$$

где $B_a(t)$ – ограничивающий бокс агента в момент времени t .

2. Критерий временной экспозиции (ΔT): контакт признается потенциально опасным, если агент находится в зоне взаимодействия дольше порогового значения τ :

$$E_{time} = \int_{t_{start}}^{t_{end}} [K_{(ovl)(t)} > \theta] dt \geq \tau,$$

где θ – порог чувствительности, а τ – минимальное время контакта, отсекающее кратковременные графические шумы.

3. Аномалия вектора скорости (ΔV): физический контакт почти всегда сопровождается изменением кинематики. Мы вычисляем модуль изменения скорости агента в момент пересечения маски:

$$\Delta V = |v(t_{contact}) - v(t_{pre-contact})|,$$

если $\Delta V > \gamma$, это показатель для остановки и совершения действия или о факте столкновения.

Итоговый алгоритм формирует список событий $L = \{e_1, e_2, \dots, e_m\}$, где каждое событие – это короткий видеофрагмент (слайс), в котором выполнены все три условия:

$$Event = \{t \mid K_{(ovl)} > \theta \wedge E_{(time)} \geq \tau \wedge \Delta V_{(anomaly)}\},$$

Это позволяет трансформировать 24-часовой архив в 2-минутную нарезку наиболее вероятных моментов повреждения.

Для оценки эффективности предложенных решений был сформирован тестовый датасет их 12 видео. Были включены сценарии с инцидентами (касание авто, удары дверью) и сценарии проходящих пешеходов

В качестве базовой линии (Baseline) были взяты стандартный детектор движения и фильтр объектов «Человек/Транспорт» (как в системах Ivideon/Trassir).

Таблица 2

Сравнение метрик эффективности поиска инцидентов

Метрика	Стандартный детектор движения	Объектная фильтрация (YOLO)	Предложенный метод (STI)
Precision (Точность)	0.12	0.45	0.89
Recall (Полнота)	0.98	0.91	0.93
F1-Score	0.21	0.60	0.91
Коэфф. сжатия архива	1.5x	5x	45x

При использовании предложенного метода время, затрачиваемое оператором на идентификацию одного инцидента в 24-часовом архиве, сократилось с 140 минут (ручная перемотка) до 3 минут (просмотр автоматически сформированных фрагментов).

В данной работе решена актуальная задача автоматизации ретроспективного поиска инцидентов в больших видеоархивах. Основные научные и практические результаты:

Обоснована неэффективность существующих коммерческих VMS-решений для поиска микро-инцидентов (царапин, касаний) из-за избыточного количества ложных срабатываний.

Разработан гибридный метод, использующий пространственно-временной анализ взаимодействий, что позволило повысить точность фильтрации событий до 89%.

Доказано, что предложенный подход обеспечивает сжатие времени анализа архива в десятки раз, сохраняя при этом высокую полноту выборки.

Разработанная методика может быть интегрирована в существующие системы облачного видеонаблюдения в качестве дополнительного интеллектуального слоя для сервисов страхования и мониторинга безопасности парковок.

Список литературы

1. Антонов С.В. Методы глубокого обучения в задачах интеллектуального анализа видеопотоков высокого разрешения / С.В. Антонов, Д.И. Петров // Вестник компьютерных и информационных технологий. – 2024. – №3.

2. Гоял Р. Детекция взаимодействия людей и транспортных средств в городской среде с использованием пространственно-временных графов / Р. Гоял [и др.] // Журнал обработки изображений в реальном времени. – 2024.

3. Джочер Г. Документация Ultralytics YOLOv8 / Г. Джочер, А. Чаурасия, Дж. Цю. – URL: <https://docs.ultralytics.com/> (дата обращения: 14.04.2026).

4. Иванов А.А. Автоматизация мониторинга объектов парковочного пространства на базе нейросетевых детекторов / А.А. Иванов // Системы безопасности и связи. – 2025.

5. Мишра С. Оптимизация эффективности оператора в крупномасштабных системах видеонаблюдения / С. Мишра // Международный журнал информационной безопасности и киберпреступности. – 2025.

6. Николов И.А. Надежное обнаружение аномалий в системах долгосрочного видеонаблюдения / И.А. Николов // Материалы 19-й международной совместной конференции по теории и приложениям компьютерного зрения. – 2024.

7. Сингх А. Интеллектуальная система видеонаблюдения: отслеживание подозрительной активности с помощью YOLOv8 и Vision Transformer / А. Сингх [и др.] // Материалы международной конференции по исследованиям и разработкам. – 2025.

8. Тервен Х. Всесторонний обзор YOLO: от YOLOv1 до YOLOv8 и далее / Х. Тервен, Д. Кордова-Эспарза // Препринт arXiv. – 2023.

9. Уайлс А. Сжатое зрение для эффективного понимания видеоданных / А. Уайлс [и др.] // Сборник открытого доступа CVF. – 2022.

10. ByteTrack: Мультиобъектный трекинг путем сопоставления каждого бокса детекции / Ю. Чжан, П. Сунь, Ю. Цзян [и др.] // Материалы конференции IEEE/CVF по компьютерному зрению и распознаванию образов. – 2022.

11. Ivideon: Облачное видеонаблюдение – официальный сайт и техническая документация системы. – URL: <https://ru.ivideon.com/> (дата обращения: 14.04.2026).