
ИНФОРМАТИКА, УПРАВЛЕНИЕ И СИСТЕМНЫЙ АНАЛИЗ

УДК 004.932

DOI 10.46960/1816-210X_2025_3_7

EDN QKUUBE

**МНОГОКАМЕРНОЕ СОПРОВОЖДЕНИЕ ОБЪЕКТОВ
С УЧЕТОМ ГЛОБАЛЬНЫХ И ЛОКАЛЬНЫХ ПРИЗНАКОВ****М.Б. Багиров**ORCID: 0000-0003-1656-0849 e-mail: bagirov_mirabbas@mail.ruНижегородский государственный технический университет им. Р.Е. Алексеева
*Нижний Новгород, Россия***Э.С. Соколова**ORCID: 0000-0003-0860-2463 e-mail: essokolowa@gmail.comНижегородский государственный технический университет им. Р.Е. Алексеева
Нижний Новгород, Россия

Предложен подход к организации кросс-камерного трекинга объектов в условиях сложных динамических сцен, включающий решение задач детектирования, распознавания и идентификации при сложных траекториях движения, наложении объектов друг на друга в рамках одного кадра, временном исчезновении объекта из области обзора видеокамер и последующей повторной идентификации при его появлении на видеопотоке в пределах панорамы. Вводится новая концепция составного дескриптора, который включает в себя как глобальные, так и локальные характеристики объектов. Для оценки эффективности работы алгоритма использованы показатели F-меры и достоверности сопровождения множества объектов на панораме видеопотоков. Выполнена настройка параметров для обеспечения максимального значения показателей, проведено сравнение разработанного метода с широко используемым в настоящее время методом сопровождения людей *DeepSortV2*.

Ключевые слова: многокамерные системы, видеоаналитика, сопровождение объектов, повторная идентификация, глобальные и локальные признаки.

ДЛЯ ЦИТИРОВАНИЯ: Багиров, М.Б. Многокамерное сопровождение объектов с учетом глобальных и локальных признаков / М.Б. Багиров, Э.С. Соколова // Труды НГТУ им. Р.Е. Алексеева. 2025. № 3. С. 7-14. DOI: 10.46960/1816-210X_2025_3_7 EDN: QKUUBE

**MULTI-CAMERA OBJECT TRACKING
WITH CONSIDERATION OF GLOBAL AND LOCAL FEATURES****M.B. Bagirov**ORCID: 0000-0003-1656-0849 e-mail: bagirov_mirabbas@mail.ruNizhny Novgorod State Technical University n.a. R.E. Alekseev
*Nizhny Novgorod, Russia***E.S. Sokolova**ORCID: 0000-0003-0860-2463 e-mail: essokolowa@gmail.comNizhny Novgorod State Technical University n.a. R.E. Alekseev
Nizhny Novgorod, Russia

Abstract. This paper describes the proposed approach to cross-camera object tracking in complex dynamic scenes, which includes solving the problems of detection, recognition and identification in complex motion trajectories, overlapping of objects within one frame, temporary disappearance of an object from the video cameras viewing area and

subsequent re-identification when it appears on the video stream within the panorama. A new concept of composite descriptor is introduced, which includes both global and local object characteristics. The F-measure and reliability indicators of tracking multiple objects on a panorama of video streams are used to evaluate the efficiency of the algorithm. The parameters were adjusted to ensure the maximum value of the indicators, and the developed method was compared with the widely used DeepSortV2 people tracking method.

Key words: multi-camera systems, video analytics, object tracking, re-identification, global and local features.

FOR CITATION: M.B. Bagirov, E.S. Sokolova. Multi-camera object tracking with consideration of global and local features. Transactions of NNSTU n. a. R.E. Alekseev. 2025. № 3. Pp. 7-14. DOI: 10.46960/1816-210X_2025_3_7 EDN: QKUUBE

Введение

В современных условиях наблюдается повсеместная интеграция видеокамер в различные сферы человеческой деятельности, где они стали неотъемлемым элементом цифровой инфраструктуры, внедряясь в разнообразные социально-технические и интеллектуальные системы: умные города, автономные транспортные средства, промышленную робототехнику и др. Прогнозируемый к 2027 г. рост объема рынка интеллектуального видеонаблюдения демонстрирует растущую потребность в автоматизированной обработке видеопотоков. Особенно востребованными становятся технологии, способные анализировать данные в реальном времени для задач управления трафиком и предотвращения инцидентов на производстве.

Прорыв в области искусственного интеллекта, особенно с появлением трансформерных архитектур и методов самообучения, кардинально изменил подходы к компьютерному зрению. Современные алгоритмы, такие как *YOLOv8* и *RT-DETR*, демонстрируют рекордную точность детекции объектов даже в условиях низкой освещенности или частичных окклюзий. Примечательно, что данный технологический скачок совпал по времени с резким увеличением вычислительных мощностей доступного оборудования, что создало уникальные условия для практической реализации теоретических разработок. Автоматизированный анализ данных во многом позволяет значительно расширить функционал систем видеонаблюдения. Однако перечисленные ранее достижения в этой предметной области в основном применимы для анализа одиночных кадров, одиночных камер или простых конфигураций двухкамерных систем, а не для комплексной обработки многокамерных сцен с динамически меняющимися условиями. Таким образом, до сих пор остается актуальной проблема разработки систем многокамерного сопровождения объектов, в особенности из-за перекрытий и взаимодействия отслеживаемых объектов, непредсказуемости их траектории движения, асинхронности видеопотоков и распространенной гетерогенности сенсоров (спектральные характеристики, угол обзора).

Особую актуальность приобретают исследования в области многокамерного обнаружения и отслеживания объектов, где требуется координация данных между несколькими устройствами для создания единой картины наблюдения, в том числе, для панорам, созданных путем синтеза видеопотоков нескольких камер. Несмотря на значительный потенциал этого направления, описаний готовых и проверенных решений в открытых источниках крайне мало. Это обстоятельство создает уникальную нишу для научных разработок и внедрения инновационных подходов, способных преодолеть существующие ограничения и обеспечить качественный скачок в эффективности многокамерных систем. Помимо этого, современные системы видеоаналитики сталкиваются с постоянно растущими требованиями к точности и функциональности. Развитие технологий, увеличение объемов данных и усложнение задач наблюдения стимулируют запросы к более совершенным алгоритмам обработки информации. В частности, требуется не только повышение качества отслеживания объектов, но и расширение возможностей систем за счет интеграции дополнительных функций, таких как прогнозирование движений, анализ поведения и детектирование аномалий. Все это делает необходимым углубленное исследование новых методологий и технологических решений, способных удовлетворить возросшие ожидания от систем видеоаналитики.

Подходы к обнаружению и трекингу людей в кросс-камерных сценариях

В основе работы современных систем многокамерного сопровождения объектов должна лежать задача сопоставления объектов на изображениях с разных камер: например, человек идентифицируется в поле зрения одной камеры и далее повторно идентифицируется в базе данных (БД) при его переходе в поле зрения другой камеры. Данная задача является достаточно сложной, так как меняется ракурс человека, а, следовательно, его внешний вид, поза, освещенность и т.д. [1, 2]. Особую сложность представляют ситуации, когда система работает с многообъектными сценами, характеризующимися высокой степенью динамики и сложностью структуры. Например, если несколько объектов движутся одновременно, то их траектории могут пересекаться, что создает дополнительные трудности для алгоритмов детектирования и сопровождения. В таких случаях возникает необходимость учитывать нелинейные траектории движения, резкие изменения скорости объектов, а также возможное частичное или полное перекрытие одного объекта другим. Все эти факторы могут привести к снижению точности работы системы.

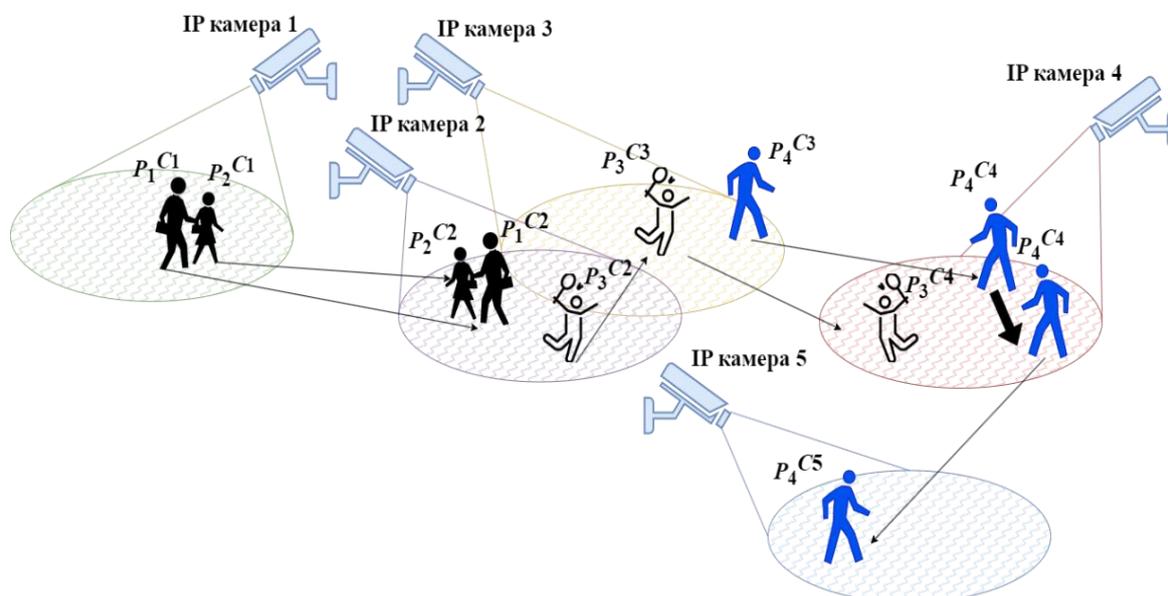


Рис. 1. Структура схемы сопровождения и повторной идентификации людей в многокамерной системе видеонаблюдения

Fig. 1. Structure of the scheme for people tracking and re-identification in a multi-camera video surveillance system

Кроме того, важным аспектом является влияние фонового окружения, которое может быть крайне разнообразным и сложным. Наличие динамических элементов фона, таких как движущиеся автомобили, колеблющиеся ветви деревьев или меняющаяся интенсивность света, добавляет еще один уровень сложности в процесс анализа видеопотока. Эти факторы требуют от систем видеоаналитики использования прогрессивных методов фильтрации и сегментации, чтобы отделить значимые объекты от незначительных деталей окружающей среды. На рис. 1 представлена схема работы распределенной системы видеоаналитики, реализующей непрерывный трекинг объектов в сети из пяти IP-камер с частично перекрывающимися и независимыми зонами наблюдения.

Каждому обнаруженному объекту (обозначаемому как P_n , где n – уникальный цифровой идентификатор) присваивается также и идентификатор камеры, представленный как C_k , в зоне видимости которой он находится. При переходе объекта из зоны видимости камеры C_k в зону камеры C_{k+1} система анализирует пространственно-временные ограничения, сравнивает

уникальный набор признаков объекта, а также сохраняет изначально присвоенный идентификатор объекта P_n в рамках всей многокамерной системы, производя таким образом его повторную идентификацию. Возможны варианты непрерывного сопровождения людей при перемещении на территории, без выхода их из поля зрения камер, а также прерывание сопровождения при выходе человека из зоны наблюдения с возобновлением сопровождения при следующем появлении в полях зрения камер с его повторной идентификацией.

Разработанный метод многокамерного сопровождения людей на панораме видеопотоков [3] содержит алгоритмы, реализующие обнаружение, распознавание, повторную идентификацию человека при его очередном входе в поля зрения камер, а также построение траекторий движения людей. Метод включает следующие этапы.

1. *Детектирование людей с помощью модели сверточной нейронной сети (СНС) YOLO*, обладающей высокой скоростью обработки данных и способной с большой точностью идентифицировать людей [4, 5]. Механизм работы данного алгоритма представлен на рис. 1 в рамках IP камеры 1, где происходит распознавание объектов и присвоение им определенных и отличных друг от друга ID (или P_n) в рамках всей многокамерной системы, которые остаются неизменными и позволяют отслеживать конкретный объект на всех последующих изображениях вне зависимости от его перемещений или изменений условий съемки. В ходе экспериментальных исследований были протестированы различные версии модели YOLO для определения их эффективности в условиях практического применения. Результаты показали, что использование современной версии YOLOv8 позволяет существенно улучшить качество детектирования по сравнению с предыдущими реализациями. Так, например, анализ статистики ложных срабатываний продемонстрировал снижение ошибок на 7 % в среднем по сравнению с YOLOv5. Это достижение объясняется как оптимизацией архитектуры сети, так и внедрением новых методов обучения, которые делают YOLOv8 более устойчивой к различным помехам и сложностям в данных. Экспериментальные данные также подтвердили способность YOLOv8 работать в режиме реального времени без потери точности, что делает данную версию особенно привлекательной для интеграции в системы видеоаналитики. Благодаря высокой производительности и гибкости настройки, YOLOv8 представляет собой перспективное решение для разработки современных многокамерных систем отслеживания объектов.

2. *Многокамерное сопровождение людей на панораме видеопотоков с камер*. Ключевые особенности реализации включают в себя оптимизированную нейросетевую сверточную модель ResNet-50, обладающую улучшенной обобщающей способностью для различных условий съемки (различных входных данных), меньшим количеством параметров и снижающую вычислительную сложность при сохранении точности. Во многом это достигнуто путем добавления *skip connections* после каждого слоя и внедрения иерархического *softmax* на выходе всей сети. Основные изменения включают в себя исключение начального сверточного слоя с фильтром $[7 \times 7]$ и уменьшение количества выходов последнего полносвязного слоя до 512 нейронов. Также для надежного сопровождения людей в условиях перекрытий, сложных траекторий и временного исчезновения из поля зрения камер используется динамически формируемый дескриптор уникальных признаков объектов, включающий комплекс следующих компонентов:

- выходные данные сверточной нейросети при последнем корректном обнаружении объекта;
- цветовые и текстурные характеристики, рассчитанные на момент последнего корректного обнаружения (гистограммы);
- пространственные параметры объекта на предыдущем кадре;
- уникальный цифровой идентификатор (ID).

Усовершенствованный алгоритм сопровождения множества объектов в условиях сложной визуальной среды, характерной для современных многокамерных систем видеомониторинга, демонстрирует высокую эффективность при работе с перекрывающимися объектами,

нелинейными траекториями движения и случаями временного исчезновения из поля зрения и включает в себя следующие составляющие.

1. Процедура сопоставления визуальных образов на новых и уже проанализированных кадров видеопотока. Для установления соответствия между трекируемыми объектами (Ob_tr) и новыми обнаружениями (Ob_det) в рамках одного кадра разработана комбинированная метрика оценки схожести. Формула вычисления интегрального показателя различия имеет вид:

$$d_{Ob_tr, Ob_det} = \alpha * \sqrt{\sum_{j=1}^{512} (f_{curr} - f_{prev})^2} + \beta * (\sqrt{(l_{curr} - l_{prev})^2} + \sqrt{(w_{curr} - w_{prev})^2} + \sqrt{(h_{curr} - h_{prev})^2}) \quad (1)$$

где: f_{curr}, f_{prev} – векторы визуальных признаков объекта (дескрипторы);

l_{curr}, l_{prev} – координаты центров bounding box; w_{curr}, w_{prev} (ширина);

h_{curr}, h_{prev} (высота) – габариты размеров объекта (его ограничительной рамки);

α и β – адаптивные весовые коэффициенты, подбираемые экспериментально для разных условий использования системы.

Первый компонент формулы отвечает за семантическую схожесть объектов на основе нейросетевых признаков, второй – за пространственную согласованность их положения и габаритов. В ходе анализа данных система выполняет параллельную обработку двух категорий объектов, результатом которой является вычисление оценки и матрицы схожести. К данным объектам относятся вновь обнаруженные люди в рамках текущего кадра, а также уже находящиеся под мониторингом системы индивидуумы, получившие свой идентификационный номер в рамках предыдущих кадров. Для определения наиболее точного соответствия между новыми обнаружениями и существующими идентификаторами в матрице схожести используется венгерский алгоритм, который обеспечивает глобальное оптимальное распределение соответствий, минимизируя суммарную метрику различий между сопоставляемыми объектами. Входные данные представляются в виде прямоугольной матрицы аффинитета, где по вертикали отображаются треки (объекты сопровождения с уникальными идентификаторами), а по горизонтали размещаются новые детекции текущего кадра. Каждый элемент полученной матрицы содержит вычисленное значение (весовой коэффициент) метрики сходства между конкретной парой «трек-детекция».

2. Соблюдение контроля проверки присутствия сопровождаемых людей в поле зрения камеры и сохранение присвоенных им ранее ID при их обнаружении. По последнему корректному обнаружению человека происходит обновление координат (ширины и высоты ограничительной рамки объекта) с целью уменьшения вероятности ложного срабатывания алгоритма и изменения ID в комплексном дескрипторе. Для определения степени совпадения изображений используется коэффициент Жаккарда, который представляет собой меру сходства двух множеств, показывающую долю совпадений общих элементов относительно остальных уникальных значений этого множества.

Алгоритм предназначен для исследования ситуации, приводящей к ошибкам определения системой людей в кадре. К примеру, может происходить присвоение идентификатора человека, который вышел из поля зрения камер и не фиксируется на панораме изображения, другому человеку, который появляется в текущем кадре почти в том же месте. Для уменьшения вероятности образования данной ситуации алгоритм исследует признаки дескриптора, в которые входят ширина, высота и присвоенные на прошлых и нынешних кадрах координаты. Еще одна ситуация для рассмотрения, которая может вызвать ошибку, заключается в том, что человек покидает кадр, и необходимо прекратить его сопровождение, не «перепутав» его с другим человеком, находящимся рядом с местом его выхода из поля зрения. Во избежание данной ошибки предлагается ввести показатель «видимости сопровождаемого объекта», который

включает взаимообусловленные показатели сильных (признаки при последнем идентифицированном корректном обнаружении) и слабых (схожесть показателей цветowych гистограмм) характеристик. Проведен комплексный вычислительный эксперимент по тонкой настройке алгоритма сопровождения, направленный на максимизацию показателя *MOTA* (Multi-Object Tracking Accuracy). Этот интегральный показатель учитывает точность идентификации объектов, минимизацию переключений идентификаторов, способность восстановления траекторий объектов после временных окклюзий, корректность обработки случаев повторного появления объектов.

3. *Алгоритм реконструкции и визуализации траектории движения объектов.* Для визуализации и анализа перемещений объектов разработан метод построения траекторий, основанный на накоплении пространственных координат объекта со всех кадров видеопотока, на которых он присутствует, сохраненных в дескрипторе признаков (центры ограничительных прямоугольников). Формальное представление траектории движения объекта $Tr(Ob)$ имеет вид:

$$Tr(Ob) = \left\{ (x_{Ob_q}^{F_1}, y_{Ob_q}^{F_1}), (x_{Ob_q}^{F_2}, y_{Ob_q}^{F_2}), \dots, (x_{Ob_q}^{F_n}, y_{Ob_q}^{F_n}) \right\}, \quad (2)$$

где $x_{Ob_q}^i, y_{Ob_q}^i$ – координаты центра объекта сопровождаемого объекта с уникальным идентификатором Ob_q на видеопоследовательности i .

Система осуществляет постоянное сопровождение объекта и построение его траектории до наступления одного из терминальных событий: полного выхода объекта за границы зоны видимости камер или длительной окклюзии (перекрытия другими объектами), превышающей установленный временной порог. При возобновлении видимости объекта после временного отсутствия реализован механизм восстановления траектории. Особенностью алгоритма является интеграция предиктивной модели, использующей исторические (априорные) данные о перемещениях объекта.

4. *Детектирование лиц людей для повышения точности обнаружения, повторной идентификации и сопровождения.* Разработан усовершенствованный алгоритм сопровождения объектов, интегрирующий анализ антропометрических и биометрических характеристик. Метод демонстрирует повышенную надежность в условиях частичных окклюзий и высокой плотности объектов. Для повышения точности трекинга в сложных условиях предложена расширенная модель дескриптора, включающая следующие локальные признаки:

- геометрические параметры лица: нормированные координаты центральной точки лицевой области, пространственные характеристики ограничивающего прямоугольника;
- вектор признаков последнего правильного распознавания объекта, извлеченный нейросетевым энкодером;
- метрика соответствия с эталонными признаками из базы данных (расстояние);
- динамические параметры сопровождения: счетчик последовательных успешных детекций лица без верификации, последние успешные композитные признаки верхней части тела (тор, плечевой пояс) из нейросети.

После тщательного сравнительного анализа современных алгоритмов биометрической идентификации была отобрана и модифицирована компактная и эффективная нейросетевая архитектура *MobileFaceNet*. К ее ключевым техническим преимуществам можно отнести размер входного изображения (112x112 пикселей), 128-мерное пространство признаков (характеристики лица объекта) и использование depthwise-сверток. Во время обучения нейросети была применена модифицированная функция потерь *ArcFace*. Экспериментальная оценка проводилась на стандартном датасете *MegaFace* [6] с использованием восьмиядерного процессора *Intel Xeon E3-1270v3* с частотой 3.50 ГГц. В результате значение точности распознавания составило 94 % с временем обработки изображения 42 мс.

Процесс сопоставления людей на соседних кадрах видеопотока с распознаванием лиц начинается с детекции человека в кадре с последующим выделением его *bounding box*. Затем локализуется лицо, которое также обрамляется прямоугольной областью. Далее выполняется

сравнение полученного лица с записями в базе данных. Если совпадение найдено, фиксируется факт идентификации, а дескрипторы лица обновляются. В случае отсутствия соответствия в БД проводится дополнительная верификация, после которой новый образ сохраняется в базу для последующего распознавания.

Ключевым этапом является определение зоны поиска лица в пределах обнаруженного человека. Если соотношение высоты и ширины *bounding box* превышает 3:1, анализируется только верхняя треть изображения, где с высокой вероятностью располагается лицо. В остальных случаях обработке подвергается весь фрагмент, что повышает точность алгоритма. Идентификация лица осуществляется путем сравнения его признаков с шаблонами из базы данных. При этом используется метрика расстояния между дескрипторами: если минимальное расстояние не превышает заданного порога, распознавание считается успешным. В таком случае система обновляет соответствующий комплексный дескриптор в базе данных, что позволяет адаптироваться к изменениям во внешности человека (например, смена ракурса, освещения или мимики). Экспериментальные испытания полученного алгоритма, представленные на рис. 2, продемонстрировали высокую эффективность. В тестовой сцене внутри помещения с участием шести человек модель *MobileFaceNet* корректно идентифицировала все лица, занесенные в БД (рис. 2а). В условиях сложных сцен (на примере представлена оживленная улица) с частыми перекрытиями объектов *ResNet-50* и *MobileFaceNet* сохраняли стабильную точность, несмотря на динамику сцены и помехи (рис. 2б).

Данный подход обеспечивает надежное сопровождение людей в реальном времени и может быть интегрирован в системы видеонаблюдения, контроля доступа и биометрической аутентификации.



Рис. 2. Результаты работы предложенного метода при различных сценариях

Fig. 2. Results of the proposed method under different scenarios

Для комплексной оценки производительности предложенного метода был проведен сравнительный анализ с эталонным алгоритмом *DeepSORT V2* на базе стандартизированных метрик (*F-мера* и *MOTA*), учитывающих полноту трекинга объектов, временную согласованность идентификаторов и достоверность множественного сопровождения. Эксперимент показал, что при работе метода *DeepSORT V2* значение показателей *F-меры* равно 0,82, *MOTA* равно 0,75. При реализации разработанного метода значение *F-меры* равно 0,92, а *MOTA* –

0,89. Такой результат получен вследствие того, что в методе *DeepSORT V2* нейросеть натренирована на внешний вид людей, а разработанный метод учитывает локальные признаки людей, в частности, распознавание их лиц. Помимо этого, важной особенностью является то, что предложенный метод способен проводить анализ панорамного изображения, сформированного из видеопотоков сразу нескольких камер, тогда как метод *DeepSORT V2* предназначен для обработки кадров лишь одного видеопотока.

Заключение

Разработан, обучен и протестирован метод детектирования, распознавания и повторной идентификации людей в многокамерных системах видеоаналитики, основанный на гибридной архитектуре адаптированных сверточных нейронных сетей *Resnet-50* и *YOLOv8*. Экспериментальная валидация проведена на расширенном наборе данных, включающем в себя стандартные бенчмарки *Market-1501* и *CUHK03*, специализированные датасеты *MARS* и *iLIDS-VID* и собственный набор аннотированных данных людей (*MirITeam500K*). Данный метод адаптируется к обнаружению и сопровождению транспортных средств, где в качестве локальных признаков используются государственные регистрационные номера. Отметим, что данный метод сопровождения людей применим также и для режима реального времени, посредством оптимизации количества пула кадров видеопотока, в результате которого анализ происходит с учетом скорости изменения сцен.

Библиографический список

1. **Игнатъева, С.А.** Принципы организации и анализ подходов к повышению точности повторной идентификации людей в распределенных системах видеонаблюдения // Вестник Полоцкого государственного университета. Серия С. Фундаментальные науки. 2022. № 4. С. 13-25.
2. **Николаев, Д.А.** Архитектура программного обеспечения для многокамерного сопровождения в системах видеонаблюдения с децентрализованной структурой. Моделирование, оптимизация и информационные технологии. 2020;8(4). [Электронный ресурс] // Режим доступа: <https://moitvvt.ru/ru/journal/pdf?id=856>
3. **Багиров, М.Б.** Метод автоматизированной калибровки камер и сшивки видеоизображений в многокамерных системах видеонаблюдения / М.Б. Багиров, Э.С. Соколова // Труды НГТУ им. Р.Е. Алексеева. 2023. № 4. С. 7-19.
4. **Зотов, С.С.** Обнаружение объектов в реальном времени с помощью алгоритмов распознавания YOLO / С.С. Зотов, А.А. Яковлев, Д.А. Колчинцев // Международный научный журнал «Синергия наук». 2018. № 26.
5. **Mehra, A.** Understanding YOLOv8 Architecture, Applications & Features // [Электронный ресурс] – Режим доступа: <https://www.labellerr.com/blog/understanding-yolov8-architecture-applications-features>
6. **Narayan, N.** Person Re-identification for Improved Multi-person Multi-camera Tracking by Continuous Entity Association. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 64-70.

*Дата поступления
в редакцию: 19.07.2024*

*Дата принятия
к публикации: 23.04.2025*