

## Автоматический захват и сопровождение объектов интереса в видеоданных с глобальным движением

Н. А. Обухова<sup>✉</sup>, А. А. Мотыко, А. А. Чиркунова, А. А. Поздеев, Е. А. Литвинов

Санкт-Петербургский государственный электротехнический университет "ЛЭТИ"  
им. В. И. Ульянова (Ленина), Санкт-Петербург, Россия

<sup>✉</sup> natalia172419@yandex.ru

### Аннотация

**Введение.** Автоматический захват и сопровождение движущихся объектов в видеоданных, получаемых видеокamerой, установленной на подвижном носителе, является сегодня одной из самых востребованных задач. К факторам, затрудняющим ее успешное решение, относятся сложная фоновая обстановка, наличие ситуаций перекрытия объектов фоном и друг другом, существенное и быстрое изменение размеров объекта интереса, существенно нелинейная траектория с резкими изменениями направления движения подвижного носителя, на котором установлен сенсор.

**Цель работы.** Разработать метод автоматического захвата и сопровождения движущихся объектов в видеоданных, полученных в сложных условиях наблюдения. Дополнительным требованием к методу на этапе сопровождения является ограничение на вычислительные ресурсы.

**Материалы и методы.** Автоматический захват основан на сверточной нейронной сети с архитектурой YOLOv8. Сопровождение объектов реализовано без нейросетевых решений. Для обеспечения устойчивого сопровождения использованы одновременно несколько детекторов с последующим анализом получаемых ими данных. Применен детектор на основе гистограмм ориентированных градиентов (HOG), дополненный детектором на основе корреляционной фильтрации и предсказанием траектории движения на основе фильтра Калмана.

**Результаты.** На этапе автоматического захвата значение оценки вероятности правильного обнаружения TPR равно 0.81, оценка вероятности ложной тревоги параметра FPR соответствует 0.10. На этапе сопровождения интенсивность отказов (срывов сопровождения)  $6 \cdot 10^{-5}$ .

**Заключение.** Предложенный метод позволяет обнаруживать и успешно сопровождать объекты на расстоянии 1500 м при размере проекции объекта на плоскость кадра  $5 \times 5$  пикселей в условиях глобального движения, сложного фона и существенной динамики свойств объекта интереса.

**Ключевые слова:** автоматический захват, сопровождение объектов интереса, глобальное движение, фильтр Калмана, гистограмма ориентированных градиентов, корреляционное сопровождение

**Для цитирования:** Автоматический захват и сопровождение объектов интереса в видеоданных с глобальным движением / Н. А. Обухова, А. А. Мотыко, А. А. Чиркунова, А. А. Поздеев, Е. А. Литвинов // Изв. вузов России. Радиоэлектроника. 2024. Т. 27, № 5. С. 24–40. doi: 10.32603/1993-8985-2024-27-5-24-40

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

**Источник финансирования.** Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (Минобрнауки России) в рамках реализации комплексного проекта по созданию высокотехнологичного производства по теме "Мультимодальный комплекс контроля воздушного пространства аэропорта" (Соглашение о предоставлении субсидии федерального бюджета на развитие кооперации государственного научного учреждения и организации реального сектора экономики в целях реализации комплексного проекта по созданию высокотехнологичного производства № 075-11-2023-007 от 10.02.2023 г.) и в рамках Постановления Правительства РФ от 9 апреля 2010 г. № 218. Работа выполнена на базе Федерального государственного автономного образовательного учреждения высшего образования "Санкт-Петербургский государственный электротехнический университет "ЛЭТИ" им. В. И. Ульянова (Ленина)" (СПбГЭТУ "ЛЭТИ").

Статья поступила в редакцию 12.09.2024; принята к публикации после рецензирования 18.10.2024; опубликована онлайн 29.11.2024

## Automatic Detection and Tracking of Objects of Interest in Video Data with Global Motion

Natalia A. Obukhova<sup>✉</sup>, Alexander A. Motyko,  
Anastasia A. Chirkunova, Alexander A. Pozdeev, Elisey A. Litvinov

Saint Petersburg Electrotechnical University, St Petersburg, Russia

<sup>✉</sup> natalia172419@yandex.ru

### Abstract

**Introduction.** At present, automatic capture and tracking of moving objects in video data obtained by a video camera mounted on a mobile carrier represents a relevant research task. Its successful solution is challenged by such factors, as a non-uniform background, object overlapping between one another and the background, significant and rapid changes in the size of the object of interest, abrupt changes in the movement trajectory of the mobile carrier.

**Aim.** To develop an automatic method for detecting moving objects followed by their tracking in video data obtained under difficult observation conditions. An additional requirement imposed on the tracking stage consists in the restriction of computing resources.

**Materials and methods.** The method is based on a convolutional neural network with a YOLO architecture. Due to the restriction of computing resources, object tracking is implemented without neural network solutions. In order to ensure stable tracking, several detectors are used simultaneously with the subsequent analysis of the data obtained. The tracking stage involves a detector based on histograms of oriented gradients (HOG), supplemented by a detector based on correlation filtering and motion trajectory prediction based on the Kalman filter.

**Results.** At the automatic detection stage, the TPR, averaged over all video files participating in the experiments, was equal to 0.81, with the FPR corresponding to 0.10. At the tracking stage, the failure rate (tracking failures) was  $6 \cdot 10^{-5}$ .

**Conclusion.** The proposed method can be successfully used to detect and track objects at a distance of 1500 m with an object projection size on the frame of  $5 \times 5$  pixels under the conditions of global motion, a non-uniform background, and significant changes in the properties of the object of interest.

**Keywords:** automatic detection, objects of interest tracking, global motion, Kalman filter, histogram of oriented gradients, correlation tracking

**For citation:** Obukhova N. A., Motyko A. A., Chirkunova A. A., Pozdeev A. A., Litvinov E. A. Automatic Detection and Tracking of Objects of Interest in Video Data with Global Motion. Journal of the Russian Universities. Radioelectronics. 2024, vol. 27, no. 5, pp. 24–40. doi: 10.32603/1993-8985-2024-27-5-24-40

**Conflict of interest.** The authors declare no conflicts of interest.

**Source of funding.** The work was carried out with the financial support of the Ministry of Science and Higher Education of the Russian Federation as part of the implementation of a comprehensive project to create high-tech production on the topic "Multimodal complex for airport airspace control" (Agreement on the provision of subsidies from the federal budget for the development of cooperation between a state scientific institution and organization of the real sector of the economy in order to implement a comprehensive project for the creation of high-tech production № 075-11-2023-007 dated 02.10.2023) and within the framework of the Decree of the Government of the Russian Federation of April 9, 2010 № 218. The work was carried out on the basis of the Federal State Autonomous educational institution of higher education " Saint Petersburg Electrotechnical University (ETU).

Submitted 12.09.2024; accepted 18.10.2024; published online 29.11.2024

**Введение.** В последние годы телевизионным системам и оптоэлектронным комплексам автоматического захвата и сопровождения движущихся объектов уделяется повышенное внимание. В значительной мере это обусловле-

но появлением высококачественных видеокамер, а также новых аппаратных решений с большой вычислительной мощностью и высоким быстродействием. Новые возможности позволили в реальном масштабе времени реа-

лизовать сложные методы обработки изображений, повысившие тактико-технические характеристики телевизионных систем. В то же время проблема автоматического захвата и сопровождения движущихся объектов в видеоданных остается по-прежнему одной из самых сложных и не полностью решенных.

В данной статье предлагается метод автоматического захвата и сопровождения неточечных объектов интереса в условиях сложной фоновой обстановки, при наличии ситуаций перекрытия объектов фоном и друг другом (окклюзии объектов), а также при пониженной освещенности. Существенными особенностями решаемой задачи являются:

- размещение видеокамеры на подвижном носителе (движение последнего имеет сложную траекторию с резкими изменениями направления);
- существенное и быстрое изменение размеров объекта интереса от  $5 \times 5$  элементов разложения до  $200 \times 200$  элементов разложения (при разрешении  $1920 \times 1080$ ).

Дополнительным требованием на этапе сопровождения является ограничение на вычислительные ресурсы.

**Автоматический захват.** Под результатом захвата (обнаружением объекта интереса) будем понимать выделение объекта стробом, центр которого соответствует центру масс пикселей проекции объекта в плоскости кадра. Границы строба представляют собой прямоугольник, описанный вокруг изображения объекта.

Автоматическое обнаружение объектов интереса может быть реализовано методами, основанными на дискриминантных признаках (яркость, цвет, текстура и движение), и методами глубокого обучения.

При обнаружении объектов по яркостному признаку предполагают наличие однородного фона. Обнаружение объектов интереса на сложном фоне вызывает существенные трудности [1, 2].

Текстурный признак и совокупность разрабатываемых на его основе алгоритмов [3–7] позволяют решать задачи автоматического захвата и устойчивого сопровождения на неоднородном фоне, наиболее эффективным представляется метод на основе гистограммы ориентированных градиентов [8].

В телевизионных системах обнаружения и сопровождения информация об объекте представле-

на в виде набора кадров. Это позволяет использовать для обнаружения признаков движения, а также проследить динамику свойств объекта интереса.

В видеоряде детектирование движущихся объектов может быть решено путем вычитания статичного фона [9–11]. Методы на основе вычитания фона сегментируют изображение на передний план и фон. Передний план содержит движущиеся объекты, такие как люди, автомобили, в то время как фон содержит статические объекты, такие как дорога, здания, деревья, стоящие автомобили и т. д.

В этом подходе сначала захватывают опорное фоновое изображение, когда объекты интереса отсутствуют в сцене. Движущийся объект обнаруживают путем вычитания текущего кадра изображения из опорного фонового изображения. На практике фон любой сцены постепенно меняется со временем, поэтому опорное фоновое изображение периодически обновляют, чтобы избежать ложного обнаружения объектов. В этом случае критичным становится принятие решения об обновлении фона. Более эффективными являются методы, автоматически обновляющие оценку фона в процессе работы, например на основе фильтра Калмана.

Детектирование объектов на основе признака достаточно часто реализуют на основании абсолютной межкадровой разности ТВ-сигналов – так называемой энергии движения.

Абсолютная межкадровая разность – это скалярная оценка. С ее помощью невозможно разделить изображения объектов, находящихся в непосредственной близости друг к другу, а также разрешить ситуацию их перекрытия (окклюзию объектов). Существенную проблему вызывает присутствие в кадре глобального движения [12].

Альтернативным способом оценки признака движения являются векторы движения [12–14]. После проецирования на плоскость кадра реальное трехмерное движение объектов отображается двумерным движением, которое может быть оценено дискретными смещениями фрагментов изображения – полем оптического потока или полем векторов движения. Наличие информации о направлении и величине смещения позволяет:

- обнаруживать объекты на сложном фоне, находящиеся в непосредственной близости друг к другу;

– разрешать ситуации окклюзии при сопровождении путем выявления объекта, находящегося на переднем плане.

Главной проблемой применения векторов оптического потока является высокая вероятность получения аномальных, не отражающих истинное движение, векторов, которые возникают на слаботекстурированных фрагментах изображения [15, 16].

На современном этапе наилучшие характеристики при решении задач обнаружения демонстрируют методы, основанные на применении сверточных нейронных сетей [17].

Сверточную нейронную сеть можно представить в виде двух последовательно выполняемых блоков: автоматического экстрактора признаков из изображений и решающего блока, который реализует целевой анализ (обнаружение, классификацию, идентификацию и пр.). Иллюстрация для типовой архитектуры сети (на примере YOLOv8) приведена на рис. 1.

С помощью сверток с многочисленными ядрами фильтров выделяются признаки (backbone часть), которые затем подаются в оконечную часть сети, решающую целевую задачу.

Для автоматической оценки параметров фильтров экстрактора признаков и решающего блока (РБ) необходима обучающая выборка.

Она представляет собой размеченные видеоданные, в которых для серии входных изображений заранее известны целевые ответы. Следует отметить, что для обучения современных сверточных нейронных сетей объем обучающей выборки должен быть достаточно большим: обычно решение типовой промышленной задачи требует минимум 50–100 тысяч размеченных изображений.

Среди нейросетевых архитектур, предназначенных для обнаружения на данный момент, выделяются архитектуры типа YOLO (последняя версия – 11, активно используются версии 8–11). Архитектура YOLO обеспечивает наилучший компромисс между точностью обнаружения и скоростью работы [18]. На рис. 2 представлены сравнительные характеристики различных поколений моделей, полученные на аппаратном обеспечении GPU T4, с использованием программного обеспечения TensorRT 10.

Из приведенных положений следует, что наиболее эффективным в рамках решаемой задачи является использование сверточных нейронных сетей, текстурного признака с оценкой на основе гистограммы ориентированных градиентов и признака движения, оцениваемого на основе оптического потока.

Эффективность перечисленных подходов была оценена для реализации автоматического

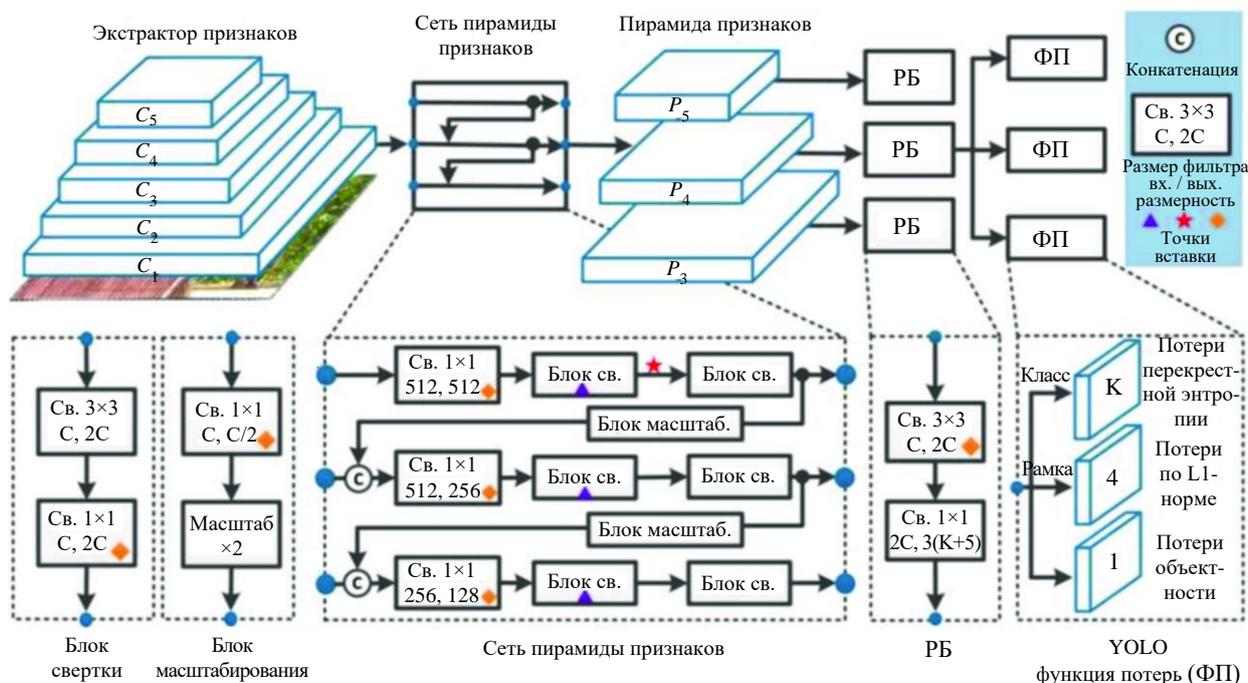


Рис. 1. Общая схема сверточной нейронной сети на примере архитектуры YOLOv8

Fig. 1. General scheme of a convolutional neural network using the YOLOv8 architecture as an example

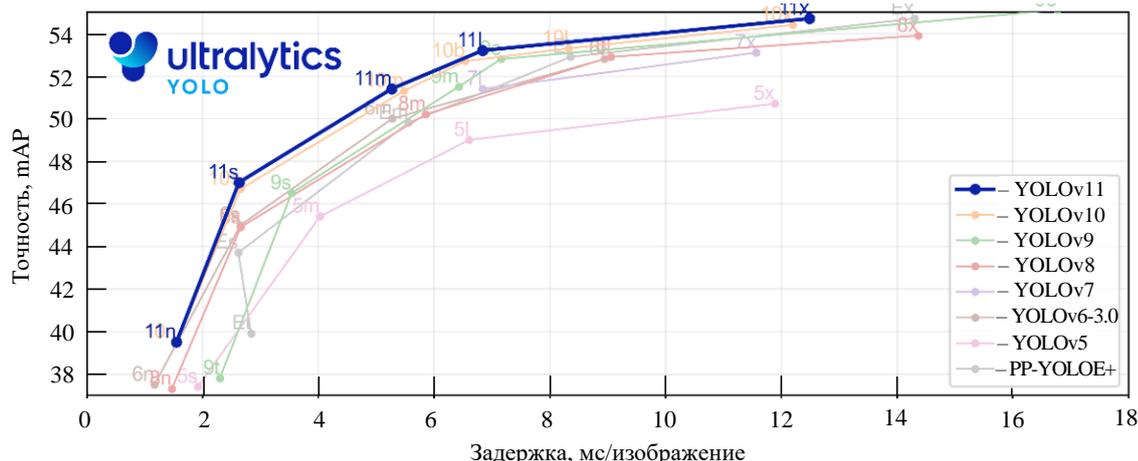


Рис. 2. Характеристики (точность/быстродействие) различных архитектур YOLO

Fig. 2. Characteristics (accuracy/performance) of various YOLO architectures

захвата в рамках решаемой задачи. В анализе не использовались методы, основанные на яркостном признаке, в силу существенной неоднородности фона в решаемой задаче и, соответственно, априорно низкой эффективности их применения. Присутствие выраженного глобального движения, обусловленного резкими маневрами носителя, на котором установлен сенсор, позволяет исключить из рассмотрения методы, основанные на вычитании фона и на энергии движения. Таким образом, сравнительный анализ был выполнен для методов, основанных на текстурном признаке, а именно на гистограмме ориентированных градиентов, векторах оптического потока и сверточных нейронных сетях.

На основе полученных результатов моделирования (оценивалась метрика mAP) для автоматического обнаружения объекта интереса было выбрано нейросетевое решение – архитектура YOLOv8.

Разрешение кадров, получаемых от видеокамеры, соответствует стандарту FHD. Использование изображений большого размера для обучения модели неэффективно: возникает проблема с нехваткой памяти, процедура осуществляется медленно, результирующая модель получается "тяжелой". Масштабирование снимков, традиционно используемое в подобных случаях, невозможно, так как в результате теряются малоразмерные объекты интереса и впоследствии модель не сможет их обнаруживать.

В этой связи модель обнаружения была обучена на снимках стандартного для архитектуры YOLOv8 размера –  $320 \times 320$  пикселей.

Данные изображения были получены нарезкой исходных снимков на фрагменты. Алгоритм фрагментирования содержал специальную логику для недопущения "разрезания" целевых объектов. Таким образом, фрагменты формировались вокруг имевшихся в обучающей выборке размеченных объектов интереса. Дополнительно для обеспечения сбалансированности выборки подготавливались негативные фрагменты, содержащие фон различного вида при отсутствии объектов интереса.

Исходный набор содержал более 120 000 снимков, включающих данные от сенсоров различного типа. После процедуры декорреляции объем данных тренировочной выборки составил 43 000 снимков, объем валидационной выборки – 10 000.

После фрагментации под размер  $320 \times 320$  объем тренировочной выборки составил 55 000, объем валидационной выборки – 15 000 снимков.

После автоматизированного анализа выборки и очистки от данных с некорректной или неточной разметкой выход составил 27 000 изображений для тренировки модели и 8000 – для валидации. После добавления негативных примеров итоговый объем составил: 87 000 – для тренировки, 20 000 – для валидации.

В результате оценки точностных параметров модели по валидационным выборкам были получены следующие результаты. Значение F-меры (гармоническое среднее между точностью и полнотой, одна из основных метрик для оценки качества модели) соответствует 0.8. На рис. 3 приведен график, иллюстрирующий зависимость точность–

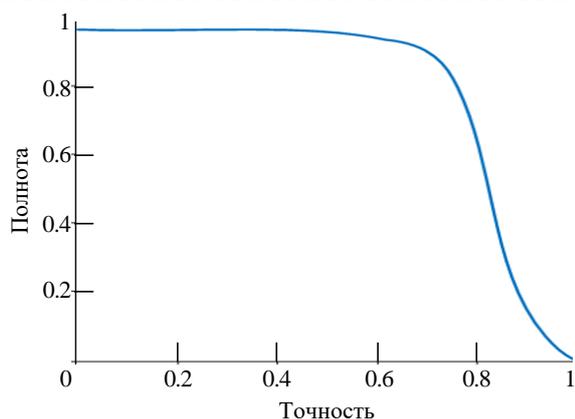


Рис. 3. Кривая зависимости точность–полнота для модели, осуществляющей обнаружение

Fig. 3. The accuracy–completeness dependence curve for the model performing the detection

полнота для модели [19]. Следует отметить, что рассматриваемая модель представляет собой одноклассовый детектор объектов интереса.

Полученные точностные характеристики достаточно высокие с учетом того, что результат детектирования не является "финальным". На основе результатов нейросетевого детектора формируется строб для дальнейшего автоматического сопровождения объекта интереса.

**Сопровождение объектов интереса. Основные принципы метода автоматического сопровождения объектов интереса на сложном фоне.** Анализ работ в области видеоаналитики позволяет выделить два основных подхода к задаче сопровождения: алгоритмы, построенные на сверточных нейронных сетях, и не нейросетевые "классические" алгоритмы [20]. Алгоритмы, использующие нейронные сети, показывают хорошие результаты, однако их выигрыш по сравнению с аналогами не столь существенен, как для задач обнаружения и классификации. При этом сверточные нейронные сети требуют значительных вычислительных ресурсов, что является проблемой при развертывании подобных решений в реальных системах с ресурсными ограничениями. Поэтому для решения сформулированной задачи в рамках ограничений на вычислительную сложность были использованы подходы, не использующие нейронные сети.

Среди таких подходов можно выделить подходы на основе корреляционной фильтрации и гистограмм ориентированных градиентов.

При корреляционной фильтрации сопровождение осуществляется сравнением текущего изображения с эталонным. Эталонное изображение записывается в момент формирования оператором (или алгоритмом автоматического детектирования) строба для объекта интереса и содержит видеоинформацию не только об объекте, но и об окружающем его фоне в пределах строба. Отклонение текущего положения объекта от предшествующего и сигнал ошибки определяют сравнением эталонного и текущего изображений с помощью корреляционной фильтрации (корреляционной решающей функции).

При реализации сопровождения необходимо не только выполнить детектирование объекта, но и построить модель изменения параметров объекта интереса. Данная модель позволяет предсказать изменение параметров объекта во времени, а также выполнить сопровождение "по памяти" в случае окклюзии. Наиболее эффективным инструментом для реализации сопровождения является фильтр Калмана. Широкое применение фильтра Калмана в задачах сопровождения можно объяснить способностью извлечения информации из зашумленных данных. Математически в данной модели реализовано два этапа вычислений. На первом этапе (этап прогнозирования) предсказывают значение положения объекта на основе предыдущих значений. Второй этап (этап корректировки) предполагает уточнение предсказанного значения и оценку ошибки. Уточнение происходит на основе измерения текущих данных об объекте интереса.

Для решаемой задачи в двумерном пространстве вектор состояния системы (объекта интереса) и вектор управления имеют вид

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix};$$
$$\mathbf{u}_k = \begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix},$$

где  $x, y$  – координаты центра тяжести объекта интереса;  $\dot{x}, \dot{y}$  – скорость объекта интереса;  $\ddot{x}, \ddot{y}$  – ускорение объекта интереса.

Уравнения этапа прогнозирования имеют вид

$$\hat{\mathbf{x}}_{k|k-1} = F_k \mathbf{x}_{k-1|k-1} + B_k \mathbf{u}_k;$$

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k,$$

где  $\hat{\mathbf{x}}_k$  – прогнозируемый вектор состояния системы;  $F_k$  – матрица состояния системы;  $B_k$  – матрица управления;  $P_k$  – ковариационная матрица, которая описывает связь между положением объекта и его скоростью;  $Q_k$  – ковариационная матрица процесса.

Матрицы состояния системы  $F_k$  и управления  $B_k$  определены следующим образом:

$$F_k = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix};$$

$$B_k = \begin{bmatrix} \Delta t^2/2 & 0 \\ 0 & \Delta t^2/2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix},$$

где  $\Delta t$  – временной интервал между соседними состояниями системы.

Уравнения на этапе корректировки предсказания имеют вид

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k \tilde{\mathbf{y}}_k;$$

$$P_{k|k} = [1 - K_k H_k] P_{k|k-1};$$

$$\tilde{\mathbf{y}}_k = \mathbf{z}_k - H_k \hat{\mathbf{x}}_{k|k-1},$$

где  $K_k$  – коэффициент усиления Калмана;  $\mathbf{z}_k$  – полученное положение объекта интереса в результате детектирования;  $H$  – матрица, характеризующая поступление информации с разных датчиков. В рамках решаемой задачи:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Коэффициент усиления Калмана:

$$K_k = P_{k|k-1} H_k^T S_k^{-1};$$

$$S_k = H_k P_{k|k-1} H_k^T + R_k.$$

Матрицу шума измерений  $R$  инициализируют квадратами отклонений по соответствующей координате:

$$R = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}.$$

Особенностями решаемой задачи являются сложный фон, существенная динамика свойств объекта интереса, глобальное движение в кадре, вызванное размещением видеокамеры на подвижном носителе. Одновременно с этим существуют ограничения на вычислительные ресурсы.

При условии ограничений на вычислительные ресурсы целесообразно отказаться от нейросетевого детектора, а для обеспечения устойчивого сопровождения использовать несколько детекторов одновременно. В качестве базы на этапе сопровождения использован детектор на основе гистограмм ориентированных градиентов (HOG), дополненный детектором на основе корреляционной фильтрации и предсказанием траектории движения на основе фильтра Калмана. Необходимость дополнения детектора на основе HOG корреляционным детектором обусловлена неустойчивой работой HOG-детектора при резкой динамике размеров объекта интереса.

Таким образом, метод сопровождения на каждом кадре имеет три набора данных для анализа: результат корреляционного совмещения, результат выделения на основе гистограмм ориентированных градиентов (HOG-детектора) и предсказанные оценки параметров объекта интереса на основе фильтра Калмана. Каждый набор данных включает в себя положение центра тяжести и размеры объекта интереса.

Сравнение результатов детектирования и предсказания по модели на основе фильтра Калмана проводят в соответствии с мерой Intersection Over Union (IOU):

$$\text{IOU} = \frac{|A \cap B|}{|A \cup B|},$$

где  $A$  и  $B$  – прямоугольные области, представляющие собой результат детектирования или предсказания;  $|\cdot|$  – мощность соответствующего множества (площадь прямоугольной области);  $\cap$ ,  $\cup$  – пересечение и объединение множеств соответственно.

Для обеспечения устойчивого сопровождения в условиях резкого изменения размеров объекта интереса и динамичного маневрирования носителя, на котором установлен видео-

датчик, предложен следующий алгоритм анализа полученных данных.

Если совпадают данные обоих детекторов и оценки, предсказанные фильтром Калмана, то проводится переинициализация корреляционного детектора. Под переинициализацией будем понимать обновление содержания эталонного изображения на основе текущих видеоданных. Если данные, полученные детектором на основе корреляционного совмещения и HOG-детектором, совпадают, но существенно расходятся с оценками, предсказанными фильтром Калмана, то это означает присутствие выраженного глобального движения (резкий маневр носителя) и требует переинициализации параметров фильтра Калмана. Если совпадают данные, полученные с фильтра Калмана и корреляционного детектора, но при этом результаты HOG и корреляционного детекторов имеют расхождение, это означает приближение объекта интереса со значительным увеличением его размеров. В этом случае для предотвращения потери объекта интереса сопровождение реализуют только на основе корреляционного фильтра и проводят переинициализацию HOG-детектора. Переход на корреляционное сопровождение реализуют по пороговому ограничению:

$$\text{area\_thr} \cdot S_c > S_{tr},$$

где  $\text{area\_thr}$  – пороговое ограничение;  $S_c$  – площадь отклика корреляционного детектора;  $S_{tr}$  – площадь отклика HOG-детектора.

Значение  $\text{area\_thr}$  определено из следующих соображений. Высокое значение порога  $\text{area\_thr}$  позволяет повысить качество сопровождения при приближении цели, но существенно снижает устойчивость сопровождения на сложном фоне. В этом случае часто происходит переключение на корреляционный трекер и переинициализация CSRT-трекера. Поэтому величину порога рекомендуется сделать равной 0.6–0.7, чтобы обеспечить баланс между качеством сопровождения при быстром приближении объекта и на сложном фоне.

Условием прекращения сопровождения является отсутствие совпадений наборов данных от детекторов и фильтра Калмана более чем в  $N$  кадрах. Для фиксации этой ситуации преду-

смотрен специальный счетчик  $N_k$ . Значение счетчика  $N_k$  увеличивается, если все три набора данных для анализа (результат работы HOG и корреляционного детекторов, а также фильтра Калмана) не совпадают и сопровождение выполняется на основе фильтра Калмана (сопровождение "по памяти"). Если обнаружено совпадение и для дальнейшего сопровождения возвращен результат одного из детекторов, то счетчик  $N_k$  обнуляется.

Структурная схема предложенного алгоритма приведена на рис. 4.

**Особенности программной реализации предложенного метода сопровождения.** В программном решении предложенного алгоритма для реализации HOG-детектора использован Channel and Spatial Reliability Tracker (CSRT) [8]. Для описания детектируемого объекта в CSRT помимо HOG используются признаки Color Names. HOG отражает информацию о текстуре и форме, а Color Names – о цвете, что помогает использовать более полное представление об объекте. Применение Color Names вместо классического трехкомпонентного описания позволяет добиться большей устойчивости к изменениям в цвете фона или условиям освещения.

Основная идея CSRT лежит в расчете пространственной и канальной надежности. Под каналами в данном случае понимаются каналы признаков HOG и Color Names [21]. Авторы [22] используют несколько HOG-каналов для более точного представления текстурных и контурных особенностей объекта. Разные каналы могут включать разные ориентации градиентов и различные размеры ячеек. Для каждого канала рассчитывается весовой коэффициент, характеризующий его степень надежности для отслеживания объекта.

Карта пространственной надежности отражает "уверенность" в том, что определенные области объекта являются надежными для отслеживания. Эта карта помогает алгоритму фокусироваться на тех областях объекта, которые лучше соответствуют оригинальным признакам, и игнорировать области, которые могут быть искажены или скрыты, а также лучше отслеживать объекты непрямоугольной формы.

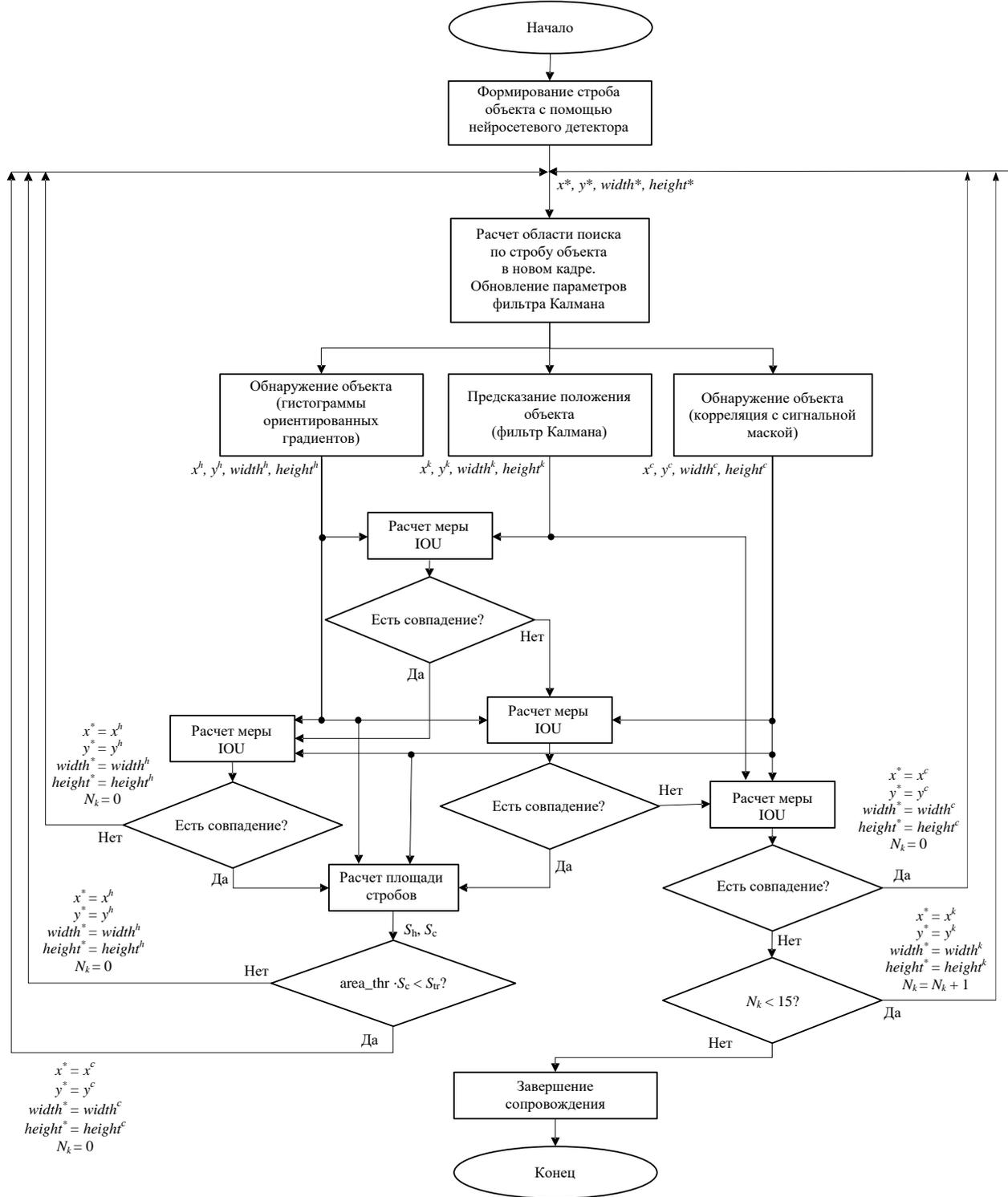


Рис. 4. Структурная схема алгоритма сопровождения

Fig. 4. Block diagram of the tracking algorithm

В качестве корреляционного детектора при реализации предложенного метода был использован KCF (Kernelized Correlation Filters) [23]. Работа KCF организована на применении корреляционных фильтров, которые обучаются на основе примеров объекта в разных положениях и

масштабах. В отличие от классического корреляционного метода KCF обладает высокой вычислительной эффективностью за счет использования свойств циклической матрицы и быстрого преобразования Фурье (FFT) [23]. Одной из ключевых особенностей KCF является использование

"ядерного трюка", который позволяет преобразовывать входные данные в более высокоразмерное пространство, где объекты легче различимы.

Проведенные эксперименты показали, что лучшее качество сопровождения при увеличении размеров цели обеспечивает KCF-трекер, а сопровождение объекта на сложном фоне – CSRT-трекер. Высокая надежность сопровождения объекта на сложном фоне обусловлена встроенным в CSRT-трекер алгоритмом обучения. Такой подход, в свою очередь, приводит к тому, что переинициализация CSRT-трекера снижает точность на нескольких первых кадрах, поэтому данную процедуру следует выполнять только при существенной необходимости.

Для повышения надежности сопровождения после переинициализации CSRT-трекера предлагается следующая процедура. В качестве основного принимается результат корреляционного трекера без переинициализации CSRT-трекера. В случае если результаты KCF и CSRT совпадают при обработке последующих кадров, основным снова становится CSRT-трекер. В противном случае происходит его повторная переинициализация.

**Экспериментальное исследование.** Основная задача экспериментального исследования метода автоматического захвата – оценить вероятность правильного обнаружения на заданной дальности. Предельная дальность, для которой выполнялась оценка вероятности правильного обнаружения, – 1500 м.

Были подготовлены видеоданные со следующими характеристиками:

- размер объекта интереса  $30 \times 30 \times 30$  см;
- площадь объекта интереса от  $5 \times 5$  до  $200 \times 200$  пикселей для формата  $1920 \times 1080$  пикселей при отношении сигнал/шум не менее 40 дБ, яркостном контрасте объекта интереса к фону не менее 15 %;

- метеорологическая дальность видимости не менее 1500 м;

- уровень освещенности не менее 500 лк.

При съемке видеоданных использовалась тестовая камера с разрешением  $1920 \times 1080$  с переменным фокусным расстоянием и углами поля зрения от  $2.3$  до  $63^\circ$ . Объект интереса имел размеры  $20 \times 18 \times 5.5$  см.

При съемке видеоданных фокусное расстояние объектива камеры изменялось таким образом, чтобы угол поля зрения соответствовал значению из табл. 1. При этом размеры объекта интереса в пикселях на дистанции 100 м соответствуют размерам цели (не менее  $30 \times 30 \times 30$  см) на дистанции 1500 м при использовании камеры с узким углом обзора.

Съемка видеоданных для задачи предельной дистанции обнаружения проводится на двух дистанциях согласно табл. 1. Примеры кадров из тестовых видеопоследовательностей представлены на рис. 5.

Для каждого видеофайла выполнялось автоматическое обнаружение объекта интереса в каждом кадре. По результатам анализа кадра записывались следующие результаты: номер кадра и наличие/отсутствие фактов обнаружения. При обнаружении объекта интереса записываются его координаты.

По полученным данным рассчитывается оценка вероятности правильного обнаружения TPR, оценка вероятности ложной тревоги FPR на заданной дальности обнаружения:

$$TPR = TP/QP;$$

$$FPR = FP/Q,$$

где TP – обнаруженные объекты интереса; QP – общее число объектов интереса; FP – число ложных срабатываний детектора; Q – общее число детектированных объектов. Указанные параметры считаются нарастающим итогом для кадров видеоряда.

Табл. 1. Параметры дальности и размеров объекта для экспериментального исследования

Tab. 1. Parameters of the range and size of the object for experimental research

| Параметр                                  | Параметры камеры обнаружения | Параметры тестовой камеры в режиме обнаружения |
|---|------------------------------|--|
| Угол поля зрения, ...°                    | 6                            | 27.5   |
| Дальность № 1, м                          | 1500                         | 100  |
| Размер объекта на дистанции № 1, пикселей | 7.3                          | 7.35   |
| Дальность № 2, м                          | 2230                         | 150  |
| Размер объекта на дистанции № 2, пикселей | 4.9                          | 4.9  |



Рис. 5. Примеры кадров из тестовых видеопоследовательностей

Fig. 5. Examples of frames from test video sequences

Полученные результаты приведены в табл. 2. На этапе сопровождения в качестве основной характеристики рассматривалось число срывов сопровождения. Под срывом в данном исследовании будем понимать потерю объекта с последующей необходимостью выполнения

Табл. 2. Данные тестов для оценки вероятности правильного обнаружения на предельной дальности 1500 м

Tab. 2. Test data to assess the probability of correct detection at a maximum range of 1500 m

| Видео   | Объект | TP   | FP  | QP   | Q    | TPR  | FPR  |
|---|--------|------|-----|------|------|------|------|
| 1. День, объект на фоне облачного неба                      | 1      | 1149 | 20  | 1151 | 1169 | 0.99 | 0.02 |
| 2. Сумерки, объект на фоне облачного неба и леса            | 1      | 2506 | 278 | 2784 | 2784 | 0.90 | 0.10 |
| 3. Сумерки, объект на фоне растительности                   | 1      | 46   | 2   | 74   | 48   | 0.62 | 0.04 |
| 4. Сумерки, 2 объекта на динамическом фоне (ветер, листва): |        |      |     |      |      |      |      |
| объект 1  | 1      | 49   | 36  | 56   | 85   | 0.88 | 0.42 |
| объект 2  | 2      | 76   | 9   | 88   | 85   | 0.86 | 0.11 |
| 5. День, объект на низкодетальном фоне, дымка               | 1      | 339  | 6   | 699  | 345  | 0.48 | 0.02 |
| 6. День, 2 объекта на высокодетальном фоне, дымка:          |        |      |     |      |      |      |      |
| объект 1  | 1      | 1601 | 114 | 1722 | 1715 | 0.93 | 0.07 |
| объект 2  | 2      | 1388 | 79  | 1722 | 1567 | 0.81 | 0.05 |
| 7. День, объект пересекает линию горизонта (небо и лес)     | 1      | 234  | 21  | 298  | 261  | 0.78 | 0.08 |

Табл. 3. Результаты экспериментального исследования

Tab. 3. Experimental results

| № видео | Описание   | Число кадров | Число срывов (предложенный алгоритм) | Число срывов (SAM) |
|---------|--|--------------|--------------------------------------|--------------------|
| 1       | Объект на простом фоне (небо)  | 2987         | 0                                    | 0                  |
| 2       | Объект на простом фоне (небо), присутствует кратковременный перелет объекта через линию горизонта на сложный фон   | 9108         | 0                                    | 0                  |
| 3       | Объект на сложном фоне (лес, дома)   | 1906         | 1                                    | 1                  |
| 4       | Объект на сложном фоне (лес, поле, дорога)   | 633          | 0                                    | 0                  |
| 5       | Объект на простом фоне (небо), существенное изменение размеров объекта   | 1160         | 0                                    | 0                  |
| 6       | Объект на сложном фоне (лес, поле), присутствуют моменты, когда объект сливается с фоном   | 1237         | 0                                    | 1                  |
| 7       | Объект на сложном фоне (небо, лес, дорога), присутствуют моменты с резким глобальным движением, изменением размеров объекта  | 1110         | 4                                    | 3                  |
| 8       | Объект на сложном фоне (лес, небо), присутствуют моменты, когда объект сливается с фоном   | 926          | 2                                    | 1                  |
| 9       | Объект на сложном фоне (лес, поле, дорога), присутствует глобальное движение, значимое изменение размеров объекта, повороты объекта интереса и носителя камеры   | 702          | 2                                    | 2                  |
| 10      | Объект на простом фоне (небо), низкое значение ОСШ, присутствуют помехи, появление "теней" объектов в результате рассогласования каналов, глобальное движение, существенное изменение размеров объекта | 675          | 4                                    | 2                  |

автоматического захвата с нейросетевым детектором. В качестве конкурирующего решения для проведения сравнительного исследования был использован новейший алгоритм на базе технологии глубокого обучения – SAM [24]. Данный нейросетевой алгоритм на сегодняшний день является одним из лидеров по точностным характеристикам и пользуется популярностью у исследователей.

Результаты экспериментального исследования 10 видеороликов приведены в табл. 3. Как видно из описания, на видеороликах присутствуют ситуации, возникающие в реальных усло-

виях и затрудняющие процесс сопровождения.

Примеры кадров с существенным увеличением размеров объекта интереса и кадров с резкой сменой направления движения из-за маневра носителя приведены на рис. 6 и 7.

Полученные данные табл. 3 позволяют рассчитать показатель интенсивности отказов (срывов сопровождения) [25]:

$$F = Ft / N,$$

где  $Ft$  – число срывов сопровождения;  $N$  – общее число кадров видеоряда.

Соответственно, по данным экспериментов,



Рис. 6. Пример изменения размеров объекта

Fig. 6. Example of resizing the object



Рис. 7. Пример значительного глобального движения в результате поворота и смены направления носителя

Fig. 7. Example of significant global motion as a result of turning and changing the direction of the carrier

предложенный алгоритм имеет интенсивность отказов  $6 \cdot 10^{-4}$ . При этом аналогичный параметр у конкурирующего решения (SAM) составил  $5 \cdot 10^{-4}$ . Как видно по полученным значениям и табл. 3, оба алгоритма демонстрируют близкие результаты. SAM показывает относительно лучшую точность, однако с практической точки зрения характеристики двух алгоритмов (достигнутые значения интенсивности отказов) сопоставимы.

**Заключение.** Предложенный метод автоматического захвата и сопровождения позволяет обнаруживать объект на дальности 1500 м при размере проекции объекта на плоскость кадра от  $5 \times 5$  элементов разложения. Усредненное значение по всем видеофайлам серии оценки вероятности правильного обнаружения TPR соответствует значению 0.81, оценка вероятности ложной тревоги параметра FPR соответствует значению 0.10.

Экспериментальное исследование показало, что алгоритм сопровождения демонстрирует хорошие результаты на простом и сложном фоне. Срывы сопровождения в основном возникают в случаях значительных поворотов

объекта интереса при изменении его размеров, на границах перехода с одного фона на другой. Срывы могут происходить при наличии нескольких сложных ситуаций одновременно, которые не критичны по отдельности, например резкое глобальное движение и помеха. Предложенный алгоритм сопровождения обладает сопоставимыми с практической точки зрения точностными характеристиками с передовым современным аналогом, построенным на базе технологий глубокого обучения. При этом предложенный алгоритм существенно проще в развертывании на целевой аппаратуре: не требует графических ускорителей, отличается высоким быстродействием, простотой и гибкостью настройки под специфику и характеристики системы.

Достигнутые результаты обусловлены использованием многопризнаковой детекции с последующей обработкой, позволяющей выявить моменты маневрирования носителя и другие сложные ситуации. К достоинствам предложенного метода также следует отнести низкую вычислительную емкость (нейросетевой детектор присутствует только на этапе автоматического захвата).

#### Авторский вклад

**Обухова Наталия Александровна** – разработка метода и алгоритмов автоматического захвата и траекторного сопровождения объектов интереса.

**Мотыко Александр Александрович** – разработка метода и алгоритмов автоматического захвата и траекторного сопровождения объектов интереса.

**Чиркунова Анастасия Анатольевна** – разработка метода и алгоритмов автоматического захвата и траекторного сопровождения объектов интереса.

**Поздеев Александр Анатольевич** – разработка метода и алгоритмов автоматического захвата и траекторного сопровождения объектов интереса.

**Литвинов Елисей Александрович** – моделирование алгоритмов автоматического захвата и траекторного сопровождения объектов интереса.

### Author's contribution

**Natalia A. Obukhova**, development of a method and algorithms for automatic capture and trajectory tracking of objects of interest.

**Alexander A. Motyko**, development of a method and algorithms for automatic capture and trajectory tracking of objects of interest.

**Anastasia A. Chirkunova**, development of a method and algorithms for automatic capture and trajectory tracking of objects of interest.

**Alexander A. Pozdeev**, development of a method and algorithms for automatic capture and trajectory tracking of objects of interest.

**Elisey A. Litvinov**, implementation of the algorithms for automatic capture and trajectory tracking of objects of interest.

### Список литературы

1. Yu Zhang, Xiangzhi Bai, Tao Wang. Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure // *Information Fusion*. 2017. Vol. 35. P. 81–101. doi: 10.1016/j.inffus.2016.09.006
2. Hossain M. D., Chen D. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective // *ISPRS J. of Photogrammetry and Remote Sensing*. 2019. Vol. 150. P. 115–134. doi: 10.1016/j.isprsjprs.2019.02.009
3. Segmentation of Natural Images by Texture and Boundary Compression / H. Mobahi, S. Rao, A. Yang, S. Sastry, Y. Ma // *Intern. J. of Computer Vision*. 2011. Vol. 95. P. 86–98.
4. Natural Image Segmentation with Adaptive Texture and Boundary Encoding / S. R. Rao, H. Mobahi, A. Y. Yang, S. S. Sastry, Y. Ma; Ed. by H. Zha, R.-i. Taniguchi, S. Maybank // *Computer Vision – ACCV 2009. Lecture Notes in Computer Science*. Berlin: Springer, 2010. Vol. 5994. P. 135–146. doi: 10.1007/978-3-642-12307-8\_13
5. Texture image segmentation using fused features and active contour / M. Gao, H. Chen, Sh. Zheng, B. Fang, L. Zhang // *23<sup>rd</sup> Intern. Conf. on Pattern Recognition, Cancun, Mexico, 04–08 Dec. 2016*. IEEE, 2016. P. 520–526. doi: 10.1109/ICPR.2016.7899935
6. Bouaynaya N., Charif-Chefchaoui M., Schonfeld D. Spatially Variant Morphological Restoration and Skeleton Representation // *IEEE Transactions on Image Processing*. 2006. Vol. 15, iss. 11. P. 3579–3591. doi: 10.1109/TIP.2006.877475
7. Reconfigurable architecture for computing histograms in real-time tailored to FPGA-based smart camera / L. Maggiani, C. Salvadori, M. Petracca, P. Pagano, R. Saletti // *IEEE 23<sup>rd</sup> Intern. Symp. on Industrial Electronics (ISIE), Istanbul, Turkey, 01–04 June 2014*. IEEE, 2014. P. 1042–1046. doi: 10.1109/ISIE.2014.6864756
8. Dalal N., Triggs B. Histograms of oriented gradients for human detection // *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05), San Diego, USA, 20–25 June 2005*. IEEE, 2005. Vol. 1. P. 886–893. doi: 10.1109/CVPR.2005.177
9. Kumar S., Yadav J. S. Video Object Extraction and its Tracking using Background Subtraction in Complex Environments // *Perspectives in Science*. 2016. Vol. 8. P. 317–322. doi: 10.1016/j.pisc.2016.04.064
10. Shaikh S. H, Saeed K., Chaki N. Chapter-2, Moving Object Detection Approaches, Challenges and Object Tracking // *In: Moving Object Detection using Background Subtraction*. SpringerBriefs in Computer Science. Cham: Springer, 2014. P. 5–11. doi: 10.1007/978-3-319-07386-6\_2
11. Barnich O., Droogenbroeck M. V. ViBe: A Universal Background Subtraction Algorithm for Video Sequences // *IEEE Transactions on Image Processing*. 2011. Vol. 20, iss. 6. P. 1709–1724. doi: 10.1109/TIP.2010.2101613
12. Обухова Н. А. Сегментация объектов интереса на основе признака движения в видеокomпьютерных системах // *Инфокоммуникационные технологии*. 2007. № 1. С. 77–85.
13. Aslani S., Mahdavi-Nasab H. Optical Flow Based Moving Object Detection and Tracking for Traffic Surveillance // *Intern. J. of Electrical, Computer, Energetic, Electronic and Communication Engineering*. 2013. Vol. 7, № 9. P. 1252–1256.
14. Kale K., Pawar S., Dhulekar P. Moving Object Tracking Using Optical Flow And Motion Vector Estimation // *4<sup>th</sup> Intern. Conf. on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, India, 02–04 Sept. 2015*. IEEE, 2015. P. 1–6. doi: 10.1109/ICRITO.2015.7359323
15. Обухова Н. А. Априорная оценка достоверности векторов оптического потока (векторов движения) // *Изв. вузов России. Радиоэлектроника*. 2006. № 3. С. 30–36.
16. Обухова Н. А. Векторы оптического потока в задачах сегментации и сопровождения подвижных объектов // *Изв. вузов России. Радиоэлектроника*. 2006. № 2. С. 42–51.
17. Object detection with deep learning: a review / Z.-Q. Zhao, P. Zheng, S.-T. Xu, X. Wu // *IEEE transaction on neural networks and learning systems*. 2019. Vol. 30, iss. 11. P. 3212–3232. doi: 10.1109/TNNLS.2018.2876865
18. Terven J., Córdova-Esparza D.-M., Romero-González J.-A. A. Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS // *Machine Learning and Knowledge Extraction*. 2023. № 5. P. 1680–1716. doi: 10.3390/make5040083
19. On model evaluation under non-constant class

imbalance / J. Brabec, T. Komarek, V. Franc, L. Machlica // Intern. Conf. on Computational Science. Lecture Notes in Computer Science. Cham: Springer, 2020. P. 74–87. doi: 10.1007/978-3-030-50423-6\_6

20. Verma R. A Review of Object Detection and Tracking Methods // Intern. J. of Advance Engineering and Research Development. 2017. Vol. 4, iss. 10. P. 569–578.

21. Discriminative correlation filter tracker with channel and spatial reliability / A. Lukezic, T. Vojir, L. Cehovin Zajc, J. Matas, M. Kristan // Intern. J. of Computer Vision. 2018. Vol. 126, iss. 8. P. 671–688. doi: 10.1007/s11263-017-1061-3

22. Learning color names for real-world applications / J. van de Weijer, C. Schmid, J. Verbeek, D. Lar-

lus // IEEE Trans. Image Proc. 2009. Vol. 18, iss. 7. P. 1512–1523. doi: 10.1109/TIP.2009.2019809

23. High-Speed Tracking with Kernelized Correlation Filters / J. F. Henriques, R. Caseiro, P. Martins, J. Batista // IEEE Trans on PAMI. 2015. Vol. 37, iss. 3. P. 583–596. doi: 10.1109/TPAMI.2014.2345390

24. Segment Anything / A. Kirillov, E. Mintun, N. Ravi, H. Mao, Ch. Rolland, L. Gustafson // IEEE/CVF Intern. Conf. on Computer Vision (ICCV), Paris, France, 01–06 Oct. 2023. IEEE, 2023. P. 3992–4003. doi: 10.1109/ICCV51070.2023.00371

25. Метрики оценки алгоритмов автоматического сопровождения / А. Е. Щелкунов, В. В. Ковалев, К. И. Морев, И. В. Сидько // Изв. ЮФУ. Технические науки. 2020. № 1. С. 233–245. doi: 10.18522/2311-3103-2020-1-233-245

### Информация об авторах

**Обухова Наталия Александровна** – доктор технических наук (2009), декан факультета радиотехники и телекоммуникаций, зав. кафедрой телевидения и видеотехники Санкт-Петербургского государственного электротехнического университета "ЛЭТИ" им. В. И. Ульянова (Ленина). Автор более 130 научных работ. Сфера научных интересов – компьютерное зрение и видеоаналитика; машинное обучение и цифровая обработка изображений; видеосистемы и системы поддержки принятия решений; smart-технологии формирования изображений.

Адрес: Санкт-Петербургский государственный электротехнический университет "ЛЭТИ" им. В. И. Ульянова (Ленина), ул. Проф. Попова, д. 5 Ф, Санкт-Петербург, 197022, Россия

E-mail: natalia172419@yandex.ru

<https://orcid.org/0000-0003-1953-2085>

**Мотыко Александр Александрович** – кандидат технических наук (2012), доцент кафедры телевидения и видеотехники Санкт-Петербургского государственного электротехнического университета "ЛЭТИ" им. В. И. Ульянова (Ленина). Автор более 80 научных работ. Сфера научных интересов – компьютерное зрение; колориметрия; глубокое обучение.

Адрес: Санкт-Петербургский государственный электротехнический университет "ЛЭТИ" им. В. И. Ульянова (Ленина), ул. Проф. Попова, д. 5 Ф, Санкт-Петербург, 197022, Россия

E-mail: motyko.alexandr@yandex.ru

<https://orcid.org/0000-0003-4241-4298>

**Чиркунова Анастасия Анатольевна** – кандидат технических наук (2017), доцент кафедры телевидения и видеотехники Санкт-Петербургского государственного электротехнического университета "ЛЭТИ" им. В. И. Ульянова (Ленина). Автор более 40 научных работ. Сфера научных интересов – телевизионные системы космического и специального назначения; цифровая обработка изображений; системы формирования, обработки и отображения видеоинформации.

Адрес: Санкт-Петербургский государственный электротехнический университет "ЛЭТИ" им. В. И. Ульянова (Ленина), ул. Проф. Попова, д. 5 Ф, Санкт-Петербург, 197022, Россия

E-mail: aachirkunova@etu.ru

<https://orcid.org/0009-0001-7550-2887>

**Поздеев Александр Анатольевич** – кандидат технических наук (2023), доцент кафедры телевидения и видеотехники Санкт-Петербургского государственного электротехнического университета "ЛЭТИ" им. В. И. Ульянова (Ленина). Автор более 20 научных работ. Сфера научных интересов – цифровая обработка изображений; машинное обучение и интеллектуальный анализ данных; колориметрия.

Адрес: Санкт-Петербургский государственный электротехнический университет "ЛЭТИ" им. В. И. Ульянова (Ленина), ул. Проф. Попова, д. 5 Ф, Санкт-Петербург, 197022, Россия

E-mail: puches4@gmail.com

<https://orcid.org/0000-0003-0003-4051>

**Литвинов Елисей Александрович** – магистр по направлению "Радиотехника" (2024), аспирант (2024) кафедры телевидения и видеотехники Санкт-Петербургского государственного электротехнического университета "ЛЭТИ" им. В. И. Ульянова (Ленина). Автор трёх научных публикаций. Сфера научных интересов – цифровая обработка изображений; прикладные телевизионные системы; разработка встраиваемых систем.

Адрес: Санкт-Петербургский государственный электротехнический университет "ЛЭТИ" им. В. И. Ульянова (Ленина), ул. Проф. Попова, д. 5 Ф, Санкт-Петербург, 197022, Россия

E-mail: lelisej@mail.ru

<https://orcid.org/0009-0001-1387-989X>

## References

1. Yu Zhang, Xiangzhi Bai, Tao Wang. Boundary Finding Based Multi-Focus Image Fusion through Multi-Scale Morphological Focus-Measure. *Information Fusion*. 2017, vol. 35, pp. 81–101. doi: 10.1016/j.inffus.2016.09.006
2. Hossain M. D., Chen D. Segmentation for Object-Based Image Analysis (OBIA): A review of Algorithms and Challenges from Remote Sensing Perspective. *ISPRS J. of Photogrammetry and Remote Sensing*. 2019, vol. 150, pp. 115–134. doi: 10.1016/j.isprsjprs.2019.02.009
3. Mobahi H., Rao S., Yang A., Sastry S., Ma Y. Segmentation of Natural Images by Texture and Boundary Compression. *Intern. J. of Computer Vision*. 2011, vol. 95, pp. 86–98.
4. Rao S. R., Mobahi H., Yang A. Y., Sastry S. S., Ma Y. Natural Image Segmentation with Adaptive Texture and Boundary Encoding. *Computer Vision – ACCV 2009. Lecture Notes in Computer Science*. Ed. by H. Zha, R.-i. Taniguchi, S. Maybank. Berlin, Springer, 2010, vol. 5994, pp. 135–146. doi: 10.1007/978-3-642-12307-8\_13
5. Gao M., Chen H., Zheng Sh., Fang B., Zhang L. Texture Image Segmentation Using Fused Features and Active Contour. 23<sup>rd</sup> Intern. Conf. on Pattern Recognition, Cancun, Mexico, 04–08 Dec. 2016. IEEE, 2016, pp. 520–526. doi: 10.1109/ICPR.2016.7899935
6. Bouaynaya N., Charif-Chefchaoui M., Schonfeld D. Spatially Variant Morphological Restoration and Skeleton Representation. *IEEE Transactions on Image Processing*. 2006, vol. 15, iss. 11, pp. 3579–3591. doi: 10.1109/TIP.2006.877475
7. Maggiani L., Salvadori C., Petracca M., Pagano P., Saletti R. Reconfigurable Architecture for Computing Histograms in Real-Time Tailored to FPGA-Based Smart Camera. *IEEE 23<sup>rd</sup> Intern. Symp. on Industrial Electronics (ISIE)*, Istanbul, Turkey, 01–04 June 2014. IEEE, 2014, pp. 1042–1046. doi: 10.1109/ISIE.2014.6864756
8. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, USA, 20–25 June 2005. IEEE, 2005, vol. 1, pp. 886–893. doi: 10.1109/CVPR.2005.177
9. Kumar S., Yadav J. S. Video Object Extraction and its Tracking using Background Subtraction in Complex Environments. *Perspectives in Science*. 2016, vol. 8, pp. 317–322. doi: 10.1016/j.pisc.2016.04.064
10. Shaikh S. H., Saeed K., Chaki N. Chapter-2, Moving Object Detection Approaches, Challenges and Object Tracking. In: *Moving Object Detection using Background Subtraction*. SpringerBriefs in Computer Science. Cham, Springer, 2014, pp. 5–11. doi: 10.1007/978-3-319-07386-6\_2
11. Barnich O., Droogenbroeck M. V. ViBe: A Universal Background Subtraction Algorithm for Video Sequences. *IEEE Transactions on Image Processing*. 2011, vol. 20, iss. 6, pp. 1709–1724. doi: 10.1109/TIP.2010.2101613
12. Obuhova N. A. Segmentation of Objects of Interest Based on Motion Feature in Video Computer Systems. *Infocommunication Technologies*. 2007, no. 1, pp. 77–85. (In Russ.)
13. Aslani S., Mahdavi-Nasab H. Optical Flow Based Moving Object Detection and Tracking for Traffic Surveillance. *Intern. J. of Electrical, Computer, Energetic, Electronic and Communication Engineering*. 2013, vol. 7, no. 9, pp. 1252–1256.
14. Kale K., Pawar S., Dhulekar P. Moving Object Tracking Using Optical Flow And Motion Vector Estimation. 4<sup>th</sup> Intern. Conf. on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, India, 02–04 Sept. 2015. IEEE, 2015, pp. 1–6. doi: 10.1109/ICRITO.2015.7359323
15. Obukhova N. A. Apriority Priori Estimation of Optical Flow Vectors (Motion Vectors) Reliability. *J. of the Russian Universities. Radioelectronics*. 2006, no. 3, pp. 30–36. (In Russ.)
16. Obukhova N. A. Optical Flow Vectors in Tasks of Moving Objects Segmentation and Tracking. *J. of the Russian Universities. Radioelectronics*. 2006, no. 2, pp. 42–51. (In Russ.)
17. Zhao Z.-Q., Zheng P., Xu S.-T., Wu X. Object Detection with Deep Learning: a Review. *IEEE Transaction on Neural Networks and Learning Systems*. 2019, vol. 30, iss. 11, pp. 3212–3232. doi: 10.1109/TNNLS.2018.2876865
18. Terven J., Córdova-Esparza D.-M., Romero-González J.-A. A. Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*. 2023, no. 5, pp. 1680–1716. doi: 10.3390/make5040083
19. Brabec J., Komarek T., Franc V., Machlica L. On Model Evaluation under Non-Constant Class Imbalance. *Intern. Conf. on Computational Science. Lecture Notes in Computer Science*. Cham, Springer, 2020, pp. 74–87. doi: 10.1007/978-3-030-50423-6\_6
20. Verma R. A Review of Object Detection and Tracking Methods. *Intern. J. of Advance Engineering and Research Development*. 2017, vol. 4, iss. 10, pp. 569–578.
21. Lukezic A., Vojir T., Cehovin Zajc L., Matas J., Kristan M. Discriminative Correlation Filter Tracker with Channel and Spatial Reliability. *Intern. J. of Computer Vision*. 2018, vol. 126, iss. 8, pp. 671–688. doi: 10.1007/s11263-017-1061-3
22. Van de Weijer J., Schmid C., Verbeek J., Larlus D. Learning Color Names for Real-World Applications. *IEEE Trans. Image Proc.* 2009, vol. 18, iss. 7, pp. 1512–1523. doi: 10.1109/TIP.2009.2019809
23. Henriques J. F., Caseiro R., Martins P., Batista J. High-Speed Tracking with Kernelized Correlation Filters. *IEEE Trans on PAMI*. 2015, vol. 37, iss. 3, pp. 583–596. doi: 10.1109/TPAMI.2014.2345390
24. Kirillov A., Mintun E., Ravi N., Mao H., Rolland Ch., Gustafson L. Segment Anything. *IEEE/CVF Intern. Conf. on Computer Vision (ICCV)*, Paris,

France, 01–06 Oct. 2023. IEEE, 2023, pp. 3992–4003.  
doi: 10.1109/ICCV51070.2023.00371

25. Shchelkunov A. E., Kovalev V. V., Morev K. I.,  
Sidko I. V. The Metrics for Tracking Algorithms Evalu-

ation. *Izvestiya SFEDU. Engineering Sciences*. 2020,  
no. 1, pp. 233–245. doi: 10.18522/2311-3103-2020-1-  
233-245 (In Russ.)

### **Information about the authors**

**Natalia A. Obukhova** – Dr Sci. in Engineering (2009), Dean of the Faculty of Radio Engineering and Telecommunications, Head of the Department of Television and Video Engineering of Saint Petersburg Electrotechnical University. The author of more than 130 scientific publications. Area of expertise: computer vision and video analytics; machine learning and digital image processing; video systems and decision support systems; smart imaging technologies.

Address: Saint Petersburg Electrotechnical University, 5 F, Professor Popov St., St Petersburg 197022, Russia

E-mail: natalia172419@yandex.ru

<https://orcid.org/0000-0003-1953-2085>

**Alexander A. Motyko** – Cand. Sci. (Eng.) (2012), Associate Professor of Television and Video Engineering of Saint Petersburg Electrotechnical University. The author of more than 80 scientific publications. Area of expertise: computer vision; colorimetry; deep learning.

Address: Saint Petersburg Electrotechnical University, 5 F, Professor Popov St., St Petersburg 197022, Russia

E-mail: motyko.alexandr@yandex.ru

<https://orcid.org/0000-0003-4241-4298>

**Anastasia A. Chirkunova** – Cand. Sci. (Eng.) (2017), Associate Professor of Television and Video Engineering of Saint Petersburg Electrotechnical University. The author of more than 40 scientific publications. Area of expertise: space and special purpose television systems; digital image processing; video information generation, processing and display systems.

Address: Saint Petersburg Electrotechnical University, 5 F, Professor Popov St., St Petersburg 197022, Russia

E-mail: aachirkunova@etu.ru

<https://orcid.org/0009-0001-7550-2887>

**Alexander A. Pozdeev** – Cand. Sci. (Eng.) (2023), Associate Professor of Television and Video Engineering of Saint Petersburg Electrotechnical University. The author of more than 20 scientific publications. Area of expertise: digital image processing; machine learning and data mining; colorimetry

Address: Saint Petersburg Electrotechnical University, 5 F, Professor Popov St., St Petersburg 197022, Russia

E-mail: puches4@gmail.com

<https://orcid.org/0000-0003-0003-4051>

**Elisey A. Litvinov**, Master's degree in Radio Engineering, postgraduate student of Television and Video Engineering of Saint Petersburg Electrotechnical University. The author of three scientific publications. Area of expertise: digital image processing; applied television systems; development of embedded systems.

Address: Saint Petersburg Electrotechnical University, 5 F, Professor Popov St., St Petersburg 197022, Russia

E-mail: lelisej@mail.ru

<https://orcid.org/0009-0001-1387-989X>

---