

# ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ, РЕЧИ, ТЕКСТА И РАСПОЗНАВАНИЕ ОБРАЗОВ

## SIGNAL, IMAGE, SPEECH, TEXT PROCESSING AND PATTERN RECOGNITION



УДК 004.932  
<https://doi.org/10.37661/1816-0301-2021-18-1-43-60>

Оригинальная статья  
Original Paper

### Обнаружение и сопровождение объектов на видеопоследовательностях: формализация, критерии и результаты

Р. П. Богущ<sup>1</sup>✉, С. В. Абламейко<sup>2,3</sup>

<sup>1</sup>Полоцкий государственный университет,  
ул. Блохина, 29, Новополоцк, 211440, Беларусь  
✉E-mail: bogushr@mail.ru

<sup>2</sup>Белорусский государственный университет,  
пр. Независимости, 4, Минск, 220030, Беларусь

<sup>3</sup>Объединенный институт проблем информатики  
Национальной академии наук Беларуси,  
ул. Сурганова, 6, Минск, 220012, Беларусь

**Аннотация.** Одним из перспективных направлений развития и внедрения искусственного интеллекта является автоматическое обнаружение и отслеживание движущихся объектов в системах видеонаблюдения. В работе представлена формализация обнаружения и сопровождения одного и множества объектов на видеопоследовательностях. Рассмотрены критерии, характеризующие качество обнаружения сопровождаемых объектов, точность определения местоположения объекта на кадре, траекторию движения и точность сопровождения множества объектов. На основе рассмотренного обобщения разработан алгоритм сопровождения людей, использующий сверточные нейронные сети для детектирования людей и формирования признаков. Нейросетевые признаки включены в составной дескриптор, содержащий также геометрические и цветовые характеристики для описания каждого обнаруженного человека в кадре. Приведены результаты экспериментов на основе рассмотренных критериев, экспериментально подтверждено, что улучшение работы детектора позволяет повысить точность сопровождения объектов. Представлены примеры кадров обработанных видеопоследовательностей с визуализацией траекторий движения людей.

**Ключевые слова:** видеонаблюдение, движущийся объект, сверточная нейронная сеть, сопровождение через обнаружение, траектория движения

**Для цитирования.** Богущ, Р. П. Обнаружение и сопровождение объектов на видеопоследовательностях: формализация, критерии и результаты / Р. П. Богущ, С. В. Абламейко // Информатика. – 2021. – Т. 18, № 1. – С. 43–60. <https://doi.org/10.37661/1816-0301-2021-18-1-43-60>

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

---

Поступила в редакцию | Received 11.01.2021  
Принята к публикации | Accepted 10.02.2021  
Опубликована | Published 26.03.2021

## Object detection and tracking in video sequences: formalization, metrics and results

Rykhard P. Bohush<sup>1✉</sup>, Sergey V. Ablameyko<sup>2,3</sup>

<sup>1</sup>*Polotsk State University,  
st. Blokhin, 29, Novopolotsk, 211440, Belarus*  
✉E-mail: bogushr@mail.ru

<sup>2</sup>*Belarusian State University,  
av. Nezaliezhnasti, 4, 220030, Minsk, Belarus*

<sup>3</sup>*The United Institute of Informatics Problems  
of the National Academy of Sciences of Belarus,  
st. Surganova, 6, Minsk, 220012, Belarus*

**Abstract.** One of the promising areas of development and implementation of artificial intelligence is the automatic detection and tracking of moving objects in video sequence. The paper presents a formalization of the detection and tracking of one and many objects in video. The following metrics are considered: the quality of detection of tracked objects, the accuracy of determining the location of the object in a frame, the trajectory of movement, the accuracy of tracking multiple objects. Based on the considered generalization, an algorithm for tracking people has been developed that uses the tracking through detection method and convolutional neural networks to detect people and form features. Neural network features are included in a composite descriptor that also contains geometric and color features to describe each detected person in the frame. The results of experiments based on the considered criteria are presented, and it is experimentally confirmed that the improvement of the detector operation makes it possible to increase the accuracy of tracking objects. Examples of frames of processed video sequences with visualization of human movement trajectories are presented.

**Keywords:** video surveillance, moving object, convolutional Neural Network, tracking by detection, motion trajectory

**For citation.** Bohush R. P., Ablameyko S. V. Object detection and tracking in video sequences: formalization, metrics and results. *Informatics*, 2021, vol. 18, no. 1, pp. 43–60 (in Russian). <https://doi.org/10.37661/1816-0301-2021-18-1-43-60>

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

**Введение.** Обнаружение и сопровождение объектов на видеопоследовательностях относятся к основным задачам в «компьютерном зрении», которые имеют различное количество технических применений и все шире будут использоваться для распознавания типа активности человека в системах мониторинга и охраны, сопровождения объектов в системах технического зрения на производстве, оценки движения людей при реабилитации в медицине, анализа окружающей обстановки в автоматизированных системах вождения транспортных средств [1]. В отличие от изображений видеопоследовательности содержат значительно больший объем информации, который изменяется и в пространстве, и во времени. Поэтому обработка и анализ последовательностей изображений позволяют выявлять не только статические, но и динамические признаки объектов, что приводит к повышению результативности автоматизированной работы систем видеонаблюдения в целом.

Наиболее изученной является задача обнаружения людей и лиц на видеопоследовательностях [2]. В работе [3] предложен алгоритм сопровождения людей на видеоряде на основе метода Монте-Карло для Марковских цепей, в работе [4] – алгоритм обнаружения и сопровождения людей в интеллектуальных детекторах внештатных ситуаций на базе метода опорных векторов. В настоящее время для обнаружения объектов широкое развитие и применение получили алгоритмы на основе сверточных нейронных сетей, которые устойчивы к изменениям освещенности, динамическому заднему фону и позволяют осуществлять детектирование даже в случае существенных перекрытий объектов [5, 6].

Сопровождение множества людей в системах видеонаблюдения представляет собой актуальную прикладную задачу, однако в настоящее время она не решена в полной мере. Это обусловлено следующими факторами: отображением на видеопоследовательности огромного множества различных вариантов изображений одного и того же человека из-за его временных переходов на сцене; неоднородным задним фоном, фрагменты которого могут быть схожи по форме, текстуре или цвету с изображениями людей; низким уровнем освещенности в помещениях; наличием теней; множественными перекрытиями движущихся людей между собой или с другими объектами; значительной схожестью признаков разных сопровождаемых людей; достаточно быстрым их движением и, в ряде случаев, изменяющимся ускорением; нелинейной траекторией движения.

Существует ряд подходов для решения задачи сопровождения, однако из-за указанных проблем их результативность является недостаточной. Этап формирования эффективного набора признаков, по которым будет осуществляться обнаружение и сопровождение объектов на видеопоследовательности, является одним из наиболее сложных, так как для него имеются ограничения: следует использовать признаки, которые заранее могут быть получены для описания объектов; требуется определить ограниченный набор признаков, который позволит получить максимальную результативность, т. е. необходимо исключить малоинформативные признаки; возможно применение алгоритмов, которые удовлетворяют вычислительным требованиям прикладных систем. Авторами разработан ряд решений, которые предназначены для обнаружения и сопровождения динамических объектов на видеопоследовательностях [7], людей [8], дыма и пламени [9]. Можно говорить о том, что используемый набор признаков в значительной мере связан с применяемыми алгоритмами обнаружения и сопровождения для решения поставленных задач. Поэтому для разработки эффективных методов, методик и алгоритмов обнаружения и сопровождения объектов на видеопоследовательностях необходимо провести четкую формализацию указанных задач. Необходимо определить объекты, которые будут обнаруживаться и сопровождаться, установить основные этапы данного процесса, а также критерии, позволяющие оценить качество обработки и показать, как это реализуется на практике.

В настоящей работе предлагается формализация задач обнаружения и сопровождения объектов на видеопоследовательностях. На основе рассмотренного обобщения разработан алгоритм сопровождения множества людей, использующий метод «сопровождение через обнаружение» (tracking by detection) и сверточные нейронные сети для детектирования и формирования признаков каждого человека. Соответствие между людьми на кадрах устанавливается с использованием составного дескриптора, описывающего каждого человека и содержащего нейросетевые, геометрические и цветовые признаки. Приводятся результаты экспериментов на основе рассмотренных критериев, позволяющих оценить качество работы алгоритмов. Экспериментально подтверждается, что улучшение работы детектора повышает точность сопровождения объектов.

**Формализация задачи обнаружения объектов.** Видеопоследовательность, или видеопоток, – это последовательность цифровых изображений (кадров)  $V = \{F_k\}$ , где  $k$  – номер изображения в последовательности.

Объект на изображении (Ob) представляет собой локальную область, отличающуюся от окружающего фона и отображающую некоторую часть признаков объекта реального мира.

На каждом кадре последовательности, полученной со стационарной видеокамеры, как правило, отображается множество объектов:

$$OB_{F_k} = \{Ob_q^{F_k}\}, q = 1, \dots, Q.$$

По критерию движения каждый из них может быть отнесен к двум основным классам:

– стационарный (неподвижный) объект на последовательности изображений, который описывается множеством признаков  $Ft_{Ob_q}^S$  и его координатами  $x_{Ob_q}$ ,  $y_{Ob_q}$ , не изменяющимися в течение интервала времени  $t$ . Такой объект может быть представлен формальной моделью

$$\text{Ob}_q^S = \left( Ft_{\text{Ob}_q}^S, x_{\text{Ob}_q}, y_{\text{Ob}_q}, Ns_{\text{Ob}_q}^{F_k} \right),$$

где  $\left( Ft_{\text{Ob}_q}^S, x_{\text{Ob}_q}, y_{\text{Ob}_q} \right) = \text{const} \forall F_k, k \in t$ ;  $Ns_{\text{Ob}_q}^{F_k}$  – множество возможных шумовых воздействий на объект;

– движущийся объект на последовательности изображений, характеризующийся изменением одного или нескольких основных параметров (формы, размеров, координат) в течение интервала времени  $t$ . Трансформация формы и (или) размеров объекта приводит к изменению его признаков на кадрах  $ft_{\text{Ob}_q}^{F_k}$ . Такой объект может быть представлен формальной моделью

$$\text{Ob}_q^D = \left( ft_{\text{Ob}_q}^{F_k}, x_{\text{Ob}_q}^{F_k}, y_{\text{Ob}_q}^{F_k}, Ns_{\text{Ob}_q}^{F_k} \right),$$

где  $x_{\text{Ob}_q}^{F_k}, y_{\text{Ob}_q}^{F_k}$  – координаты объекта;  $Ft_{\text{Ob}_q}^D$  – множество признаков движущегося объекта,  $Ft_{\text{Ob}_q}^D \supseteq ft_{\text{Ob}_q}^{F_k}, \forall k \in t$ . При этом  $ft_{\text{Ob}_q}^{F_k} \cap ft_{\text{Ob}_q}^{F_{k+i}}$ , т. е. для одного и того же движущегося объекта на последовательности кадров характерно изменение его признаков.

Обнаружение объекта – это определение местоположения заданного объекта  $\text{Ob}^e$  на изображении  $F$ , причем его размеры меньше размеров изображения, а количество объектов на изображении заведомо неизвестно.

В общем случае процесс обнаружения объекта  $\text{Ob}^e$  реализуется путем сравнения признаков эталона  $Ft^e$  и всех возможных фрагментов на плоскости изображения с использованием метода  $M$  с учетом правила

$$S \left( Ft^e, Ft_{\text{Ob}_q}^F \right) \xrightarrow[Z]{M} \max,$$

где  $S$  – точность обнаружения,  $Z$  – множество ограничений.

При обнаружении стационарного объекта на последовательности изображений необходимо учитывать изменчивость динамической сцены, так как кроме статических на ней присутствуют и движущиеся объекты, причем их количество может изменяться. Объекты  $\text{Ob}_q^D$  могут перекрывать стационарный объект, что приведет к изменению его признаков на последовательности кадров:

$$Ft_{\text{Ob}_q}^{S, F_k} \neq Ft_{\text{Ob}_q}^{S, F_{k+1}}.$$

Следовательно, для обнаружения стационарного объекта на последовательности изображений, отображающих динамическую сцену, необходимо использовать метод  $M_{STV}$ , учитывающий изменение признаков во времени:

$$S \left( Ft^e, Ft_{\text{Ob}_q}^{S, F_k} \right) \xrightarrow[Z_{STV}]{M_{STV}} \max,$$

где  $Z_{STV}$  – множество ограничений при обнаружении стационарного объекта на последовательности изображений.

Вместе с тем использование видеопоследовательностей позволяет более эффективно решать задачи, не доступные при обработке одного изображения. Условия съемки динамической сцены могут привести к тому, что объект будет сильно искажен на текущем кадре и его обнаружение в заданный момент времени невозможно, например, из-за существенного перекрытия движущимся объектом, которое через несколько кадров может уменьшиться до значения, позволяющего правильно обнаружить объект на изображении.

Реальное движение, которое происходит на трехмерной динамической сцене, на видеопоследовательности представляется в виде двухмерной проекции. Перемещение этой проекции обуславливает изменение яркости пикселей областей движения на кадрах.

Обнаружение движущегося объекта – это определение местоположения объекта  $Ob^D$  на текущем кадре видеопоследовательности  $F^k$  на основе заданного  $F^e$ :

$$SD(Ob_{F^k}^D, Ob_{F^e}^D) \xrightarrow{MD} \xrightarrow{ZD} \max,$$

где  $SD$  – точность обнаружения движущегося объекта,  $MD$  – применяемый метод,  $ZD$  – множество ограничений.

В качестве  $F^e$  может использоваться предыдущий кадр или кадр, не содержащий всех возможных движущихся объектов на динамической сцене. Второй подход позволит обнаруживать движущиеся объекты, которые неподвижны в течение некоторого интервала времени.

С учетом ряда прикладных задач, которые могут быть решены с использованием анализа последовательностей изображений, множества различных объектов и множества возможных комбинаций признаков для них, общие эффективные дескрипторы и метод для обнаружения определить практически невозможно. Наиболее результативными являются специализированные методы обнаружения и анализа объектов на последовательностях изображений, ориентированные на применение для конкретных прикладных задач.

**Формализация задачи сопровождения объектов.** Сопровождение движущегося объекта – определение местоположения одного и того же объекта на каждом кадре видеопоследовательности в течение интервала времени  $t$ . Благодаря этому возможно построение траектории движения объекта, анализ которой необходим при решении определенных практических задач. Например, резкое изменение траектории движения человека в помещении может свидетельствовать о его падении, что важно для системы типа «умный дом».

Для осуществления сопровождения необходимо выполнить процедуры обнаружения и локализации. Используются различные способы описания объекта наблюдения:

- одной точкой, характеризующей центр масс объекта или центр минимально возможного прямоугольника, описанного вокруг объекта;
- набором ключевых точек, по которым объект может быть однозначно опознан на последующих кадрах;
- геометрическим примитивом, описанным вокруг объекта (чаще всего прямоугольником, реже эллипсом);
- внешним контуром объекта;
- набором областей, максимально устойчивым при движении, или всей областью объекта;
- инвариантными характеристиками объекта, например текстурой, цветовой гаммой и др.

Траектория движения объекта представляет собой последовательное отображение движения (или перемещения) данного объекта на видеопоследовательности:

$$Tr(Ob^D) = (Ob_{F_k}^D), \forall k \in t.$$

В физике траекторией движения называют линию, которую описывает частица при своем движении. Очевидно, что для решаемой задачи в итоге необходимо определить линию на требуемом кадре видеопоследовательности, которая покажет, как передвигался наблюдаемый объект за определенный промежуток времени. Данная линия может быть построена на основе множества точек, в которых находился объект при своем перемещении на предыдущих кадрах видеопоследовательности, т. е. траектория всего движения вычисляется как последовательное перемещение объекта от кадра к кадру. Могут быть разные варианты определения координат объекта на кадре, но наиболее используемый подход предполагает нахождение его центра (один пиксел на кадре) с координатами  $x_{Ob_q}^{F_k}, y_{Ob_q}^{F_k}$ . Как правило, движение рассматривается в системе координат кадра. Тогда траектория движения объекта на видеопоследовательности описывается набором координат центра объекта на каждом кадре:

$$Tr(\text{Ob}^D) = (\text{Ob}_{F_k}^D) = \left( (x_{\text{Ob}_q}^{F_1}, y_{\text{Ob}_q}^{F_1}), (x_{\text{Ob}_q}^{F_2}, y_{\text{Ob}_q}^{F_2}), (x_{\text{Ob}_q}^{F_3}, y_{\text{Ob}_q}^{F_3}), \dots, (x_{\text{Ob}_q}^{F_n}, y_{\text{Ob}_q}^{F_n}) \right).$$

Желательно, чтобы траектория отображалась тонкой линией или линией единичной ширины поперечного сечения на результирующем кадре. Следует учитывать, что объекты на видеопоследовательности могут появляться не с первого кадра и выходить за пределы сцены не на последнем кадре, а в некоторый момент времени на  $n$ -м кадре. Тогда траектория движения может быть построена с кадра  $m$ , на котором обнаружен объект, и до  $n$ -го кадра. На время кратковременного скрытия движущегося объекта за другими или при невозможности его обнаружения по иным причинам траектория прерывается, но восстанавливается через несколько кадров, т. е. наблюдается ее фрагментация.

Принято различать сопровождение одного объекта (visual object tracking, VOT) и сопровождение множества объектов (multiple object tracking, MOT). Первый тип сопровождения характеризуется тем, что объект обнаруживается и локализуется на первом кадре, другие объекты не детектируются. Соответственно, требуется найти траекторию его движения  $Tr'$  с использованием метода *MTS* и ограничений *ZTS*:

$$STS(Tr') \xrightarrow{\frac{MTS}{ZTS}} \max,$$

где *STS* – точность сопровождения движущегося одиночного объекта.

Несмотря на разработку многочисленных алгоритмов, за последнее десятилетие задача VOT не завершена в полной мере из-за возможности значительного визуального изменения объекта и освещенности, помех на заднем плане, окклюзий.

При сопровождении множества объектов требуется определить множество траекторий движения  $TR' = \{Tr'_q\}$  объектов на кадрах с последующим сопоставлением их между собой для определения перемещения всех объектов между кадрами:

$$STM(TR, TR') \xrightarrow{\frac{MTM}{ZTM}} \max,$$

где *TR* – множество реальных траекторий движения объектов; *STM* – критерий, характеризующий точность сопровождения множества объектов.

В кадре могут присутствовать несколько одновременно движущихся объектов. При этом объекты могут иметь практически идентичные визуальные признаки, выходить за пределы сцены на непродолжительное время или совсем ее покидать, могут появляться новые объекты, в том числе в местах выхода предыдущих, например в дверном проеме при входе в помещение, т. е. возможна потеря сопровождаемого объекта из-за его пересечения со схожим объектом или перекрытия элементом заднего плана. Сопровождение множества объектов выполняется на длительных временных интервалах и допускается прогнозирование местоположения на последующих кадрах.

Оба типа сопровождения, VOT и MOT, на первом шаге требуют обнаружения объектов. В практических приложениях эта задача может быть выполнена и оператором путем выделения заданного объекта на первом кадре, но такой подход используется, как правило, только при необходимости сопровождения одного объекта, так как при MOT на кадрах могут появляться новые объекты, требующие трекинга.

Одной из основных проблем для практического использования алгоритма сопровождения является обеспечение высокой точности при ограниченных аппаратных ресурсах и входных данных. В общем, точное сопровождение может быть достигнуто на основе решения задачи глобальной оптимизации, которая требует наличия сразу всей последовательности кадров, что в реальных системах видеонаблюдения реализовать невозможно. В существующих методах задача сопровождения объектов часто формулируется как задача оптимизации с использованием графовых алгоритмов [10]. Каждый детектированный объект представляется в виде вершины, а переход от одной вершины к другой определяется используемой функцией сходства.

Установление ассоциации на графах может быть решено методом нахождения пути с минимальной стоимостью, в котором задача сопровождения объектов наиболее эффективно решается путем глобальной оптимизации. Алгоритмы сопровождения на основе выделения и анализа особых точек требуют обязательного наличия углов в контурах изображения. При их малом количестве эффективность сопровождения будет низкой. Использование цветовых характеристик объектов для сопровождения рассмотрено в методе из работы [11]. Вероятностные подходы используют утверждение о том, что движущийся объект имеет определенное состояние, которое измеряется на каждом кадре, и для оценки его положения на следующем кадре необходимо обобщить значения из предыдущих. Для этого применяют методы на основе фильтра Калмана [12] или фильтра частиц [13]. Однако объекты могут иметь выраженную нелинейную траекторию движения (например, в ограниченном помещении). В этом случае оценка нового состояния на основе предыдущих будет определяться с высокой погрешностью, поэтому для решения различных прикладных задач используются разные подходы.

**Критерии эффективности обнаружения и сопровождения.** Одним из важнейших вопросов при оценке результатов обнаружения и сопровождения объектов является выбор критериев оценки качества выделенных объектов. При решении указанных двух задач требуется определить их местоположение на кадрах. Однако для задачи сопровождения имеется информация о расположении объектов на всех или некоторых предыдущих кадрах. Поэтому результативность сопровождения должна определяться не только такими известными характеристиками, как точность и полнота, но и другими критериями, которые учитывают сходство объектов на разных кадрах при нахождении траектории движения. Очевидно, что при оценке алгоритмов сопровождения множества объектов для последующего их сравнения следует применять набор различных критериев. В работах [14, 15] рассмотрены метрики, характеризующие точность и эффективность алгоритмов обнаружения и сопровождения. В статье [16] приводится обзор существующих критериев для определения качества сопровождения объектов на видеопоследовательностях различными алгоритмами, рассматриваются интегральные метрики, которые направлены на получение общей оценки алгоритма сопровождения.

Для сравнения эффективности алгоритмов сопровождения множества можно выделить:

1. *Критерии, характеризующие обнаружение сопровождаемых объектов:*

точность – доля правильно обнаруженных объектов относительно всех обнаруженных объектов при сопровождении:

$$PRC = \frac{TP}{TP + FP},$$

где  $TP$  – количество правильных обнаружений сопровождаемых объектов на всех кадрах видеопоследовательности,  $FP$  – количество ложных обнаружений сопровождаемых объектов на всех кадрах видеопоследовательности;

полнота – доля правильно обнаруженных объектов относительно всех объектов, которые должны быть обнаружены при сопровождении:

$$RCL = \frac{TP}{TP + FN},$$

где  $FN$  – количество пропусков сопровождаемых объектов на всех кадрах видеопоследовательности.

Комплексный критерий обобщает точность и полноту путем вычисления их гармонического среднего за период сопровождения и позволяет таким образом оценить работу алгоритмов сопровождения по единому показателю, который может иметь максимальное значение, равное единице, и рассчитывается как

$$IDF_1 = \frac{2TP}{2TP + FP + FN}.$$

При оценке качества сопровождения используется метрика  $IoU$  (intersection over union), которая вычисляется для обнаруженной ( $T_{tr}$ ) и известной ( $T_e$ ) прямоугольных областей объекта в кадре:

$$IoU = \frac{T_{tr} \cap T_e}{T_{tr} \cup T_e}.$$

Значение  $IoU$  должно быть не менее 0,5 и индекс человека не должен изменяться при сопровождении для принятия решения о правильном сопровождении объекта ( $TP$ ).

2. *Критерий, характеризующий точность определения местоположения (точность локализации) сопровождаемых объектов на кадрах видеопоследовательности.* Определяется как средняя точность локализации объектов в кадре (multi-object tracking precision,  $MOTP$ ) при сопровождении, может иметь максимальное значение, равное единице, и рассчитывается по формуле

$$MOTP = \frac{\sum_k IOU(Ob_i, Ob_i^e)_k}{\sum_k NOb_k},$$

где  $Ob_i, Ob_i^e$  – сопровождаемый объект и соответствующий эталонный объект из базы данных;  $NOb_k$  – количество сопровождаемых объектов в кадре.

3. *Критерии, оценивающие траектории движения объектов:*

количество отслеживаемых траекторий (mostly tracked,  $MT$ ) – общее число траекторий для объектов на видеопоследовательностях, для которых полученная траектория движения соответствует заданной не менее чем на 80 % ее продолжительности;

количество частично отслеживаемых траекторий (partially tracked,  $PT$ ) – общее число траекторий для объектов на видеопоследовательностях, для которых полученная траектория движения соответствует заданной в пределах от 20 до 80 % ее продолжительности;

количество потерянных траекторий (mostly lost,  $ML$ ) – общее число объектов на видеопоследовательностях, для которых полученная траектория движения соответствует заданной менее чем на 20 % ее продолжительности;

количество фрагментаций траектории (track fragmentations,  $FM$ ) – общее число случаев, когда траектория меняет свой статус с отслеживаемой на неотслеживаемую, но отслеживание той же самой траектории возобновляется позже.

Критерии  $MT$ ,  $PT$ ,  $ML$  и  $FM$  не учитывают изменение индекса сопровождаемого объекта данной траектории.

Общее количество срывов за счет перезахвата ложного объекта при сопровождении (mismatches,  $MM$ ) – количество переключений для траекторий движения сопровождаемых объектов на другие.

На рис. 1 приведены примеры ошибочных ситуаций при сопровождении множества объектов. На третьем и седьмом кадрах показаны примеры обмена траекториями двух движущихся объектов в каждом случае. Соответственно, суммарное значение  $MM = 4$ . При  $k = 5$ , т. е. на пятом кадре, объекты  $Ob_1$  и  $Ob_4$  покидают сцену, однако рядом с  $Ob_4$  на ней появляется объект  $Ob_5$ , который по своим признакам схож с  $Ob_4$ , что приводит к ложному обнаружению, т. е. для данного кадра  $FP = 1$ . При  $k = 3$  и  $k = 4$  объект  $Ob_3$  не обнаружен, значит,  $FN = 2$ . Также наблюдается потеря его траектории и фрагментация, поскольку траектория для данного объекта восстанавливается на пятом кадре.

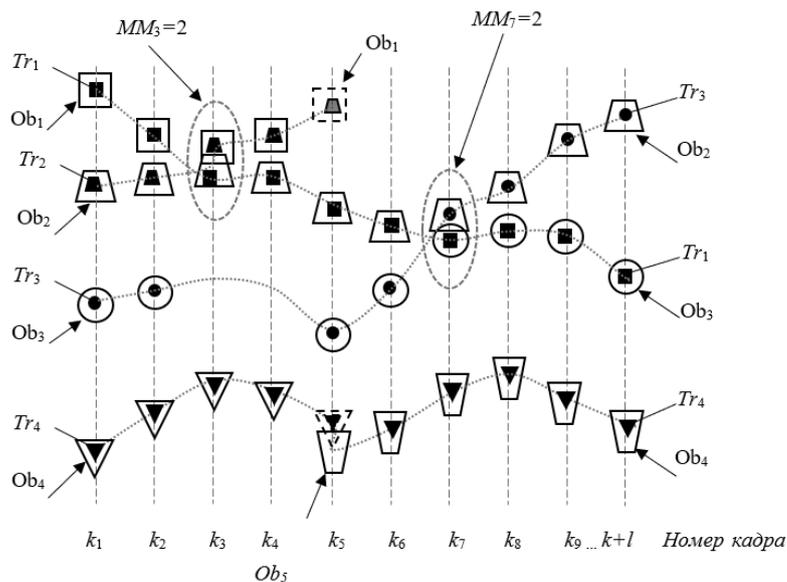


Рис. 1. Примеры ошибок при сопровождении объектов

4. *Критерий, характеризующий точность сопровождения множества объектов* (multiple object tracking accuracy, *MOTA*). Учитывает количество ложноположительных и ложноотрицательных результатов обнаружений, число обменов траекториями и характеризует точность сопровождения объектов во времени с учетом восстановления траектории при кратковременном отсутствии объекта:

$$MOTA = 1 - \frac{\sum_k FN_k + FP_k + MM_k}{\sum_k G_k},$$

где  $MM_k$  – количество срывов сопровождения за счет перезахвата ложного объекта при сопровождении на  $k$ -м кадре,  $G_k$  – количество объектов на  $k$ -м кадре.

Значение данного критерия должно стремиться к единице (максимально возможное), но может быть отрицательным в тех случаях, когда количество ошибок при сопровождении превышает количество всех объектов на сцене.

5. *Критерий, характеризующий быстродействие*. Скорость  $F_{reg}$  работы алгоритма сопровождения без учета детектирования объектов определяется максимально возможным числом кадров, которые могут быть обработаны за секунду.

**Этапы алгоритма сопровождения через обнаружение.** Для обнаружения и одновременного сопровождения людей с построением их траектории движения разработан алгоритм на основе метода «сопровождение через обнаружение». Основные этапы алгоритма показаны на рис. 2. Ключевым критерием является точность сопровождения множества объектов, которую следует максимизировать. На первом этапе выполняется задача обнаружения людей, находящихся в кадре. В последнее время наиболее часто используемым для этого инструментом являются сверточные нейронные сети (СНС). При решении данной задачи необходимо применять быстродействующую СНС, которая также должна обеспечивать высокую точность: чем точнее работает детектор, тем качественнее может быть сопровождение. Среди существующих СНС предъявляемым требованиям удовлетворяет модель YOLO [17], которая и применяется для данного шага. Выходным результатом этапа обнаружения является выделенная прямоугольная область кадра, как правило, включающая изображение человека.

Затем для выделенных областей кадра осуществляется процедура формирования признаков. Для этого предлагается составной дескриптор, включающий геометрические и СНС-признаки

всего изображения человека и верхней его части, получаемые при помощи представленной на рис. 2 архитектуры CHC FTrack, состоящей из 29 сверточных и одного полносвязного слоев. Составной дескриптор включает также индекс человека, который на последующих кадрах при правильном сопровождении не должен изменяться.

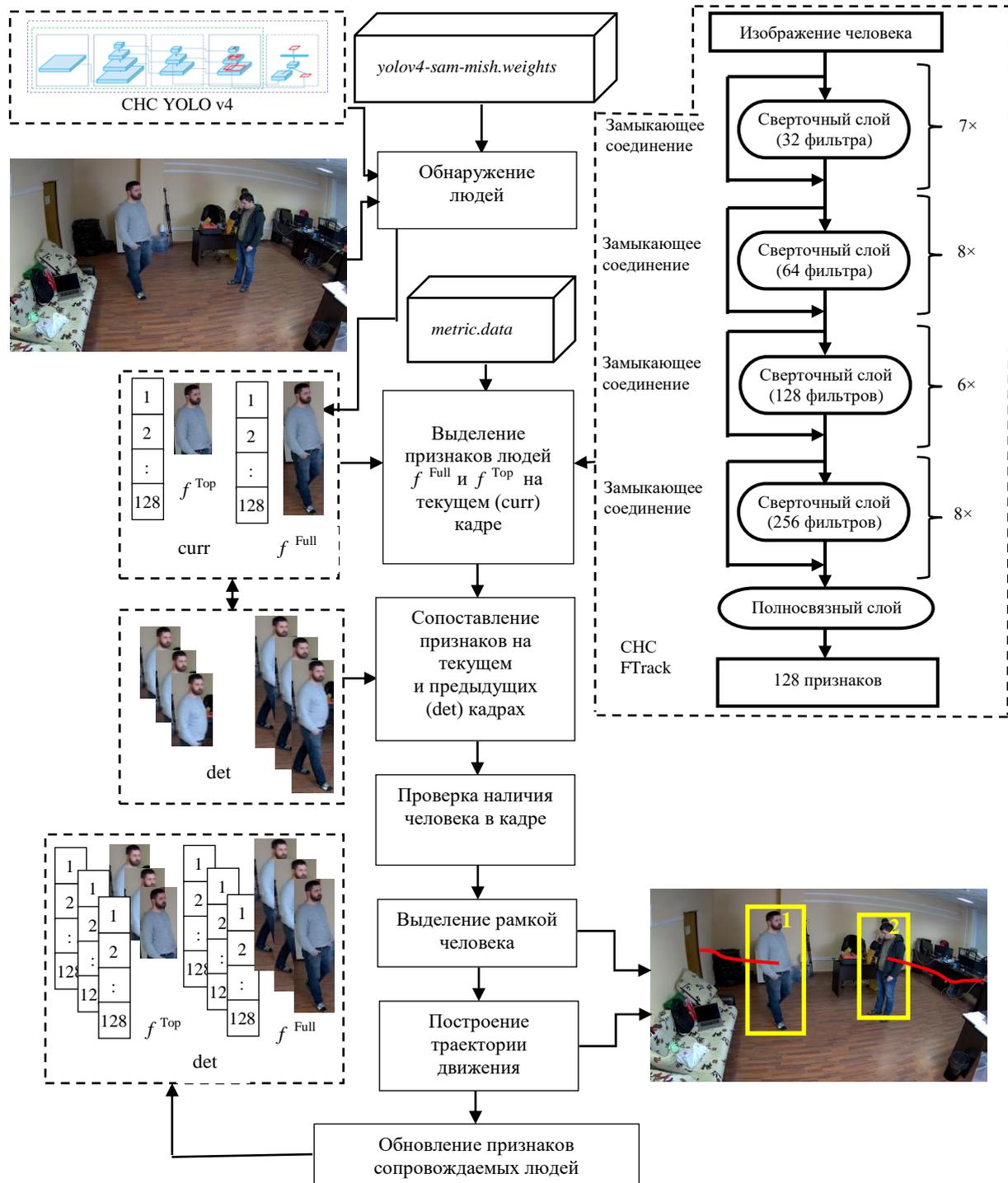


Рис. 2. Общая схема работы алгоритма, где *yolov4-sam-mish.weights* и *metric.data* – весовые коэффициенты для CHC YOLO v4 и FTrack

При движении в помещении человек может зайти за объект фона и, соответственно, признаки будут вычислены для верхней части его фигуры. Поэтому СНС-признаки вычисляются для всей и верхней половины фигуры человека, если ширина выделенного объекта меньше его высоты; в противном случае принимается решение, что полученные СНС-признаки характеризуют верхнюю часть фигуры. Далее выполняется сопоставление составных дескрипторов сопровождаемых объектов с обнаруженным детектором на текущем кадре [18].

С целью повышения результативности сопровождения сравниваются признаки для последних  $n$  верных обнаружений человека, после чего формируется матрица соответствий, которая поступает на вход венгерского алгоритма [19], применяемого для решения задачи о назначениях. Для этого осуществляется построение матрицы весовых коэффициентов. Ее элементами являются значения схожести между дескрипторами всех обнаруженных объектов на текущем кадре, количество которых определяет число столбцов, и сопровождаемых объектов с предыдущих кадров, количество которых определяет число строк. Венгерский алгоритм вычитает максимальное значение схожести из каждого элемента матрицы коэффициентов так, чтобы преобразовать ее в матрицу оптимальных назначений, в каждой строке и каждом столбце которой содержится только по одному нулю. Координаты нулей устанавливают соответствие объектов на кадрах. Эффективное решение данного этапа позволяет корректно соотносить обнаруженные объекты на текущем кадре и сопровождаемые объекты с предыдущих кадров, а также формировать правильные траектории движения для каждого из них.

Визуализация фрагмента траектории выполняется на основе вычисления кратчайшего расстояния между координатами обнаруженного объекта  $x_{Ob_q}^{F_k}, y_{Ob_q}^{F_k}$  на текущем кадре и его координатами  $x_{Ob_q}^{F_{k-1}}, y_{Ob_q}^{F_{k-1}}$  на кадре  $(k - 1)$  при последнем правильном обнаружении:

$$tr_k = \sqrt{(x_{Ob_q}^{F_k} - x_{Ob_q}^{F_{k-1}})^2 + (y_{Ob_q}^{F_k} - y_{Ob_q}^{F_{k-1}})^2}. \quad (1)$$

При обнаружении объекта на предыдущем кадре  $l = 1$ . Применение выражения (1) для каждого кадра, начиная с последующего после обнаружения объекта, позволяет получить и визуализировать всю траекторию его движения на сцене.

В случае если детектор не обнаруживает человека, для подтверждения его выхода со сцены используется анализ признаков цветowych гистограмм областей человека на кадрах. Для уменьшения влияния изменения освещенности цветовой пространство изображения преобразовывается из RGB в HSV, далее для оценки схожести используются только данные цветовой тона по условию из работы [18]. Человек считается вышедшим со сцены и его сопровождение прекращается, если значение схожести гистограммных признаков цветовой тона, вычисленное для изображений данного человека на текущем кадре и при последнем правильном его обнаружении, меньше заданного.

**Результаты исследований.** Для разметки видеопоследовательностей использовался инструмент `labelImg`, который предоставляет возможность выбора двух форматов разметки: `Yolo` и `PascalVOC`. В связи с тем что `labelImg` поддерживает только формат для детектирования

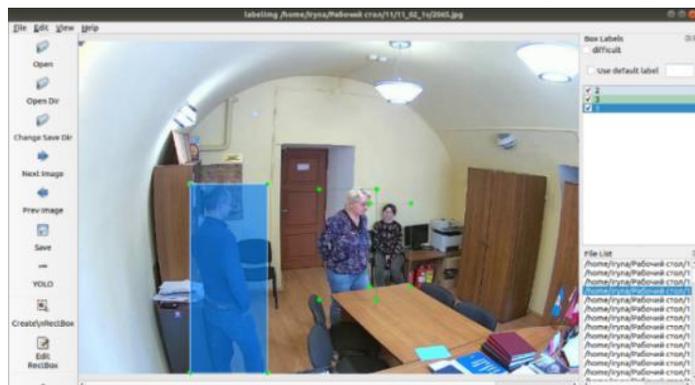


Рис. 3. Пример аннотирования видеокadra инструментом `labelImg`

объектов, видеопоследовательность разбивается на отдельные кадры. После этого на них выделяются объекты, а в качестве классов объектов используются индексы, соответствующие появлению людей на видеоряде. Затем следует объединение полученных текстовых файлов для оценки алгоритма отслеживания. На рис. 3 показан пример аннотирования кадра тестовой видеопоследовательности.

Тестирование выполнено на шести видеопоследовательностях (рис. 4) с суммарным количеством кадров 11 890, которые были получены со стационарной видекамеры в помещениях с различным освещением, нелинейной траекторией движения, полным и частичным перекрытием людей, с очень схожими внешними характеристиками, выходом людей из помещения с последующим их возвращением в кадр.

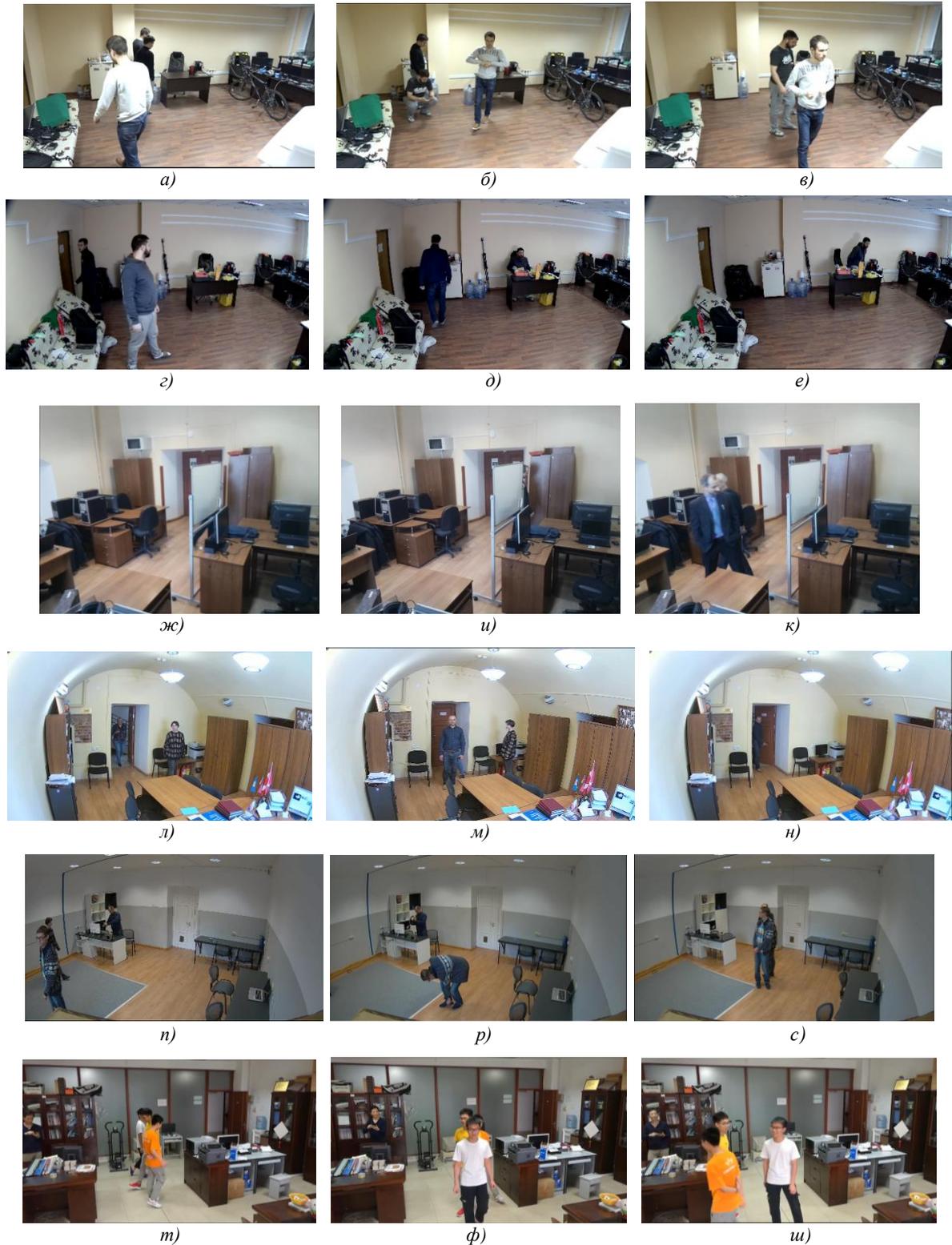


Рис. 4. Примеры кадров видеопоследовательностей

Первая тестовая видеопоследовательность (рис. 4, *a–в*) включает 2320 кадров с неравномерностью освещения и наличием теней. Влияние неравномерности освещения очевидно, если сравнить уровни яркости верхней половины фигуры человека в белом свитере на представленных изображениях из видеоряда. В кадре количество людей изменяется от одного до трех. При этом двое из них имеют очень схожие внешние характеристики: одежда практически идентична по цветовым характеристикам; рост, телосложение и цвет волос также подобны. На рис. 4, *a* и *в* видно, что два человека с наиболее схожей внешностью на многих кадрах перемещаются по сцене, перекрывая друг друга, причем на рис. 4, *в* показан пример практически полного их перекрытия в движении. На одном из кадров данной видеопоследовательности (рис. 4, *б*) человек изменил свое положение – присел.

На рис. 4, *г–е* показаны кадры второго видеоряда из 1350 изображений, который получен в том же помещении. Во время съемки на сцене присутствовали два человека, признаки изображений которых достаточно схожи. Данный видеоряд отличается более низким уровнем освещения и неоднородностью. Соответственно, когда один человек переходит в зону очень слабого освещения по направлению к двери (рис. 4, *г*) или другой человек – из ближней зоны съемки в дальнюю (рис. 4, *д*), цветовые характеристики их изображений значительно изменяются. Кроме того, на рис. 4, *д* большая часть фигуры человека скрыта объемом фона (столом). На данной видеопоследовательности достаточно четко видны тени движущихся людей, например, на стене в углу возле двери (рис. 4, *д*) и в момент выхода человека из-за стола (рис. 4, *е*). Также спинка стула на рис. 4, *е* по своим внешним признакам схожа с сидящим человеком, что является существенным шумовым фактором.

Видеокадр, представленный на рис. 4, *ж*, свидетельствует о том, что на третьем тестовом видеоряде (рис. 4, *ж–к*) люди передвигаются по помещению, скрываясь практически полностью за объектами фона (на данном кадре – за доской) и значительно удаляясь от видеокамеры. Тестовый видеоряд состоит из 1280 кадров. Неравномерность освещения также приводит к наличию теней на видеоизображениях. Объекты фона, стулья и куртка на спинке стула являются шумовыми факторами, так как их изображения могут быть ошибочно приняты за изображения людей в черной одежде, т. е. их цветовые характеристики схожи с объектами фона. Пример сложной ситуации показан на рис. 4, *и*, когда человек выходит из-за доски, но при этом вначале появляется его тень, которая достаточно хорошо видна на шкафу и может быть ошибочно определена детектором как человек. Из рис. 4, *к* видно, что на видеоряде траектории движения людей пересекаются и их изображения недостаточно четкие. В данный момент это является дополнительным шумом.

Четвертая видеопоследовательность (рис. 4, *л–н*) состоит из 3450 кадров, на которых присутствуют от одного до трех человек. Шумовыми факторами на кадрах являются: тени фигур людей на стене (рис. 4, *л, м*); пересечение траекторий движущихся людей с полным скрытием одного человека за другим (рис. 4, *м*); ситуация, когда при движении нижняя половина фигуры человека часто скрывается объектом фона, столом; множественные входы в помещение и выходы из него людей в разной их очередности (например, рис. 4, *л, н*). Вместе с тем изображения людей на данной видеопоследовательности имеют хорошее качество.

Примеры кадров пятого тестового видеоряда, который получен при более низком уровне освещения сцены и состоит из 1850 кадров, показаны на рис. 4, *п–с*. На сцене присутствуют в основном три человека, несколько людей выходят из кадра и затем возвращаются, человек совершает наклоны (рис. 4, *р*). Движения людей происходят по сложным траекториям со значительным удалением от видеокамеры и множественными перекрытиями, пример перекрытия изображений двух человек показан на рис. 4, *п*, а пример кадра, на котором два человека практически полностью скрыты третьим, приведен на рис. 4, *с*.

Шестая видеопоследовательность включает 1640 кадров (рис. 4, *т–ш*) и отличается недостаточной четкостью изображений людей, постоянным их движением по нелинейным траекториям с резкими изменениями направлений, а также множественными пересечениями. На многих кадрах присутствуют четыре человека, однако изображения только трех из них были размечены для тестирования.

На рис. 5 показаны примеры обнаружения и сопровождения людей на кадрах видеопоследовательности, которая не использовалась для сравнения алгоритмов. На рис. 5, б человек 4 (рис. 5, а) не обнаруживается детектором и его визуальные признаки в это время существенно изменяются за счет того, что он надевает куртку, поэтому на данном кадре сопровождение для него не выполняется. Однако, как видно на рис. 5, в, в дальнейшем сопровождение человека 4 восстанавливается с искомым индексом. В течение практически полного перекрытия двух изображений людей 3 и 4 (рис. 5, з) детектор обнаруживает только одного человека. Так как признаки данных людей на последних кадрах перед их перекрытием вычисляются не для всей фигуры, а только для ее верхних видимых частей, которые схожи, это приводит к обмену индексами. Анализ кадров на рис. 5, д, е показывает, что при кратковременном перекрытии видеоизображений двух движущихся людей (рис. 5, д) со схожими визуальными признаками их индексация не изменяется, т. е. сопровождение выполняется правильно (рис. 5, е).

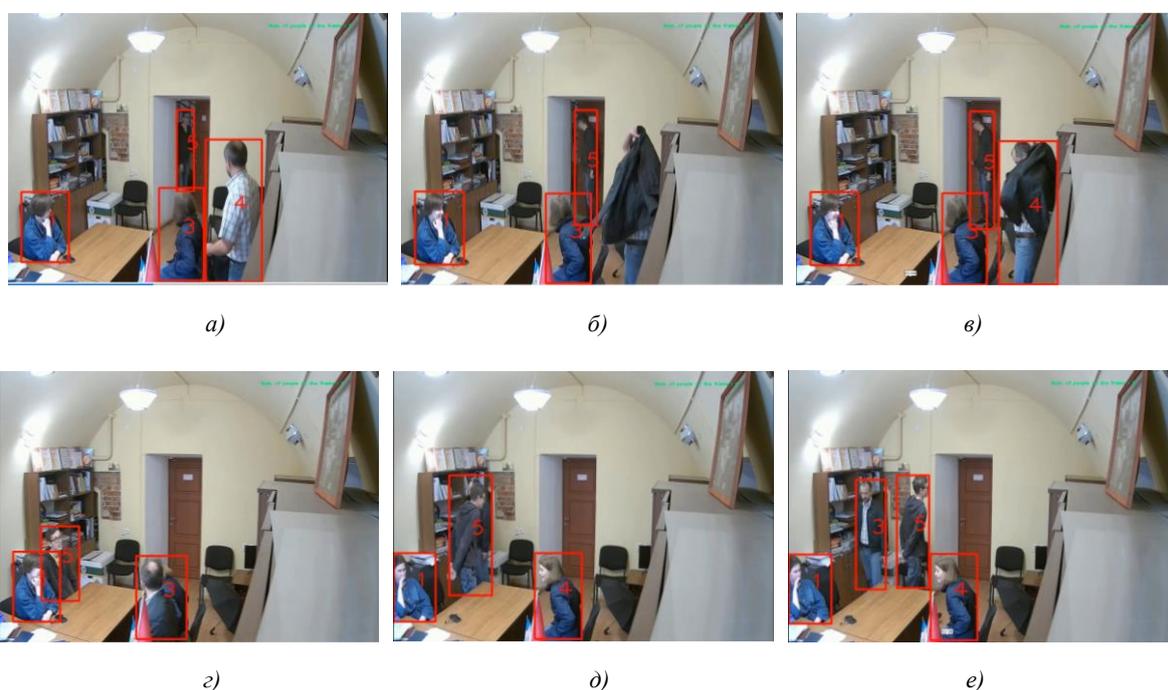


Рис. 5. Примеры сопровождения людей на видеопоследовательности

В таблице представлены результаты экспериментов по оценке результативности сопровождения людей с использованием рассмотренных критериев, лучшие по каждой группе результаты выделены жирным шрифтом. Были реализованы две модификации разработанного алгоритма, отличающиеся архитектурами СНС, которые используются для обнаружения людей. Модификация 1 применяет СНС YOLO v3, модификация 2 – СНС YOLO v4. Данные СНС характеризуются различной точностью детектирования объектов, сравнение их представлено в работе [17]. Для всех алгоритмов критерий  $ML$  равен нулю.

Результаты экспериментов по оценке эффективности алгоритмов сопровождения

Алгоритмы	Критерии										
	$IDF_1$	$MOTP$	$PRC$	$RCL$	$FP$	$FN$	$FM$	$MT$	$PL$	$MM$	$MOTA$
DS_v2 [20]	0,545	0,819	0,0153	0,8914	<b>135</b>	1759	215	14	3	61	0,8793
Модификация 1	<b>0,833</b>	0,824	0,0286	0,9408	252	959	195	<b>16</b>	<b>1</b>	<b>21</b>	0,9237
Модификация 2	0,807	<b>0,829</b>	<b>0,0448</b>	<b>0,9535</b>	395	<b>754</b>	<b>194</b>	15	2	39	<b>0,9266</b>

Анализ данных таблицы показывает, что поставленная задача повышения точности сопровождения множества объектов выполнена, так как для двух модификаций предложенного алгоритма критерий  $MOTA$  выше по сравнению с подходом из работы [20]. Кроме того, со-

гласно полученным результатам повышение результативности обнаружения позволяет улучшить  $MOTA$ , так как при использовании для обнаружения людей СНС YOLO v3  $MOTA = 0,9237$ , а при СНС YOLO v4  $MOTA = 0,9266$ .

Пример применения разработанного алгоритма для видеонаблюдения вне помещений показан на рис. 6. На рис. 6, а представлен 23-й кадр видеоряда, полученный видеосистемой охраны периметра здания, на котором в правом нижнем углу обнаружен и выделен рамкой входящий в зону контроля человек. На рис. 6, б показан 346-й кадр видеоряда данной видеопоследовательности с отображением всей траектории движения сопровождаемого человека по направлению ко входу в арку здания. Визуализация траектории выполняется на основе последовательности его положений на кадрах, определяемых координатами центра рамки в течение заданного времени. Анализ рис. 6, в позволяет идентифицировать на кадрах перемещения каждого из трех человек на контролируемой сцене по их траекториям: вышедший из арки человек 2, пройдя вправо за опору уличного освещения, пошел вглубь сцены, отдаляясь от видеокамеры; человек 3 вошел в кадр в правой нижней его части и двигался по направлению к арке ближе к стене здания по сравнению с траекторией человека, который первым появился на динамической сцене (рис. 6, б); человек 4 пошел из арки в правую сторону кадра. На рис. 6, г показана траектория человека 2, которая свидетельствует о том, что он сменил направление движения после 828-го кадра по направлению к видеокамере и вышел со сцены в правом нижнем углу кадра. Таким образом, анализ траектории на кадре рис. 6, г показывает, как передвигался человек 2 с момента его появления на контролируемой сцене до выхода из нее.

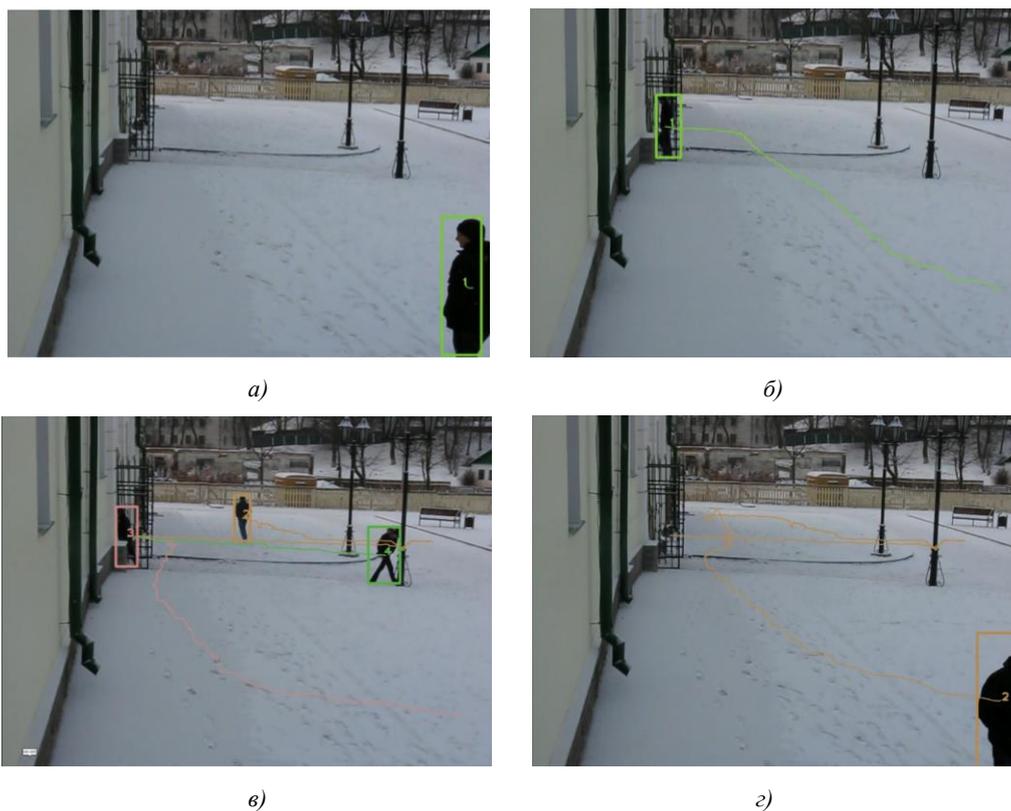


Рис. 6. Отображение траекторий движения людей на кадрах видеопоследовательности:  
 а) на 23-м кадре; б) на 346-м кадре; в) на 828-м кадре; г) на 1755-м кадре

**Заключение.** В работе представлена формализация задачи обнаружения стационарных и движущихся объектов на видеопоследовательностях с учетом их особенностей. Рассмотрены и описаны два типа сопровождения: одного и множества объектов на видеопоследовательностях. Представлены критерии, которые могут быть использованы для оценки качества обнаружения сопровождаемых объектов, точности определения местоположения объекта на кадре,

результативности построения траектории движения, точности сопровождения множества объектов. Рассмотрены основные ошибки, возникающие при сопровождении множества объектов. На основе представленной формализации решения задач обнаружения и сопровождения разработан и приведен в статье алгоритм сопровождения людей. Данный алгоритм использует СНС для детектирования людей и формирования их признаков. Комплексный дескриптор описания обнаруженного изображения человека в кадре включает нейросетевые, геометрические и цветовые признаки. На основе рассмотренных критериев реализованы и исследованы две модификации представленного алгоритма, отличающиеся архитектурами СНС, которые применяются для обнаружения людей. Результаты экспериментов подтвердили, что повышение результативности обнаружения позволяет улучшить и качество сопровождения. Точность сопровождения множества объектов при использовании для обнаружения YOLO v3 оценивается величиной 0,9237, при применении СНС YOLO v4 – величиной 0,9266. Приведены примеры обнаружения и сопровождения людей на кадрах видеопоследовательности при существенном изменении их признаков и полном перекрытии, а также примеры визуализации траекторий движения.

### Список использованных источников

1. Maggio, E. Video Tracking: Theory and Practice / A. Cavallaro, E. Maggio. – Wiley, 2011. – 295 p.
2. Datta, A. K. Face Detection and Recognition: Theory and Practice / A. K. Datta, M. Datta, P. K. Banerjee. – CRC Press, 2016. – 350 p.
3. Купляков, Д. А. Алгоритм сопровождения людей в видео на основе метода Монте-Карло для Марковских цепей / Д. А. Купляков, Е. В. Шальнов, А. С. Конушин // Программирование. – 2017. – № 4. – С. 13–21. <https://doi.org/10.1134/s0361768817040053>
4. Кирпичников, А. П. Обнаружение и сопровождение людей в интеллектуальных детекторах внештатных ситуаций / А. П. Кирпичников, С. А. Ляшева, М. П. Шлеймович // Вестник Казанского технологического университета. – 2014. – Т. 17. – С. 351–356.
5. Ravish Aradhya, H. V. Object detection and tracking using deep learning and artificial intelligence for video surveillance applications / H. V. Ravish Aradhya // Intern. J. of Advanced Computer Science and Applications. – 2019. – Vol. 10, no. 12. – P. 517–530. <https://doi.org/10.14569/IJACSA.2019.0101269>
6. Mane, S. Moving object detection and tracking using convolutional neural networks / S. Mane, S. Mangale // Second Intern. Conf. on Intelligent Computing and Control Systems (ICICCS). – Madurai, India, 2018. – P. 1809–1813. <https://doi.org/10.1109/ICCONS.2018.8662921>
7. Motion detection and tracking algorithms in video streams / R. Bogush [et al.] // VNU J. of Science, Mathematics – Physics. – 2009. – Vol. 25, no. 3. – P. 143–151.
8. Богуш, Р. П. Алгоритм сопровождения людей на видеопоследовательности с использованием идентификации по лицам для наблюдения внутри помещений / Р. П. Богуш, И. Ю. Захарова, С. В. Абламейко // Вестник информационных и компьютерных технологий. – 2020. – № 7. – С. 3–14. <https://doi.org/10.14489/vkit.2020.07>
9. An effective algorithm to detect both smoke and flame using color and wavelet analysis / S. Ye. [et al.] // Pattern Recognition and Image Analysis. – 2017. – Vol. 27, no. 1. – P. 131–138. <https://doi.org/10.1134/S1054661817010138>
10. Залесский, Б. А. Отслеживание динамических объектов и их распознавание с помощью графовых алгоритмов / Б. А. Залесский, А. И. Кравчонок // Информатика. – 2006. – № 2(10). – С. 7–26.
11. Залесский, Б. А. Алгоритм отслеживания объектов движущейся видеокамерой / Б. А. Залесский // Докл. Нац. акад. наук Беларуси. – 2020. – Т. 64, № 2. – С. 144–149. <https://doi.org/10.29235/1561-8323-2020-64-2-144-149>
12. Агафонов, В. Ю. Использование фильтра Калмана в задачах трекинга объектов / В. Ю. Агафонов, В. Л. Розалиев, А. В. Заболеева-Зотова // Интеллектуальные системы. Теория и приложения. – 2016. – Т. 20, № 4. – С. 13–17.
13. Gustafsson, F. Particle filters for positioning, navigation and tracking / F. Gustafsson, F. Gunnarsson, N. Bergman // IEEE Transactions on Signal Processing. – 2002. – Vol. 50(2). – P. 425–437. <https://doi.org/10.1109/78.978396>
14. Performance measures and a data set for multi-target, multi-camera tracking / E. Ristani [et al.] // Computer Vision ECCV 2016 Workshops. Lecture Notes in Computer Science. – 2016. – Vol. 9914. – P. 17–35.
15. MOTChallenge: A benchmark for single-camera multiple target tracking [Electronic resource] / P. Dendorfer [et al.] // Intern. J. of Computer Vision. – 2020. – Mode of access: <https://link.springer.com/article/10.1007/s11263-020-01393-0>. – Date of access: 12.06.2020. <https://doi.org/10.1007/s11263-020-01393-0>

16. Метрики оценки алгоритмов автоматического сопровождения / А. Е. Щелкунов [и др.] // Известия ЮФУ. Техн. науки. – 2020. – № 1(211). – С. 233–245.
17. Bochkovskiy, A. YOLOv4: Optimal Speed and Accuracy of Object Detection [Electronic resource] / A. Bochkovskiy, Ch.-Y. Wang, H.-Y. M. Liao. – Mode of access: <https://arxiv.org/abs/2004.10934>. – Date of access: 12.06.2020.
18. Богуш, Р. П. Алгоритм сопровождения людей на видеопоследовательностях с использованием сверточных нейронных сетей для видеонаблюдения внутри помещений / Р. П. Богуш, И. Ю. Захарова // Компьютерная оптика. – 2020. – Т. 44, № 1. – С. 109–116. <https://doi.org/10.18287/2412-6179-CO-565>
19. Kuhn, H. W. The Hungarian method for the assignment problem / H. W. Kuhn // *Naval Research Logistics Quarterly*. – 1955. – No. 2. – P. 83–97.
20. Wojke, N. Simple online and realtime tracking with a deep association metric / N. Wojke, A. Bewley, D. Paulus // *Proc. of the IEEE Intern. Conf. on Image Processing (ICIP)*. – Beijing, China, 2017. – P. 3645–3649. <https://doi.org/10.1109/ICIP.2017.8296962>

## References

1. Maggio E., Cavallaro A. *Video Tracking: Theory and Practice*. Wiley, 2011, 295 p.
2. Datta A. K., Datta M., Banerjee P. K. *Face Detection and Recognition: Theory and Practice*. CRC Press, 2016, 350 p.
3. Kuplyakov D. A., Shalnov E. V., Konushin A. S. Algoritm soprovozhdenija ljudej v video na osnove metoda Monte-Karlo dlja Markovskih cepej [Markov chain Monte Carlo based video tracking algorithm]. *Programirovanie [Programming and Computer Software]*, 2017, no. 4, pp. 13–21 (in Russian). <https://doi.org/10.1134/s0361768817040053>
4. Kirpichnikov A. P., Lyasheva S. A., Shleymovich M. P. Obnaruzhenie i soprovozhdenie ljudej v intellektual'nyh detektorah vneshtatnyh situacij [Detection and tracking of people in intelligent detectors of emergency situations]. *Vestnik Kazanskogo tehnologicheskogo universiteta [Kazan Technological University Bulletin]*, 2014, vol. 17, pp. 351–356 (in Russian).
5. Ravish Aradhya H. V. Object detection and tracking using deep learning and artificial intelligence for video surveillance applications. *International Journal of Advanced Computer Science and Applications*, 2019, vol. 10, no. 12, pp. 517–530. <https://doi.org/10.14569/IJACSA.2019.0101269>
6. Mane S., Mangale S. Moving object detection and tracking using convolutional neural networks. *Second International Conference on Intelligent Computing and Control Systems (ICICCS)*. Madurai, India, 2018, pp. 1809–1813. <https://doi.org/10.1109/ICCONS.2018.8662921>
7. Bogush R., Maltsev S., Kastyuk A., Brovko N., Gluhau D. Motion detection and tracking algorithms in video streams. *VNU Journal of Science, Mathematics – Physics*, 2009, vol. 25, no. 3, pp. 143–151.
8. Bogush R. P., Zakharova I. Yu., Ablameyko S. V. Algoritm soprovozhdenija ljudej na videoposledovatel'nosti s ispol'zovaniem identifikacii po licam dlja nabljudenija vnutri pomeshhenij [Algorithm for person tracking on video sequences using face identification for indoor surveillance]. *Vestnik informacionnyh i komp'yuternyh tehnologij [Herald of Computer and Information Technologies]*, 2020, no. 7, pp. 3–14 (in Russian). <https://doi.org/10.14489/vkit.2020.07>
9. Ye S., Zhican B., Chen C., Bohush R., Ablameyko S. An effective algorithm to detect both smoke and flame using color and wavelet analysis. *Pattern Recognition and Image Analysis*, 2017, vol. 27, no. 1, pp. 131–138. <https://doi.org/10.1134/S1054661817010138>
10. Zalesky B. A., Kravchonok A. I. Otslezhivanie dinamicheskikh objektov i ih raspoznavanie s pomoshh'ju grafovyh algoritmov [Tracking dynamical objects and their recognition by graph algorithms]. *Informatika [Informatics]*, 2006, no. 2(10), p. 7–26 (in Russian).
11. Zalesky B. A. Algoritm otslezhivaniya objektov dvizhushhejsja videokameroy [Object tracking algorithm by moving video camera]. *Doklady Nacional'noj akademii nauk Belarusi [Doklady of the National Academy of Sciences of Belarus]*, 2020, vol. 64, no. 2, pp. 144–149 (in Russian). <https://doi.org/10.29235/1561-8323-2020-64-2-144-149>
12. Agafonov V. Yu., Rozaliev V. L., Zaboleeva-Zotova A. V. Ispol'zovanie fil'tra Kalmana v zadachah trekinga objektov [Using the Kalman filter in object tracking tasks. Intelligent systems]. *Intellektual'nye sistemy. Teoriya i prilozheniya [Theory and Applications]*, 2016, vol. 20, no. 4, pp. 13–17 (in Russian).
13. Gustafsson F., Gunnarsson F., Bergman N. Particle filters for positioning, navigation and tracking. *IEEE Transactions on Signal Processing*, 2002, vol. 50(2), pp. 425–437. <https://doi.org/10.1109/78.978396>
14. Ristani E., Solera F., Zou R., Cucchiara R., Tomasi C. Performance measures and a data set for multi-target, multi-camera tracking. *Computer Vision ECCV 2016 Workshops. Lecture Notes in Computer Science*, 2016, vol. 9914, pp. 17–35.

15. Dendorfer P., Ošep A., Milan A., Schindler K., Cremers D., ..., Leal-Taixé L. MOTChallenge: A benchmark for single-camera multiple target tracking. *International Journal of Computer Vision*, 2020. Available at: <https://link.springer.com/article/10.1007/s11263-020-01393-0> (accessed 12.06.2020). <https://doi.org/10.1007/s11263-020-01393-0>
16. Shchelkunov A. E., Kovalev V. V., Morev K. I., Sidko I. V. Metriki ocenki algoritmov avtomaticheskogo soprovozhdenija [The metrics for tracking algorithms evaluation]. *Izvestija Juzhnogo federal'nogo universiteta. Tehnicheskie nauki [News of the Southern Federal University. Engineering Sciences]*, 2020, no. 1(211), pp. 233–245 (in Russian).
17. Bochkovskiy A., Wang Ch.-Y., Liao H.-Y. M. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. Available at: <https://arxiv.org/abs/2004.10934> (accessed 12.06.2020).
18. Bohush R. P., Zakharava I. Y. Algoritm soprovozhdenija ljudej na videoposledovatel'nostjah s ispol'zovaniem svetochnyh nejronnyh setej dlja videonabljudenija vnutri pomeshhenij [Person tracking algorithm based on convolutional neural network for indoor video surveillance]. *Komp'yuternaja optika [Computer Optics]*, 2020, vol. 44, no. 1, pp. 109–116 (in Russian). <https://doi.org/10.18287/2412-6179-CO-565>
19. Kuhn H. W. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 1955, no. 2, pp. 83–97.
20. Wojke N., Bewley A., Paulus D., Wojke N. Simple online and realtime tracking with a deep association metric. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. Beijing, China, 2017, pp. 3645–3649. <https://doi.org/10.1109/ICIP.2017.8296962>

### Информация об авторах

*Богуш Рихард Петрович*, кандидат технических наук, доцент, заведующий кафедрой вычислительных систем и сетей, Полоцкий государственный университет.

E-mail: bogushr@mail.ru

*Абламейко Сергей Владимирович*, академик Национальной академии наук Беларуси, доктор технических наук, профессор, главный научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси; профессор механико-математического факультета, Белорусский государственный университет.

E-mail: ablameyko@bsu.by

### Information about the authors

*Rykhard P. Bohush*, Cand. Sci. (Eng.), Associate Professor, Head of the Department of Computer Systems and Networks, Polotsk State University.

E-mail: bogushr@mail.ru

*Sergey V. Ablameyko*, Academician of the National Academy of Sciences of Belarus, Dr. Sci. (Eng.), Professor, Chief Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus; Professor of the Faculty of Mechanics and Mathematics, Belarusian State University.

E-mail: ablameyko@bsu.by