

МЕТОД ВИЗУАЛЬНОГО МУЛЬТИТРЕКИНГА В РЕАЛЬНОМ ВРЕМЕНИ НА ОСНОВЕ КОРРЕЛЯЦИОННЫХ ФИЛЬТРОВ

Аннотация. В статье предложен подход к отслеживанию множества объектов, применяющий трекинг на основе детектирования с использованием корреляционных фильтров для отслеживания цели между ключевыми кадрами. Это позволяет повысить надежность трекинга объектов, движущихся нелинейно, сохранив при этом скорость работы, достаточную для обработки видео в реальном времени.

Ключевые слова. Компьютерное зрение, визуальный трекинг, корреляционный фильтр, мультитрекинг, трекинг на основе детектирования.

Вступление

Задача отслеживания множества объектов на видеов реальном времени является неотъемлемой частью многих прикладных областей компьютерного зрения, таких как построение систем видеонаблюдения, отслеживания дорожного трафика. Особый интерес данная задача представляет для систем способных активно взаимодействовать с окружающим миром, в частности для анализа сцены самоуправляемым автомобилем, беспилотным летательным аппаратом, промышленным или бытовым роботом. Данные системы имеют более высокие требования к скорости и точности работы алгоритма отслеживания. Также, поскольку автономные робототехнические агенты имеют более ограниченные вычислительные ресурсы, по сравнению со стационарными системами, для них очень важно уменьшение ресурсозатратности системы отслеживания.

По сравнению с задачей отслеживания одного объекта, при отслеживании множества объектов возрастает объем вычислений необходимых для сопровождения цели, также появляется необходимость определения количества объектов, которое обычно изменяется с течением времени, и поддержания их идентичности.

Постановка задачи мультитрекингов реальном времени

Задана последовательность кадров, получаемая в реальном времени с помощью цветной монокамеры и класс отслеживаемых объектов. Необходимо для каждого объекта, построить его траекторию на протяжении всего видеоряда. То есть, необходимо обнаружить появление новых объектов в кадре, локализовать обнаруженные объекты на последующих кадрах, а также обнаружить исчезновение объектов из кадра.

Анализ существующих методов мультитрекинга

Задача мультитрекинга изучается уже долгое время, за которое было предложено множество подходов, такие как применение марковских моделей [1], байесовских сетей [2], трекары на основе нейронных сетей [3], и трекары на основе детектирования [4].

В последнее время в связи с ускорением работы детекторов на основе нейронных сетей [5] [6] трекинг на основе детектирования (TBD - Tracking By Detection) набрал большую популярность. Большая часть современных мультитрекаров в той или иной мере использует этот подход [7].

Трекинг на основе детектирования заключается в том, что на каждом кадре объекты обнаруживаются детектором, а затем связываются в траектории.

Для того что бы связать объекты на двух соседних кадрах, сначала для объектов с предыдущего кадра предсказываются на текущем кадре. Затем предсказанные объекты сопоставляются с объектами, которые обнаружил детектор при помощи признаков схожести (пространственных или визуальных). Траектории обновляются с координатами соответствующих сопоставленных объектов обнаруженных, детектором. Для не сопоставленных объектов инициализируются новые траектории.

Использование TBD позволяет автоматически обнаруживать появление новых целей в кадре и исчезновение старых, а также позволяет корректировать параметры модели за счет исправления ошибки между состоянием, полученным с детектора и предсказанным состоянием.

Прогнозирование положения объекта на основе его динамической модели

Наиболее распространённым методом для прогнозирования положения цели на следующем кадре является применение динамической модели объекта. Динамическая модель предсказывает траекторию движения объекта в кадре путём аппроксимации уравнения динамики. Как правило, используется линейная динамическая модель, которая описывает движение тела с постоянной скоростью, либо с постоянным ускорением.

В большинстве подходов для коррекции параметров уравнения динамики применяется фильтр Калмана [8] [9]. Трекер поочередно прогнозирует новое положение объекта, и корректирует параметры модели для минимизации ошибки между предсказанной позицией и позицией, полученной от детектора. Основной причиной применения динамической модели является её низкая вычислительная сложность, что очень важно для достижения необходимой скорости работы алгоритма, поскольку количество вычислений пропорционально увеличивается с количеством отслеживаемых объектов.

Однако у данного подхода есть существенные недостатки. При использовании трекинга на основе детектирования предполагается, что каждый кадр будет обработан детектором и, соответственно, на каждом кадре будет корректироваться динамическая модель. Однако, в большинстве систем скорости работы детектора недостаточно для этого. Поэтому между кадрами, на которых срабатывает детектор (ключевыми кадрами) траектория объекта обновляется только при помощи нескорректированной модели, что значительно снижает точность и надежность трекинга. К тому же, из-за того, что фильтр Калмана плохо справляется с нелинейным движением [10], если цель между ключевыми кадрами изменит скорость или направление движения, ошибка между предсказанным и реальным положением будет слишком большой и из-за чего объект будет потерян (рис. 1). Поэтому, этот подход не применим к маневренным объектам, которые могут резко изменить скорость или направление движения таким как пешеход или автомобиль.

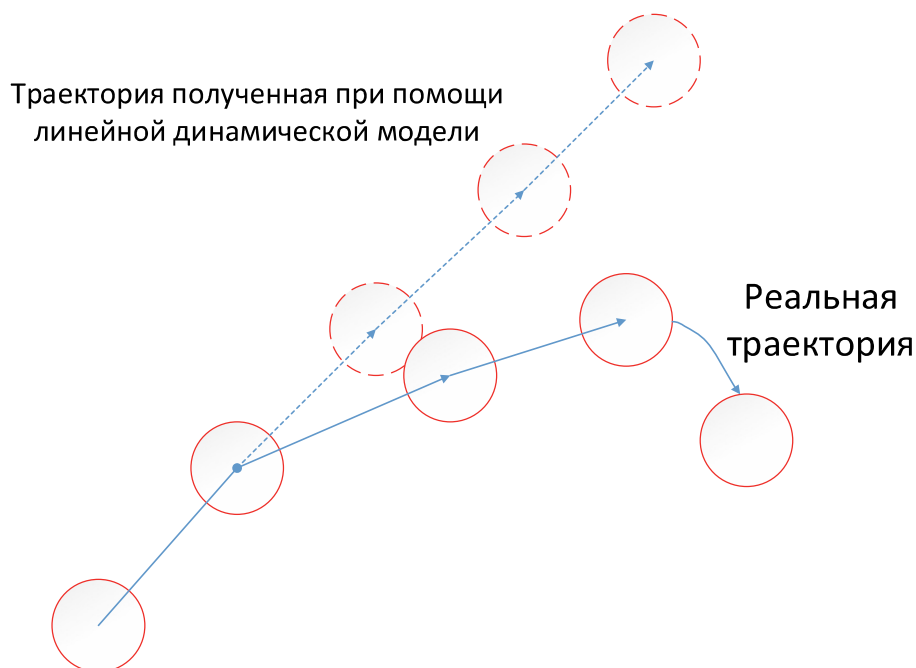


Рис. 1. – Потеря объекта, движущегося нелинейно, при трекинге с помощью линейной динамической модели

Данную проблему можно решить путем замены фильтра Калмана на трекер, использующий визуальные признаки, для построения траектории между ключевыми кадрами.

К сожалению, скорость большинства современных методов визуального трекинга таких как метод скользящего среднего [11], нейронные сети [12], бустинг [13]

недостаточна для сопровождения множества объектов в реальном времени. Однако семейство трекеров, основанных на корреляционных фильтрах способно сопровождать объект с высокой точностью на скорости более 250 FPS и на данный момент являются наиболее эффективными [14].

Мультитрекинг на основе корреляционных фильтров

Для решения задачи мультитрекинга предлагается использовать модификацию подхода трекинга на основе детектирования, в которой для построения траектории каждого объекта между ключевыми кадрами используются независимые трекеры на основе корреляционных фильтров.

В данной системе можно выделить три основные операции: обнаружение, матчинг, трекинг.

На ключевом кадре детектор определяет координаты обнаруженных объектов. После этого обнаруженные объекты сопоставляются с объектами, которые отслеживаются на данный момент (рис. 2).

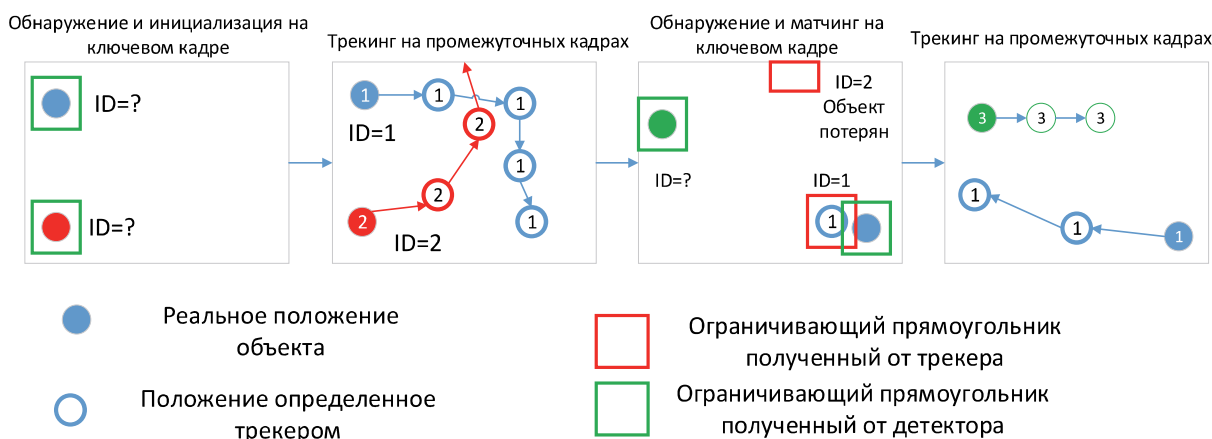


Рис. 2. – Пример работы системы мультитрекинга

- Если два объекта сопоставлены, трекер обновляет свою модель используя позицию и визуальные признаки соответствующего объекта, обнаруженного детектором.
- Если отслеживаемый объект не был сопоставлен, он считается потерянным.
- Если обнаруженный детектором объект не был сопоставлен, считается, что в кадре появился, новый объект для которого создается новый трекер с новым уникальным идентификатором.

После матчинга до следующего ключевого кадра цели отслеживаются с помощью корреляционных трекеров.

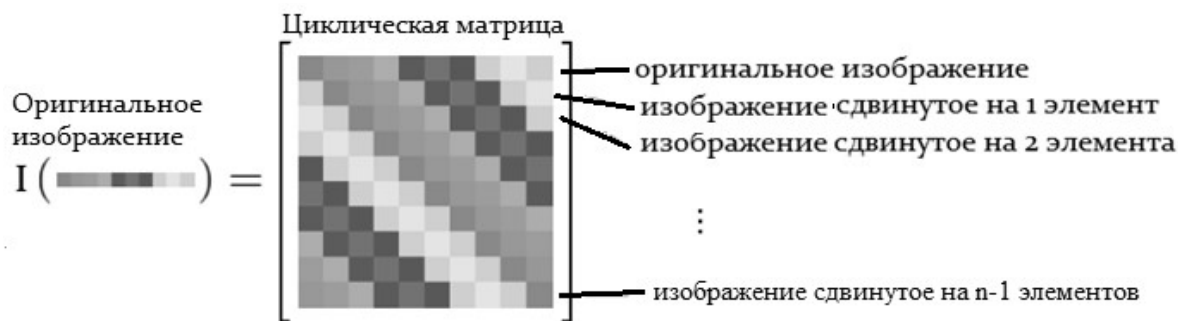


Рис. 3. - Матрица, образованная циклическим сдвигом одномерного изображения

Принцип работы корреляционного трекера

Основная идея корреляционного трекинга заключается в обучении корреляционного фильтра с помощью которого можно на кадре отделить цель от заднего фона. Параметры корреляционного фильтра H подбираются таким образом, чтобы его свёртка с изображением представляла собой распределение Гаусса с пиком в центре отслеживаемого объекта. Для обучения такого фильтра используется обучающая выборка, на которой минимизируется ошибка между реальным результатом свёртки и ожидаемым (Гауссианой), при помощи метода наименьших квадратов [15] по формуле (1).

$$H = \arg \min \sum_{i=0}^n (F_i \odot H - G_i)^2 \# \quad (1)$$

Вобщем случае для обучения фильтра по формуле (1) приходится решать систему уравнений, что значительно затрудняет обработку кадров в реальном времени.

В работе Энрикеза[16] было предложено эффективное с точки зрения сложности вычислений решение этой задачи, основанное на свойствах циклических матриц. Для этого предлагается генерировать обучающую выборку с помощью циклического сдвига изображения (рис. 3). Матрица, которая образована циклическим сдвигом изображения содержит позитивный пример для обучения в первой строке и набор негативных примеров в остальных строках. Более того, циклическая матрица в пространстве Фурье образует диагональную матрицу, а использование диагональной матрицы позволяет применять поэлементные операции вместо матричных, что снижает сложность вычислений с $O(n^2)$ до $O(n \log(n))$, благодаря чему значительно повышается скорость работы трекера.

Исходя из этих соображений в работе [17] был разработан быстрый алгоритм корреляционного трекинга. В основе алгоритма лежат три функции. Функция корреляции ядра (2) – вспомогательная функция, которая используется двумя другими функциями.

$$k(x, x') = \exp\left(-\frac{1}{\sigma^2} (\|x^2\| + \|x'^2\| - 2 \text{FFT}(\widehat{x} \odot \widehat{x}'))\right) \# \quad (2)$$

Функция для нахождения оптимальных параметров корреляционного фильтра:

$$H = \frac{G}{k(x, x) + \lambda}, \# \quad (3)$$

где: G – 2D распределение гаусса, λ – параметры регуляризации.

Функция для определения положения цели на кадре (4). Результатом данной функции является матрица корреляции между визуальными признаками цели признаками области поиска, при всех возможных сдвигах.

$$f(x, z) = k(x, z) \odot \alpha, \# \quad (4)$$

где: α – корреляционный фильтр, z – сохранённый образец визуальных признаков цели, x – признаки, извлеченные из области поиска.

Операции в функциях (3), (4) производятся в пространстве Фурье, поэтому для получения окончательного результата из функций (3) и (4) необходимо произвести обратное преобразование Фурье (IFFT). К тому же производится свёртка вектора визуальных признаков с оконной функцией Хеннинга (5), для того чтобы произвести оконное преобразование Фурье (FFT).

$$w(n) = \left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right) \# \quad (5)$$

Рассмотрим алгоритм сопровождения объекта при помощи адаптивного корреляционного трекера. Сначала из изображения извлекаются визуальные признаки, которые используются для обучения корреляционного фильтра. Далее на последующих кадрах поочередно происходит процесс локализации цели и обновления визуальной модели с помощью изображения, полученного в новой позиции (рис. 4).

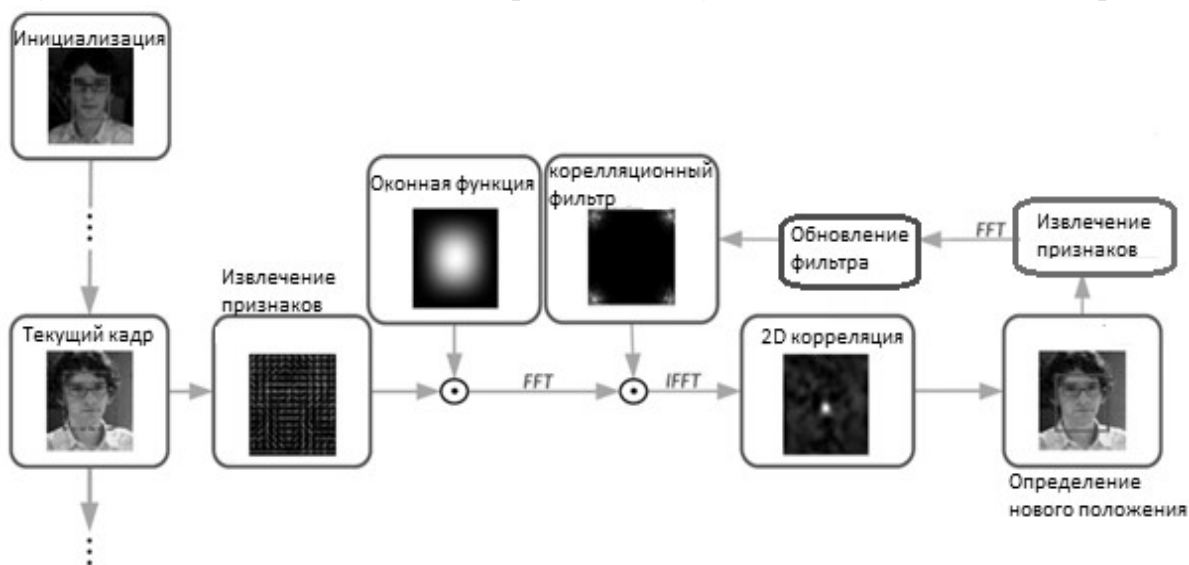


Рис. 4. – Схема работы корреляционного трекера

Инициализация трекера

Входными данными для инициализации являются кадр видео I_t и ограничивающий прямоугольник B , который определяет расположение цели на кадре P и её размер S .

В системе применяется подход, который позволяет определить не только перемещение объекта, но и изменение его масштаба[18]. На выделенном регионе рассчитывается вектор признаков и инициализируются две модели: трансляционная и масштабирующая. Трансляционная модель применяется для определения перемещения объекта, масштабирующая – для определения изменения размера.

Трансляционная модель состоит из вектора визуальных признаков Z_t и корреляционного фильтра H . Вектор признаков представляет собой гистограмму ориентированных градиентов (HOG), а H рассчитывается с помощью уравнения (3).

Масштабирующая модель представляет собой массив M из N_s масштабирующих коэффициентов, которые рассчитываются по формуле (6) и матрицы масштабированных визуальных признаков.

$$M = \alpha^n, \# \tag{6}$$

где: α – шаг масштабирования, а $n \in \left\{ -\frac{N_s - 1}{2}, \dots, \frac{N_s - 1}{2} \right\}$

Матрица масштабированных визуальных признаков представляет собой N_s гистограмм, которые получаются путём расчета HOG для разных масштабов изображения (рис. 5), полученных и последующего расчета HOG для каждого масштаба изображения. Номер строки в матрице определяет масштаб изображения, при котором были извлечены признаки.

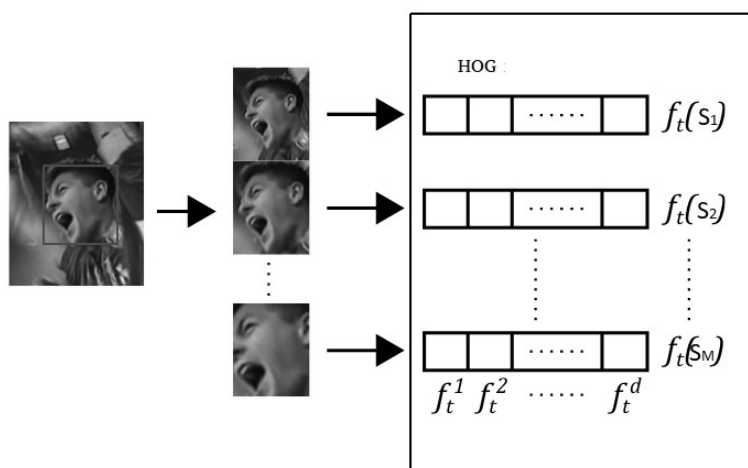


Рис. 5. – Извлечение визуальных признаков для масштабирующей модели

Определение изменения позиции

Определение позиции происходит в окрестности около предыдущего расположения цели. Окрестность поиска получается путем увеличения размера текущего ограничивающего прямоугольника B . Для нахождения цели рассчитываются НОГ признаки X_t в области поиска. Затем по формуле (4) рассчитывается корреляционная матрица R . В матрице по формуле (7) определяется точка с максимальным значением корреляции $P_{max} = \{x, y\}$, координаты которой соответствуют смещению центра цели относительно предыдущего кадра. Новые координаты цели определяются по формуле (8).

$$P_{max} = \arg \max R(P) \# \quad (7)$$

$$P_t = P_{t-1} + P_{max} \# \quad (8)$$

Определение изменения размера

Для определения изменения размера в новой позиции цели рассчитывается матрица масштабированных визуальных признаков. Затем по формуле (4) находится её корреляция с матрицей, сохранённой в масштабной модели. Аналогично, по формуле (7) определяется точка с максимальным значением корреляции. Координата x этой точки, соответствует индексу в массиве масштабирующих коэффициентов M , а значение расположенного по этому индексу масштабирующего коэффициента, определяет отношение размера цели на текущем кадре, к размеру цели на предыдущем кадре.

Обучение модели

Поскольку со временем визуальное представление цели изменяется, необходимо постоянно обновлять прогнозирующие модели.

После получения новой позиции и размера цели, снова извлекаются признаки, которые используются для обновления трансляционной и масштабной модели.

Обновленная модель является взвешенной суммой текущей модели и модели, рассчитанной на новом кадре (9).

$$F_{t+1} = (1 - \eta)F_t + \eta F_u \# \quad (9)$$

где: F_{t+1} – обновленный вектор признаков, F_t – старый вектор признаков, F_u – признаки, извлеченные на новом кадре, а η – коэффициент обучения.

В трансляционной модели обучается вектор визуальных признаков и корреляционный фильтр. В масштабирующей модели индивидуально обновляется каждый вектор признаков из массива.

Обнаружение потери

При извлечении визуальных признаков объекта для обновления визуальной модели, также извлекаются признаки из заднего фона, попавшего в ограничительную рамку объекта. Влияние заднего фона может привести к постепенному смещению трекера относительно цели и в итоге к потере отслеживаемого объекта.

Для оценки качества корреляционной матрицы R применяется метрика Peak

to Sidelobe Ratio (PSR), которая рассчитывается по формуле (10). Для этого корреляционная матрица разбивается на две области – центральный и боковой лепесток. К центральной области относят область 11×11 вокруг пика, к боковой – всё остальное. Если значение PSR меньше порогового, то считается что положение определено неточно и обновление визуальной модели не производится.

$$PSR = \frac{R_{max} - \mu}{\sigma}, \# \quad (10)$$

где: μ – среднее значение бокового лепестка, а σ – СКО бокового лепестка.

Матчинг

Для корректного построения траектории необходимо определить какие из объектов, обнаруженных детектором уже отслеживаются. Для нахождения этого необходимо сопоставить друг с другом объекты, обнаруженные детектором и отслеживаемые трекером. Для решения этой задачи составляется матрица, в которой попарно оценивается схожесть, а затем объекты сопоставляются с помощью Венгерского алгоритма.

Для сравнения двух объектов используется комбинированный критерий, который учитывает, как пространственные, так и визуальные критерии схожести.

В качестве пространственного критерия схожести используется отношение площадей пересечения к объединению (Intersection Over Union) ограничительных рамок объектов, которое учитывает, как и расположение так и размер объектов и находится по формуле (11).

$$IOU = \frac{S_i}{S_u} \# \quad (11)$$

В качестве критерия визуальной схожести используется корреляция цветовых гистограмм, которое рассчитывается по формуле(12).

$$d(H_1, H_2) = \frac{\sum(H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum(H_1(I) - \bar{H}_1)^2 \sum(H_2(I) - \bar{H}_2)^2}}, \# \quad (12)$$

где: $\bar{H} = \frac{1}{N} \sum_j H(j)$, а N – количество корзин в гистограмме.

Итоговая функция схожести имеет вид (13):

$$C = IOU \cdot dH \# \quad (13)$$

Выводы

В работе предложен подход к решению задачи мультитрекинга, который позволяет повысить точность и надёжность трекинга объектов, движущихся нелинейно. Это достигается за счёт использования визуальных признаков для отслеживания цели, вместо применения динамической модели. Использование трекинга на основе корреляционных фильтров, позволило достичь производительности, достаточной для сопровождения множества целей в реальном времени.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Kratz L., Nishino K. Tracking with local spatio-temporal motion patterns in extremely crowded scenes // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010. С.693-700.
2. Nillius P., Sullivan J., Carlsson. S. Multi-target tracking linking identities using baesian network inference // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006. С.2187-2194.
3. Target identity-aware network flow for online multiple target tracking / A. Dehghan и др. // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2015. С.1146-1154.
4. Bochinski E., Eiselein V., Sikora T. High-Speed tracking-by-detection without using image information // IEEE International Conference on Advanced Video and Signal Based Surveillance.2017.С.1-6.
5. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications / A. Howard и др. // URL: <https://arxiv.org/pdf/1704.04861.pdf>
6. Faster R-CNN: Towards real-time object detection with region proposal networks / S. Ренд и др. // Advances in neural information processing systems. 2015.С.91–99.
7. Multiple Object Tracking: A Literature Review /Luo W. и др.// URL: <https://arxiv.org/pdf/1409.7618.pdf>
8. Reid D. An algorithm for tracking multiple targets // IEEE Transactions on Automatic Control. 1980. С. 843-854.
9. Simple online and realtime tracking / A. Bewley и др.// IEEE International Conference. on Image Procesing. 2016.С.3464–3468.
10. Survey of Video Object Tracking./ M. Li и др. // International Journal of Control and Automation. 2015. С.303-312.
11. Vojir T., Noskova J.,Matas J. Robust scale-adaptive mean-shift for tracking. // Pattern Recognition Letters. 2014.С.250-258.
12. Nam, H., Han, B. Learning multi-domain convolutional neural networks for visual tracking //URL: <https://arxiv.org/pdf/1510.07945.pdf>
13. Kalal Z., Mikolajczyk K., Matas J. Tracking learning detection // URL: <http://www.tldvision.com/pdf/2011-Kalal-PhD-Thesis.pdf>
14. The Visual Object Tracking VOT2016 challenge results/ M. Kristan и др. //European Conference on Computer Vision Workshops.2016. С.777-823.
15. Visual object tracking using adaptive correlation filters / D. Bolme и др. // Computer Vision and Pattern Recognition.2010. С.2544–2550.
16. Exploiting the circulant structure of tracking-by-detection with kernels/ J. Henriques и др. // Computer Vision–ECCV. 2012. С.702–715.
17. High-speed tracking with kernelized correlation filters/ J. Henriques и др. // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015. С.583–596.
18. Discriminative scale space tracking / M. Danelljan и др. // IEEE transactions on pattern analysis and machine intelligence. 2017. С.1561–1575.