

Комбинирование нескольких алгоритмов для увеличения точности обнаружения и локализации пешеходов

В.В. Стротов¹, П.Е. Жгутов¹

¹ ФГБОУ ВО РГРТУ им. В.Ф. Уткина, ул. Гагарина, 59/1, Рязань, 390005, Россия

Аннотация

В данной работе рассмотрена задача разработки алгоритма обнаружения пешеходов в последовательности изображений. Предложена концепция комбинированного алгоритма, объединяющего детектор пешеходов с функцией агрегирования нескольких каналов, вместе с алгоритмом выделения движения и функцией слежения. В работе представлено описание этапов работы алгоритма, а также экспериментальных исследований, которые проводились на наборе тестовых видеосюжетов. В ходе исследования комбинированного алгоритма проводилось сравнение качества обнаружения комбинированного алгоритма и детектора пешеходов по отдельности. Проводилось исследование зависимости результатов работы алгоритмов от изменений параметров детектора. На основе результатов экспериментальных исследований был сделан вывод о лучших показателях точности и качества обнаружения у комбинированного алгоритма, способного найти применение в системах компьютерного зрения.

Ключевые слова

Детектор пешеходов с функцией агрегирования каналов (ACF), метод вычисления признаков при мультимасштабном представлении сигнала (Fast Feature Pyramids), метод выделения движения, функция слежения, область интереса (ROI).

Combining Several Algorithms to Increase the Accuracy of Pedestrian Detection and Localization

V.V. Strotov¹, P.E. Zhgutov¹

¹ Ryazan State Radio Engineering University named after V.F. Utkin, Gagarina str., 59/1, Ryazan, 390005, Russia

Abstract

In this paper, the problem of developing an algorithm for detecting pedestrians in a sequence of images is considered. The concept of a combined algorithm combining a pedestrian detector with an aggregation function of several channels, together with a motion detection algorithm and a tracking function is proposed. The paper describes the stages of the algorithm, as well as experimental studies that were conducted on a set of test videos. During the study of the combined algorithm, the quality of detection of the combined algorithm and the pedestrian detector was compared separately. The study of the dependence of the results of the algorithms on changes in the detector parameters was carried out. Based on the results of experimental studies, a conclusion was made about the best indicators of accuracy and quality of detection in a combined algorithm that can find application in computer vision systems.

Keywords

People detector with aggregate channel feature (ACF), a method for calculating features in a multiscale representation of a signal (Fast Feature Pyramids), formatting examples, move extract method, track function, region of interesting (ROI).

ГрафиКон 2022: 32-я Международная конференция по компьютерной графике и машинному зрению, 19-22 сентября 2022 г., Рязанский государственный радиотехнический университет им. В.Ф. Уткина, Рязань, Россия

EMAIL: hedgehog.str@gmail.com (В.В. Стротов); zhgutovp@mail.ru (П.Е. Жгутов)
ORCID: 0000-0003-1691-7094 (В.В. Стротов); 0000-0001-9778-2026 (П.Е. Жгутов)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Введение

Компьютерные технологии в сфере обработки изображений и компьютерного зрения повсеместно применяются в современном мире. С их помощью решаются множественные задачи обнаружения, локализации, слежения, мониторинга и регулирования.

Зачастую системы компьютерного зрения решают задачи обнаружения и слежения в рамках более глобальных задач мониторинга и анализа дорожного движения. Подобные системы способны не только оптимально управлять дорожным движением (камеры подсчета трафика на перекрестке), но и обеспечивать высокий уровень безопасности (системы предупреждения водителей на пешеходных переходах, расположенных в опасных и аварийных местах и т.д.). Слежение за человеком является неотъемлемой частью подобных систем. Система, обеспечивающая обнаружение, локализацию и слежение за человеком, должна обладать высокой точностью и быстродействием, чтобы соответствовать критериям безопасности.

В современных реалиях одновременное достижение точности и быстродействия является задачей сложной и требовательной к аппаратному обеспечению. Внедрение системы компьютерного зрения в городскую среду подразумевает использование большого количества устройств, рассчитанных на самостоятельную работу без помощи глобальных вычислительных центров. Массовость применения подразумевает использование устройств относительно простых и дешевых, чаще использующих в качестве вычислителя процессоры для встраиваемых систем, реже – ПЛИС (программируемые логические интегральные схемы) начального уровня. Данные условия накладывают сильные ограничения на использование сложных алгоритмов по типу сверточных нейросетей с большим количеством слоев. Использование классических и обучаемых алгоритмов позволяет успешно разместить программное обеспечение на устройстве, однако у таких алгоритмов имеются проблемы с точностью. Они проявляются в виде ложных срабатываний и пропусков объекта. Оптимальным путем создания системы компьютерного зрения, отвечающую заданным требованиям, является разработка простых алгоритмов обнаружения и слежения, настраиваемых вручную, способных выполняться непосредственно на устройствах с ограниченными вычислительными мощностями. Основной проблемой в таком подходе является поиск алгоритма, соответствующего показателям качества.

2. Разработка комбинированного алгоритма

В настоящее время системы предупреждения участников дорожного движения, системы мониторинга и аналитики, построенные на использовании компьютерного зрения, используются уже достаточно долго. В мире системы видеоаналитики используются для мониторинга трафика, подсчета участников дорожного движения. Системы видеоаналитики помогают развивать городскую инфраструктуру, создавать благоприятную и безопасную среду для жителей города. Информация, полученная в результате анализа мониторингов, помогает спроектировать новые современные районы, проложить новые или обустроить старые дороги и тротуары.

Однако при всех достоинствах таких систем используются они далеко не повсеместно по ряду причин:

- Система должна быть массовой и обязана охватывать большую часть городского пространства;
- Система должна обладать высокой точностью и быстродействием;
- Система должна обладать приемлемой стоимостью, т.к. она является массовой.

Поэтому при желании использовать систему компьютерного зрения в городской среде, нужно внимательно изучить задачи и потребности, которые перед ней ставятся. Задачи могут быть комплексными, как одновременный мониторинг пешеходного и автомобильного трафика, или узкоспециализированными, как поиск пешеходов вблизи пешеходных переходов и предупреждение водителей о возможной опасности.

При разработке системы видеоаналитики выбор лежит между подходами, основанными на нейросетевых технологиях и на обучаемых или классических алгоритмах.

Нейросети обладают высокими показателями точности и быстродействия и не менее высокими требованиями к аппаратному обеспечению, что сильно усложняет создание массовой и простой системы видеоналитики. Классические и обучаемые алгоритмы просты в реализации и гораздо менее требовательны к аппаратному обеспечению, однако при увеличении быстродействия их точность существенно уменьшается. При решении узкоспециализированных задач выгоднее использовать комбинацию взаимодополняющих классических и обучаемых алгоритмов.

Целью данной работы является комбинирование нескольких относительно несложных алгоритмов для улучшения точности обнаружения пешеходов.

Для решения задачи обнаружения пешеходов был выбран детектор человека на базе функции ACF (aggregate channel function). Детектор обладает достаточной для решения поставленной задачи точностью и высоким быстродействием, однако имеет ряд серьезных недостатков. При работе детектора наблюдается повышенный уровень ложных срабатываний. Детектор может производить выделение человекоподобных объектов (фотоаппаратные треноги, рекламные баннеры и др.). Для устранения ложных срабатываний необходимо сократить область интереса детектора до зоны, в которой обнаружено движение. Также в силу особенностей алгоритма ему свойственны пропуски объектов. Обуславливается это тем, что алгоритму требуется предварительное обучение, а подобрать всеобъемлющую обучающую выборку невозможно. Для решения данных проблем было решено скомбинировать детектор пешеходов вместе с алгоритмами выделения движения и слежения за объектом. Структурная схема предложенного алгоритма представлена на рисунке 1.

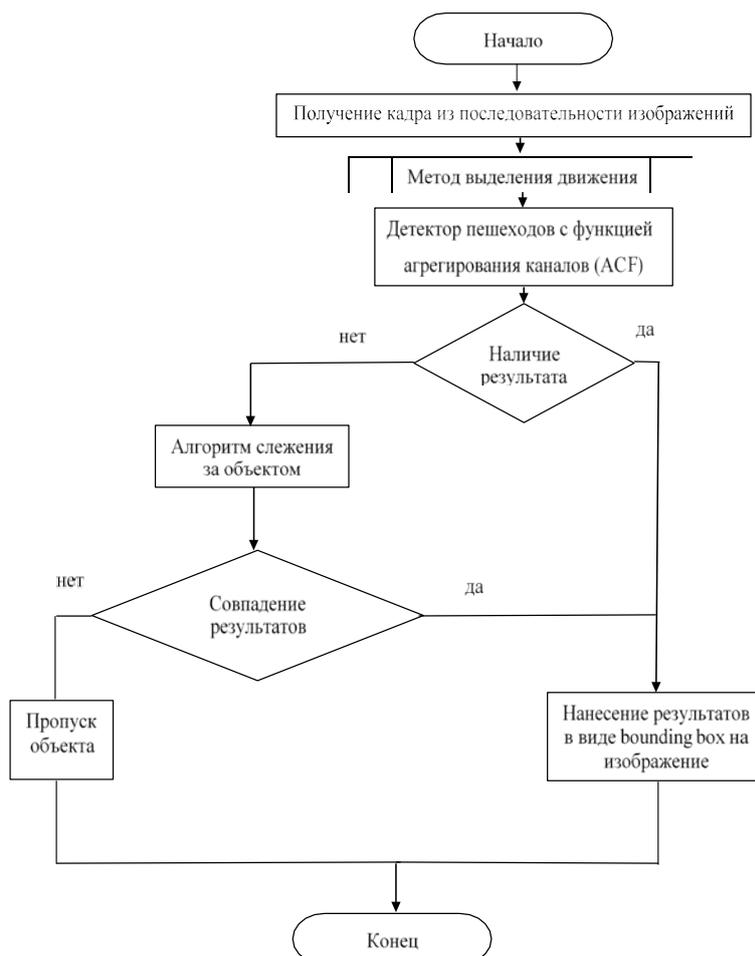


Рисунок 1 – Структурная схема комбинированного алгоритма

2.1. Детектор пешеходов с функцией агрегирования каналов (ACF)

Детектор людей с функцией ACF является одной из последних итераций детекторов, разработанных командой Петра Доллара (Piotr Dollar) [1]. Основными отличиями от детектора, разработанного в 2000-х годах, является замена метода обучения на базе SVM (Support Vector Machine, метод опорных векторов) на метод обучения на основе деревьев, использование функции Fast Feature Pyramid для вычисления признаков при пирамидальном (мульти-масштабном) представлении сигнала, а также применение уже упомянутой функции ACF. Функция ACF позволяет формировать вектор признаков [2], состоящий из 10 каналов, из которых 6 занимают направления гистограммы направленных градиентов (HOG) [3] (6 направлений по 30°), 3 канала занимают признаки, извлеченные с помощью цветовой модели LUV, еще один канал занимает нормированный модуль градиента. Также одним из самых важных преимуществ данного детектора является метод Fast Feature Pyramids, разработанный Петром Долларом, который позволяет избежать прямого вычисления признаков на каждом уровне пирамиды. В работе [1] описан способ, который позволяет вычислять признаки на уровнях внутри октавы за счет масштабирующих коэффициентов. Такой подход существенно увеличивает скорость работы алгоритма. Также в работе [1] показано, что предложенный детектор достигает скорости обработки около 30 кадров в секунду. При этом использование классического метода вычисления HOG на пирамидальном изображении дает всего 12 кадров в секунду на том же вычислительном устройстве.

2.2. Алгоритм выделения движения

Описание предлагаемого алгоритма выделения движения приведено в [3,4]. Работа алгоритма начинается с выбора опорного кадра и начальной оценки фона. После выбора опорного кадра изображение из цветного переводится в полутоновое. Затем происходит накопление информации о яркости точек.

$$F_{\text{нак}} = \sum_{i=k_n}^{k_k} I_i, \quad (1)$$

где I_i – i -й кадр последовательности, k_n – опорный кадр задающийся оператором, k_k – заданное количество кадров для оценки фона.

Оценка фона $F_{\text{оценка}}$ осуществляется путем применения усредняющего фильтра для заданного числа кадров

$$F_{\text{оценка}} = F_{\text{нак}} / k_k, \quad (2)$$

После получения начальной оценки фона следует выделение движения методом нахождения разности $I_{\text{разнст}}$ между текущим кадром и оценкой фона, затем происходит переход к бинарному изображению bwI путем сравнения разности с порогом бинаризации T .

$$I_{\text{разнст}} = I_i - F_{\text{оценка}} \quad (3)$$

$$bwI = I_{\text{разнст}} > T \quad (4)$$

Для улучшения качества бинарного изображения последовательно применяются последовательные морфологические операции открытия и закрытия. Исходя из знаний о размере минимальных объектов на кадре в операции открытия используется структурирующий элемент с размерами меньшими, чем искомый объект, что позволяет устранить искажения бинарного изображения объекта, вызванного влиянием аддитивного шума без значительного искажения формы объекта. В операции морфологического закрытия

применяется структурирующий элемент приблизительно в 2 раза больший, нежели в операции открытия, что позволяет связать близлежащие области объекта и более точно его отобразить. Исходное изображения приведено слева, результат обработки – справа. Область, в которой зафиксировано движение, выделена прямоугольником.

Результат обработки тестового изображения приведен на рисунке 2.



Рисунок 2 – Пример выделения объекта по признаку движения:
а – исходное изображение, б – результирующее бинарное изображение

2.3. Алгоритм слежения за объектом

Для того, чтобы минимизировать пропуски объекта, было решено использовать алгоритм слежения, основанный на оценке межкадрового сдвига объекта. Работа алгоритма заключается в оценке скорости передвижения объекта по межкадровому сдвигу и прогнозирования местоположения объекта на следующих кадрах на основе полученной информации.

Межкадровый сдвиг вычисляется нахождением разности между координатами центров ограничивающих прямоугольников (bounding box) объекта, полученных в результате работы алгоритма выделения движения на текущем и предыдущем кадрах. Для отслеживания изменений динамики движения человека используется усреднение методом скользящего окна. Благодаря этому происходит регулярное обновление данных, по которым вычисляется межкадровый сдвиг. Это позволяет учитывать изменения в движении человека – увеличивается точность прогнозирования.

Для увеличения надежности алгоритма слежения используется проверка точности прогнозирования. Когда предложенный алгоритм сталкивается с пропуском, происходит сравнение результатов, полученных от детектора и алгоритма слежения на предыдущем кадре. Оценивается расстояние между центрами ограничивающих прямоугольников, и на основе оценки принимается решение об использовании результатов алгоритма слежения на текущем кадре. Пример работы алгоритма слежения представлен на рисунке 3.



Рисунок 3 – Пример работы алгоритма слежения: а – кадр видеопоследовательности, обработанный алгоритмом без функции слежения; б – кадр видеопоследовательности, обработанный алгоритмом с включенной функцией слежения

2.4. Комбинирование детектора пешеходов с функцией ACF и метода выделения движения

Как упоминалось ранее, самостоятельная работа детектора дает средние показатели точности из-за того, что генерирует множество ложных срабатываний. Так как детектор работает не с последовательностью изображений, а обрабатывает каждое изображение независимо от предыдущих, то ложные срабатывания в подавляющем количестве случаев появляются на неподвижных объектах, частично напоминающих облик человека (рисунок 4).

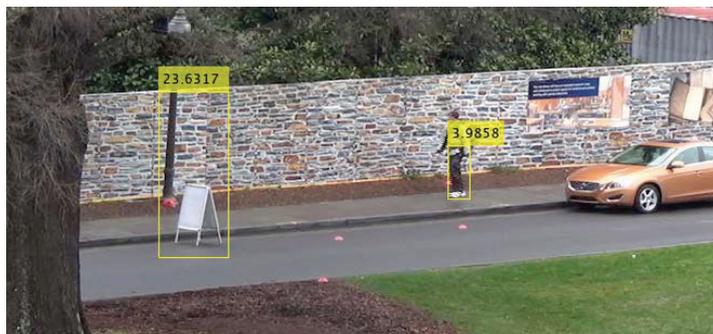


Рисунок 4 – Пример ложного срабатывания

Чтобы исключить подобные ложные срабатывания, необходимо сократить область интереса детектора (ROI) до зоны, в которой выявлено движение. Для решения данной задачи был выбран алгоритм выделения движения (его описание приведено выше). Перед тем, как подать изображение на вход детектору пешеходов, оно предварительно обрабатывается детектором движения. Если такая зона существует, то она описывается с помощью ограничивающего прямоугольника, после этого информация передается в детектор в качестве области интереса.

Благодаря такому решению достигнуто сразу несколько положительных результатов. Во-первых, удалось исключить ложные срабатывания на неподвижных человекоподобных объектах. Во-вторых, повысилась степень уверенности детектора в принятии решения. В-третьих, несмотря на добавление дополнительного этапа обработки удалось увеличить скорость работы. Это объясняется тем, что вычисление НОГ-дескрипторов происходит на площади существенно меньшей, чем площадь всего изображения.

3. Экспериментальные исследования

Для оценки работы предложенного комбинированного алгоритма были проведены экспериментальные исследования, в которых сравнивались результаты самостоятельной работы детектора пешеходов с результатами работы комбинированного алгоритма.

Чтобы оценить точность выделений объекта с помощью тестируемых алгоритмов, для каждой видеопоследовательности, используемой в исследованиях, были созданы маски объекта, представляющие собой ограничивающие прямоугольники, заданные вручную. Мерами качества были выбраны стандартные метрики для задач машинного обучения – частота правильных обнаружений (TP, True Positive), частота ложных выделений (FP, False Positive) и частота пропусков объекта (FN, False Negative). Использование метрики TN (True Negative) является нецелесообразным, так как в данной работе нет задачи разделения выделенных объектов на разные классы.

В ходе эксперимента результаты работы тестируемых алгоритмов сравнивались с масками на предмет совпадения. Основопологающим фактором в принятии решения о правильности выделения является расстояние между центрами маски и центром ограничивающего прямоугольника, выдаваемого алгоритмом. Допустимой считается ошибка равная 25% от размеров маски.

Также исследовалась зависимость метрик от параметра алгоритма, отвечающего за площадь прямоугольника. Данный параметр отвечает за максимальные размеры

прямоугольника, который может полностью описать человека на изображении. Правильный выбор параметра – важная часть настройки детектора. Подбор прямоугольника под точные размеры человека дает высокие показатели точности относительно FP и TP, однако детектор при такой настройке допускает множество пропусков объекта. С увеличением площади снижается количество пропусков объекта, но с тем же увеличивается количество ложных срабатываний.

На основе проведенных исследований были построены графики зависимости метрик от площади прямоугольника. Графики представлены на рисунках 5 – 6.

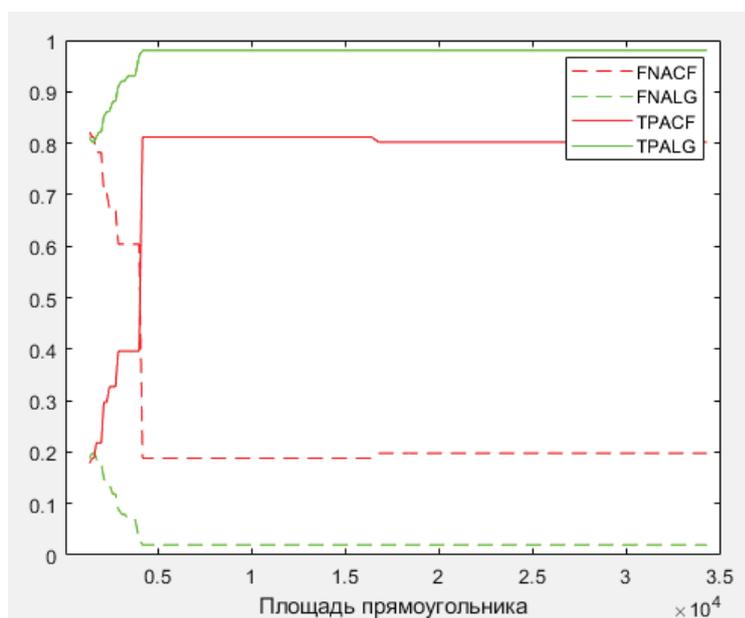


Рисунок 5 – Графики зависимости показателей TP и FN от площади прямоугольника для самостоятельной работы детектора на основе функции ACF (красные линии) и для комбинированного алгоритма (зеленые линии)

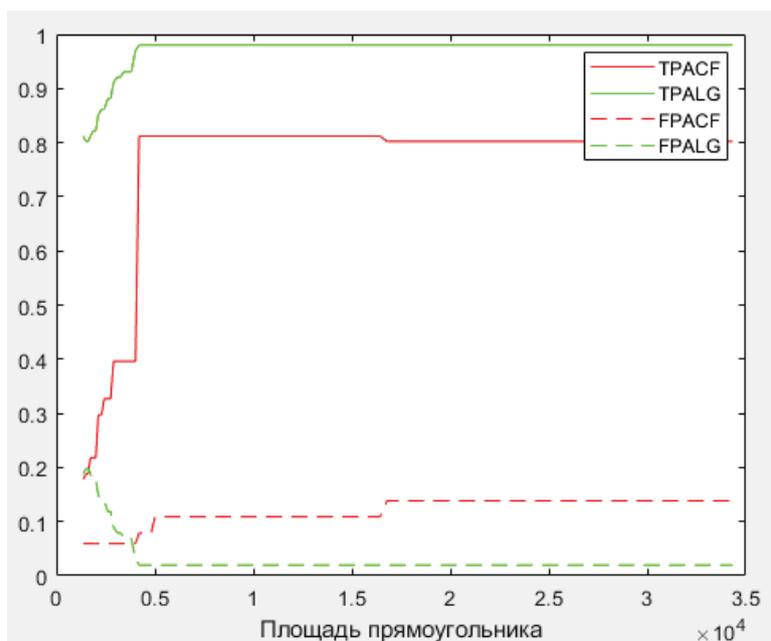


Рисунок 6 – Графики зависимости показателей TP и FP от площади прямоугольника для самостоятельной работы детектора на основе функции ACF (красные линии) и для комбинированного алгоритма (зеленые линии)

На графиках, представленных на рисунке 5, видно, что частота правильных обнаружений TP выше у комбинированного алгоритма и близка к единице. При этом частота пропусков FN также ближе к нулю, чем при самостоятельной работе детектора. Можно отметить тот факт, что при увеличении площади прямоугольника, частота правильных выделений TP увеличивается, а частота ложных срабатываний FP уменьшается.

По графикам зависимостей TP и FP от площади прямоугольника (рисунок 5) можно отметить, что при использовании комбинированного алгоритма с ростом площади прямоугольника частота ложных срабатываний уменьшается, а неувеличивается, как в случае самостоятельной работы детектора. Также были проанализированы данные об ошибках, на основе которых результат работы относили к той или иной метрике. В результате анализа было выявлено, что ложные срабатывания, выявленные в работе комбинированного алгоритма вызваны самим объектом, за которым ведется слежение.

По результатам экспериментальных исследований можно сделать вывод, что комбинированный алгоритм обладает лучшими показателями выбранных метрик – TP, FP и FN, чем самостоятельно работающий детектор. Стоит отметить, что при использовании комбинированного алгоритма удалось повысить степень уверенности детектора примерно на 15%.

4. Заключение

В данной работе рассматривается решение задачи обнаружения пешеходов на видеопоследовательности [5]. Предложен комбинированный алгоритм обнаружения пешеходов, представляющий собой комбинацию детектора пешеходов, алгоритма выделения движения и функции слежения.

По проведенным экспериментальным исследованиям сделан вывод о работоспособности предложенного комбинированного алгоритма. В данных исследованиях проводилось сравнение результатов самостоятельной работы детектора пешеходов с функцией ACF и результатов, полученных от комбинированного алгоритма. На основе экспериментальных исследований можно сделать вывод, что предложенный комбинированный алгоритм обладает лучшими показателями точности, на используемых в экспериментах видеопоследовательностях исключен факт ложных срабатываний, возросла уверенность детектора в принимаемом решении. В сравнении с детектором пешеходов точность комбинированного алгоритма увеличилась на 20%.

5. Список источников

- [1] Dollar, P., R. Appel, S. Belongie, and P. Perona. Fast Feature Pyramids for Object Detection. // IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 36, Issue 8, 2014, pp. 1532–1545.
- [2] Fajar A., Handayani T., Nanik S. Combination of Aggregated Channel Features (ACF) Detector and Faster R-CNN to Improve Object Detection Performance in Fetal Ultrasound Images. // International Journal of Intelligent Engineering and Systems, Vol.11, No.6, 2018. pp. 65-74.
- [3] N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. // Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pages 886–893. IEEE, 2005.
- [4] Алпатов Б. А., Бабаян П. В., Ершов М. Д. Подходы к обнаружению и оценке параметров движущихся объектов на видеопоследовательности применительно к транспортной аналитике // Компьютерная оптика. – 2020. – Т. 44. – №. 5, стр. 746-758.
- [5] Ершов М.Д., Шубин Н.Ю. Алгоритмы обработки изображений для решения задач анализа дорожной обстановки // Цифровая обработка сигналов. №3. 2017. – С.63-64.