

Телевидение и обработка изображений

УДК 004.931; 004.932

Оригинальная статья

<https://doi.org/10.32603/1993-8985-2021-24-3-49-59>

Метод автоматического определения трехмерной траектории транспортных средств на изображении

И. Г. Зубов^{1✉}, Н. А. Обухова²

¹ООО "НЕКСТ", Москва, Россия

²Санкт-Петербургский государственный электротехнический
университет "ЛЭТИ" им. В. И. Ульянова (Ленина), Санкт-Петербург, Россия

✉ZubovIG@gmail.com

Аннотация

Введение. Важной составной частью системы управления беспилотным транспортным средством (ТС) является модуль анализа окружающего пространства. Традиционно его реализуют на основе датчиков разного типа, включая видеокамеры, лидары и радары. Развитие вычислительной и телевизионной техники позволяет реализовать модуль анализа окружающего пространства используя в качестве датчиков только видеокамеры, что снижает себестоимость модуля в целом. Основной задачей при обработке видеоданных является анализ окружающего пространства как трехмерной сцены. Трехмерная траектория объекта, в которой наряду с его локализацией на изображении учтены также габаритные размеры, ракурс (угол поворота) и вектор движения, предоставляет исчерпывающую информацию для анализа реального взаимодействия объектов. Основой построения трехмерной траектории является оценка ракурса ТС.

Цель работы. Разработка метода автоматической оценки ракурса ТС на основе анализа видеоданных от одной видеокамеры.

Методы и материалы. Предложен автоматический метод оценки ракурса ТС на изображении на основе каскадного подхода. Метод включает локализацию ТС, определение его ключевых точек, сегментацию ТС и оценку ракурса. Локализация ТС и определение его ключевых точек решены на основе сверточной нейронной сети. Сегментацию ТС и формирование маски объекта выполняют с переходом в полярную систему координат и поиском внешнего контура с помощью алгоритмов теории графов. Целевой ракурс ТС определяют сопоставлением Фурье-образа сигнатур маски ТС и шаблонов, полученных на основе трехмерных моделей.

Результаты. Эксперименты подтвердили корректность определения ракурса ТС на основе предложенного метода. Точность определения ракурса ТС на открытом наборе изображений Carvana составила 89 %.

Заключение. Предложен новый подход к задаче оценки ракурса ТС, предполагающий переход от сквозного обучения нейронных сетей для решения сразу нескольких задач, таких, как локализация, классификация, сегментация и определение ракурса, к каскадному анализу информации. Обеспечение высокой точности оценки ракурса при сквозном обучении требует больших репрезентативных наборов данных, что затрудняет масштабируемость решений для условий российских дорог. Предложенный метод позволяет определять ракурс ТС с высокой точностью без больших затрат на ручную аннотацию данных и обучение.

Ключевые слова: сверточные нейронные сети, анализ карт активаций, локализация ключевых точек, сегментация изображений, сопоставление с шаблоном

Для цитирования: Зубов И. Г., Обухова Н. А. Метод автоматического определения трехмерной траектории транспортных средств на изображении // Изв. вузов России. Радиоэлектроника. 2021. Т. 24, № 3. С. 49–59. doi: 10.32603/1993-8985-2021-24-3-49-59

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию 09.04.2021; принята к публикации после рецензирования 10.05.2021; опубликована онлайн 29.06.2021

© Зубов И. Г., Обухова Н. А., 2021



Контент доступен по лицензии Creative Commons Attribution 4.0 License
This work is licensed under a Creative Commons Attribution 4.0 License

Method for Automatic Determination of a 3D Trajectory of Vehicles in a Video Image

Ilya G. Zubov^{1✉}, Natalya A. Obukhova²

¹ Ltd "Next", Moscow, Russia

² Saint Petersburg Electrotechnical University, St Petersburg, Russia

✉ ZubovIG@gmail.com

Abstract

Introduction. An important part of an automotive unmanned vehicle (UV) control system is the environment analysis module. This module is based on various types of sensors, e.g. video cameras, lidars and radars. The development of computer and video technologies makes it possible to implement an environment analysis module using a single video camera as a sensor. This approach is expected to reduce the cost of the entire module. The main task in video image processing is to analyse the environment as a 3D scene. The 3D trajectory of an object, which takes into account its dimensions, angle of view and movement vector, as well as the vehicle pose in a video image, provides sufficient information for assessing the real interaction of objects. A basis for constructing a 3D trajectory is vehicle pose estimation.

Aim. To develop an automatic method for estimating vehicle pose based on video data analysis from a single video camera.

Materials and methods. An automatic method for vehicle pose estimation from a video image was proposed based on a cascade approach. The method includes vehicle detection, key points determination, segmentation and vehicle pose estimation. Vehicle detection and determination of its key points were resolved via a neural network. The segmentation of a vehicle video image and its mask preparation were implemented by transforming it into a polar coordinate system and searching for the outer contour using graph theory.

Results. The estimation of vehicle pose was implemented by matching the Fourier image of vehicle mask signatures and the templates obtained based on 3D models. The correctness of the obtained vehicle pose and angle of view estimation was confirmed by experiments based on the proposed method. The vehicle pose estimation had an accuracy of 89 % on an open Carvana image dataset.

Conclusion. A new approach for vehicle pose estimation was proposed, involving the transition from end-to-end learning of neural networks to resolve several problems at once, e.g., localization, classification, segmentation, and angle of view, towards cascade analysis of information. The accuracy level of end-to-end learning requires large sets of representative data, which complicates the scalability of solutions for road environments in Russia. The proposed method makes it possible to estimate the vehicle pose with a high accuracy level, at the same time as involving no large costs for manual data annotation and training.

Keywords: convolutional neural networks, analysis of activation maps, detection of key points, image segmentation, pattern matching

For citation: Zubov I. G., Obukhova N. A. Method for Automatic Determination of a 3D Trajectory of Vehicles in a Video Image. Radioelectronics. 2021, vol. 24, no. 3, pp. 49–59. doi: 10.32603/1993-8985-2021-24-3-49-59

Conflict of interest. The author declares no conflict of interest.

Submitted 09.04.2021; accepted 10.05.2021; published online 29.06.2021

Введение. Системы технического зрения используют сегодня для решения широкого круга задач. Среди них одной из наиболее приоритетных является задача обеспечения безопасности на дороге, решение которой предполагает развитие беспилотных транспортных средств (ТС).

Важная составная часть системы управления беспилотным ТС – модуль анализа окружающего пространства. Традиционно его реализуют на основе датчиков разного типа, включая видеокамеру, лидары и радары. Датчики последних двух типов являются достаточно дорогостоящими.

Развитие вычислительной и телевизионной техники позволяет реализовать модуль анализа окружающего пространства, используя в качестве датчика только видеокамеру, что существенно снижает себестоимость модуля в целом.

Внедрение датчиков на основе видеокамеры (монокулярной камеры) доказало свою эффективность в таких задачах, как анализ дорожного трафика и система помощи водителю [1, 2]. Эти системы позволяют уменьшить риск столкновения с пешеходом и иным ТС, а также предотвратить сход с полосы движения и т. д.

подавляющее большинство существующих систем на основе монокулярных камер ориентировано на анализ окружающего ТС пространства как двумерного изображения [3–9]. Однако объекты взаимодействуют в физическом пространстве сцены, которое является трехмерным. Следовательно, основной задачей систем технического зрения является анализ окружающего пространства как трехмерной сцены. В силу этого для получения исчерпывающей информации об объектах и для анализа их реального взаимодействия необходимо определить их трехмерные траектории. Под такой траекторией подразумевается локализация объекта на изображении, определение его габаритных размеров, ракурса (угол поворота) α [10] (рис. 1) и вектора движения.

Базой для построения трехмерной траектории является определение ракурса ТС. Современные методы определения ракурса на основе видеоданных, как правило, основаны на глубоком обучении (Deep Learning) нейросетевых моделей [11–14]. В подавляющем большинстве случаев сверточные нейронные сети (НС) для задачи оценки ракурса ТС обучаются сквозным способом для решения сразу нескольких

задач, таких, как локализация, классификация, сегментация, определение ракурса и т. д. Точность такого подхода в основном зависит от репрезентативности обучающей выборки. Открытые наборы данных, как правило, не включают в себя изображения, полученные при различных условиях наблюдения (различные погодные условия, разное время суток и точка наблюдения и т. д.), поэтому разработанные методы являются слабо масштабируемыми.

Определение ракурса ТС на основе видеоданных – задача, слабо изученная по сравнению с такими задачами, как локализация, классификация и сегментация объектов. Для последних перечисленных задач существуют большие наборы открытых размеченных данных. Представляется целесообразным разбить задачу оценки ракурса на подзадачи, уровень изученности которых намного больше.

В настоящей статье предложен новый метод определения ракурса ТС, основанный на каскадном анализе информации. Метод позволяет определить ракурс ТС без больших затрат на ручную аннотацию данных и обучение. Достигнутая точность определения ракурса ТС составляет 89 % на открытом наборе изображений Carvana.

Структурная схема предложенного метода представлена на рис. 2. На первом шаге анализируют изображение нейросетевым детектором объектов, в результате чего получают информацию о местоположении ТС на изображении. Далее с помощью предложенного алгоритма анализа внутренних сверточных слоев НС и изменения ее архитектуры локализуют ключевые точки, принадлежащие объекту интереса.

На третьем шаге, используя априорную информацию о местоположении ТС и ключевых точек на изображении, проводят сегментацию изображения и формируют маску объекта. Из полученной маски ТС извлекают сигнатуры – зависимости расстояния от ключевой точки до границ объекта как функции угла. Фурье-образ сигнатур является ключевым признаком, на основе которого определяют ракурс ТС.

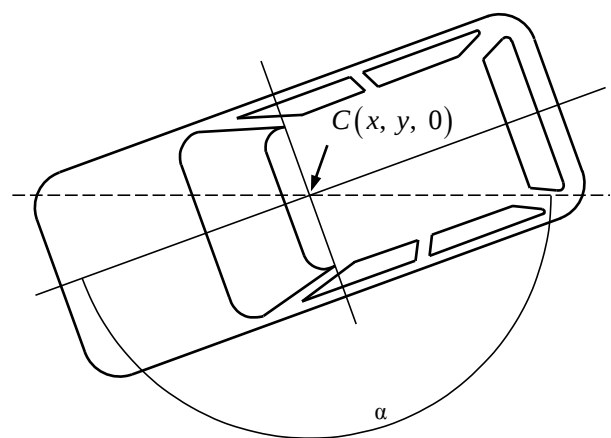


Рис. 1. Автомобиль, описываемый ориентацией α и центроидом на плоскости земли C

Fig. 1. Vehicle, described by orientation α and centroid on the ground plane C

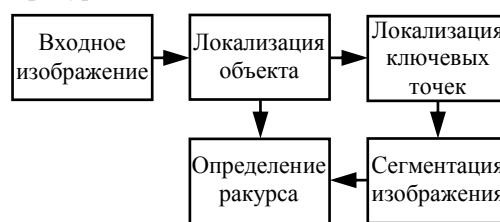


Рис. 2. Функциональная схема метода определения ракурса ТС

Fig. 2. Block scheme of the method for determining the vehicle pose

Для оценки ракурса предварительно на основе трехмерных моделей формируют шаблоны возможных ракурсов ТС. Целевой ракурс ТС определяют сопоставлением Фурье-образа сигнатур маски ТС и шаблонов, полученных на основе трехмерных моделей. Ракурс целевого ТС определяют как ракурс шаблона, имеющего наибольший коэффициент корреляции Пирсона [15].

Далее в настоящей статье подробно рассмотрены описанные шаги предложенного каскадного метода.

Локализация ТС. ТС являются одним из основных объектов, присутствующих во многих классических наборах данных для обнаружения и распознавания, таких, как PASCAL-VOC [16], ImageNet [17] и COCO [18], поэтому на шаге локализации ТС целесообразно использовать известную модель. Приняв на вход изображение, НС выдает координаты ограничивающих объекты прямоугольников и вероятности принадлежности выделенных изображений объектов классам. При выборе НС учтены следующие критерии: высокое быстродействие и точность локализации, достаточная для дальнейшего эффективного анализа. В результате использована сеть YOLOv4 [9], обладающая высокой скоростью анализа (около 65 кадров в секунду на NVIDIA Tesla V100) и достаточной средней точностью (43.5 % по набору данных COCO).

Локализация ключевых точек. Как показано в [19], промежуточные выходы сверточных НС часто связаны с семантическими частями объектов, причем несколько промежуточных сверточных слоев могут быть ассоциированы с одной и той же частью [20]. В последнем случае объединение выходных сигналов этих слоев позволяет локализовать ключевую точку объекта интереса.

Для извлечения ключевых точек, принадлежащих отдельным частям ТС, была проанализирована НС YOLOv4 с базовой структурой CSPDarknet53 [9]. Чтобы произвести анализ каждого слоя НС, необходимо изменить ее архитектуру. Для анализа внутренних слоев был добавлен выход к каждому активационному слою с последующим масштабированием выходной матрицы до разрешения входного изображения (рис. 3).

Измененная структура имеет 2 выхода: один, локализуя объекты на изображении, а второй в виде матрицы всех промежуточных сверточных слоев базовой структуры с размерами $W \times H \times N$, где W и H – высота и ширина входного изображения соответственно; N – количество промежуточных сверток НС.

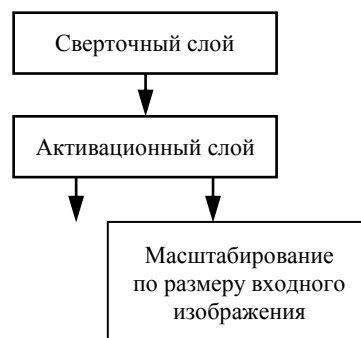


Рис. 3. Добавление выходного канала в базовую структуру
Fig. 3. Adding an output channel to the basic structure

Для дальнейшего анализа необходим набор аннотированных данных X – множество изображений (I) и масок (A_k). Набор должен содержать информацию о местоположении k -й интересующей части объекта. Пример аннотации представлен на рис. 4.

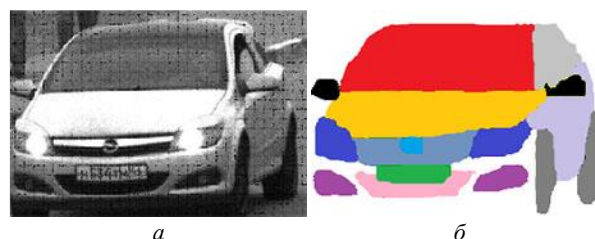


Рис. 4. Пример аннотации: a – исходное изображение; b – маски частей ТС

Fig. 4. An example of annotation: a – the original image; b – masks of vehicle constructive parts

Для определения ключевой точки необходимо выделить сверточные слои и каналы, удовлетворяющие двум условиям:

– активирующиеся в интересующей области

$$\hat{l}_{ijk} = \{Q_{ijk} \geq th_1\};$$

– не активирующиеся в интересующей области

$$\tilde{l}_{ijk} = \{Q_{ijk} \leq th_2\},$$

где

$$Q_{ijk} = \sum_{\{I, A_k \in X\}} \frac{\sum_{\substack{y \in 1, H \\ x \in 1, W}} f_{ij}[I(y, x)] A_k(y, x)}{\sum_{\substack{y \in 1, H \\ x \in 1, W}} f_{ij}[I(y, x)]}$$

– доля сосредоточения активации в области интереса в j -м канале i -го слоя НС; th_1 , th_2 – пороги, причем $f_{ij}(I)$ – результат функции активации i -го слоя, j -го канала, масштабированный до размера входного изображения. Порог th_1 выбирается экспериментально.

Порог th_2 выбирается в зависимости от функции активации, используемой в анализируемой архитектуре. Например, при использовании функции активации ReLU [21], $th_2 = 0$ будет указывать на полное отсутствие активации нейрона в области интереса.

Элементы, не входящие в множества $\hat{l}_{ijk}$ и $\tilde{l}_{ijk}$, могут быть удалены из дополнительного выходного слоя измененной архитектуры НС (см. рис. 3).

Агрегируя данные с выхода базовой структуры НС, получим новое изображение I' , элементы которого, локализованные в интересующей части объекта, обладают наибольшими значениями интенсивности (рис. 5):

$$F(x, y, k) = \sum_{\substack{i, j, k \in \hat{l}_{ijk} \\ y \in \hat{H} \\ x \in \hat{W}}} f_{ij}[I'(x, y)] - \sum_{\substack{i, j, k \in \tilde{l}_{ijk} \\ y \in \tilde{H} \\ x \in \tilde{W}}} f_{ij}[I'(x, y)].$$

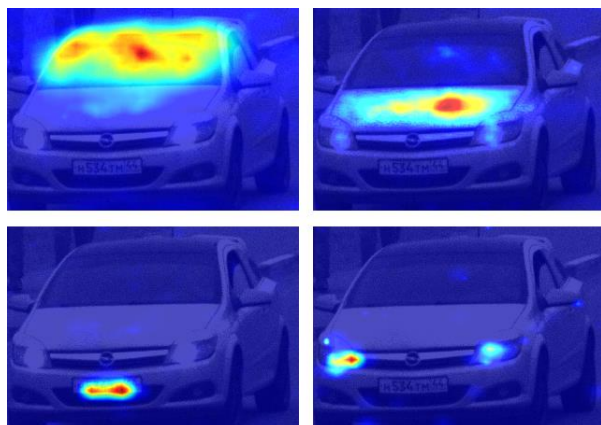


Рис. 5. Пример визуализации агрегированных выходов базовой структуры нейронной сети YOLOv4

Fig. 5. Visualization sample of the intermediate outputs of the convolutional layers of YOLOv4 neural network

НС принимает на вход изображение I' и выдает координаты ограничивающего прямоугольника: левого верхнего угла $x_{\min}, y_{\min}$ и правого нижнего угла $x_{\max}, y_{\max}$. Чтобы получить координаты ключевой точки этого изображения, воспользуемся выражением

$$x_{\text{кл}}, y_{\text{кл}} = \arg \max_{\substack{x' \in x_{\min}, x_{\max}; \\ y' \in y_{\min}, y_{\max}}} F(x', y').$$

Сегментация изображения. Сегментация изображения производится на основе модификации ранее предложенного одним из авторов настоящей статьи метода [22].

Основные шаги метода сегментации.

1. *Масштабирование изображения.* Изображение (рис. 6, а) масштабируется с помощью метода билинейной интерполяции [23] таким образом, чтобы соотношение сторон составляло 1:1 (рис. 6, б).

2. *Детектирование границ.* На данном шаге выделяют контуры с помощью фильтра Кэнни [24] (рис. 6, в).

3. *Преобразование в полярную систему координат.* Каждый пиксел $M(x', y')$ полученного множества описывается в полярной системе координат, используя центр масштабированного изображения как начало координат:

$$r = \sqrt{(x' + w/2)^2 + (y' + h/2)^2}; \quad (1)$$

$$\text{tg } \varphi = (2y' + h)/(2x' + w), \quad (2)$$

где r – расстояние от пиксела до центра ограничивающего объект параллелепипеда $O = (w/2, h/2)$;

$h = y_{\max} - y_{\min}$; $w = x_{\max} - x_{\min}$; φ – угол, образованный лучом OM с полярной осью.

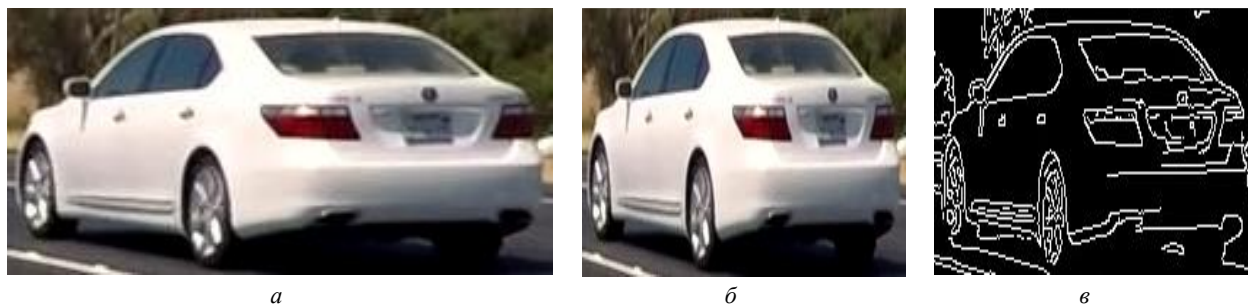


Рис. 6. Предобработка изображения: а – исходное изображение выделенного объекта; б – масштабированное изображение; в – результат обработки детектором Кэнни

Fig. 6. Image preprocessing: а – the original image of the selected object; б – the scaled image; в – the result of processing using the Canny detector

Преобразование элементов множества в полярную систему координат позволяет описать распределение внешнего контура объекта во всей области полярного угла $0 < \varphi \leq 2\pi$.

4. *Поиск кратчайшего пути в графе.* Исходя из того, что внешний контур объекта определен для всего диапазона полярного угла от 0 до 2π , задачу выделения внешнего контура объекта на изображении можно представить как поиск кратчайшего пути во взвешенном графе. Таким образом, проблема нахождения внешнего контура сводится к поиску пути наименьшего веса.

Представим результат полярного преобразования на шаге 3 в прямоугольной системе координат (рис. 7) с высотой h_φ и шириной w_φ . Полученное изображение R представляется в виде ориентированного взвешенного графа G , ребрами на котором соединены только ближайшие соседи. Под ближайшим соседом пиксела q понимаются пикселы p , каждая из координат которых отличается от координат q не больше чем на единицу:

$$(\varphi'_p - \varphi'_q = 1) \vee (|r'_p - r'_q| \leq 1),$$

где $\varphi'_q, r'_q; \varphi'_p, r'_p$ – координаты пикселов q и p соответственно в представлении графа в прямоугольной системе координат.

Существенно влияет на результаты сегментации определение веса ребра, характеризующего меру сходства ближайших соседей. С учетом того, что внешний контур объекта расположен вблизи наибольшего радиуса, вес ребер графа, разделяю-

щего 2 соседних пиксела p и q , выбран следующим образом:

$$d(p, q) = 2h_\varphi - R(\varphi'_q, r'_q)r'_q - R(\varphi'_p, r'_p)r'_p.$$

Из существующего разнообразия методов, использующих теорию графов, выбран алгоритм поиска кратчайшего пути A^* [25]. Этот алгоритм находит путь наименьшей стоимости от заданного начального до целевого узла (из одной или нескольких возможных целей). A^* использует оценку наименьших затрат вида

$$f(v) = g(v) + j(v),$$

где v – текущая вершина; $g(v)$ – наименьшее расстояние от стартовой вершины до текущего положения; $j(v)$ – эвристическая функция (манхэттенское расстояние) приближения расстояния от текущего местоположения до конечной цели.

В качестве начальной и конечной вершин были выбраны точки с координатами $(0, h_\varphi/2)$ и $(w_\varphi, h_\varphi/2)$ соответственно.

На рис. 8 представлен результирующий путь с наименьшим весом, рассчитанный по алгоритму A^* для изображения на рис. 7.

5. *Формирование маски объекта.* Используя обратное преобразование выражений (1), (2), полученный путь переводится в декартову систему координат. В результате формируется замкнутый контур, описывающий объект интереса, – маска

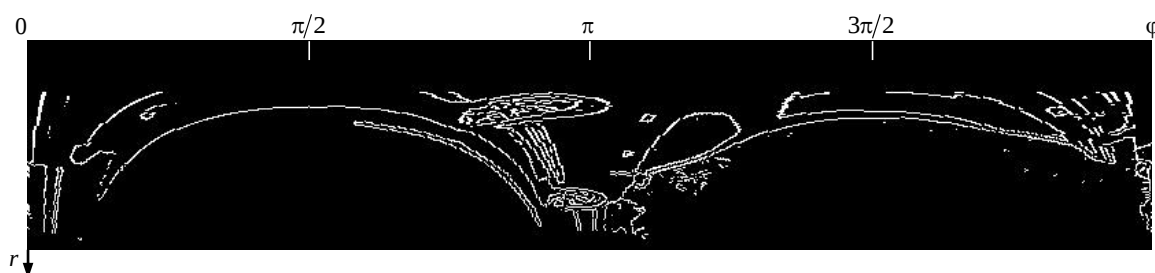


Рис. 7. Результат полярного преобразования, представленный в прямоугольной системе координат

Fig. 7. The result of a polar transformation represented in a rectangular coordinate system

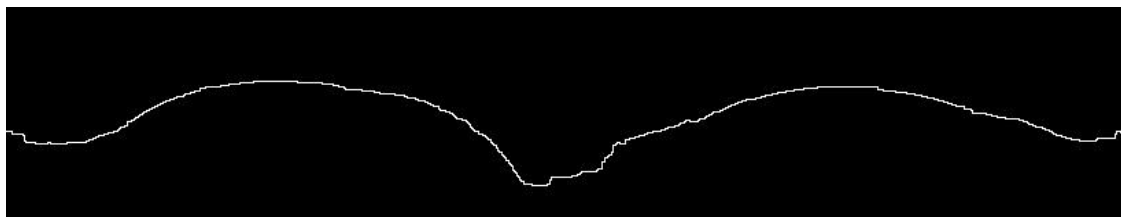


Рис. 8. Результат выделения кратчайшего пути по графу

Fig. 8. The result of the selection of the shortest path through the graph



Рис. 9. Результат работы алгоритма

Fig. 9. The result of the algorithm operation

объекта. Расположив маску на изображении, пиксели, попавшие в ее область, выделяются в качестве изображения объекта (рис. 9). Тем самым решается задача формирования сегментированного изображения.

Определение ракурса. ТС является жестким объектом, внешний вид которого зависит от марки автомобиля и точки наблюдения. В связи с этим для классификации ракурса воспользуемся методом, основанным на сопоставлении признаков исходного объекта и шаблонов, полученных на основе трехмерных моделей [15].

Опишем полученный на предыдущем шаге контур с помощью сигнатуры (одномерной функции). Зададим сигнатуру как зависимость расстояния от ключевой точки объекта до его границы, нормированного на его максимальное значение, в функции угла (рис. 10):

$$r_n(\varphi) = r(\varphi) / \max r(\varphi).$$

Такие сигнатуры инвариантны к параллельному переносу, однако зависят от поворота и изменения масштаба. Чтобы получить инвариантность к изменению масштаба, воспользуемся дискретным преобразованием Фурье [24]. Вычислим ДПФ:

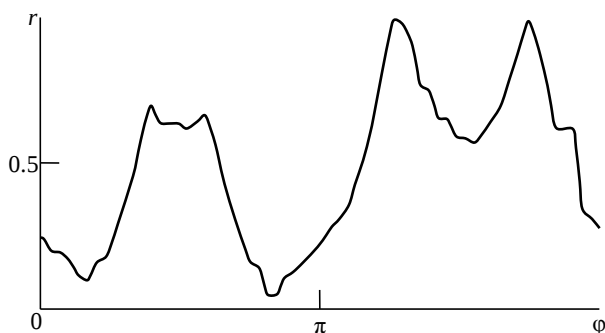


Рис. 10. Пример сигнатуры ТС

Fig. 10. An example of the vehicle signature

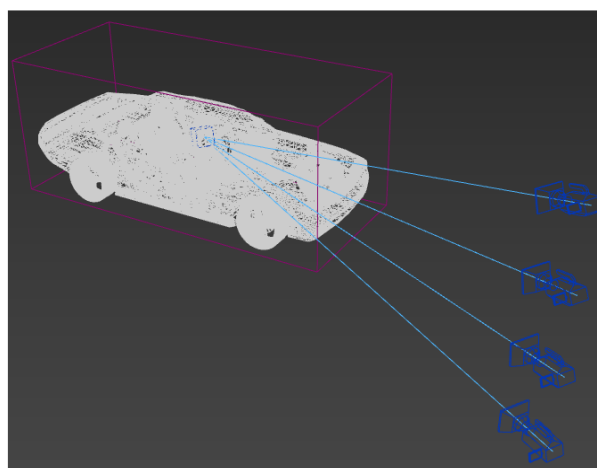


Рис. 11. Пример формирования шаблонов

Fig. 11. An example of forming templates

$$X_k = \sum_{\varphi=0}^{2\pi} r_n(\varphi) e^{-ik\varphi}.$$

В качестве набора эталонных данных предлагается использовать изображения разных ракурсов ТС, полученных из трехмерных моделей (рис. 11).

Трехмерные модели выбраны таким образом, чтобы представить внутриклассовые вариации определенной категории. Например, в качестве категории автомобиля выбраны седан, купе, внедорожник, хэтчбек и т. д. Высота точки наблюдения и количество ракурсов ТС выбираются в зависимости от условий эксплуатации.

Ракурс целевого ТС α (см. рис. 1) определяется как ракурс шаблона, с которым ракурс целевого ТС имеет наибольшее значение коэффициента корреляции Пирсона:

$$R_{xy} = \frac{\sum_{i=1}^m (\varphi_i - \bar{\varphi})(r_{ni} - \bar{r}_n)}{\sqrt{\sum_{i=1}^m (\varphi_i - \bar{\varphi})^2 \sum_{i=1}^m (r_{ni} - \bar{r}_n)^2}},$$

где $\bar{\varphi}$, $\bar{r}_n$ – математические ожидания полярного угла и полярного радиуса соответственно.

Локализованные ранее ключевые точки позволяют производить анализ ракурса в ограниченном диапазоне, что снижает количество необходимых сравнений с шаблоном и увеличивает точность метода.

Результаты. Экспериментальная проверка предложенного метода проводилась с использованием набора изображений Carvana [26], в котором содержится 318 ТС различных классов, каждое из которых представлено 16 ракурсами (пример на рис. 12).

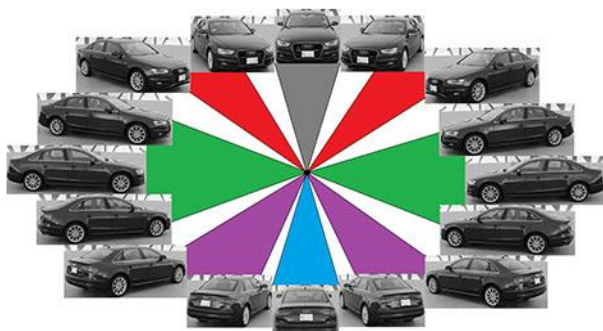


Рис. 12. Дискретизация ракурса ТС в зависимости от расположения ключевых точек

Fig. 12. Discretization of the vehicle pose estimation depending on key points collocation

Точность определение ракурса ТС для каждого класса

The accuracy level when determining the vehicle pose and angle of view for each class

Метрика	Класс															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
precision	0.97	0.92	0.84	0.94	0.74	0.86	0.78	0.92	0.93	0.80	0.96	0.98	0.86	0.95	0.95	0.99
recall	0.92	0.97	0.97	0.55	0.56	0.58	0.94	1.00	0.97	1.00	1.00	0.99	0.99	0.97	0.99	0.98
F1	0.95	0.94	0.90	0.69	0.64	0.69	0.86	0.96	0.95	0.89	0.98	0.98	0.92	0.96	0.97	0.98

Указанная база изображений была получена с помощью настраиваемой вращающейся фотостудии.

Для определения качества предложенного метода используются следующие метрики точности:

$$\text{precision} = \frac{TP}{TP + FP}; \text{ recall} = \frac{TP}{TP + FN};$$

$$F1 = \frac{TP}{TP + 0.5(FP + FN)},$$

где TP – правильно назначенные положительные классификации; FP – ложно назначенные положительные классификации; FN – ложно назначенные отрицательные классификации.

В таблице представлены результаты эксперимента для каждого отдельного класса, а на рис. 13 представлена матрица ошибок [27].

Для сравнения разработанного метода определения ракурса с аналогами была использована нейросетевая модель AAVR [28] в реализации, представленной в [29]. Средняя точность предложенного метода на наборе изображений Carvana составила 89 % против 68 % у аналога.

Вывод. В статье представлен новый метод автоматического определения трехмерной траектории ТС на изображении. Разработанный метод позволяет извлекать информацию о местоположении ключевых точек, сегментировать изображение и определять ракурс ТС. Работоспособность предложенного метода проверена тестированием на открытой базе изображений Carvana. Разработанный метод оценки ракурса ТС не требует трудоемкой разметки данных и обучения нейросетевой модели. Для определения ракурса ТС метод использует трехмерную модель ТС. Точность предложенного метода на базе изображений Carvana составила 89 %.

1	0.92	0	0	0	0	0	0	0	0.08	0	0	0	0	0	0	0
2	0	0.97	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0
3	0	0.02	0.97	0	0	0	0	0	0	0	0.02	0	0	0	0	0
4	0	0.03	0.15	0.55	0.08	0	0.03	0.03	0	0.04	0	0.02	0.05	0	0	0
5	0	0.02	0.02	0.03	0.56	0.07	0.02	0	0	0.18	0	0	0.09	0	0	0
6	0	0.02	0	0	0.11	0.58	0.20	0.02	0	0.01	0.02	0	0.01	0.04	0	0
7	0	0	0	0	0	0	0.94	0.02	0	0	0	0	0	0	0.04	0
8	0	0	0	0	0	0	0	1.00	0	0	0	0	0	0	0	0
9	0.03	0	0	0	0	0	0	0	0.97	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1.00	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	1.00	0	0	0	0	0
12	0	0	0	0.01	0	0	0	0	0	0	0	0.99	0	0	0	0
13	0	0	0	0	0.01	0	0	0	0	0	0	0	0.99	0	0	0
14	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0.97	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.99	0
16	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0.98
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Предсказанный ракурс

Рис. 13. Матрица ошибок

Fig. 13. Confusion matrix

Список литературы

1. Forward Collision Warning with a Single Camera / E. Dagan, O. Mano, G. P. Stein, A. Shashua // Proc. of the IEEE Intelligent Vehicles Symp., Parma, Italy, 14–17 Jun. 2004. Piscataway: IEEE, 2004. P. 37–42. doi: 10.1109/IVS.2004.1336352

2. Комплекс видеофиксации нарушений ПДД "ИСКРА-ВИДЕО-2" КР. URL: http://www.simicon.ru/rus/product/gun/archive/iv2_k.html (дата обращения 29.08.2020)

3. MMDetection: open MMLab detection Toolbox and Benchmark / K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong,

X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, D. Lin. URL: <https://arxiv.org/pdf/1906.07155.pdf> (дата обращения 29.08.2020)

4. SSD: Single Shot MultiBox Detector / W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg // Europ. Conf. on Computer Vision, ECCV 2016, Amsterdam, The Netherlands, 8–16 Oct. 2016. P. 21–37. doi: 10.1007/978-3-319-46448-0_2.

5. Focal loss for dense object detection / T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar // IEEE Trans. on Pattern Analysis and Machine Intelligence. 2018. Vol. PAMI-42, iss. 2. P. 318–327.

6. Li B., Liu Y., Wang X. Gradient Harmonized Single-stage Detector // Thirty-Third AAAI Conf. on Artificial Intelligence, AAAI-19, Jan. 27 – Feb. 1, 2019. P. 8577–8584.

7. FCOS: Fully Convolutional One-Stage Object Detection / Z. Tian, C. Shen, H. Chen, T. He. 2019. URL: <https://arxiv.org/abs/1904.01355> (дата обращения 29.08.2020)

8. Mask R-CNN / K. He, G. Gkioxari, P. Dollar, R. Girshick // IEEE Intern. Conf. on Computer Vision, ICCV 2017, Venice, Italy, Oct. 22–29, 2017. URL: <https://arxiv.org/pdf/1703.06870.pdf> (дата обращения 29.08.2020)

9. Bochkovskiy A., Wang C.-Y., Mark Liao H.-Y. YOLOv4: Optimal Speed and Accuracy of Object Detection. URL: <https://arxiv.org/pdf/2004.10934.pdf> (дата обращения 21.02.2021)

10. Classification and Pose Estimation of Vehicles in Videos by 3D Modeling within Discrete-Continuous Optimization / M. Hoedlmoser, B. Micusik, M.-Y. Liu, M. Pollefeys, M. Kampel // 2nd Intern. Conf. on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), Zurich, Switzerland, 13–15 Oct. 2012. Piscataway: IEEE, 2012. P. 198–205. doi: 10.1109/3DIMPVT.2012.23

11. Subcategory-aware Convolutional Neural Networks for Object Proposals and Detection / Y. Xiang, W. Choi, Y. Lin, S. Savarese, 2017. URL: <https://arxiv.org/abs/1604.04693>. (дата обращения 02.06.2020)

12. 3D Bounding Box Estimation using Deep Learning and Geometry / A. Mousavian, D. Anguelov, J. Flynn, J. Kosecka. 2017. URL: <https://arxiv.org/abs/1612.00496> (дата обращения 02.06.2020)

13. Monocular 3D Object Detection for Autonomous Driving / X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler, R. Urtasun // IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, USA, June 27–30, 2016. Piscataway: IEEE, 2016. P. 2147–2156.

14. 6-DoF Object Pose from Semantic Keypoints / G. Pavlakos, X. Zhou, A. Chan, K. G. Derpanis, K. Daniilidis // IEEE Intern. Conf. on Robotics and Automation (ICRA), Singapore, 29 May – 3 June, 2017. Piscataway: IEEE, 2017. P. 2011–2018. doi: 10.1109/ICRA.2017.7989233

15. Zubov I. G. Vehicle Pose Estimation Based on Object Contour // IEEE Conf. of Russian Young Researchers

in Electrical and Electronic Engineering, ELCONRUS 2020, St Petersburg and Moscow, 27–30 Jan. 2020. Piscataway: IEEE, 2020. P. 1452–1454. doi: 10.1109/ElConRus49466.2020.9039472

16. The PASCAL Visual Object Classes Challenge (VOC2007). URL: <http://www.pascal-network.org/challenges/VOC/voc2007/index.html> (дата обращения 01.06.2020)

17. ImageNet: A Large-Scale Hierarchical Image Database / J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei // Computer Vision and Pattern Recognition. 2009, CVPR 2009, Miami, USA, Jun. 20–25, 2009. P. 248–255. doi: 10.1109/CVPR.2009.5206848

18. Microsoft COCO: Common Objects in Context / T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, P. Dollar // Europ. Conf. on Computer Vision. ECCV 2014, Zurich, Switzerland, Sept. 6–12, 2014. P. 740–755. doi: 10.1007/978-3-319-10602-1_48

19. Zeiler M. D., Fergus R. Visualizing and Understanding Convolutional Networks // Proc. of the ECCV Europ. Conf. on Computer Vision, Zurich, Switzerland, 6–12 Sept. 2014. Piscataway: IEEE, 2014. P. 818–833. doi: 10.1007/978-3-319-10590-1_53

20. Зубов И. Г. Метод автоматического определения ключевых точек объекта на изображении // Изв. вузов России. Радиоэлектроника. 2020. Т. 23, № 6. С. 6–16. doi: 10.32603/1993-8985-2020-23-6-6-16

21. Deep Residual Learning for Image Recognition / K. He, X. Zhang, S. Ren, J. Sun // IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, USA, Jun. 27–30, 2016. Piscataway: IEEE, 2016. P. 770–778. doi: 10.1109/CVPR.2016.90

22. Зубов И. Г. Метод автоматической сегментации транспортных средств на изображении // Изв. вузов России. Радиоэлектроника. 2019. Т. 22, № 5. С. 6–16. doi: 10.32603/1993-8985-2019-22-5-6-16

23. Щерба Е. В. Анализ применимости методов интерполяции и экстраполяции для решения задачи восстановления изображения // Компьютерная оптика. 2009. Т. 33, № 3. С. 336–339. URL: <http://www.computeroptics.smr.ru/KO/PDF/KO33-3/33313.pdf> (дата обращения 20.08.2020)

24. Гонсалес Р., Вудс Р. Цифровая обработка изображений. 3-е изд. М.: Техносфера, 2012. 834 с.

25. Dechter R., Pearl J. Generalized Best-First Search Strategies and the Optimality of A* // J. of the ACM (JACM). 1985. Vol. 32, № 3. P. 505–536. URL: <https://www.ics.uci.edu/~dechter/publications/r0.pdf> (дата обращения 29.08.2020)

26. Carvana Image Masking Challenge. URL: <https://www.kaggle.com/c/carvana-image-masking-challenge> (дата обращения 29.08.2020)

27. Vehicle Key-Point & Orientation Estimation. URL: https://github.com/Pirazh/Vehicle_Key_Point_Orientation_Estimation (дата обращения 29.08.2020)

28. Flusser J. On the Independence of Rotation Moment Invariants // Pattern Recognition. 2000. Vol. 33, iss. 9. P. 1405–1410. doi: 10.1016/S0031-3203(99)00127-2

29. A Dual-Path Model with Adaptive Attention for Vehicle Re-Identification / P. Khorramshahi, A. Kumar, N. Peri,

S. S. Rambhatla, J.-C. Chen, R. Chellappa. P. 6132–6141. URL: https://openaccess.thecvf.com/content_ICCV_2019/papers/Khorramshahi_A_Dual-Path_Model_With_Adaptive_Attention_for_Vehicle_Re-Identification_ICCV_2019_paper.pdf (дата обращения 29.08.2020)

Информация об авторах

Зубов Илья Геннадьевич – магистр техники и технологий (2016), программист-алгоритмист компании ООО "НЕКСТ". Автор 11 научных работ. Сфера научных интересов – цифровая обработка изображений; машинное обучение; видеоаналитика и прикладные телевизионные системы.

Адрес: ООО "НЕКСТ", Пресненская наб., д. 12, этаж 35, пом. № 3, Москва, 123317, Россия

E-mail: ZubovIG@gmail.com

<https://orcid.org/0000-0003-0407-5651>

Обухова Наталия Александровна – доктор технических наук (2009), профессор (2004), заведующая кафедрой телевидения и видеотехники Санкт-Петербургского государственного электротехнического университета "ЛЭТИ" им. В. И. Ульянова (Ленина). Автор более 130 научных работ. Сфера научных интересов – цифровая обработка изображений; машинное обучение; видеоаналитика и прикладные телевизионные системы.

Адрес: Санкт-Петербургский государственный электротехнический университет "ЛЭТИ" им. В. И. Ульянова (Ленина), ул. Профессора Попова, д. 5, Санкт-Петербург, 197376, Россия

E-mail: natalia172419@yandex.ru

References

1. Dagan E., Mano O., Stein G. P., Shashua A. Forward Collision Warning with a Single Camera. Proc. of the IEEE Intelligent Vehicles Symp., Parma, Italy, 14–17 Jun. 2004. Piscataway, IEEE, 2004, pp. 37–42. doi: 10.1109/IVS.2004.1336352

2. Complex of video recording of traffic violations "ISKRA-VIDEO-2" KR. Available at: http://www.simicon.ru/rus/product/gun/archive/iv2_k.html (accessed 29.08.2020)

3. Chen K., Wang J., Pang J., Cao Y., Xiong Y., Li X., Sun S., Feng W., Liu Z., Xu J., Zhang Z., Cheng D., Zhu C., Cheng T., Zhao Q., Li B., Lu X., Zhu R., Wu Y., Dai J., Wang J., Shi J., Ouyang W., Loy C. C., Lin D. MMDetection: open MMLab detection Toolbox and Benchmark. Available at: <https://arxiv.org/pdf/1906.07155.pdf> (accessed 29.08.2020)

4. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A. C. SSD: Single Shot MultiBox Detector. Europ. Conf. on Computer Vision, ECCV 2016, Amsterdam, The Netherlands, 8–16 Oct. 2016. doi: 10.1007/978-3-319-46448-0_2. P. 21–37

5. Lin T.-Y., Goyal P., Girshick R., He K., Dollar P. Focal loss for dense object detection. IEEE Trans. on Pattern Analysis and Machine Intelligence. 2018, vol. PAMI-42, iss. 2, pp. 318–327.

6. Li B., Liu Y., Wang X. Gradient Harmonized Single-stage Detector. Thirty-Third AAAI Conf. on Artificial Intelligence, AAAI-19, Jan. 27 – Feb. 1, 2019, pp. 8577–8584.

7. Tian Z., Shen C., Chen H., He T. FCOS: Fully Convolutional One-Stage Object Detection. 2019. Available at: <https://arxiv.org/abs/1904.01355> (accessed 29.08.2020)

8. He K., Gkioxari G., Dollar P., Girshick R. Mask R-CNN. IEEE Intern. Conf. on Computer Vision, ICCV 2017, Venice,

Italy, Oct. 22–29, 2017. Available at: <https://arxiv.org/pdf/1703.06870.pdf> (accessed 29.08.2020)

9. Bochkovskiy A., Wang C.-Y., Mark Liao H.-Y. YOLOv4: Optimal Speed and Accuracy of Object Detection. Available at: <https://arxiv.org/pdf/2004.10934.pdf> (accessed 21.02.2021)

10. Hoedlmoser M., Micusik B., Liu M.-Y., Pollefeys M., Kampel M. Classification and Pose Estimation of Vehicles in Videos by 3D Modeling within Discrete-Continuous Optimization. 2nd Intern. Conf. on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), Zurich, Switzerland, 13–15 Oct. 2012. Piscataway, IEEE, 2012, pp. 198–205. doi: 10.1109/3DIMPVT.2012.23

11. Xiang Y., Choi W., Lin Y., Savarese S. Subcategory-aware Convolutional Neural Networks for Object Proposals and Detection, 2017. Available at: <https://arxiv.org/abs/1604.04693>. (accessed 02.06.2020)

12. Mousavian A., Anguelov D., Flynn J., Kosecka J. 3D Bounding Box Estimation using Deep Learning and Geometry. 2017. Available at: <https://arxiv.org/abs/1612.00496> (accessed 02.06.2020)

13. Chen X., Kundu K., Zhang Z., Ma H., Fidler S., Urtasun R. Monocular 3D Object Detection for Autonomous Driving. IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, USA, Jun. 27–30, 2016. Piscataway, IEEE, 2016, pp. 2147–2156.

14. Pavlakos G., Zhou X., Chan A., Derpanis K. G., Daniilidis K. 6-DoF Object Pose from Semantic Keypoints. IEEE Intern. Conf. on Robotics and Automation (ICRA), Singapore, 29 May – 3 June, 2017. Piscataway, IEEE, 2017, pp. 2011–2018. doi: 10.1109/ICRA.2017.7989233

15. Zubov I. G. Vehicle Pose Estimation Based on Object Contour. IEEE Conf. of Russian Young Researchers in Electrical and Electronic Engineering, ELCONRUS 2020, St Petersburg and Moscow, 27–30 Jan. 2020. Piscataway, IEEE, 2020, pp. 1452–1454. doi: 10.1109/ElConRus49466.2020.9039472
16. The PASCAL Visual Object Classes Challenge (VOC2007). Available at: <http://www.pascal-network.org/challenges/VOC/voc2007/index.html> (accessed 01.06.2020)
17. Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L. ImageNet: A Large-Scale Hierarchical Image Database. Computer Vision and Pattern Recognition 2009, CVPR 2009, Miami, USA, Jun. 20–25, 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848
18. Lin T.-Y., Maire M., Belongie S., Bourdev L., Girshick R., Hays J., Perona P., Ramanan D., Zitnick C. L., Dollar P. Microsoft COCO: Common Objects in Context. Europ. Conf. on Computer Vision. ECCV 2014, Zurich, Switzerland, Sept. 6–12, 2014, pp. 740–755. doi: 10.1007/978-3-319-10602-1_48
19. Zeiler M. D., Fergus R. Visualizing and Understanding Convolutional Networks. Proc. of the ECCV Europ. Conf. on Computer Vision, Zurich, Switzerland, 6–12 Sept. 2014. Piscataway, IEEE, 2014, pp. 818–833. doi: 10.1007/978-3-319-10590-1_53
20. Zubov I. G. An Automatic Method for Interest Point Detection. J. of the Russian Universities. Radioelectronics. 2020, vol. 23, no. 6, pp. 6–16. doi: <https://doi.org/10.32603/1993-8985-2020-23-6-6-16> (In Russ.)
21. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, USA, Jun. 27–30, 2016. Piscataway, IEEE, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90
22. Zubov I. G. Method for Automatic Segmentation of Vehicles in Digital Images. J. of the Russian Universities. Radioelectronics. 2019, vol. 22, no. 5, pp. 6–16. doi: 10.32603/1993-8985-2019-22-5-6-16 (In Russ.)
23. Shcherba E. V. Application Analysis of Interpolation and Extrapolation Methods as Used for Image Restoration. Computer Optics. 2009, vol. 33, no. 3, pp. 336–339. Available at: <http://www.computeroptics.smr.ru/KO/PDF/KO33-3/33313.pdf> (accessed 20.08.2020). (In Russ.)
24. Gonzalez R., Woods R. Digital Image Processing. 3^d ed. Prentice Hall, 2012, 834 p. (In Russ.)
25. Dechter R., Pearl J. Generalized Best-First Search Strategies and the Optimality of A*. J. of the ACM (JACM). 1985, vol. 32, no. 3, pp. 505–536. Available at: <https://www.ics.uci.edu/~dechter/publications/r0.pdf> (accessed 29.08.2020)
26. Carvana Image Masking Challenge. Available at: <https://www.kaggle.com/c/carvana-image-masking-challenge> (accessed 29.08.2020)
27. Vehicle Key-Point & Orientation Estimation. Available at: https://github.com/Pirazh/Vehicle_Key_Point_Orientation_Estimation (accessed 29.08.2020)
28. Flusser J. On the Independence of Rotation Moment Invariants. Pattern Recognition. 2000, vol. 33, iss. 9, pp. 1405–1410. doi: 10.1016/S0031-3203(99)00127-2
29. Khorramshahi P., Kumar A., Peri N., Rambhatla S. S., Chen J.-C., Chellappa R. A Dual-Path Model with Adaptive Attention for Vehicle Re-Identification. pp. 6132–6141. Available at: https://openaccess.thecvf.com/content_ICCV_2019/papers/Khorramshahi_A_Dual-Path_Model_With_Adaptive_Attention_for_Vehicle_Re-Identification_ICCV_2019_paper.pdf (accessed 29.08.2020)

Information about the authors

Ilya G. Zubov, Master of Engineering and Technology (2016), Ltd "Next" algorithm programmer. The author of 11 scientific publications. Area of expertise: digital image processing; applied television systems.
Address: Ltd "Next", 12, Presnenskaya Nab., floor 35, room № 3, Moscow 123317, Russia
E-mail: ZubovIG@gmail.com
<https://orcid.org/0000-0003-0407-5651>

Natalia A. Obukhova, Dr. of Sci. (Engineering) (2009), Professor (2004), the Chief of the Department of Television and Video Equipment of Saint Petersburg Electrotechnical University. The author of more than 130 scientific publications. Area of expertise: digital image processing; applied television systems.
Address: Saint Petersburg Electrotechnical University, 5 Professor Popov St., St Petersburg 197376, Russia
E-mail: natalia172419@yandex.ru