

АЛГОРИТМ АНАЛИЗА ЗАШУМЛЕННЫХ ВИДЕОДАНЫХ ДЛЯ НАВИГАЦИИ АВТОНОМНЫХ РОБОТОВ

А.А. Кузнецов

Южно-Уральский государственный университет, г. Челябинск, Российская Федерация

E-mail: kuznetsova@susu.ru

Аннотация. Автономные роботы широко используются для различных целей, включая мониторинг, исследования, чрезвычайные ситуации и т. д. Главными преимуществами автономных роботов являются мобильность, простота развертывания и эффективность в реальных сценариях. Актуальной является задача навигации автономных роботов на основе данных, полученных из видеоканала. Целью исследования является реализация алгоритма анализа зашумленных видеоданных для детекции ключевых точек навигационных объектов. Традиционные подходы детекции ключевых точек навигационных объектов используют итерационные алгоритмы, без применения нейронных сетей. Предлагается использовать нейросетевую модель YOLOv11 для решения задачи. В качестве входных данных нейросети будут выступать RGB-изображения. Результатом работы будут координаты ключевых точек навигационных объектов. Приведены данные тестов алгоритма на данных датасета и данных, полученных с испытаний демонстратора двигательной установки с центральным телом. Используется метрика OKS (Object Keypoint Similarity). Результатом исследования является нейронная сеть, разработанная и обученная для детекции навигационных объектов и их ключевых точек. Определяя ключевые точки объектов, можно корректировать инерциальные навигационные системы. Полученная нейронная сеть устойчива к различным помехам и шумам, в том числе перекрытиям.

Ключевые слова: навигация; нейронные сети; YOLOv11; автономные роботы.

Введение

Навигация автономного робота может осуществляться через анализ видеоканала с использованием технологии компьютерного зрения. В [1] приводится обзор современных методов компьютерного зрения для автономных транспортных средств. Для обеспечения навигации с использованием компьютерного зрения основной задачей является обнаружение на изображении объектов [1], на основании положения и типа которых можно строить маршруты следования автономного робота. Используются такие методы, как классический конвейер обнаружения, так и использование глубокого обучения.

Тем не менее видеоканал может содержать различные шумы. Это могут быть цифровой шум, сниженный контраст, размытие, перекрытие объектов другими объектами и т. д.

Цифровой шум [2] – это нежелательные искажения или помехи в цифровом изображении или видео, которые возникают из-за различных факторов, таких как ограничения оборудования, условия съемки или обработка сигнала. Визуальными проявлениями такого шума является зернистость, цветовые артефакты, потеря деталей. Причинами могут быть высокое значение ISO, слабое освещение, ограничения сенсора, сжатие данных и т. д. Ухудшает качество изображения путем снижения резкости, ухудшения цветопередачи и потери деталей.

Сниженный контраст и размытие [3, 4] – это два распространенных дефекта изображения, которые могут возникать по разным причинам. Они ухудшают качество изображения, делая его менее четким и выразительным. Сниженный контраст характеризуется низкой разницей между светлыми и темными областями, потерей деталей в тенях и цветах. Причинами могут служить плохое освещение, атмосферные условия, ограничения динамического диапазона. Размытие характеризуется потерей резкости, смазыванием границ объектов. Причинами могут служить движение камеры или объекта, неправильная фокусировка, дифракция и т. д.

При перекрытии отслеживаемых объектов другими объектами может теряться информация, на основании которой автономному роботу нужно строить маршрут. Так, перекрытие навигационных объектов может вызвать ошибки в расчете положения автономного робота, что повлечет за собой некорректно рассчитанный маршрут. Навигационная система должна быть устойчива ко всем типам шумов в видеопотоке, чтобы объекты, по которым строится маршрут, всегда определялись. На сегодняшний день существуют различные навигационные системы для автономных роботов. В [5] описывается ORB-SLAM2. Это полная система SLAM (одновременная локализация и построение карты) для монокулярных, стерео- и RGB-D-камер. Система работает в режиме реального времени на стандартных процессорах в самых разных условиях: от небольших ручных последовательностей в помещениях до дронов, летающих в промышленных условиях, и автомобилей, едущих по городу. Из недостатков можно отметить высокие требования к вычислительным ресурсам бортового вычислителя, зависимость от зашумления и отсутствие семантической информации. ORB-SLAM2 работает только на уровне геометрических особенностей, не распознавая объекты. В [6] описывается фильтр Калмана с ограничением по нескольким состояниям. Фильтр Калмана с ограничением по нескольким состояниям – это расширенный вариант фильтра Калмана, разработанный специально для визуально-инерциальной навигации. Он эффективно комбинирует данные от камеры и инерциального измерительного блока, позволяя роботам и дронам оценивать свое положение без накопления ошибок (дрейфа). Поскольку фильтр Калмана с ограничением по нескольким состояниям не строит глобальную карту, со временем накапливается дрейф (особенно в монотонном окружении). Отсутствие распознавания объектов ограничивает использование метода.

Алгоритм анализа зашумленного видеопотока на основе нейронных сетей

Предлагаемый алгоритм анализа зашумленного видеопотока использует для обнаружения навигационных объектов и их ключевых точек нейронную сеть YOLOv11 [7]. Эта нейронная сеть использует однопроходный подход, использующий все изображения в качестве входных данных, для прогнозирования ограничивающих рамок и вероятностей классов. YOLOv11 оптимизирована для баланса между точностью (mAP) и скоростью (FPS). Данную нейронную сеть предполагается использовать на автономных роботах, следовательно, она должна показывать хорошую производительность на одноплатных компьютерах, например, на NVIDIA Jetson Orin NX или Orange Pi 5. В качестве навигационных объектов будут использоваться маркеры ArUco [8]. Они представляют собой черно-белые квадратные изображения с уникальным идентификатором и внутренним кодом, который легко распознается алгоритмами. На данный момент нейронная сеть обучалась обнаруживать три маркера ArUco (рис. 1).



Рис. 1. Маркеры ArUco

Получив ключевые точки маркера, можно решать задачу Perspective-n-Point (PnP) [9]. PnP – это классическая задача компьютерного зрения, которая заключается в определении положения и ориентации камеры (ее позы, т. е. 3D-положения и вращения) относительно сцены по n известным 3D-точкам и их 2D-проекциям на изображении. В данной работе задача (PnP) будет решаться методом Direct Least-Squares (DLS) [10].

Полученная информация о положении камеры в пространстве может использоваться для коррекции инерциальных датчиков, которые при двойном интегрировании могут накапливать значительные погрешности [11]. Так как в данной работе в качестве таких объектов используются маркеры ArUco, следовательно, навигационным объектом будет до трех заранее заданных

маркеров и углы маркера в качестве ключевых точек. На рис. 2 приведен пример такого маркера с выделенными ключевыми точками: ключевые точки выделены разными цветами, так как нейронная сеть должна также определять, какая конкретно ключевая точка находится в данных координатах.

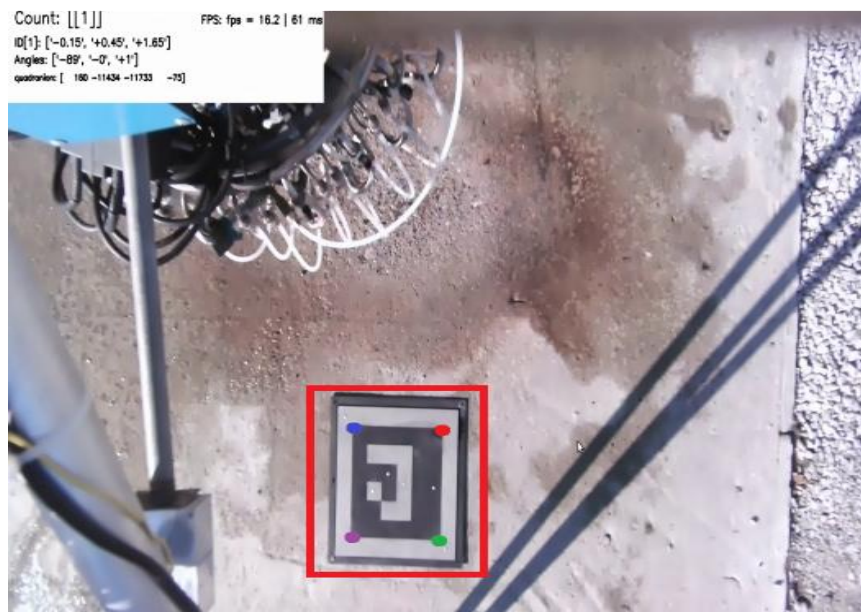


Рис. 2. Маркер на испытательной площадке

Создание обучающей выборки проходило в два шага. На первом шаге была реализована программа [12] на языке C++ для генерации синтетических данных. С помощью реализованной программы было сгенерировано 4 000 изображений. Пример такого синтетического изображения представлен на рис. 3. На втором шаге собирались реальные данные. Реальные данные получались путем записи видео в лаборатории. Полученное видео затем разбивалось на кадры (рис. 4). Полученные кадры затем размечались в формате, понятном для нейронной сети.



Рис. 3. Пример сгенерированного синтетического изображения



Рис. 4. Пример изображения, полученного в лаборатории

Численные эксперименты

Для примера работы нейронной сети можно использовать изображение без шумов и с большим маркером. Как изображено на рис. 5, стандартные алгоритмы и нейронная сеть справились с задачей. Тем не менее если на маркер начинает падать тень, как на рис. 6, стандартные алгоритмы не могут обнаружить маркер.

Для тестирования работы с зашумленными изображениями на незашумленное изображение будут накладываться различные шумы. На рис. 7 показано исходное изображение и его зашумленная версия. Для зашумления изображения применялся аддитивный гауссовский шум. К исходным пикселям изображения добавлялись независимые значения, каждое из которых является случайной величиной, имеющей нормальное распределение с нулевым средним и стандартным

отклонением 800. Этот шум – «белый» (некоррелированный во времени/пространстве) и одинаково воздействует на все изображения. Для тестов будет использоваться фрагмент видео размером в 240 кадров, полученного с испытаний.

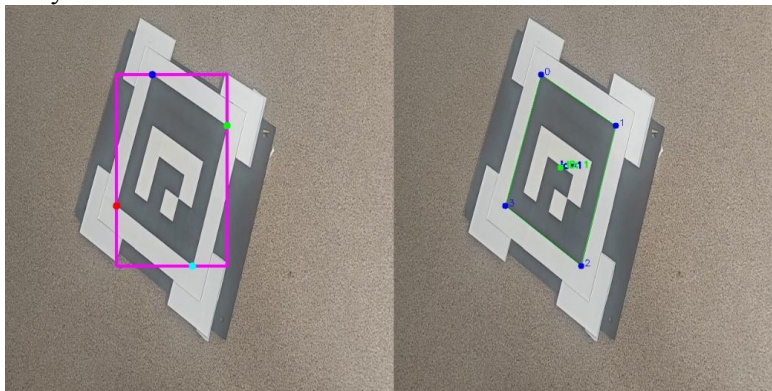


Рис. 5. Пример работы нейронной сети (слева) и алгоритмов, реализованных в OpenCV (справа)

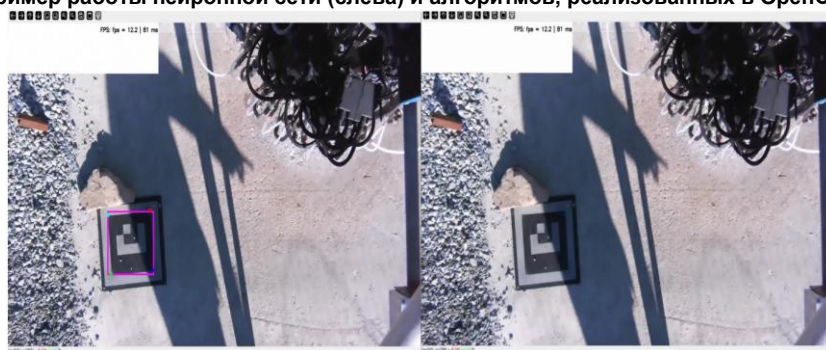


Рис. 6. Пример работы нейронной сети (слева) и алгоритмов, реализованных в OpenCV (справа), на изображении с зашумлением

В качестве метрики для ключевой точки i будет использоваться *Object Keypoint Similarity* (*OKS*)

$$OKS_i = \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right), \quad (1)$$

где d_i – евклидово расстояние между предсказанной и настоящей ключевой точкой, s – площадь объекта, $k_i = 0,5$ для всех i . Метрика точности для всего объекта будет средним значением точности всех ключевых точек. *OKS* – это стандартная метрика для оценки качества моделей компьютерного зрения, работающих с ключевыми точками объектов (например, обнаружение позы человека, трекинг животных, анализ жестов). Она широко используется в таких наборах данных, как COCO (*Common Objects in Context*) и MPII. *OKS* измеряет сходство между предсказанными ключевыми точками и эталонными точками, учитывая масштаб объекта и сложность распознавания каждой точки. *OKS* обеспечивает сбалансированную оценку, учитывающую нюансы локализации ключевых точек, что делает её золотым стандартом в задачах обнаружения ключевых точек объекта.

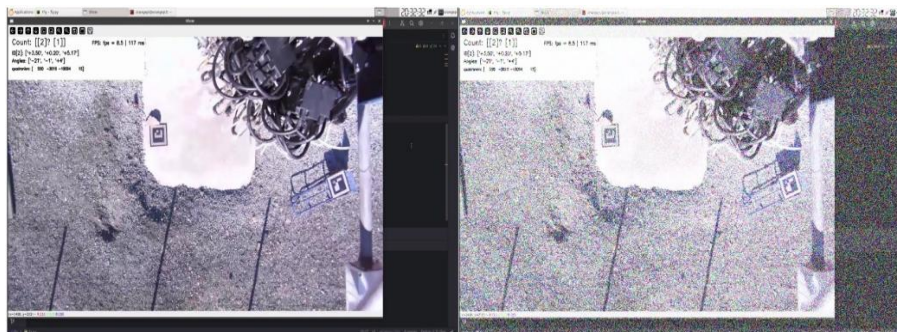


Рис. 7. Исходное изображение (слева) и зашумленное изображение (справа)

На рис. 8 (слева) показана точность определения ключевых точек при уровне шума 150. Показано, что классические методы справляются с задачей точнее в большинстве случаев. На рисунке 8 (справа) показана точность определения ключевых точек при уровне шума 800. Показано, что точность нейронной сети снизилась не более чем на 7 %. Классические методы не обнаружили маркер.

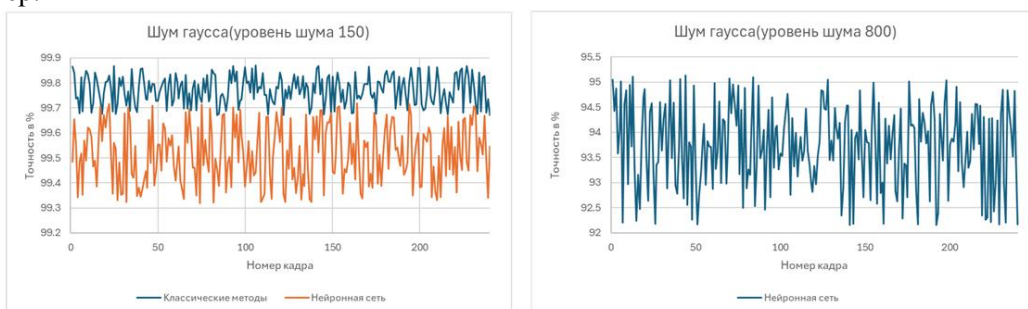


Рис. 8. Точность работы стандартных алгоритмов и нейронной сети в зависимости от уровня шума

В табл. 1 указано среднее значение точности обнаружения ключевых точек на видео для классических методов и нейронной сети.

Таблица 1

Средние значения точности для шума Гаусса для классических методов и нейронной сети

Классические методы (уровень шума 150)	Нейронная сеть (уровень шума 150)	Нейронная сеть (уровень шума 800)
99,77129962	99,51717543	93,73195243

На рис. 9 представлен пример незашумленного изображения и этого же изображения с наложенным медианным размытием с размером ядра 14. Для тестирования работы с медианным размытием на незашумленное изображение будет накладываться медианное размытие с различными ядрами (размерами от 1 до 14). Для тестов будет использоваться фрагмент видео размером в 240 кадров, полученный с испытаний. В качестве метрики будет использоваться OKS (1).

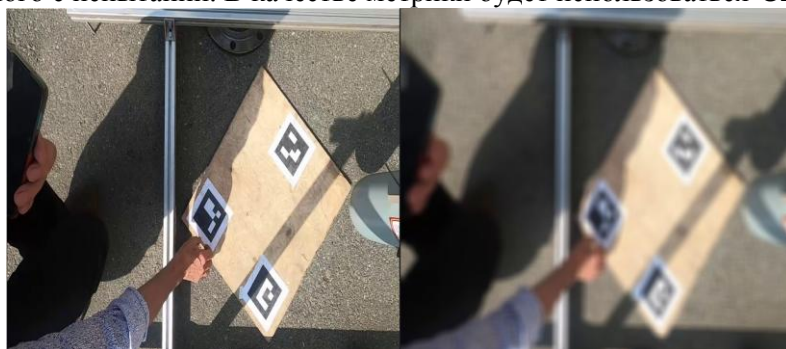


Рис. 9. Пример изображения без шума (слева) и с шумом в виде размытия (справа)

На рис. 10 (слева) показана точность работы определения ключевых точек классическими методами и нейронной сетью при размере ядра 3. Показано, что классические методы справляются с задачей точнее нейронной сети при малом размере ядра (в данном случае 3). На рис. 10 (справа) показано, что при высоких значениях ядра (в данном случае 14) нейронная сеть показывает снижение точности в среднем на 5 %. Классические методы не обнаруживают маркер.

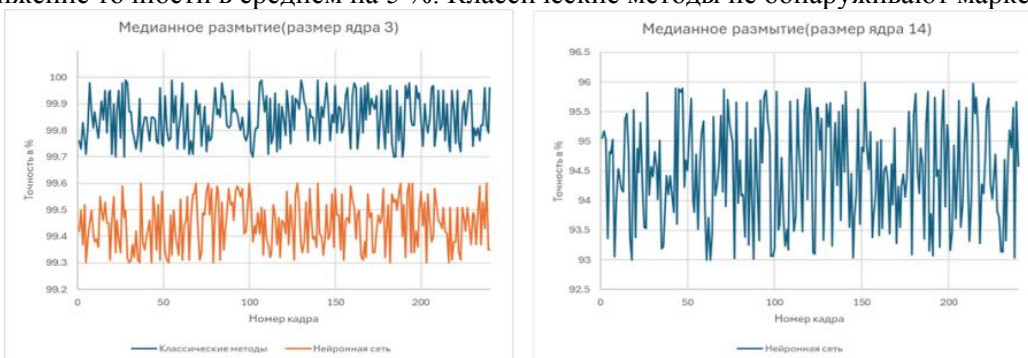


Рис. 10. Точность работы стандартных алгоритмов и нейросети в зависимости от уровня шума

Средние значения точности для медианного размытия для классических методов и нейронной сети		
Классические методы (размер ядра 3)	Нейронная сеть (размер ядра 3)	Нейронная сеть (размер ядра 14)
99,84641667	99,44591667	94,51533333

В табл. 2 указано среднее значение точности обнаружения ключевых точек на видео для классических методов и нейронной сети.

Навигационные объекты могут перекрываться другими объектами. Навигационная система даже в таких случаях должна уметь определять ключевые точки навигационного объекта. Так как в данной работе используются маркеры ArUco в качестве навигационных объектов, то в случае их перекрытия или перекрытия их ключевых точек (углов маркера) необходимо все равно определить их углы. На рис. 11 и 12 показаны изображения с перекрытым маркером. Нейронная сеть определяет углы маркера с точностью 90 % в среднем.

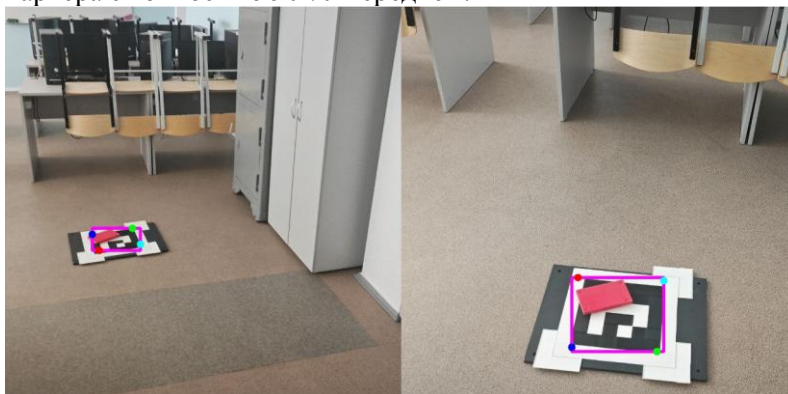


Рис. 11. Пример работы нейронной сети на изображениях с перекрытым маркером

На рис. 13 показано изображение с перекрытым маркером в ключевой точке. В данном случае, хоть нейронная сеть и определила корректно маркер и его ключевые точки, точность составила 80 % в среднем.



Рис. 12. Пример работы нейронной сети на изображениях с перекрытым маркером



Рис. 13. Пример работы нейронной сети на изображениях с перекрытым маркером

Нейронная сеть может работать и с другими навигационными объектами. Например, рассмотрим фидуциарные маркеры reacTIVision [13]. Так же, как и маркеры ArUco, данные маркеры можно использовать для определения положения автономного робота. Все что нужно сделать, это переобучить нейронную сеть для работы с этими маркерами. При обучении также можно использовать исходный датасет. На рис. 15 показан пример, где на место исходного маркера подставлен новый. Так как координаты исходных маркеров известны, то данный процесс можно автоматизировать.

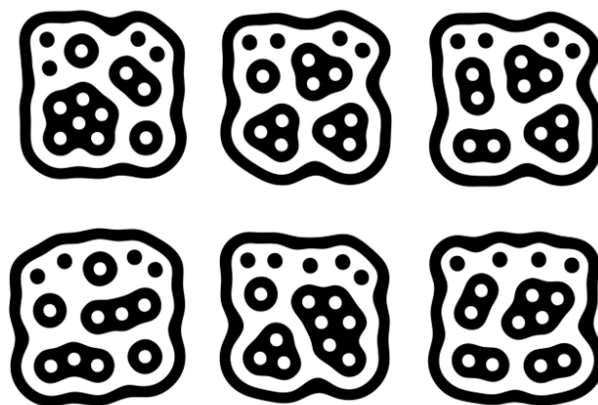


Рис. 14. Пример маркеров reacTIVision

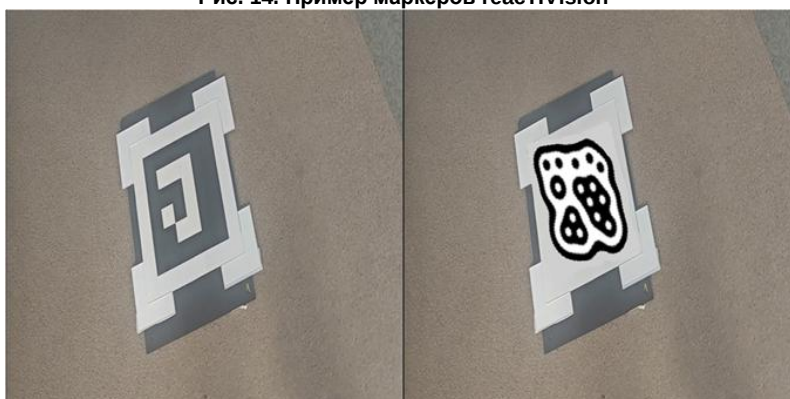


Рис. 15. Исходное изображение (слева) и изображение с наложенным новым маркером

Заключение

В рамках исследования было разработано и испытано программное обеспечение для обработки данных с бортовых камер на основе сверточных нейронных сетей, оптимизированное для выполнения на бортовых вычислителях. Разработанное программное обеспечение использует современную архитектуру нейронных сетей, что позволяет обеспечивать приемлемый уровень точности обнаружения и отслеживания объектов интереса на ограниченных ресурсах бортовых вычислителей, установленных на БЛА. Реализованные алгоритмы устойчивы к помехам на изображении, таким как цифровой шум, низкая контрастность изображения, перекрытие объектов интереса другими объектами. Детекция навигационных объектов позволяет выделять ключевые точки для коррекции систем навигации БЛА в отсутствие информации о местоположении с внешних источников, таких как спутниковая связь. Было показано, что при наложении шума на изображение, точность обнаружения объектов интереса и их ключевых точек падала не более чем на 7 %. При перекрытии объектов интереса точность составляла 80 %.

Литература

1. Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art / J. Janai, F. Güney, A. Behl, A. Geiger // Foundations and Trends in Computer Graphics and Vision. – 2020. – Vol. 12, no. 1–3. – P. 1–308.
2. Boyat, A.K. Review Paper: Noise Models in Digital Image Processing / A.K. Boyat, B.K. Joshi // Signal & Image Processing: An International Journal (SIPIJ). – 2015. – Vol. 6, no. 2. – P. 63–75.

3. Sasi, N. Contrast Limited Adaptive Histogram Equalization for Qualitative Enhancement of Myocardial Perfusion Images / N. Sasi, V. Jayasree // *Engineering*. – 2013. – Vol. 5, no.10 B – P. 326–331.
4. Deep Image Deblurring: A Survey / K. Zhang, W. Ren, W. Luo *et al.* // *arXiv* P. 1–2 <https://arxiv.org/abs/2201.10700>
5. Mur-Artal, R. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras / R. Mur-Artal, J.D. Tardos // *IEEE Transactions on Robotics*. – 2017. – Vol. 33, no. 5. – P. 1255–1262.
6. MSCEqF: A Multi State Constraint Equivariant Filter for Vision-aided Inertial Navigation / A. Fornasier, P. van Goor, E. Allak *et al.* // *IEEE Robotics and Automation Letters*. – 2024. – Vol. 9, no. 1. – P. 731–738
7. Jocher G. Ultralytics YOLO11 [Электронный ресурс] / G. Jocher, J. Qiu. – Режим доступа: <https://github.com/ultralytics/ultralytics> [Дата обращения: 25 июня 2025]
8. Automatic generation and detection of highly reliable fiducial markers under occlusion / S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, M.J. Marín-Jiménez // *Pattern Recognition*. – 2014. – Vol. 47, no. 6. – P. 2280–2292.
9. Fischler, M.A. Random Sample Consensus: a Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography / M.A. Fischler, R.C. Bolles // *Communications of the ACM*. – 1981. – Vol. 24, no. 6. – P. 381–395.
10. Hesch J. A Direct Least-Squares (dls) Method for PnP / J. Hesch, S. I. Roumeliotis // *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. – 2011. – P. 383–390.
11. Vision-Aided Inertial Navigation for Spacecraft Entry, Descent, and Landing / A.I. Mourikis, N. Trawny, S.I. Roumeliotis *et al.* // *IEEE Transactions on Robotics*. – 2009. – Vol. 25, Iss. 2. – P. 264–280.
12. Свидетельство 2022667746 Российская Федерация. Свидетельство о государственной регистрации программы для ЭВМ. Программа генерации обучающей выборки для нейронной сети, выполняющей функции локальной навигации летательного аппарата по маркерам на основе видеопотока с камеры / А.А. Кузнецов; заявитель и правообладатель А.А. Кузнецов. – № 2022667746; заявл. 26.08.2022; опубл. 26.08.2022 // *Реестр программ для ЭВМ*. – 2022. – 1 с.
13. Kaltenbrunner M., Bencina R. reacTIVision: A Computer-Vision Framework for Table-Based Tangible Interaction / M. Kaltenbrunner, R. Bencina // *Proc. 1st International Conference on Tangible and Embedded Interaction*. – 2007. – P. 69–74.

Поступила в редакцию 7 июля 2025 г.

Сведения об авторе

Кузнецов Антон Андреевич – аспирант, кафедра прикладной математики и программирования, Южно-Уральский государственный университет, г. Челябинск, Российская Федерация, email: kuznetsovaa@susu.ru.

*Bulletin of the South Ural State University
Series "Mathematics. Mechanics. Physics"
2025, vol. 17, no. 3, pp. 38–46*

DOI: 10.14529/mmph250304

ALGORITHM FOR ANALYZING NOISY VIDEO DATA FOR AUTONOMOUS ROBOTS NAVIGATION

A.A. Kuznetsov

*South Ural State University, Chelyabinsk, Russian Federation
E-mail: kuznetsovaa@susu.ru*

Abstract. Autonomous robots are widely used for various purposes, including monitoring, research, defense, emergency situations, etc. The main advantages of autonomous robots are mobility, ease of deployment and real efficiency. The problem of autonomous robot navigation based on data obtained from a video channel is relevant. The study aims to implement an algorithm for analyzing noisy video data to

detect key points of navigation objects. Traditional approaches to detecting key points of navigation objects apply iterative algorithms without the use of neural networks. YOLOv11 neural network model and RGB images as input data can solve the problem. The work results in finding the coordinates for the key points of navigation objects. The algorithm test data on the dataset and data obtained from testing the demonstrator of an engine with a central body are provided. The OKS (Object Keypoint Similarity) metric is used. The study results in developing a neural network, trained to detect navigation objects and their key points. Defining the key points of objects enables to correct inertial navigation systems. The obtained neural network is resistant to various noises, including overlaps.

Keywords: navigation; neural networks; YOLOv11; autonomous robots.

References

1. Janai J., Güney F., Behl A., Geiger A. Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art. *Foundations and Trends in Computer Graphics and Vision*, 2020, Vol. 12, no. 1–3, pp. 1–308. DOI: 10.1561/06000000079
2. Boyat A.K., Joshi B.K. Review Paper: Noise Models in Digital Image Processing. *Signal & Image Processing: An International Journal (SIPIJ)*, 2015, Vol. 6, no. 2, pp. 63–75.
3. Sasi N., Jayasree V. Contrast Limited Adaptive Histogram Equalization for Qualitative Enhancement of Myocardial Perfusion Images. *Engineering*, 2013, Vol. 5, no.10 B, pp. 326–331. DOI: 10.4236/eng.2013.510B066
4. Zhang K., Ren W., Luo W., Lai W.-S., Stenger B., Yang M.-H., Li H. Deep Image Deblurring: A Survey. *arXiv*, pp. 1–2. <https://arxiv.org/abs/2201.10700>
5. Mur-Artal R., Tardós J.D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Transactions on Robotics*, 2017, Vol. 33, no. 5, pp. 1255–1262, DOI: 10.1109/TRO.2017.2705103
6. Fornasier A., van Goor P., Allak E., Mahony R., Weiss S. MSCEqF: A Multi State Constraint Equivariant Filter for Vision-Aided Inertial Navigation. *IEEE Robotics and Automation Letters*, 2024, Vol. 9, no. 1, pp. 731–738. DOI: 10.1109/LRA.2023.3335775
7. Jocher G., Qiu J. Ultralytics YOLO11. <https://github.com/ultralytics/ultralytics>
8. Garrido-Jurado S., Muñoz-Salinas R., Madrid-Cuevas F.J., Marín-Jiménez M.J. Automatic Generation and Detection of Highly Reliable Fiducial Markers under Occlusion. *Pattern Recognition*, 2014, Vol. 47, no. 6, pp. 2280–2292. DOI: 10.1016/j.patcog.2014.01.005
9. Fischler M.A., Bolles R.C. Random Sample Consensus: a Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 1981, Vol. 24, Iss. 6, pp. 381–395. DOI: 10.1145/358669.358692
10. Hesch J., Roumeliotis S.I. A Direct Least-Squares (DLS) Method for PnP. *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 383–390.
11. Mourikis A.I., Trawny N., Roumeliotis I.S., Johnson E.A., Ansar A., Matthies L. Vision-Aided Inertial Navigation for Spacecraft Entry, Descent, and Landing. *IEEE Transactions on Robotics*, 2009, Vol. 25, no. 2, pp. 264–280.
12. Certificate 2022667746 Russian Federation. Certificate of State Registration of a Computer Program. *Program for Generating a Training Sample for a Neural Network Performing the Functions of Local Navigation of an Aircraft using Markers based on a Video Stream From a Camera*. Applicant and Copyright Holder A.A. Kuznetsov. – No. 2022667746; Declared 26.08.2022; Published 26.08.2022 // Register of Computer Programs. – 2022. – 1 p.
13. Kaltenbrunner M., Bencina R. reacTIVision: a Computer-Vision Framework for Table-Based Tangible Interaction. *Proc. 1st International Conference on Tangible and Embedded Interaction*, 2007, pp. 69–74. DOI: 10.1145/1226969.1226983

Received July 7, 2025

Information about the author

Kuznetsov Anton Andreevich is Post-graduate Student, Department of Applied Mathematics and Programming, South Ural State University, Chelyabinsk, Russian Federation, e-mail: kuznetsovaa@susu.ru.