

УДК 004.89

DOI: [10.26102/2310-6018/2024.45.2.045](https://doi.org/10.26102/2310-6018/2024.45.2.045)

Интеллектуальная система управления движением квадрокоптера с помощью жестов рук

Н.Н. Чернышев✉, М.А. Шевченко, Т.В. Ниженец

МИРЭА – Российский технологический университет, Москва, Российская Федерация

Резюме. Актуальность исследования обусловлена тем, что управление квадрокоптером с помощью жестов рук более естественно и интуитивно, чем использование традиционных пультов управления. Это позволяет пользователям с легкостью освоить управление и сосредоточиться на выполнении задачи, а не на технических аспектах управления. В свою очередь, разработка системы распознавания жестов требует совершенствования алгоритмов обработки изображений на основе машинного обучения. Данная статья направлена на исследование возможности реализации управления движением квадрокоптера с использованием жестов рук, в сочетании с современными нейросетевыми технологиями. Основным подходом в исследовании данной проблемы является применение сверточных искусственных нейронных сетей для обработки изображений и выполнения задач компьютерного зрения. В работе также рассматриваются методы оптимизации гиперпараметров с помощью инструмента Optuna, использования TensorFlow Lite для реализации моделей машинного обучения на устройствах с ограниченными ресурсами и применение библиотеки MediaPipe для анализа жестов. Такие технологии, как Dropout и L2-регуляризация, используются для повышения эффективности моделей. Материалы статьи представляют практическую ценность для исследователей в области искусственного интеллекта и робототехники, разработчиков программного обеспечения и компаний, занимающихся разработкой беспилотных летательных аппаратов.

Ключевые слова: квадрокоптер, жесты рук, компьютерное зрение, сверточные нейронные сети, искусственные нейронные сети, оптимизация гиперпараметров, управление.

Для цитирования: Чернышев Н.Н., Шевченко М.А., Ниженец Т.В. Интеллектуальная система управления движением квадрокоптера с помощью жестов рук. *Моделирование, оптимизация и информационные технологии*. 2024;12(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1603> DOI: 10.26102/2310-6018/2024.45.2.045

Intelligent hand gesture-based quadcopter motion control system

N.N. Chernyshev✉, M.A. Shevchenko, T.V. Nizhenec

MIREA – Russian Technological University, Moscow, the Russian Federation

Abstract. The relevance of this research stems from the fact that controlling a drone using hand gestures is more natural and intuitive than using traditional joysticks. This allows users to easily learn control and focus on task execution rather than technical aspects of operation. In turn, developing a gesture recognition system requires advancements in machine learning-based image processing algorithms. This paper aims to investigate the feasibility of implementing drone motion control using hand gestures in conjunction with modern neural network technologies. The main approach in addressing this problem involves the application of convolutional artificial neural networks for image processing and computer vision tasks. The work also explores methods for hyperparameter optimization using the Optuna tool, the use of TensorFlow Lite for implementing machine learning models on resource-constrained devices, and the application of the MediaPipe library for gesture analysis. Technologies such as Dropout and L2-regularization are used to enhance model efficiency. The materials presented in this paper hold practical value for researchers in the fields of artificial intelligence and robotics, software developers, and companies involved in the development of unmanned aerial vehicles.

Keywords: quadcopter, hand gestures, computer vision, convolutional neural networks, artificial neural networks, hyperparameter optimization, control.

For citation: Chernyshev N.N., Shevchenko M.A., Nizhenec T.V. Intelligent hand gesture-based quadcopter motion control system. *Modeling, Optimization and Information Technology*. 2024;12(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1603> DOI: 10.26102/2310-6018/2024.45.2.045 (In Russ.).

Введение

В последние годы наблюдается стремительный рост интереса к использованию дронов в различных областях, таких как доставка товаров, мониторинг окружающей среды, военный сектор и др. Одной из важных сфер применения квадрокоптеров становится их интеграция с технологиями машинного обучения и нейронных сетей, что позволяет существенно расширить функциональные возможности этих устройств [1–3]. Например, квадрокоптер DJI Tello EDU – один из таких дронов, обладающий широкими возможностями для использования в образовательных и исследовательских целях благодаря своей доступности и гибкости в настройке.

Многочисленные исследования подтверждают эффективность применения сверточных нейронных сетей и искусственных нейронных сетей в задачах компьютерного зрения и обработки изображений [4–6]. Разработка и использование библиотек и фреймворков, таких как MediaPipe, позволяют значительно повысить производительность и точность моделей, работающих на устройствах с ограниченными ресурсами. Методы оптимизации гиперпараметров, такие как Optuna, также находят широкое применение в сфере автоматизации машинного обучения и показали свою эффективность в улучшении результата.

Несмотря на успехи в данной области, остаются нерешенными вопросы, связанные с оптимизацией скорости и точности обработки данных на устройствах с ограниченными вычислительными возможностями. Кроме того, существуют сложности интеграции различных инструментов для создания комплексных решений, способных адаптироваться под различные задачи и условия использования [7–9].

Целью данного исследования является разработка интеллектуальной системы управления движением квадрокоптера с помощью жестов рук с использованием технологий машинного обучения и нейронных сетей. Использование этой системы позволит оператору освободить руки и выполнять другие задачи во время полета, например, фотографировать, снимать видео или управлять другими устройствами. Для достижения этой цели поставлены следующие задачи:

1. Изучить возможности и ограничения применения сверточных и искусственных нейронных сетей при работе с дронами.
2. Провести анализ существующих инструментов и библиотек, таких как TensorFlow Lite и MediaPipe, для их дальнейшей интеграции в разрабатываемую интеллектуальную систему.
3. Исследовать методы оптимизации гиперпараметров с использованием Optuna.
4. Разработать и провести тестирование системы на базе квадрокоптера DJI Tello EDU с целью оценки эффективности и практической значимости разработанной интеллектуальной системы управления движением квадрокоптера с помощью жестов рук.

Таким образом, исследование направлено на создание интеллектуальной системы, объединяющей последние достижения в области машинного обучения и компьютерного зрения для обеспечения эффективной работы дронов в различных сценариях управления.

Материалы и методы

В данном исследовании ключевым аспектом является интеграция дрона DJI Tello EDU с современными методами машинного обучения и технологиями искусственного интеллекта.

Для достижения поставленных целей использовалась заранее обученная глубокая сверточная нейронная сеть от компании Google под названием MediaPipe Hands. Данное решение предоставляет возможность анализировать форму и положение руки в реальном времени вне зависимости от внешних условий, таких как геометрия руки, расстояние от устройства захвата видеопотока и условий освещенности помещения при помощи построения «скелета руки» [10]. Суть данного метода основана на алгоритме построения непрерывного скелета многоугольной фигуры, краткая схема которого проиллюстрирована на Рисунке 1.

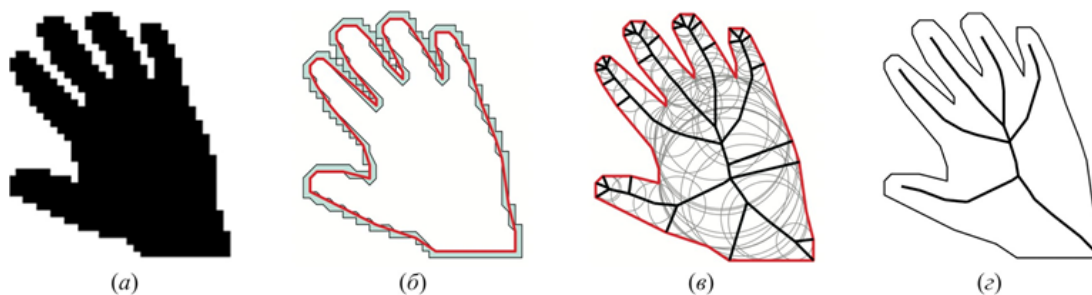


Рисунок 1 – Процесс построения скелета руки: исходное бинарное изображение (а); многоугольная аппроксимация границы объекта (б); скелет многоугольной фигуры (в); скелет после процедуры удаления шумовых, незначительных ветвей (г)
Figure 1 – Hand skeleton construction process: initial binary image (a); polygonal object boundaries approximation (b); skeleton of polygonal figure (c); skeleton after procedure of noisy, insignificant branches removal (d)

Проведен предварительный сбор данных для обучения последовательного нейросетевого классификатора, позволяющего классифицировать до 15 различных жестов руки.

Для этого, используя OpenCV совместно с MediaPipe, были собраны вектора признаков, представляющие собой нормализованные координаты 21 точки скелета руки по осям x и y , и метки жестов, представляющие собой целочисленный идентификатор жеста.

В дальнейшем собранные данные были использованы с целью синтеза глубокой нейронной сети, архитектура которой представляет собой биологически мотивированную модель, аналогичную механизмам обработки информации в мозге млекопитающих. Данная модель является последовательной и состоит из нескольких слоев искусственных нейронов, позволяющих производить нелинейные преобразования данных. Искусственные нейроны организованы в иерархическую структуру, подобную той, что наблюдается в коре головного мозга. Рассмотрим данную структуру более детально:

1. Входной слой размерностью 21×2 выполняет функцию сенсорного слоя, принимая необработанные данные в виде координат ключевых точек скелета руки, полученных при помощи обработки видеопотока (Рисунок 2).

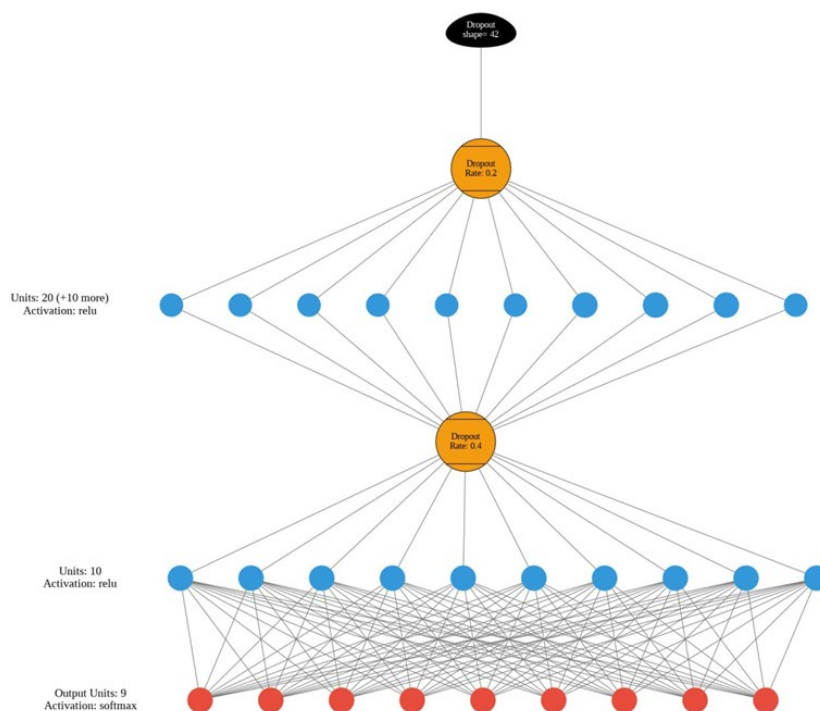


Рисунок 2 – Визуализация архитектуры нейронной сети
Figure 2 – Neural network architecture visualization

2. Далее следует слой отсева (Dropout), имитирующий процессы стохастической регуляризации в биологических нейронных сетях. Данный слой временно отключает случайные нейроны во время обучения, предотвращая излишнюю специализацию модели и способствуя более эффективному обобщению.

3. Полносвязный скрытый слой (fully connected hidden layer) в нейронной сети с функцией активации ReLU (Rectified Linear Unit) представляет собой слой, в котором каждый нейрон соединен со всеми нейронами предыдущего и следующего слоев, и активация каждого нейрона вычисляется с помощью функции ReLU. Если взвешенная сумма входов (выход предыдущего слоя) больше нуля, то активация нейрона будет равна этой взвешенной сумме, иначе активация будет равна нулю. Использование этого слоя помогает в извлечении нелинейных признаков из входных данных и улучшает способность нейронной сети к аппроксимации сложных функций.

4. Следующий слой отсева также имитирует стохастические процессы регуляризации, но с более высокой вероятностью отключения нейронов, что способствует дальнейшему обобщению модели и предотвращению переобучения.

5. Еще один полносвязный скрытый слой предоставляет возможность более высокоуровневой обработки информации, где происходит дальнейшее извлечение признаков и преобразование данных. Данный слой можно рассматривать в качестве аналогии с областями ассоциативной коры более высокого уровня, где происходит интеграция и абстрагирование информации.

6. Наконец, выходной слой содержит число нейронов, равное количеству классов жестов руки, которые должны быть распознаны и классифицированы. Выходная функция активации softmax обеспечивает нормализованную вероятностную оценку принадлежности входных данных к каждому из классов. Этот слой моделирует процессы принятия решений и формирования выходов, аналогичные тем, которые происходят в областях коры головного мозга, ответственных за моторный контроль и принятие решений.

Обучение первоначальной модели было остановлено на 249 эпохе по критерию раннего прекращения (Early stopping), который останавливает обучение, если на протяжении 20 эпох не было замечено улучшения показателей. На последней эпохе точность на обучающем наборе составила 96,61 %, а на валидационном – 76,53 %. Такая разница между обучающей и валидационной точностью свидетельствует о некоторой степени переобучения, из-за которой модель хуже адаптируется под новые входные данные (Рисунок 3).

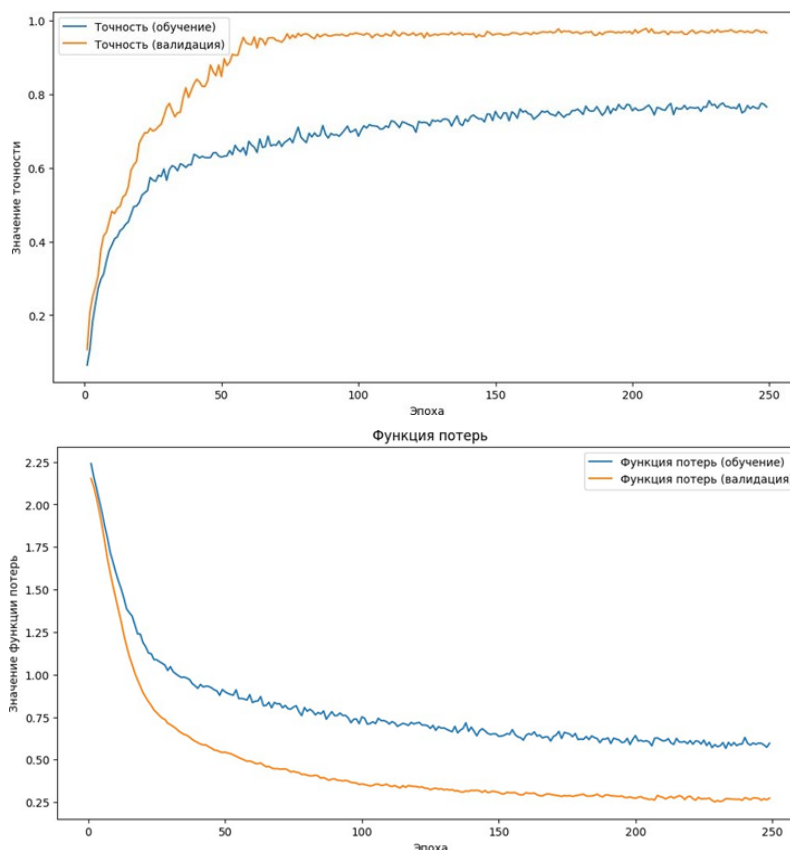


Рисунок 3 – График кривых обучения
Figure 3 – Learning curves graphs

В дальнейшем, с целью улучшения производительности и обобщающей способности модели было принято решение провести оптимизацию гиперпараметров нейронной сети. В рамках стандартной практики оптимизация, как правило, производится эмпирически: в ходе множественных экспериментов и путем подбора оптимальных параметров. Однако данный процесс может быть автоматизирован при помощи различных средств автоматического подбора гиперпараметров. Одним из таких средств, использованных в рамках данной статьи, является библиотека Optuna, реализующая эффективные алгоритмы автоматизированной настройки гиперпараметров, таких как коэффициент Dropout, коэффициент L2-регуляризации, количество нейронов в первом и втором скрытых слоях, а также коэффициент обучения.

Анализируя полученные графики (Рисунок 4), можно сделать вывод, что процесс обучения модели достиг хорошего уровня производительности:

1. Кривые обучающей и валидационной потерь демонстрируют типичную тенденцию к уменьшению, что свидетельствует об эффективном обучении модели на данных.

2. Валидационная точность достигает около 91,92 % на эпохе 156, что является достаточно высокой точностью для задачи классификации такого рода.

3. Контрольная точка раннего прекращения останавливает обучение после эпохи 156, поскольку валидационная точность не улучшалась в течение 20 последовательных эпох (на основании параметра $patience = 20$), предотвращая потенциальное переобучение.

4. Заключительный отчет о классификации жестов показывает хорошие показатели точности, полноты и F1-меры для большинства классов. Средневзвешенные значения точности, полноты и F1-меры находятся в диапазоне 0,92–0,93, что является достаточно хорошим результатом.

5. Класс с наименьшей F1-мерой – это класс 12 (0,73), что указывает на то, что этот класс может представлять наибольшую сложность для корректной классификации моделью.

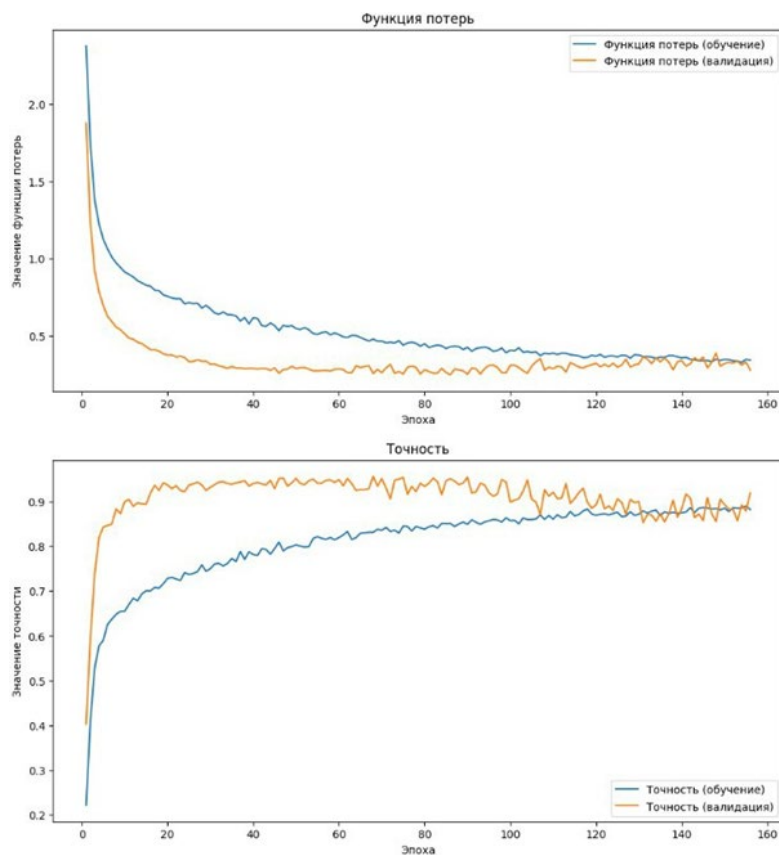


Рисунок 4 – График кривых обучения после автоматического подбора гиперпараметров
Figure 4 – Graph of learning curves after automatic selection of hyperparameters

В целом, полученные результаты свидетельствуют о том, что модель достигла хороших показателей для решаемой задачи, продемонстрировав высокую точность и сбалансированные метрики для большинства классов. Результат работы модели и перечень классифицируемых жестов приведен на Рисунке 5.

Следующим этапом работы является преобразование обученной модели в формат TensorFlow Lite, отличающийся существенно меньшим размером в памяти и более эффективными вычислениями по сравнению с полнофункциональными моделями TensorFlow. Это достигается за счет таких оптимизаций, как квантование весов и активаций, использование специализированных ядер для выполнения сверточных

операций и т. д., что позволяет использовать обученные модели на устройствах с ограниченными вычислительными ресурсами.

С целью создания интуитивно понятного интерфейса взаимодействия с оператором проведена работа над созданием программного обеспечения для интеллектуальной системы управления квадрокоптером. Работа данной системы заключается в захвате и анализе видеопотока с записывающего устройства (веб-камера) с последующей интерпретацией и отправкой команд в бортовую систему квадрокоптера DJI Tello EDU.



Рисунок 5 – Результат работы модели и перечень используемых жестов
Figure 5 – Model results and list of gestures used

Комплекс программных средств системы реализован на языке программирования Python и состоит из следующих взаимосвязанных модулей:

1. Модуль обработки видеопотока, предназначенный для распознавания жестов рук на основе видеопотока, полученного с камеры, и передачи идентификаторов распознанных жестов модулю управления квадрокоптером.

2. Модуль управления квадрокоптером, предназначенный для обработки жестов рук, полученных от системы распознавания жестов, и последующей передачи соответствующих управляющих сигналов в систему управления квадрокоптером.

3. Пользовательский интерфейс, реализованный в виде приложения с графическим интерфейсом, обеспечивающим взаимодействие пользователя с квадрокоптером посредством жестов рук.

Разработанное программное обеспечение успешно реализует функционал интеллектуальной системы управления квадрокоптером с использованием жестов рук, объединяя все компоненты в единое целое. Код программного обеспечения следует принципам SOLID и чистой архитектуры, что обеспечивает гибкость и простоту при дальнейшем добавлении новых функциональных возможностей.

Результаты

Разработанная интеллектуальная система управления была протестирована в контролируемой лабораторной среде, обеспечивающей безопасность и исключающей возможность повреждения квадрокоптера или нанесения вреда окружающим людям и объектам. Для визуализации экспериментов использовалась камера квадрокоптера и веб-камера, подключенная к персональному компьютеру.

В ходе исследований оценивались следующие аспекты работы системы: точность распознавания жестов рук, время отклика системы на выполнение жестов, стабильность работы в различных условиях освещения и фоновых сценах, корректность передачи команд управления квадрокоптеру.

Методика проведения экспериментов:

1. Калибровка системы и настройка параметров.
2. Выполнение последовательности жестов рук перед веб-камерой.
3. Оценка точности распознавания жестов путем сравнения с эталонными данными.
4. Измерение времени отклика системы от момента выполнения жеста до реакции квадрокоптера.
5. Варьирование условий освещения и фоновых сцен для оценки стабильности работы.
6. Проверка корректности передачи команд управления квадрокоптеру путем сравнения ожидаемого и фактического поведения.

Результаты экспериментов:

- точность распознавания жестов рук достигла 91 % для набора из 15 жестов в контролируемых условиях освещения и фоновых сцен;
- среднее время отклика системы на выполнение жеста составило 100 мс;
- система продемонстрировала стабильную работу в диапазоне освещенности от 200 до 3000 лк и при наличии различных фоновых сцен;
- команды управления квадрокоптером передавались корректно, квадрокоптер реагировал на жесты в соответствии с ожидаемым поведением.

Обсуждение

В рамках данного исследования была разработана интеллектуальная система управления движением квадрокоптера посредством распознавания жестов рук. Использование современных методов машинного обучения, таких как сверточные нейронные сети и искусственные нейронные сети, в сочетании с оптимизацией гиперпараметров и применением эффективных библиотек для обработки изображений и анализа жестов, позволило достичь высокой точности распознавания (91 %) и быстрого времени отклика системы (100 мс).

Одним из ключевых аспектов исследования стала интеграция различных инструментов и технологий для создания комплексного решения, адаптируемого под различные задачи. Применение TensorFlow Lite обеспечило возможность эффективного развертывания моделей машинного обучения на устройствах с ограниченными вычислительными ресурсами. Библиотека MediaPipe продемонстрировала свою эффективность в задачах анализа жестов.

Стоит отметить, что дальнейшее совершенствование системы может быть достигнуто за счет расширения набора жестов и улучшения распознавания наиболее сложных классов, а также введение возможности обработки динамических жестов для реализации более сложных сценариев полета.

Заключение

Проведенное исследование подтвердило возможность создания высокоэффективной и дешевой интеллектуальной системы управления беспилотным летательным аппаратом посредством распознавания жестов рук оператора. Применение передовых методов машинного обучения, компьютерного зрения и оптимизационных алгоритмов позволило добиться высоких показателей точности и скорости работы разработанной системы.

Ключевыми достижениями исследования стали успешная интеграция технологий сверточных и искусственных нейронных сетей, оптимизация гиперпараметров с помощью библиотеки Optuna, реализация на устройствах с ограниченными ресурсами благодаря TensorFlow Lite и эффективное распознавание жестов с использованием MediaPipe.

Полученные результаты подтверждают перспективность дальнейших исследований в области применения методов искусственного интеллекта для управления беспилотными летательными аппаратами. Разработанная система может найти применение в различных областях, таких как мониторинг, доставка грузов, а также в образовательных и развлекательных целях.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Дудин Д.Е. Система управления движением беспилотного летательного аппарата на основе распознавания жестов рук. *Юный ученый*. 2024;(5):67–72.
Dudin D.E. Sistema upravleniya dvizheniem bescpilotnogo letatel'nogo apparata na osnove raspoznavaniya zhestov ruk. *Yunyi uchenyi*. 2024;(5):67–72. (In Russ.).
2. Sanna A., Lamberti F., Paravati G., Manuri F. A Kinect-Based Natural Interface for Quadrotor Control. *Entertainment Computing*. 2013;4(3):179–186. <https://doi.org/10.1016/j.entcom.2013.01.001>
3. Zhao R., Wang K., Divekar R., Rouhani R., Su H., Ji Q. An Immersive System with Multi-Modal Human-Computer Interaction. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 15-19 May 2018, Xi'an, China*. IEEE; 2018. P. 517–524. <https://doi.org/10.1109/FG.2018.00083>
4. Мурлин А.Г., Пиотровский Д.Л., Руденко Е.А., Янаева М.В. Алгоритм и методы обнаружения и распознавания жестов руки на видео в режиме реального времени. *Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета*. 2014;(97). URL: <http://ej.kubagro.ru/2014/03/pdf/20.pdf>
Murlin A.G., Piotrovskiy D.L., Rudenko E.A., Yanaeva M.V. Algorithms and methods for detection and recognition of hand gestures on video in real time. *Politematicheskii setevoi elektronnyi nauchnyi zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta = Polythematic online scientific journal of Kuban State Agrarian University*. 2014;(97). (In Russ.). URL: <http://ej.kubagro.ru/2014/03/pdf/20.pdf>
5. Нагапетян В.Э., Хачумов В.М. Распознавание жестов руки в задаче бесконтактного управления беспилотным летательным аппаратом. *Автометрия*. 2015;51(2):103–109.
Nahapetyan V.E., Khachumov V.M. Gesture recognition in the problem of contactless control of an unmanned aerial vehicle. *Optoelectronics, Instrumentation and Data Processing*. 2015;51(2):192–197. <https://doi.org/10.3103/S8756699015020132>
6. Ярышев С.Н., Рыжова В.А. *Технологии глубокого обучения и нейронных сетей в задачах видеоанализа*. Санкт-Петербург: Университет ИТМО; 2022. 82 с.

- Yaryshev S.N., Ryzhova V.A. *Tekhnologii glubokogo obucheniya i neironnykh setei v zadachakh videoanaliza*. Saint Petersburg: ITMO University; 2022. 82 p. (In Russ.).
7. Чернышев Н.Н., Ниженец Т.В. Алгоритм планирования пути в трехмерной детерминированной среде с препятствиями на основе метода роя частиц. *Вестник Воронежского государственного технического университета*. 2022;18(6):7–14. <https://doi.org/10.36622/VSTU.2022.18.6.001>
Chernyshev N.N., Nizhenets T.V. Path planning algorithm in a three-dimensional deterministic environment with obstacles using particle swarm algorithm. *Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta = Bulletin of Voronezh State Technical University*. 2022;18(6):7–14. (In Russ.). <https://doi.org/10.36622/VSTU.2022.18.6.001>
8. Булыгин Д.А., Мамонова Т.Е. Распознавание жестов рук в режиме реального времени. *Научный вестник Новосибирского государственного технического университета*. 2020;(1):25–40. <https://doi.org/10.17212/1814-1196-2020-1-25-40>
Bulygin D.A., Mamonova T.E. Recognition of hand gestures in real time. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta = Science Bulletin of the Novosibirsk State Technical University*. 2020;(1):25–40. (In Russ.). <https://doi.org/10.17212/1814-1196-2020-1-25-40>
9. Dhawale P., Masoodian M., Rogers B. Bare-hand 3D gesture input to interactive systems. In: *CHINZ '06: Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-human Interaction: Design Centered HCI, 6-7 July 2006, Christchurch, New Zealand*. New York: Association for Computing Machinery; 2006. P. 25–32. <https://doi.org/10.1145/1152760.1152764>
10. Чудновский М.М. Алгоритм распознавания жестов руки человека на видеопоследовательности в режиме реального времени для реализации интерфейсов человеко-машинного взаимодействия. *Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева*. 2014;(3):162–167.
Chudnovsky M.M. A real-time algorithm for human's hand gesture recognition on video-sequence for human-computer interaction interfaces. *Vestnik Sibirskogo gosudarstvennogo aerokosmicheskogo universiteta imeni akademika M.F. Reshetneva*. 2014;(3):162–167. (In Russ.).

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Чернышев Николай Николаевич, кандидат технических наук, доцент, доцент кафедры автоматических систем, МИРЭА – Российский технологический университет, Москва, Российская Федерация.
e-mail: chernyshev@mirea.ru
ORCID: [0000-0003-3057-9140](https://orcid.org/0000-0003-3057-9140)

Nikolay N. Chernyshev, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Automatic Systems, MIREA – Russian Technological University, Moscow, the Russian Federation.

Шевченко Михаил Андреевич, студент бакалавриата, кафедра автоматических систем, МИРЭА – Российский технологический университет, Москва, Российская Федерация.
e-mail: shvchenko.m.a@edu.mirea.ru

Mikhail A. Shevchenko, undergraduate student, Department of Automatic Systems, MIREA – Russian Technological University, Moscow, the Russian Federation.

Ниженец Татьяна Владимировна, аспирант, старший преподаватель кафедры

Tatyana V. Nizhenets, PhD student, Senior Lecturer at the Department of Automatic

автоматических систем, МИРЭА – Systems, MIREA – Russian University of
Российский технологический университет, Technology, Moscow, the Russian Federation.
Москва, Российская Федерация.

e-mail: nizhenec@mirea.ru

ORCID: [0000-0001-6230-469X](https://orcid.org/0000-0001-6230-469X)

*Статья поступила в редакцию 10.06.2024; одобрена после рецензирования 21.06.2024;
принята к публикации 25.06.2024.*

*The article was submitted 10.06.2024; approved after reviewing 21.06.2024;
accepted for publication 25.06.2024.*