



Жестовое управление полетом малого беспилотного летательного аппарата

Николай Сергеевич **Абрамов**¹, Вита Викторовна **Саттарова**²,
Виталий Петрович **Фраленко**³, Михаил Вячеславович **Хачумов**⁴

^{1,3,4} Институт программных систем им. А. К. Айламазяна РАН, Вельское, Россия

² Российский университет дружбы народов, Москва, Россия

⁴ Российский государственный гуманитарный университет, Москва, Россия

⁴ Федеральный исследовательский центр «Информатика и управление» РАН, Москва, Россия

⁴ МИРЭА - Российский технологический университет, Москва, Россия

³ alarmod@pereslavl.ru

Аннотация. Рассмотрена задача построения жестовых команд для управления малым беспилотным летательным аппаратом, таким как квадрокоптер. Получаемые видеокamerой команды идентифицируются классификатором на основе сверточной нейронной сети, а мультимодальный интерфейс управления, оснащенный интеллектуальным решателем, преобразует их в команды управления квадрокоптером. Нейронные сети из библиотеки моделей нейронных сетей Ultralytics позволяют выделять целевые объекты в кадре в режиме реального времени. Команды управления квадрокоптером поступают в специализированную программу на смартфоне, разработанную на базе симулятора полетов DJI SDK, которая посылает команды по радиоканалу дистанционного управления.

Исследовано качество распознавания разработанных жестовых команд для квадрокоптеров DJI Phantom 3 standard edition. Представлено краткое руководство в виде сценариев работы оператора с беспилотными транспортными средствами. Раскрыты перспективы жестового управления несколькими транспортными средствами в экстремальных условиях с учётом сложностей безопасности совместного полета и взаимодействия летательных аппаратов в ограниченном пространстве.

Ключевые слова и фразы: беспилотный летательный аппарат, управление, жесты, сверточная нейронная сеть, Ultralytics, интеллектуальный интерфейс, распознавание

Благодарности: Исследование выполнено за счет гранта Российского научного фонда № 21-71-10056, <https://rscf.ru/project/21-71-10056/>

Для цитирования: Абрамов Н. С., Саттарова В. В., Фраленко В. П., Хачумов М. В. *Жестовое управление полетом малого беспилотного летательного аппарата* // Программные системы: теория и приложения. 2024. Т. 15. № 2(61). С. 21–36. https://psta.psir.ru/read/psta2024_2_21-36.pdf

Введение

В настоящее время большое количество исследований и разработок посвящено развитию беспилотных летательных аппаратов (БПЛА). Для повышения надежности и удобства взаимодействия с БПЛА в дополнение к типовым пультовым операциям управления применяют жестовые и речевые команды [1]. Для реализации такого подхода необходимы многомодальные человеко-машинные интерфейсы, адаптированные к различным типам БПЛА. Жесты подразделяются на статические и динамические и используются для передачи информации в компьютер с целью управления компьютером, летательным аппаратом, роботом и т.д.

В диссертационной работе [2] представлена комплексная методология захвата, отслеживания и распознавания динамических жестов в видеопотоке. В рамках этой методологии разработаны алгоритм захвата и отслеживания кисти человека в видеопотоке на сложном фоне; алгоритм и вычислительно-эффективная модель для распознавания жестов, основанная на нечетких конечных автоматах. Разработана методология мультимодального распознавания сцен, определяемых жестами, с использованием нечетких операторов агрегирования. Экспериментально показано, что предложенная архитектура системы распознавания динамических жестов позволяет с высокой степенью надежности распознавать в реальном времени жесты независимо от индивидуума.

В диссертационной работе [3] предложены метод извлечения изображения руки из дальностного изображения человека на основе анализа связанности точек изображения в трехмерном пространстве; метод распознавания позиции кончиков пальцев и точек соединения пальцев с ладонью руки на основе анализа контура изображения руки; метод скелетизации дальностного изображения, основанный на непрерывном скелетном представлении бинарного изображения, поиске граничных точек фигуры и создании диаграммы Вороного для этих точек; метод распознавания статических и динамических жестов рук и пальцев руки для жестовой азбуки глухонемых.

В исследовании [4] представлено решение проблемы восстановления и отслеживания трехмерного положения, ориентации и полной артикуляции человеческой руки по данным от Kinect-сенсора. Решается задача оптимизации, в которой осуществляется поиск параметров модели руки, которые минимизируют расхождение между внешним видом и трехмерной структурой данных от сенсора. 3D-трекинг движения рук выполняется в реальном времени.

В работе [5] представлена техника взаимодействия, позволяющая с помощью жестов рук управлять функциями камеры, такими как панорамирование, наклон и затвор. В основе предложенной техники — алгоритм Лукаса-Канаде.

В исследовании [6] выполнена оценка возможности нейронных сетей помогать в распознавании жестов рук в арабском жестовом языке, проведены эксперименты с нейронными сетями прямого распространения и рекуррентными нейронными сетями. Предложенная система с полностью рекуррентной архитектурой демонстрирует точность 95% при распознавании статических жестов.

В работе [7] представлен интерактивный интерфейс пользователя для распознавания жестов рук американского жестового языка с использованием буквосочетаний на пальцах. Жесты классифицируются с помощью метода случайных лесов (от англ. «random forest»). Классификатор жестов интегрирован с английским словарем для ускорения написания текстов.

В исследовании [8] представлен прототип мультимодальной системы, который интегрирует методы распознавания лица, жестов и речи для поддержки мультимодальной возможности взаимодействия человека с компьютером. Разработана многоуровневая система с несколькими камерами для наблюдения за лицом пользователя, жестами тела и пространственным расположением в комнате. Используя речевой ввод, система более точно воспринимает намерения пользователя.

В работе [9] описан механизм управления квадрокоптером с помощью жестов и поз. Этот механизм делает человеко-машинное взаимодействие более интуитивным, удобным и отзывчивым на потребности пользователя.

Управление жестовыми командами может осуществляться с использованием различных сенсорных устройств. Например, для квадрокоптера *Parrot AR.Drone 2.0* ^{URL} от компании Parrot в работе [10] была предложена бесконтактная система управления с использованием трехмерного сенсора Kinect. Однако, универсальным и распространенным видом современного устройства ввода информации является, несомненно, видеочамера. Перспективным шагом можно считать построение комбинированных интерфейсов, отличающихся большей надежностью. Например, такой интерфейс представлен в работе [11], где для распознавания жестов оператора БПЛА используется нейросетевая архитектура Yolo 5, а для обработки и распознавания речи — различные многослойные искусственные нейронные сети (ИНС).

Выделим несколько инструментальных средств распознавания образов.

Метод инвариантных моментов [12]. Здесь распознавание позиции и ориентации руки в бинарном изображении осуществляется посредством вычисления моментов изображения, при условии, что фон изображения однороден и рука является доминирующим объектом в изображении. Метод реализуется для 2D-и 3D-режимов и позволяет упростить процесс сравнения жеста с эталонами. Методы позволяют корректно, в пределах точности представления изображений, сравнивать два изображения. Инвариантный подход к распознаванию допускает точную математическую постановку задачи, позволяющую задавать классы объектов. Метод был практически применен в диссертационной работе [3].

Метод *DTW*^{URL} (англ. Dynamic Time Warping, алгоритм динамической трансформации временной шкалы). Идентификация формы осуществляется путем сравнения «скелета» руки с эталонами. Под скелетом здесь понимается набор опорных точек ладони, идентифицирующих положение ладони и пальцев. Для оценки степени схожести предварительно производится развертка скелетов, где по оси абсцисс откладываются номера точек в строгом соответствии с номерами их обхода в скелете, а по оси ординат координаты точек. После развертки расстояние между объектами оценивается за полиномиальное время с помощью алгоритма динамической трансформации шкалы времени.

Известно, что рекуррентная ИНС с архитектурой LSTM (Long Short-Term Memory) [13] показывает хорошие результаты при распознавании жестов, пример ее использования приведен в исследовании [14], где полученное значение показателя «точность» («ассигасу») не менее 0.90, а в среднем 0.93. В работе [15] было проведено обучение нейронной сети Mask-RCNN для распознавания жестов рук, проанализированы существующие способы распознавания жестов, исследованы преимущества и недостатки рассмотренных методов. Предложена собственная архитектура сверточной нейронной сети для решения задачи классификации жестов. Проведена оценка точности работы сети в зависимости от расстояния между камерой и рукой, а также в зависимости от сложности жеста.

Предлагаемый в настоящей работе интерфейс управления БПЛА обеспечивает повышенную надежность и гибкость за счет использования стандартного кнопочного управления в комбинации с голосовыми и жестовыми командами. При управлении со стационарного рабочего места управление оказывается максимально эффективным. Однако, беспилотники могут работать и в режиме передачи управления людям, находящимся в местах, где имеются пожары, задымления, высокий уровень шума. Именно в этих условиях важно получать надежно команды управления, по этой причине и вводится мультимодальность управления,

позволяющая продублировать команды, при этом с помощью ИНС определяется степень уверенности и выбирается наиболее вероятная команда.

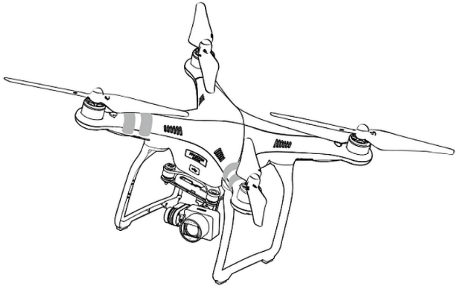
Управление БПЛА с помощью жестов включает несколько ключевых элементов:

- (1) видеокамера, способная регистрировать жесты пользователя;
- (2) алгоритмы распознавания, способные классифицировать и преобразовать жест в команду;
- (3) команды управления, осуществляющие преобразование набора жестов в команды БПЛА;
- (4) интеллектуальный интерфейс – программно-техническое средство, через которое пользователь взаимодействует с аппаратом.

Оценка эффективности управления квадрокоптером на основе системы жестовых команд

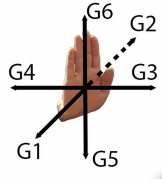
В настоящем исследовании исходным материалом служили базовые команды управления. Рассматривались движения по всем направлениям пространства, включая перемещение, поворот вокруг вертикальной оси, взлет и посадку. Для экспериментов с управлением жестами использовался квадрокоптер модели *DJI Phantom 3 Standard Edition*^[URL]. Данный аппарат оснащен многофункциональной двухповоротной видеокамерой. Общий вид квадрокоптера и его основные характеристики представлены в таблице 1.

Таблица 1. Общий вид и основные характеристики задействованного БПЛА

Квадрокоптер DJI Phantom 3 Standard Edition	Основные характеристики
	(a) стабилизированный 3-осевой подвес с камерой; (б) встроенная камера: 2.7K видео и 12MP фото; (в) радиоканал WiFi 2.4G; (г) дальность связи 1 км; (д) интеллектуальный аккумулятор 15.2 V, 480 mAh.

Предлагаемая система команд управления БПЛА приведена в таблице 2.

ТАБЛИЦА 2. Система команд для управления БПЛА

Команда	Назначение	Описание способа управления	Вид жеста
Взлет	Взлет на высоту 2.3 метра.	Жестовая команда «палец вверх».	
Приземление	Снижение и посадка в текущих координатах.	Жестовая команда «палец вниз».	
Перемещение	Перемещение по трем осям пространства.	Перемещение ладони по трем осям в шести направлениях: G1 – вперед, G2 – назад, G3 – влево, G4 – вправо, G5 – вниз, G6 – вверх.	
Домой	Взлет на высоту 30 м. и возврат по прямой в координаты начальной точки, посадка.	Жестовая команда «руки, сложенные в форме крыши», команда может выполняться автоматически в случае потери связи с БПЛА.	
Поворот	Поворот вокруг оси рыскания.	Наклон ладони по двум направлениям: G7 – поворот вокруг оси против часовой стрелки, G8 – поворот вокруг оси по часовой стрелке.	
Остановка	Прекращение движения (зависание в воздухе).	Жестовая команда «поднятый кулак».	

Инструкции для БПЛА, сформированные наземной станцией после обработки жестовых команд оператора, поступают в специализированную программу на смартфоне, разработанную на базе DJI SDK, которая посылает команды через пульт и далее радиоканал.

Для оценки возможностей актуальной версии 8.1.47 программного пакета *Ultralytics*^{URL} по локализации целевых объектов (жестов для управления БПЛА) проведен ряд экспериментов с поддерживаемыми этим пакетом конфигурациями ИНС: Yolo 5, Yolo 8, Yolo 8 Ghost, Yolo 8 Rtdetr, Yolo 9 и оригинальной архитектурой Rtdetr от Baidu [16]. Во всех случаях, кроме последнего и предпоследнего, использовались *m*-модификации ИНС, для Yolo 9 – *s*-модификация, а в последнем случае – *l*-модификация. Работа ИНС проверялась на собственном датасете с командами для управления БПЛА, содержащем жестовые команды «взлет», «приземление», «остановка» и «домой», применяемые при мультимодальном управлении БПЛА [11]; использовалась разметка с помощью *ориентированных*^{URL} и *неориентированных*^{URL} прямоугольных областей; всего 247 обучающих изображений и 122 тестовых. Тестирование выполнялось в режиме с половинной точностью (параметр *half*=True).

Повышение параметра *imgsz* (влияет на масштабирование поступающих изображений) с исходных 640 пикселей, заложенных по умолчанию в программный код библиотеки, до 1120 позволяет уверенно распознавать жесты, показываемые оператором на удалении. В процессе валидации результатов с пороговым значением *conf* (уровень уверенности ИНС) отсеиваются те зоны, потенциально содержащие жест, для которых $conf \geq 0.01$. Среди оставшихся выбираются те, у которых значение $iou \geq 0.7$ (уровень пересечения с оригинальным жестом в зоне). Значения параметров подобраны экспериментально.

Для исследования качества распознавания на вход ИНС подавались изображения, измененные до размера 1120x640 пикселей. При тестировании использовалось следующее аппаратное обеспечение: процессор общего назначения Intel Core i7-6850k (6 ядер / 12 потоков, 4 ГГц), оперативная память 32 ГБ, видеокарта Nvidia GeForce RTX 3060.

Результаты решения задачи выделения управляющих жестов (для неориентированных прямоугольников) представлены в таблице 3. Основные показатели качества – mAP50-95 и значение F1-меры, достигаемой при указанном в таблице пороге уверенности.

ТАБЛИЦА 3. Результаты работы ИНС для локализации управляющих жестов (для неориентированных прямоугольников)

Вид ИНС	Команда	Точность	Полнота	mAP50-95	F1	Порог уверенности
Yolo 5	взлет	1.000	0.965	0.871	0.99	0.723
	приземление	0.999	1.000	0.869		
	остановка	0.997	1.000	0.841		
	домой	0.997	1.000	0.802		
	все	0.998	0.991	0.846		
Yolo 8	взлет	1.000	0.924	0.749	0.96	0.583
	приземление	1.000	0.850	0.659		
	остановка	1.000	0.997	0.694		
	домой	0.928	1.000	0.591		
	все	0.982	0.943	0.673		
Yolo 9	взлет	1.000	0.999	0.868	1.00	0.242
	приземление	0.971	1.000	0.853		
	остановка	1.000	1.000	0.851		
	домой	1.000	1.000	0.854		
	все	0.993	1.000	0.856		
Yolo 8 Ghost	взлет	1.000	1.000	0.812	1.00	0.795
	приземление	0.996	1.000	0.861		
	остановка	1.000	1.000	0.829		
	домой	0.997	1.000	0.825		
	все	0.998	1.000	0.831		
Yolo 8 Rtdetr	взлет	0.998	1.000	0.857	1.00	0.861
	приземление	0.986	1.000	0.794		
	остановка	0.998	1.000	0.758		
	домой	0.997	1.000	0.864		
	все	0.995	1.000	0.818		
Rtdetr (Baidu)	взлет	0.971	1.000	0.805	0.97	0.616
	приземление	0.882	1.000	0.736		
	остановка	0.995	1.000	0.634		
	домой	0.954	1.000	0.756		
	все	0.950	1.000	0.733		

В случае использования неориентированных прямоугольников архитектуры Yolo 8 Ghost, Yolo 8 Rtdetr и Yolo 9 обеспечили безошибочное выделение всех жестовых команд без ложных срабатываний, однако, максимальную точность локализации обеспечивает новая Yolo 9, она же

имеет самое высокое значение $mAP_{50-95} = 0.856$. Детальное исследование данных, получаемых от ИНС, показало, что иногда в кадре находится несколько объектов (не более двух, как показали эксперименты), однако за счет порога уверенности можно добиться полного исключения ложного обнаружения. В данном случае Yolo 9 обрабатывает кадр видеопотока за 19.5 мс.

Результаты решения задачи выделения управляющих жестов (для ориентированных прямоугольников) представлены в таблице 4. Вариант Yolo 8 Rtdetr исключен из рассмотрения ввиду того, что слой RTDETR-Decoder в этой архитектуре нельзя заменить на ОБВ-слой.

ТАБЛИЦА 4. Результаты работы ИНС для локализации управляющих жестов (для ориентированных прямоугольников)

Вид ИНС	Команда	Точность	Полнота	mAP_{50-95}	F1	Порог уверенности
Yolo 5	взлет	1.000	1.000	0.914	1.00	0.785
	приземление	1.000	1.000	0.978		
	остановка	0.996	1.000	0.965		
	домой	0.996	1.000	0.935		
	все	0.988	1.000	0.948		
Yolo 8	взлет	0.994	1.000	0.937	1.00	0.793
	приземление	1.000	1.000	0.977		
	остановка	0.998	1.000	0.951		
	домой	1.000	1.000	0.961		
	все	0.998	1.000	0.956		
Yolo 9	взлет	1.000	1.000	0.918	1.00	0.747
	приземление	1.000	1.000	0.977		
	остановка	1.000	1.000	0.960		
	домой	0.993	1.000	0.913		
	все	0.998	1.000	0.942		
Yolo 8 Ghost	взлет	0.988	1.000	0.924	1.00	0.796
	приземление	1.000	1.000	0.961		
	остановка	0.998	1.000	0.936		
	домой	0.996	1.000	0.926		
	все	0.996	1.000	0.937		
Rtdetr (Baidu)	взлет	0.991	1.000	0.928	1.00	0.745
	приземление	0.999	1.000	0.971		
	остановка	0.998	1.000	0.964		
	домой	1.000	1.000	0.949		
	все	0.997	1.000	0.953		

Все варианты архитектур обеспечили 100%-ые полноту и F1-меру.

Однако, Yolo 8 обеспечила самое высокое значение показателя mAP50-95 = 0.956 (среднее для всех жестов) в сочетании с точностью 0.998. Если сравнивать эту ИНС с лучшей для неориентированных прямоугольников, переход к ориентированным обеспечил рост mAP50-95 на величину 0.1, что весьма существенно. Yolo 8 обрабатывает кадр видеопотока за 12.8 мс. Yolo 8 Ghost решает ту же задачу за 9.6 мс, то есть позволяет снизить временные затраты на 25%. Учитывая, что в данной задаче обе сети обеспечивают F1-меру 100%, приоритет следует отдать Yolo 8 Ghost.

Далее приведем краткий сценарий работы оператора с БПЛА:

- (1) подготовить полетный план: задать координаты доступных для полета точек (широта, долгота, высота);
- (2) активировать составляющие систему: БПЛА, пульт (для связи БПЛА с наземной станцией), мобильное приложение (клиент) и главную программу (сервер); для управления жестами к серверу должна быть подключена видеокамера;
- (3) в мобильном приложении выбрать настройки: установить порт для связи с сервером и координаты «домашней точки», флаги передачи телеметрической информации, видеопотока с бортовой камеры БПЛА и использования режима симулятора компании-разработчика квадрокоптера; после этого установится связь между сервером и БПЛА, и главная программа будет готова принимать команды оператора;
- (4) зафиксировать положение оператора в кадре видеокамеры сервера таким образом, чтобы в кадр попадали руки оператора;
- (5) для управления жестами оператор показывает в видеокамеру один из них: «взлет», «приземление», «домой», «остановка»; при этом следует показывать жест несколько секунд: это сделано для того, чтобы исключить ложные срабатывания системы распознавания;
- (6) посредством нажатия левой кнопки мыши по кнопкам интерфейса главной программы оператор может отдать команды БПЛА в соответствии с таблицей 2;
- (7) посредством зажатия левой кнопки мыши на трехмерной карте местности полета в центральной части интерфейса, в которой отображено текущее положение БПЛА, осуществляется смещение «камеры-наблюдателя» 3D-сцены; зажатие колеса мыши позволяет поворачивать сцену, а зажатие правой кнопки мыши – позволяет управлять «зумом»;
- (8) после выполнения полетного задания оператору следует дать команду «приземление» или «домой» любым описанным выше способом;
- (9) после завершения полета выключить БПЛА, пульт, мобильное приложение и остановить работу сервера.

Заклучение

Рассмотрен и прошел первичную апробацию подход к управлению малым БПЛА типа квадрокоптер на основе жестовых команд. Базовые команды управления движением летательного аппарата включают перемещения в пространстве, повороты вокруг вертикальной оси, взлет и посадку. Эксперименты показали достаточную для практического использования точность распознавания ряда жестовых команд с применением сверточных искусственных нейронных сетей, соответствующую зарубежным аналогам [4–9, 15], задача полностью решена с помощью архитектур Yolo 8 Ghost и Yolo 9 для режимов с ориентированными и неориентированными прямоугольниками для выделения зон с жестами.

Дальнейшим расширением подхода может служить управление жестами несколькими БПЛА. Выполнение сложных групповых заданий вызывает необходимость решения задач, связанных с безопасным совместным полетом и взаимодействием автономных летающих роботов в процессе функционирования на одних и тех же участках. Например, когда в условиях пожара перед несколькими БПЛА в процессе выполнения спасательных работ возникает необходимость в совместном поднятии в воздух тяжелого объекта для смены его местоположения. В этом случае каждый БПЛА должен быть способным к коллективному взаимодействию элементов группы как интеллектуальных агентов. Подобные сложные задачи интеллектуального взаимодействия автономных БПЛА требуют расширения системы команд для придания специфики каждому летательному аппарату. Кроме того, требуется наделение интерфейса и БПЛА некоторыми интеллектуальными функциями.

Список использованных источников

- [1] Абрамов Н. С., Талалаев А. А., Фраленко В. П., Хачумов М. В. *Система мультимодального управления и визуализации полета беспилотного летательного аппарата* // Авиакосмическое приборостроение.– 2023.– № 9.– С. 3–11.  ^{↑22}
- [2] Алфимцев А. Н. *Разработка и исследование методов захвата, отслеживания и распознавания динамических жестов*, Диссертация на соискание ученой степени кандидата технических наук.– М.– 2008.– 167 с. ^{↑22}
- [3] Нагапетян В. Э. *Методы распознавания жестов руки на основе анализа дальностных изображений*, Диссертация на соискание ученой степени кандидата физико-математических наук.– М.– 2008.– 117 с. ^{↑22, 24}
- [4] Oikonomidis I., Kyriazis N., Argyros A. A. *Efficient model-based 3D tracking of hand articulations using Kinect* // *Proceedings of the 22nd British Machine Vision Conference 2011*, BMVC'11 (Dundee, UK, August 29–September 2, 2011).– 2011.– ISBN 1-901725-43-X.– 11 pp.   ^{↑22, 31}

- [5] Shaowei C., Tanaka J. *Interacting with a self-portrait camera using motion-based hand gestures* // *Proceedings of the 11th Asia-Pacific Conference on Computer-Human Interaction 2013, APCHI'13* (Bangalore, India, September 24–27, 2013), New York: ACM.– 2013.– ISBN 978-1-4503-2253-9.– Pp. 93–101.  [↑23, 31](#)
- [6] Maraqa M. R., Al-Zboun F., Dhyabat M., Zitar R. A. *Recognition of Arabic Sign Language (ArSL) using recurrent neural networks* // *Journal of Intelligent Learning Systems and Applications*.– 2012.– Vol. 4.– No. 1.– Pp. 41–52.  [↑23, 31](#)
- [7] Pugeault N., Bowden R. *Spelling it out: real-time ASL fingerspelling recognition* // *Proceedings of the IEEE International Conference on Computer Vision Workshops 2011, ICCV'11* (Barcelona, Spain, 06–13 November 2011).– 2011.– Pp. 1114–1119.  [↑23, 31](#)
- [8] Zhao R., Wang K., Divekar R., Rouhani R., Su H., Ji Q. *An immersive system with multi-modal human-computer interaction* // *Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition 2018* (Xi'an, China, 15–19 May 2018).– 2018.– Pp. 517–524.  [↑23, 31](#)
- [9] Sanna A., Lamberti F., Paravati G., Manuri F. *A kinect-based natural interface for quadrotor control* // *Entertainment Computing*.– 2013.– Vol. 4.– No. 3.– Pp. 179–186.  [↑23, 31](#)
- [10] Нагапетян В. Э., Хачумов В. М. *Распознавание жестов руки в задаче бесконтактного управления беспилотным летательным аппаратом* // *Автометрия*.– 2015.– Т. 51.– № 2.– С. 103–109.  [↑23](#)
- [11] Абрамов Н. С., Емельянова Ю. Г., Талалаев А. А., Фраленко В. П., Хачумов М. В. *Архитектура мультимодального интерфейса для управления беспилотным летательным аппаратом* // *Известия высших учебных заведений. Авиационная техника*.– 2022.– № 3.– С. 55–63.  [↑23, 27](#)
- [12] Hu M. K. *Visual pattern recognition by moment invariants* // *IRE Transactions on Information Theory*.– 1962.– Vol. 8.– No. 2.– Pp. 179–187.  [↑24](#)
- [13] Hochreiter S., Schmidhuber J. *Long short-term memory* // *Neural Computation*.– 1997.– Vol. 9.– No. 8.– Pp. 1735–1780.  [↑24](#)
- [14] Жуковская В. А., Пятаева А. В. *Рекуррентная нейронная сеть для распознавания жестов русского языка с учетом языкового диалекта Сибирского региона* // *ГрафиКон 2022: материалы 32-й Международной конференции по компьютерной графике и машинному зрению* (Рязань, 19–22 сентября 2022 г.).– 2022.– С. 538–547.   [↑24](#)
- [15] Булыгин Д. А., Мамонова Т. Е. *Распознавание жестов рук в режиме реального времени* // *Научный вестник НГТУ*.– 2020.– № 1(78).– С. 25–40.  [↑24, 31](#)
- [16] Zhao Y., Lv W., Xu S., Wei J., Wang G., Dang Q., Liu Y., Chen J., *DETRs beat YOLOs on real-time object detection*.– 2023.– 14 pp. arXiv: 2304.08069 [cs.CV]  [↑27](#)

Поступила в редакцию
одобрена после рецензирования
принята к публикации
опубликована онлайн

09.02.2024;
13.03.2024;
17.03.2024;
22.04.2024.

Рекомендовал к публикации

д.ф.-м.н. А. М. Елизаров

Информация об авторах:



Николай Сергеевич Абрамов

к.т.н., ведущий научный сотрудник ИЦМС ИПС им. А.К. Айламазяна РАН. Область научных интересов: математические методы синтеза, обработки и анализа изображений и сигналов, искусственный интеллект и принятие решений, интеллектуальный анализ данных и распознавание образов, геометрия.



0000-0002-1612-3879

e-mail: n-say@nsa.pereslavl.ru



Вита Викторовна Саттарова

Студент Российского университета дружбы народов им. Патриса Лумумбы. Область научных интересов: интеллектуальный анализ данных и распознавание образов, искусственный интеллект и принятие решений, управление робототехническими системами.



0009-0008-0425-6958

e-mail: 1032201655@pfur.ru



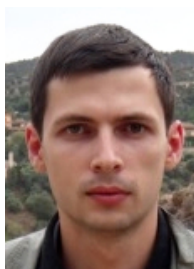
Виталий Петрович Фраленко

Кандидат технических наук, ведущий научный сотрудник ИЦМС ИПС им. А.К. Айламазяна РАН. Область научных интересов: интеллектуальный анализ данных и распознавание образов, искусственный интеллект и принятие решений, параллельно-конвейерные вычисления, сетевая безопасность, диагностика сложных технических систем, графические интерфейсы.



0000-0003-0123-3773

e-mail: alarmod@pereslavl.ru



Михаил Вячеславович Хачумов

Кандидат физико-математических наук, старший научный сотрудник ИЦМС ИПС им. А.К. Айламазяна РАН, старший научный сотрудник ФИЦ «Информатика и управление» РАН, доцент РУДН им. Патриса Лумумбы, доцент МИРЭА. Область научных интересов: интеллектуальный анализ данных и распознавание образов, искусственный интеллект и принятие решений, управление робототехническими системами.



0000-0001-5117-384X

e-mail: khmike@inbox.ru

Все авторы сделали эквивалентный вклад в подготовку публикации.

Декларация об отсутствии личной заинтересованности: благополучие авторов не зависит от результатов исследования.



Gesture control of small unmanned aerial vehicle flight

Nikolai Sergeevich **Abramov**¹, Vita Viktorovna **Sattarova**²,
Vitaly Petrovich **Fralenko**³, Mikhail Vyacheslavovich **Khachumov**⁴

^{1, 3, 4} Ailamazyan Program Systems Institute of RAS, Ves'kovo, Russia

² RUDN University, Moscow, Russia

⁴ Russian state university for the humanities, Moscow, Russia

⁴ Federal Research Center "Computer Science and Control" of RAS, Moscow, Russia

⁴ MIREA - Russian Technological University, Moscow, Russia

 alarmod@pereslavl.ru

Abstract. The problem of constructing gesture commands for controlling a small unmanned aerial vehicle, such as a quadcopter, is considered. Commands coming from a video camera are identified by a classifier based on a convolutional neural network, and the multimodal control interface equipped with an intelligent solver converts them into control commands for the quadcopter. Neural networks from the Ultralytics neural network library allow selecting targets in a frame in real-time. The commands are sent to a specialized program on a smartphone, developed on the basis of DJI SDK flight simulators, which then sends commands via the remote control channel.

The quality of recognition of developed gesture commands for DJI Phantom 3 standard edition quadcopters is investigated, and a brief guide in the form of operator work scenarios with unmanned vehicles is provided. The prospects of gesture control of several vehicles in extreme conditions have been revealed, considering the complex safety challenges of joint flight and interaction of aircraft in confined space. (*In Russian*).

Key words and phrases: unmanned aerial vehicle, control, gestures, convolutional neural network, Ultralytics, intelligent interface, recognition

2020 Mathematics Subject Classification: 68T45; 68T07, 68T40

Acknowledgments: This work was financially supported by the Russian Science Foundation, project № 21–71–10056, <https://rscf.ru/project/21-71-10056/>

For citation: Nikolai S. Abramov, Vita V. Sattarova, Vitaly P. Fralenko, Mikhail V. Khachumov. *Gesture control of small unmanned aerial vehicle flight*. Program Systems: Theory and Applications, 2024, **15**:2(61), pp. 21–36. (*In Russ.*). https://psta.pstiras.ru/read/psta2024_2_21-36.pdf

References

- [1] N. S. Abramov, A. A. Talalaev, V. P. Fralenko, M. V. Xachumov. “Multimodal control and visualization system for unmanned aerial flight”, *Aviakosmicheskoe priborostroenie*, 2023, no. 9, pp. 3–11 (in Russian). [doi](#)
- [2] A. N. Alfimcev. *Development and research of methods for capturing, tracking and recognizing dynamic gestures*, Dissertaciya na soiskanie uchenoj stepeni kandidata texnicheskix nauk, M., 2008 (in Russian), 167 pp.
- [3] V. E. Nagapetyan. *Hand gesture recognition methods based on range image analysis*, Dissertaciya na soiskanie uchenoj stepeni kandidata fiziko-matematicheskix nauk, M., 2008 (in Russian), 117 pp.
- [4] I. Oikonomidis, N. Kyriazis, A. A. Argyros. “Efficient model-based 3D tracking of hand articulations using Kinect”, *Proceedings of the 22nd British Machine Vision Conference 2011, BMVC’11* (Dundee, UK, August 29–September 2, 2011), 2011, ISBN 1-901725-43-X, 11 pp. [URL](#) [doi](#)
- [5] C. Shaowei, J. Tanaka. “Interacting with a self-portrait camera using motion-based hand gestures”, *Proceedings of the 11th Asia-Pacific Conference on Computer-Human Interaction 2013, APCHI’13* (Bangalore, India, September 24–27, 2013), ACM, New York, 2013, ISBN 978-1-4503-2253-9, pp. 93–101. [doi](#)
- [6] M. R. Maraqa, F. Al-Zboun, M. Dhyabat, R. A. Zitar. “Recognition of Arabic Sign Language (ArSL) using recurrent neural networks”, *Journal of Intelligent Learning Systems and Applications*, 4:1 (2012), pp. 41–52. [doi](#)
- [7] N. Pugeault, R. Bowden. “Spelling it out: real-time ASL fingerspelling recognition”, *Proceedings of the IEEE International Conference on Computer Vision Workshops 2011, ICCV’11* (Barcelona, Spain, 06–13 November 2011), 2011, pp. 1114–1119. [doi](#)
- [8] R. Zhao, K. Wang, R. Divekar, R. Rouhani, H. Su, Q. Ji. “An immersive system with multi-modal human-computer interaction”, *Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition 2018* (Xi’an, China, 15–19 May 2018), 2018, pp. 517–524. [doi](#)
- [9] A. Sanna, F. Lamberti, G. Paravati, F. Manuri. “A kinect-based natural interface for quadrotor control”, *Entertainment Computing*, 4:3 (2013), pp. 179–186. [doi](#)
- [10] V. E. Nagapetyan, V. M. Khachumov. “Gesture recognition in the problem of contactless control of an unmanned aerial vehicle”, *Optoelectronics, Instrumentation and Data Processing*, 51:2 (2015), pp. 192–197. [doi](#)
- [11] N. S. Abramov, Yu. G. Emel’yanova, A. A. Talalaev, V. P. Fralenko, M. V. Khachumov. “Multimodal interface architecture for unmanned aerial vehicle control”, *Russian Aeronautics*, 65:3 (2022), pp. 498–506. [doi](#)
- [12] M. K. Hu. “Visual pattern recognition by moment invariants”, *IRE Transactions on Information Theory*, 8:2 (1962), pp. 179–187. [doi](#)
- [13] S. Hochreiter, J. Schmidhuber. “Long short-term memory”, *Neural Computation*, 9:8 (1997), pp. 1735–1780. [doi](#)
- [14] V. A. Zhukovskaya, A. V. Pyataeva. “Recurrent neural network for recognition of gestures of the Russian language, taking into account the language dialect of the Siberian region”, *GrafiKon 2022: materialy 32-j Mezhdunarodnoj*

- konferencii po komp'yuternoj grafike i mashinnomu zreniyu* (Ryazan', 19–22 sentyabrya 2022 g.), 2022, pp. 538–547 (in Russia).  
- [15] D. A. Bulygin, T. E. Mamonova. “Recognition of hand gestures in real time”, *Nauchnyj vestnik NGTU*, 2020, no. 1(78), pp. 25–40 (in Russian). 
- [16] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, J. Chen, . *DETRs beat YOLOs on real-time object detection*, 2023, 14 pp. [arXiv:!\[\]\(5f28278c0ebbde07efa8ee7d80530cb7_img.jpg\) 2304.08069](https://arxiv.org/abs/2304.08069) [cs.CV] 