

ЭЛЕКТРОННАЯ КОМПОНЕНТНАЯ БАЗА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ УПРАВЛЕНИЯ

УДК 004.58

ИНТЕЛЛЕКТУАЛЬНЫЕ ИНТЕРФЕЙСЫ В СИСТЕМАХ УПРАВЛЕНИЯ БПЛА

© 2025 г. О. Г. Шишкин^{1,*}, В. М. Хачумов^{1,2,3}

¹Институт программных систем им. А.К. Айламазяна РАН, Веськово, Россия

²МИРЭА – Российский технологический университет, Москва, Россия

³Федеральный исследовательский центр “Информатика и управление” РАН, Москва, Россия

*E-mail: shishkinog@mail.ru

Поступила в редакцию 12.05.2025 г.

После доработки 22.08.2025 г.

Принята к публикации 22.08.2025 г.

Рассмотрена задача построения универсального интерфейса и интегрированной интеллектуальной системы для речевого, жестового и телекомандного управления робототехническими системами, включая транспортные средства и летательные аппараты. Интегрированная информация и средства ее обработки могут поступать и храниться в базах данных и знаний на борту и на наземных станциях управления. Рассмотрен подход к построению комбинированного интерфейса системы управления беспилотного летательного аппарата типа квадрокоптер, интегрирующего различные методы ввода, обработки и передачи информации. Создана собственная база речевых и жестовых управлений, дублирующих или взаимозаменяющих команд конкретного оператора. Набор жестов основывается на командах, принятых в авиационных системах и адаптированных под возможности предложенной реализации. Выполнена программная реализация с применением искусственных нейронных сетей и осуществлена экспериментальная проверка алгоритмов распознавания, показано, что предложенный подход обеспечивает высокую точность и полноту распознавания.

DOI: 10.56304/S2782375X25020184

ВВЕДЕНИЕ

В современных системах управления робототехническими системами используются самые разнообразные интеллектуальные датчики, предоставляющие данные в определенных форматах по отдельным каналам. К ним относятся системы голосового (речевого) управления, системы управления с помощью жестов. Применение и распознавание речи для ввода информации особенно удобно, когда руки человека заняты. Применение жеста удобно, когда человек стеснен или ограничен в произношении речи. Каждая из таких систем обладает своими особенностями, связанными с форматами представления данных и методами обработки. Так система речевого управления может использовать метод распознавания как всей команды в целом [1], так и ее формантного представления [2, 3]. Аналогично, система управления жестами может располагать методами побуквенного распознавания жеста, например, на основе русской ручной азбуки жестов [4] или специального жеста, передающего команду одним знаком руки [5, 6].

Интеллектуальные интерфейсы создаются с целью увеличения быстродействия и надежности

взаимодействия, представляет интерес решение задачи построения человеко-машинного интерфейса, способного воспринимать формирования и распознавания мультимодальной команды [7]. Совместное использование нескольких разнородных источников данных позволяет повысить точность и эффективность распознавания команды. Целесообразна раздельная фильтрация потоков данных от каждого из датчиков с последующим объединением с весами, зависящими от точности получения параметров. Приоритет при этом дается тем данным, что получены от датчиков с минимальной погрешностью оценки.

Фактически этот подход определяет стратегию построения интеллектуального человеко-машинного интерфейса. Подобная концепция не противоречит общему подходу к организации систем обработки данных от многих датчиков, которая повсеместно применяется в авиационных и космических системах. Идеи мультимодальной обработки информации изложены, например, в [8].

Создание интерфейсов, работающих непосредственно с беспилотным летательным аппаратом (БПЛА), является актуальной задачей, решение которой необходимо для оперативного

получения наиболее существенной целевой информации и принятия оперативных решений. В настоящее время имеется ряд зарубежных и отечественных систем, предназначенных для автоматической обработки аэрокосмической информации, имеющих различное предназначение: от прогнозирования погодной ситуации до выявления очагов лесных пожаров, землетрясений и других чрезвычайных ситуаций. Предлагаемые в настоящей концепции методы и технологии по сравнению с имеющимися аналогами обладают более широкой функциональностью и масштабируемостью к решению задач мониторинга за счет унификации алгоритмов.

Интеллектуальное управление заключается в реализации команд передачи мультимодальной командной информации от человека к системе. Базовыми являются системы речевого и жестового управления. Применение и распознавание речи для ввода информации особенно удобны, когда руки оператора заняты и управление осуществляется в темное время суток. Применение жестов удобно, когда при хорошей видимости речевая деятельность человека ограничена. Каждая из таких модальностей имеет свои особенности, связанные с форматами представления данных и методами обработки. Аналогично система управления жестами может располагать методами побуквенного распознавания жеста, например на основе русской ручной азбуки жестов и специальных жестов, передаваемых одной рукой. При этом должны быть определены способы обучения классификатора.

Автономные интеллектуальные роботизированные системы, как правило, должны выполнять целевые функции в условиях неопределенности, существенных внешних возмущений и наличия динамических препятствий. К актуальным задачам, решаемым БПЛА в проблемных средах, относятся мониторинг территорий на основе систем технического зрения, выполнение механических операций в опасных зонах, обеспечение мобильной связи и образование формаций для выполнения групповых целевых операций, доставка грузов в городской инфраструктуре, и различные задачи, решаемые в интересах развития экономики России.

Одной из составляющих повышения надежности управления и автономности БПЛА служит внедрение источников мультимодальной слабоструктурированной информации. Сенсорные, речевые и графические каналы информации дополняют друг друга, расширяют возможности систем управления и принятия оперативных решений в условиях естественных и искусственных помех.

Настоящая работа посвящена созданию концепции и практических решений по реализации

мультимодального интерфейса применительно к системам управления наиболее распространенных малых БПЛА посредством голосового, жестового и сенсорного ввода команд.

Применяется мультимодальный алгоритм комплексирования анализа сцен, оптимизированный для обнаружения препятствий в сельском хозяйстве с движущимися наземными транспортными средствами. Метод опирается на обработку облаков точек и ориентирован на использование пространственных, временных и мультимодальных связей между соответствующими 2D- и 3D-областями.

Важнейшей задачей в понимании сцены являются обнаружение и локализация трехмерных ограничивающих рамок объектов, принадлежащих к определенным классам. Существующие детекторы трехмерных объектов во время обучения в значительной степени полагаются на аннотированные трехмерные ограничивающие рамки, в то время как получение этих аннотаций может быть дорогостоящим и доступным только в ограниченных сценариях. Слабо контролируемое обучение (weakly supervised learning) — многообещающий подход к сокращению требований к аннотации, но существующие детекторы в основном предназначены для 2D-обнаружения, а не для 3D. В [9] предложен фреймворк VS3D со слабым контролем для обнаружения трехмерных объектов из облаков точек без использования каких-либо готовых трехмерных ограничивающих рамок для обучения. Описан неконтролируемый модуль трехмерных предположений, генерирующий предположения об объектах для составления обучающей выборки, используя нормализованные плотности облаков точек. Представлена кросс-модальная стратегия дистилляции знаний, в которой сверточная нейронная сеть учится предсказывать окончательные результаты из предложенных трехмерных объектов, используя сеть-учитель, предварительно обученную на датасете. Рассмотрены и исследованы методы комплексирования данных сенсоров видимого диапазона света с данными, полученными от 2D- или 3D-сенсоров иной природы (локационных, инфракрасных). Использование множества различных сенсоров для наблюдения за ситуацией облегчает задачу работы с объектами еще и в том смысле, что менее критичным становится ракурс, под которым объект попал в кадр; комплексирование позволяет собрать больше информации о наблюдаемом объекте в задачах слежения и мониторинга, в том числе помогает в решении задачи мониторинга структурных особенностей объектов; способствует выявлению рисков разрушения; помогает в обеспечении безопасности на предприятиях.

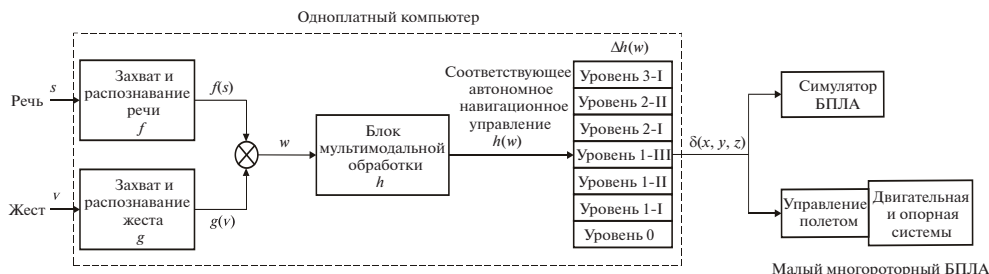


Рис. 1. Архитектура мультимодального управления БПЛА.

Метод комплексирования видеосъемки и данных от радара используется в задаче обнаружения объектов в процессе движения автомобиля с использованием нейронной сети Camera-Radar Fusion Net (CRF-Net), которая описана в [10]. Предложенный подход улучшает существующие сети обнаружения 2D-объектов за счет объединения данных камеры и проецируемых разреженных данных от радара. CRF-Net автоматически узнает, на каком уровне слияние данных от сенсоров наиболее выгодно для эффективного обнаружения.

Существующие методы комплексирования обычно полагаются на датчики глубины (например, лидары) для обнаружения и отслеживания целей в трехмерном пространстве, но только до ограниченного диапазона чувствительности из-за разреженности принимаемого сигнала. С другой стороны, камеры обеспечивают плотный и богатый визуальный сигнал, который помогает локализовать даже далекие объекты в домене изображения. Мультиобъектный трекинг (Multi-Object Tracking, MOT) позволяет мобильным роботам выполнять хорошо обоснованное планирование движения и навигацию, используя локализацию окружающих объектов в трехмерном пространстве и времени. В [11] предложен фреймворк EagerMOT, использующий для решения задачи MOT все доступные наблюдения от мультимодальных датчиков, что позволяет получить хорошо информированную интерпретацию динамики сцены.

Задача построения мультимодальных интерфейсов робототехнических систем рассмотрена в [12]. В [13] описана модель мультимодального управления БПЛА на основе голосовых и жестовых команд в формате mSVG. На рис. 1 представлена архитектура соответствующей системы управления БПЛА, выполняющего функции патрулирования, поиска и спасения людей.

В математической модели взаимодействия оператора с дроном речевой и жестовой информации поставлены в соответствие переменные s и v , а функциями обработки служат $f(s)$ и $g(v)$. Интегрированная команда управления w

проходит через модуль мультимодальной обработки, в котором генерируется управляющий сигнал $h(w)$. В свою очередь $h(w)$ подается в автономную модель для анализа величины отклонения $\Delta h(w)$ и выработки значения новых координат $\delta(x, y, z) = \Delta h(w)$. Значение $\delta(x, y, z)$ подается одновременно на контроллер и имитационную модель, представленную в виде динамического 3D-симулятора Gazebo в операционной системе Robot Operating System – ROS малого БПЛА для изменения координат в пространстве относительно его текущего положения. Речевая команда поступает через микрофон, обрабатывается и распознается с помощью инструментария CMU Sphinx ASR с пользовательским фонетическим словарем, содержащем набор командной лексики. Визуальный жест захватывает камера, подключенная к бортовому компьютеру. Распознавание выполняется с помощью комбинации алгоритма каскада Хаара (Haar Cascade classifier), основанного на машинном обучении, для отслеживания движения рук и алгоритма поиска выпуклой оболочки для подсчета количества пальцев на изображении.

Система управления БПЛА, построенная на голосовых командах, представлена в [14]. С помощью рекуррентных сенсорных сетей осуществлен семантический разбор речевых команд для дальнейшего преобразования в инструкции БПЛА. В другом подходе [15] предложено использовать анализ жестов совместно с управлением БПЛА через смартфон (пульт дистанционного управления) применительно к спасательным операциям, когда связь с оператором может быть нарушена или потеряна вовсе. БПЛА должен уметь распознавать команды человека, которого необходимо спасти. В ходе проведенных экспериментов длительностью полета 3 ч была достигнута точность распознавания команд 99.5%. При этом безопасность полета и управления осуществляется путем введения дополнительных ограничений, например виртуального ограждения допустимой зоны полета, ограничения по скорости, ускорения, высоты полета, сценариев потери данных и др.

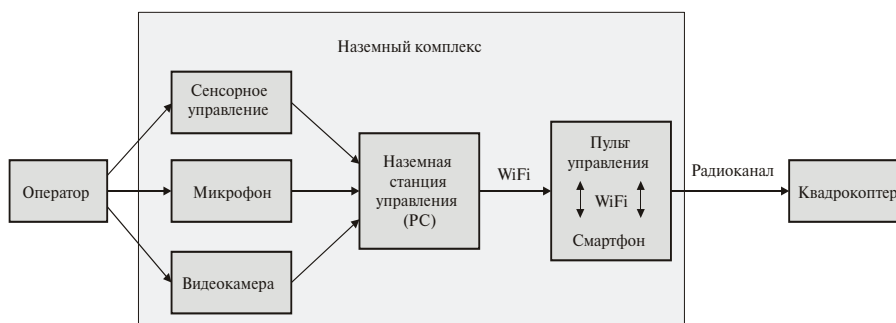


Рис. 2. Архитектура мультимодального интерфейса для управления БПЛА.

Исследования мультимодального интерфейса, который сочетает речевые команды и жесты, приведены в [16], предлагают систему управления БПЛА, в которой основным каналом являются жестовые команды, описывающие общие формы сегментов траектории, а речевыми командами дополняются геометрические параметры (например, количество метров, которые надо пролететь, или радиус окружности траектории). Для ввода жестовой информации исследователи используют специальное устройство захвата и анализа движений руки оператора.

Прототип мультимодального интерактивного квадрокоптера (робота), предложенный в [17], способен летать и реагировать на действия человека без использования внешних устройств. БПЛА оснащен двумя бортовыми камерами и микрофоном. Перчатки особого цвета помогают БПЛА найти области рук. Для обнаружения голосовых команд использован механизм Julius [18]. Используя функцию распознавания голоса, квадрокоптер поднимается на определенную высоту. Когда робот обнаруживает руки в перчатках, он следует указанному направлению, командой посадки является свист.

Взаимодействие с роботом на разных уровнях абстракции (задача, активность, путь, траектория, движение) описано в [19]. Команды, поступающие по обоим каналам управления, сначала классифицируются по отдельности. Затем результаты голосового и жестового распознавания объединяются и интерпретируются. При этом используются контекстуальные правила для устранения неоднозначности конфликтующих команд и объединения входных сигналов разной модальности. Оператор может взаимодействовать с дронами на различных уровнях абстракции с использованием жестовых метафор [20].

Решение, предложенное в [21], включает в себя носимые датчики и встроенные элементы искусственного интеллекта. Используется микрофон и инерциальный измерительный блок для “захвата” человеческого голоса и движений рук. Первый тип управления осуществляется блоком

распознавания голоса на основе рекуррентной нейронной сети (Recurrent Neural Network — RNN). Второй тип управления реализуется системой распознавания жестов на основе сверточной нейронной сети (Convolutional Neural Network — CNN). Графический блок обработки способен запускать предварительно обученную нейронную сеть на борту БПЛА. Вся система разделена на две части, отвечающие за “слабый” и “сильный” интеллект. На борту БПЛА установлена первая часть системы, в которую загружены предварительно обученные искусственные нейронные сети (ИНС). “Сильный” интеллект основан на CNN и RNN и развернут на настольном ПК/ноутбуке на земле. Человеко-машинный интерфейс состоит из четырех модулей: надеваемого на кисть устройства, процессора, пульта и дрона. Надеваемый блок на базе устройства Sensor-Tile выполняет распознавание голоса и жестов пользователя и позволяет объединять данные, полученные от бортовых датчиков.

В [22] предложена система распознавания жестов посредством трехмерного сенсора Kinect, анализирующего глубину сцены. Для управления широко распространенных недорогих дронов серии DJI Tello предлагается использовать жестовое управление через бортовую камеру [23], либо с помощью распознавания поз тела [24, 25]. Нередко в подобных работах используется аппарат ИНС сверточного типа [26].

Архитектура мультимодального интерфейса на основе видеокамеры, сенсорного управления, системы потоковой обработки речевых команд и жестов рук, разработанная в ИПС им. А.К. Айла-мазяна РАН, представлена на рис. 2 [27]. Платформой служит БПЛА типа “квадрокоптер” модели DJI Phantom 3 [28].

Как видно из рис. 2, оба потока входных управляющих команд обрабатываются на наземной станции, после чего преобразуются в понятные для БПЛА команды. Полученные команды пересылаются через смартфон и пульт в микроконтроллер квадрокоптера. В силу технических особенностей БПЛА пульт управления в режиме

Таблица 1. Точность распознавания речевых команд (формантный подход), %

Способ	Команды				
	Прямо	Вернись	Стоять	Направо	Налево
Тихо	43	78	57	87	89
Громко	92	94	83	96	97
Быстро	82	93	81	96	95
Медленно	90	97	79	93	94

удаленных команд выполняет роль роутера: он транслирует WiFi для смартфона и для наземной станции управления, объединяя их в локальную сеть. Инструкции для БПЛА, сформированные наземной станцией после обработки голосовых и жестовых команд оператора, поступают в специализированную программу на смартфоне, который посылает команды через пульт и далее радиоканал. Полетный план (маршрут) БПЛА прорабатывается заранее и задается в виде координат в конфигурационном файле на сервере. Маршрут задается последовательностью опорных точек.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Задача построения интегрированного интерфейса и интеллектуальной системы управления заключается в моделировании и реализации правил передачи командной информации от человека к системе. Необходимо вложить в искусственный интеллект способность распознавать вид передаваемой информации с подтверждением об успехе или без. При этом должен быть определен способ проведения обучения классификатора. Целесообразно изначально заложить базовый набор форм общения (команд), а затем проводить на их основе коррекцию обучения, однако любой такой базовый набор всегда будет накладывать свои ограничения.

Рассмотрим систему речевого управления робототехнической системой.

Формирование и распознавание речевого сигнала

Форматная обработка речевой команды. Имеем следующую последовательность операций.

1. Чтение команды с микрофона или из базы команд (формирование речевого сигнала, который может быть оцифрован и визуализирован).

2. Вычисление дискретного вейвлет-преобразования сигнала. Разбиение сигнала на форманты. В полученном преобразованном сигнале согласно участкам стабильности определяется количество формант (фонемоподобных единиц) и происходит их выделение.

3. Сравнение с базой данных команд, хранящей форманты (полученные форманты сверяют-

ся с эталонной фонемной базой, и каждой форманте приводится в соответствие наиболее близкая по коэффициентам фонема).

4. Распознавание на основе дискретной цепи Маркова (набор фонем сравнивается базой эталонных команд. Приводится формантный анализ в и подбор соответствия наиболее близкой эталонной фонеме, подсчитывается итоговый коэффициент отклонения распознаваемого сигнала от эталона. На основе этого коэффициента принимается решение о команде и коэффициенте уверенности распознавания команды).

5. Вывод распознанной команды (на экран выводится сигнал с формантной разметкой, соответствующей команде, приведенной в соответствие распознаваемому сигналу и степень достоверности).

Разработана программа, которая имеет дружелюбный интерфейс и легка в освоении. На рис. 3 показан пример визуализации исследуемого речевого сигнала в окне графического интерфейса.

На двух диаграммах (рис. 4, 5) показано в качестве примера, как два диктора произносят одну и ту же команду “налево”.

Команды произносятся дикторами по-разному. Соответственно их сигналы дают различное отклонение от эталона, однако оба отклонения укладываются в доверительный интервал. Экспериментально показано, что программа распознавания достаточно хорошо справляется с распознаванием ограниченного числа команд при уверенном произнесении их различными дикторами, однако возможны и ошибки распознавания. Результаты исследования с процентами успешного распознавания представлены в табл. 1. Для улучшения качества можно использовать дополнительные методы классификации сигналов (например, ввести модель скрытых Марковских процессов).

Обработка речевой команды без разбиения на форманты. В этом случае имеем следующую последовательность операций.

1. Чтение команды и ее преобразование.

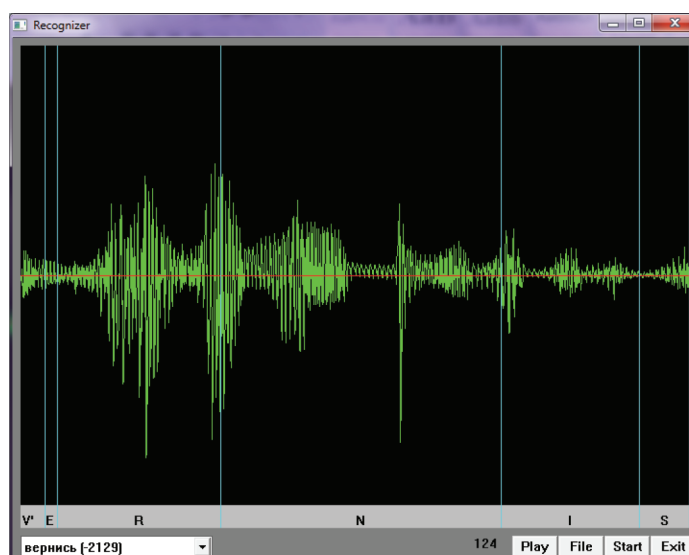


Рис. 3. Пример речевой команды.

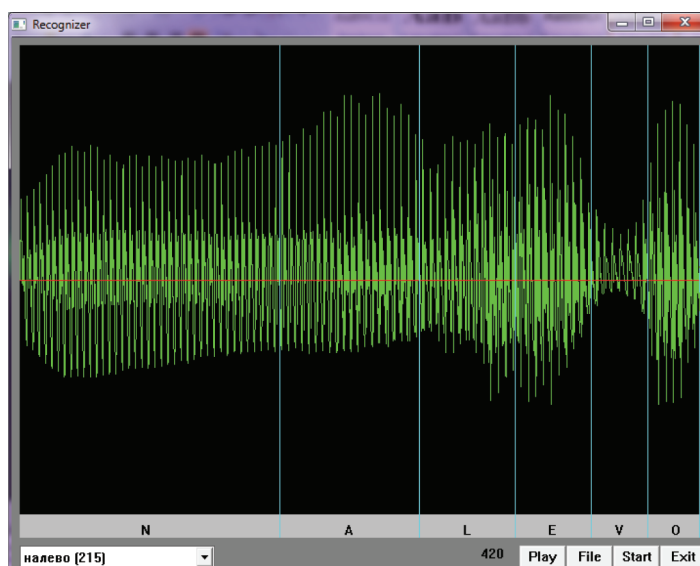


Рис. 4. Команда диктора 1.

2. Предварительная обработка (удаление постоянной составляющей, фильтрация, удаление тишины, выделение границ речевой команды).

3. Выделение признаков (вычисление кепстральных коэффициентов и их отправка в выходной канал).

4. Распознавание речевой команды (загрузка весовых коэффициентов нейронной сети, загрузка выделенных признаков; на выходе сети — идентификатор распознанной команды).

5. Вывод распознанной команды (на экран выводится сигнал с формантной разметкой, соответствующей команде, приведенной в соответ-

ствии распознаваемому сигналу и степень достоверности). Для чистоты эксперимента команды отдавались разными дикторами. Результаты экспериментов представлены в табл. 2. На основе экспериментов составлена результирующая таблица качества распознавания команд управления БПЛА (табл. 3).

Формирование и распознавание жестового сигнала

Распознавание специальных жестов управления

1. Формирование сигнала трехмерным сенсором.

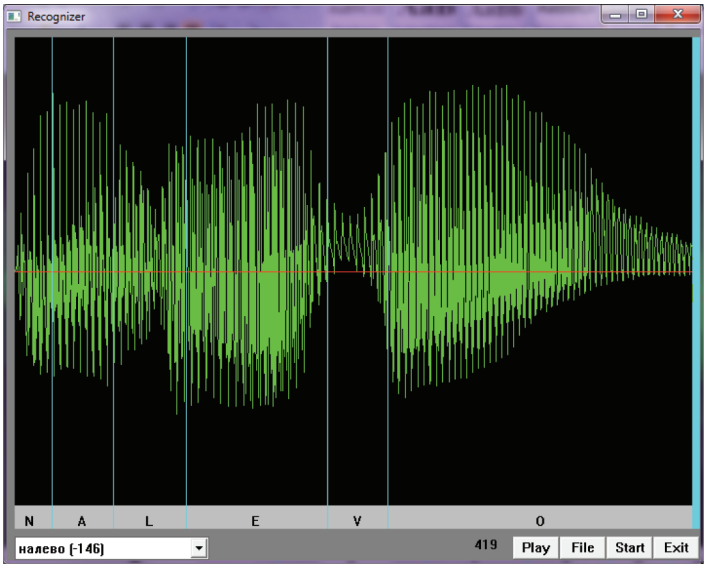


Рис. 5. Команда диктора 2.

2. Покадровая обработка видеоряда, построение дальностного изображения, выделение изображения руки (ладони) (дальностное изображение можно представить множеством точек и расстояний до сенсора $\{(x, y, d(x, y))\}$, являющимся моделью наблюдаемого объекта.)
3. Сглаживание изображения. Распознавание позиции ладони руки посредством обработки дальностного изображения. После извлечения изображения руки оно преобразовывается в бинарное изображение. За центр ладони принимается центр окружности, вписанной в фигуру руки и имеющей максимальный радиус.
4. Часть динамических жестов можно распознать посредством анализа изменений позиции центра ладони по трем осям пространства, зная

текущую позицию центра ладони (x, y, z) в каждом кадре видеоряда и начальную позицию центра ладони (x_0, y_0, z_0) . Так, например, если $z - z_0 < 0$, что соответствует приближению ладони руки к сенсору, то выполняется команда “вперед”. Другая часть жестов распознается посредством вычисления моментов бинарного изображения. Поворот ладони руки по часовой стрелке приводит к увеличению угла θ , поворот против часовой стрелки – к уменьшению. Таким образом, обнаружение характера изменений угла θ в каждом кадре видеоряда приводит к распознаванию жестов поворота “вправо” и “влево”. Для распознавания статических жестов вычисляются площади S_1 и S_2 – соответственно описанного и вписанного многоугольников бинарного изобра-

Таблица 2. Точность распознавания речевых команд (неформантный подход), %

	Диктор					
	1	2	3	4	5	6
Копировать	99.2	96	99.2	98.4	100	97.6
Вставить	100	100	100	100	100	100
Да	97.6	88	100	99.2	99.2	100
Нет	92.8	100	99.2	100	89.6	98.4
Отмена	99.2	99.2	100	100	95.2	100
Вперед	98.4	96	100	100	100	100
Назад	96	100	100	100	100	100
Удалить	99.2	100	100	100	97.6	100
Старт	100	100	100	100	99.2	96
Стоп	96.8	100	100	98.4	97.6	100
Средняя точность	97.92	97.92	99.84	99.6	97.84	99.2

Таблица 3. Точность распознавания команд управления

Команды	Способ подачи команды	
	Речь-форманты	Речь-слитно
Старт		1.00
Прямо (вперед)	0.92	0.98
Назад	0.94	0.96
Стоп	0.83	0.97
Налево	0.97	0.99
Направо	0.96	0.99

жения руки. Если значение $|S_1 - S_2|$ меньше порогового значения $\bar{S}$, то ладонь руки считается согнутой в кулак, в противном случае — открытой, что соответствует командам “стоп” и “старт”.

5. Вывод распознанной команды в систему управления. Для увеличения надежности распознавания возможно построение комитета классификаторов, в том числе на основе нейронных сетей (табл. 4).

Распознавание жеста. Геометрический скелет руки является дескриптором, который может быть использован для идентификации формы руки во множестве конфигураций. Идентификация формы осуществляется путем сравнения скелета руки с эталонными. Для оценки степени схожести выполняется развертка скелетов по специальному алгоритму после развертки расстояние между скелетами оценивается за полиномиальное время с помощью алгоритма динамической

трансформации шкалы времени (Dynamic Time Warping — DTW) [29].

Средняя точность распознавания речевых команд по всем дикторам составила 98.72% (табл. 5).

В качестве альтернативного подхода активно применялись ИНС. Обработка жестов проводилась посредством модифицированной библиотеки YOLO (ИНС “Darknet”). Для исследования звуковых команд подход заключался в использовании библиотек denoiser [30, 31], pydub [32].

В данном подходе обработка звукового потока была разделена на два этапа. Первый этап — захват звука с микрофона, удаление шумов, извлечение отдельных слов из потока. Второй этап — распознавание полученных слов с помощью сверточной нейронной сети.

В соответствии с разработанной системой команд БПЛА, некоторые речевые команды являются составными. Поэтому полнота, нормализованные точность и F1-мера этих команд зависят от соответствующих значений распознавания, входящих в них слов. В табл. 6 приведена интегрированная оценка F1-меры для различных команд, полученная в ходе серии экспериментов с жестовыми и речевыми командами.

Полученные результаты исследований показывают, что выбранные методы и программные средства позволяют распознавать речевые и жестовые команды, входящие в разрабатываемую систему управления БПЛА. Дальнейшие исследования предполагают расширение базы снимков и более объемные тесты.

Таблица 4. Жесты руки, соответствующие базовым командам управления

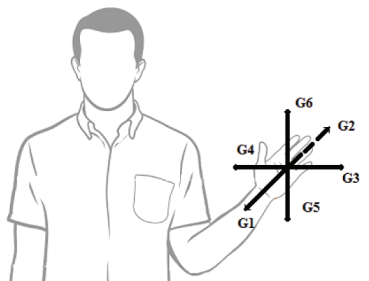
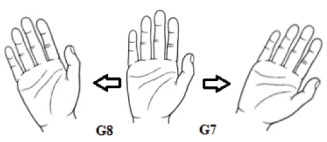
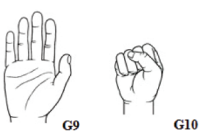
Жест	Обозначение	Базовая команда
	G1	Вперед
	G2	Назад
	G3	Влево
	G4	Вправо
	G5	Вниз
	G6	Вверх
	G7	Поворот вокруг своей оси против часовой стрелки
	G8	Поворот вокруг своей оси по часовой стрелке
	G9	Старт
	G10	Посадка

Таблица 5. Характеристики качества распознавания жестов руки

Характеристики качества распознавания										
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
Точность	0.99	1	1	1	0.98	0.99	0.98	0.98	0.9	0.92
Полнота	0.98	0.99	1	1	0.99	1	0.96	0.95	0.92	0.95

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

В работе предложено создание комбинированной системы управления робототехническими системами, содержащей мультимодальный интерфейс, основанный на использовании речевых и жестовых команд. Как показывают результаты экспериментов, разные подходы обеспечивают примерно равное качество распознавания команд. Результаты распознавания на обучающей и тестовой выборках после обучения сети следующие: точность и полнота распознавания — свыше 99%, скорость распознавания жестов — 38 кадров/с на современном персональном компьютере.

Увеличение точности и соответственно надежности системы может быть обеспечено при параллельном вводе команды разными способами с применением комитетных методов принятия решений.

Для принятия оперативных решений в условиях меняющейся обстановки предлагается использовать интеллектуальные динамические системы, основанные на правилах [33]. При данном подходе описания текущей ситуации являются предположениями некоторой базы правил, которые могут быть выполнимыми или невыполнимыми в заданной интерпретации (в текущий момент времени). Зная описание некоторого целевого состояния, с помощью алгоритмов и методов интеллектуального планирования строится такая последовательность применимых правил, которые приводят к достижению цели. Первое правило последовательности и является оперативным решением, которое необходимо принять в текущий момент времени. В дальнейшем возможно внесе-

ние изменений и уточнений в построенный план на основе полученной информации.

Для построения эффективных оперативных решений на базе БПЛА предлагается использовать следующие методы планирования:

- иерархическое планирование с декомпозицией задач и их распределением по всем уровням подведомственной иерархии,
- планирование с использованием прецедентов для парирования нестандартных ситуаций,
- эвристическое планирование для повышения быстродействия алгоритмов построения решений,
- алгоритмы планирования с отсечением по времени для построения планов, достижения цели при существенных временных ограничениях на работу планировщиков.

Используемые при построении планов базы правил могут быть пополнены с помощью экспертов или с помощью обучения (с учителем, без учителя, по прецедентам).

На основе выполненных исследований разработана программная система, которая интегрирует предлагаемые технологии управления БПЛА [34].

В общем случае для ускорения процессов решений в рамках разрабатываемых технологий целесообразно использовать аппаратную платформу многоядерных персональных компьютеров и суперкомпьютеров кластерного уровня, включающую управляющую ЭВМ, оснащенную стандартными средствами, необходимыми для работы оператора (монитор, клавиатура); вычислительные узлы кластерного уровня (базовые вычислительные модули кластерного уровня); системную сеть кластера (SCI), объединяющую вычислительные узлы; вспомогательную сеть (семейства Ethernet, с поддержкой TCP/IP), объединяющую управляющую ЭВМ и вычислительные узлы [35].

ЗАКЛЮЧЕНИЕ

Выполнены исследования в области построения мультимодального управления БПЛА и интеллектуального интерфейса, использующего пультовые, речевые и жестовые команды. Разработана и реализована система интеллектуального управления применительно к квадрокоптеру. Проведены эксперименты, показавшие высокую

Таблица 6. Статистические данные по командам

Команда	F1-мера распознавания речевой команды, %	F1-мера распознавания жестовой команды, %
Взлет	98.8	99.7
Домой	97.6	99.6
Остановка	100.0	99.7
Перемещение	96.2	99.4
Поворот	94.6	99.6
приземление	100.0	99.6

точность и полноту распознавания голосовых команд и жестов с применением ИНС сверточного типа. Совместное использование нескольких разнородных источников данных позволяет повысить точность и эффективность распознавания команд. Целесообразна раздельная фильтрация потоков данных от каждого из датчиков с последующим объединением с весами, зависящими от точности получения параметров. Приоритет при этом дается данным, полученным от датчиков с минимальной погрешностью оценки. Фактически этот подход определяет стратегию построения интеллектуального человеко-машинного интерфейса. Подобная концепция не противоречит общему подходу к организации систем обработки данных от многих датчиков, которая повсеместно применяется в авиационных и космических системах.

Ведется разработка комплексной технологии обнаружения, распознавания динамических целей, основанной на обработке аэрокосмических снимков и данных систем позиционирования с применением БПЛА. Проведены экспериментальные исследования и даны рекомендации по реализации интеллектуальной технологии. Планируются дальнейшая проработка и расширение системы команд, связанная с управлением несколькими БПЛА и командами приема-передачи управления в зоны чрезвычайных ситуаций.

Разработанные методы обработки изображений используются в учебном процессе Института комплексной безопасности РТУ МИРЭА на кафедрах КБ4, КБ14 и РУДН на кафедре “Математические методы моделирования и искусственный интеллект”.

СПИСОК ЛИТЕРАТУРЫ

1. Котомин А.В. // Информационно-телекоммуникационные технологии и матмоделирование. 2011. С. 135.
2. Мокров Е.В., Хачумов В.М., Теплоухова Н.Е. // Тезисы докладов IV Всероссийской научно-технической конференции “Актуальные проблемы ракетно-космического приборостроения и информационных технологий”. М.: Радиотехника, 2011. С. 138.
3. Мокров Е.В., Хачумов В.М., Теплоухова Н.Е. // Труды IV Всероссийской научно-технической конференции “Актуальные проблемы ракетно-космического приборостроения и информационных технологий” / Под ред. Урличича Ю.М., Романова А.А. М.: Радиотехника, 2012. С. 418.
4. Нагапетян В.Э., Хачумов В.М. // Искусственный интеллект и принятие решений. 2013. № 3. С. 59.
5. Нагапетян В.Э., Хачумов В.М. // Автометрия. 2015. Т. 51. № 2. С. 103.
6. Нагапетян В.Э., Хачумов В.М. // Сб. докладов 9-ой международной конференции “Интеллектуализация обработки информации” (Черногория, Будва). М.: Издат-во Торус Пресс, 2012. С. 445.
7. Нагапетян В.Э., Хачумов В.М. // Искусственный интеллект и принятие решений. 2015. № 1. С. 57.
8. Харламов А.А. Нейросетевой подход к интегрированному представлению и обработке информации в интеллектуальных системах. Автореферат диссертации на соискание ученой степени доктора технических наук. Москва 2008. 33 с.
9. Qin Z., Wang J., Lu Y. // Proc. 28th ACM Internat. Conf. Multimedia. 2020. P. 4144. <https://doi.org/10.1145/3394171.3413805>
10. Nobis F., Geisslinger M., Weber M. A Deep Learning-based Radar and Camera Sensor Fusion Architecture for Object Detection, 2020. <https://arxiv.org/pdf/2005.07431.pdf>
11. Kim A., Ošep A., Leal-Taixé L. EagerMOT: 3D Multi-Object Tracking via Sensor Fusion, 2021. <https://arxiv.org/pdf/2104.14682.pdf>
12. Andronas D., Apostolopoulos G., Fourtakas N., Makris S. // Proc. Manuf. 2021. V. 54. P. 197. <https://doi.org/10.1016/j.promfg.2021.07.030>
13. Abioye A.O., Prior S.D., Thomas G.T. et al. Towards Autonomous Robotic Systems. TAROS. Lecture Notes in Computer Science / Eds. Giuliani M., As-saf T., Giannaccini M. 2018. V. 10965. P. 423. https://doi.org/10.1007/978-3-319-96728-8_36
14. Oneata D., Cucu H.K. // Proc. Interspeech. 2019. P. 5. <https://doi.org/10.21437/Interspeech.2019-1390>
15. Chen B., Hua C., Li D. et al. // Appl. Sci. 2019. V. 9. P. 22. <https://doi.org/10.3390/app9163277>
16. Chandarana M., Meszaros E.L., Trujillo A., Danette A.B. // Proc. Human Factors and Ergonomics Society Annual Meeting. 2017. V. 1. № 1. P. 68. <https://doi.org/10.1177/1541931213601483>
17. Miyoshi K., Konomura R., Hori K. // Proc. 28th British HCI Conf. 2014. P. 218.
18. Chowdhary K.R. // Fund. Artif. Intel. 2020. P. 651. https://doi.org/10.1007/978-81-322-3972-7_20
19. Rossi S., Leone E., Fiore M. et al. // Proc. IROS. 2013. P. 2208. <https://doi.org/10.1109/IROS.2013.6696665>
20. Pfeil K., Koh S.L., Jr. J.J.L. // Proc. IUI. 2013. P. 257. <https://doi.org/10.1145/2449396.2449429>
21. Menshchikov A., Ermilov D., Dranitsky I. et al. // Proc. 45th Annual Conference of the IEEE Industrial Electronics Society (IECON). 2019. P. 8. <https://doi.org/10.1109/IECON.2019.8926635>
22. Nagapetyan V.E., Khachumov V.M. // Optoelectron. Instrum. Data Process. 2015. V. 51. № 2. P. 192.
23. Niu R.G.C. UAV Gesture Control. 2020. https://github.com/RobertGCNiu/UAV-Gesture-Control_Python
24. Baldini F., Anandkumar A., Murray R.M. // Proc. American Control Conference (ACC). 2020. P. 2961. <https://doi.org/10.23919/ACC45564.2020.9147400>
25. Albanis G., Zioulis N., Dimou A. et al. Photorealistic UAV-Assistant Dataset Synthesis for 3D Pose Estimation via a Smooth Silhouette Loss. 2020. 45 p. <https://arxiv.org/abs/2008.08823>

26. *Perera A.G., Law Y.W., Chahl J.* UAV-GESTURE: A Dataset for UAV Control and Gesture Recognition. 2019. 12 p. <https://arxiv.org/abs/1901.02602v1>.
27. *Абрамов Н.С., Талалаев А.А., Фраленко В.П.* Мульти模альный интерфейс для управления малыми беспилотными аппаратами v.0.5 // Свидетельство о государственной регистрации программы для ЭВМ № 2024662861, дата поступления заявки 31.05.2024, дата регистрации: 31.05.2024.
28. Phantom 3 standard. 2022. <https://www.dji.com/ru/phantom-3-standard>
29. *Müller M.* Dynamic time warping, 2007. https://doi.org/10.1007/978-3-540-74048-3_4
30. facebookresearch/denoiser: Real Time Speech Enhancement in the Wave-form Domain (Interspeech 2020), 2022. <https://github.com/facebookresearch/denoiser>
31. *Defossez A., Synnaeve G., Adi Y.* // Proc. Interspeech. 2020. P. 3291. <https://doi.org/10.21437/Interspeech.2020-2409>
32. jiaaro/pydub: Manipulate audio with a simple and easy high level interface. 2021. <https://github.com/jiaaro/pydub>
33. *Осипов Г.С.* Методы искусственного интеллекта. Физматлит, 2011. 296 с.
34. *Абрамов Н.С., Талалаев А.А., Фраленко В.П.* Мульти模альный интерфейс для управления малыми беспилотными аппаратами v.0.5 // Свидетельство о государственной регистрации программы для ЭВМ № 2024662861, дата поступления заявки 31.05.2024, дата регистрации: 31.05.2024.
35. *Абрамов С.М.* // Пути ученого. Е.П. Велихов / Под общ. ред. ак. РАН Смирнова В.П. М.: РНЦ “Курчатовский институт”, С. 325. http://omega.sp.su.ru/books/conference/PaVT2009/papers/full_papers/001.pdf