

Комбинированная система бесконтактного управления роботизированными системами на основе речевых и жестовых команд¹

Аннотация. Рассматривается задача построения универсальной интегрированной интеллектуальной системы и ее интерфейса для речевого и жестового управления робототехническими системами, включая транспортные средства и летательные аппараты. Выполнен обзор ряда зарубежных публикаций, в которых представлены комбинированные системы управления и достигнутые результаты. Предлагается подход к построению отечественной комбинированной системы управления. Система сопряжена с базой знаний и данных, использует элементы искусственного интеллекта и опирается на общность методов ввода, обработки и передачи информации разной природы.

Ключевые слова: команда, интерфейс, интеграция, речь, фонема, жест, изображение, развертка, информативные параметры, распознавание, управление.

Введение

В современных системах управления робототехническими устройствами используются разнообразные средства ввода информации, предоставляющие данные в определенных форматах по отдельным каналам. К наиболее распространенным способам, помимо традиционных систем ввода информации с клавиатуры, мыши и манипуляторов, относятся бесконтактные системы голосового (речевого) управления и управления с помощью жестов. Речевая подача команд особенно удобна, когда руки человека заняты. Жесты удобны, когда человек стеснен или ограничен в произнесении речи.

Следует выделить работу [1], в которой представлен ретроспективный обзор методов и систем, ориентированных на управление жестами. Показано, что подобный способ применим во многих областях и рекомендуется для пожилых пользователей, которые плохо знакомы со сложными устройствами ввода, или людей с ограниченными возможностями. Иссле-

дованы разные типы жестов, проанализированы приложения, технологии и интерфейсы. В работе [2] представлено решение для интеграции промышленного робота (ABB IRB140) с системами автоматического распознавания речи и компьютерного зрения. Робот предназначен для решения задач манипулирования с объектами, размещенными в случайном порядке на столе.

Компьютерная система должна распознавать их характеристики (форму, размер, цвет, положение и ориентацию) и понимать человеческую речь в качестве команд для манипулирования.

В последнее время за рубежом опубликованы работы, посвященные построению систем управления, использующих комбинированный принцип управления на основе звука и жестов. Интеграция нескольких способов подачи команд целесообразна, когда требуется достижение высокой надежности управления в условиях помех или ограничений возможности человека. В работе [3] представлена система реального времени для управления небольшим

¹ Работа выполняется в рамках проекта РФФИ № 13-07-00025 А и проекта Программы фундаментальных исследований ОНИТ 1 РАН «Интеллектуальные информационные технологии, системный анализ и автоматизация».

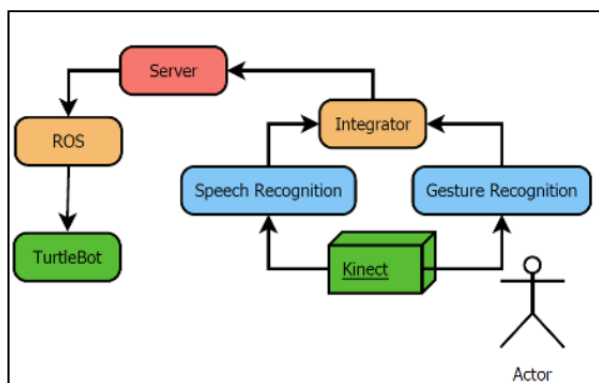


Рис. 1. Архитектура комбинированной системы

мобильным роботом с использованием аудио и видеоканалов. Архитектура системы представлена на Рис. 1.

Каждый способ ввода команды обладает своими особенностями, связанными с форматами представления данных и методами обработки. В качестве входного устройства в работе используется сенсор глубины Kinect, который позволяет получить видеопотоки из цветных и дальностных изображений. Одновременно с этим в систему подаются звуковые сигналы с помощью четырех встроенных микрофонов. Распознавание фиксированного числа команд осуществляется с применением метода динамического искажения времени DTW (Dynamic time warping). В основе распознавания жеста лежит анализ скелета человека [4]. Отдельные команды передаются в, так называемый, fusion-центр, который решает проблему противоречивых и взаимодополняющих инструкций.

Решение задачи построения комбинированных систем бесконтактного управления требует

построения специальных средств обработки. Так, в работе [5] дается описание робототехнической системы, управляемой дистанционно с помощью голоса, жестов и человеко-машинного интерфейса. Архитектура системы представлена на Рис. 2.

Система состоит из трех базовых программных компонентов, включая программу получения и распознавания команд управления из нескольких входов, клиент-сервер сети связи и программу смешения (комбинирования) команд с последующей ее обработкой роботом и его рукой. Человеко-машинный интерфейс обеспечивается применением проводных и беспроводных микрофонов, а также проводных датчиков ориентации, установленных на руках человека. Набор жестов основывается на командах, принятых на самолетах ВМС. Слияние и обработка нескольких команд выполняется на основе установленных приоритетов и анализе задержек времени от источника этих команд. В настоящей работе рассматриваются некоторые отечественные решения задачи управления голосом и жестами, предлагается подход, создающий предпосылки для построения комбинированных бесконтактных систем управления робототехническими системами, например, малыми беспилотными аппаратами.

1. Постановка задачи

Решение задачи комбинированного управления требует учета особенностей управления голосом и жестами и соответствующих методов распознавания команд. Так система речевого управления может использовать метод распо-

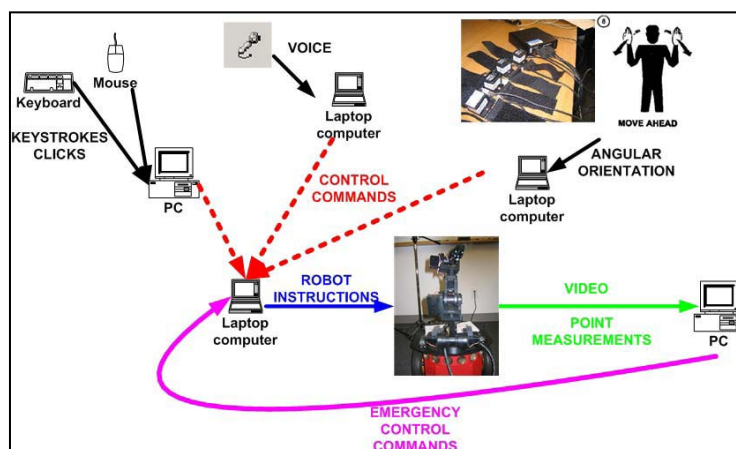


Рис. 2. Схема системы дистанционного управления с помощью звука и жеста

знания как всей команды в целом [6], так и ее формантного представления [7]. Аналогично, система управления жестами может располагать методами побуквенного распознавания жестов, например, на основе русской ручной азбуки (РРА) [8] или специального жеста, передающего команду единым знаком руки [9]. Для увеличения надежности бесконтактного управления необходимо построение человеко-машинного интерфейса, способного воспринимать независимо любой из перечисленных способов формирования и распознавания команды или одновременную подачу речевой и жестовой команды. 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 команд с весами, зависящими от точности получения параметров. Приоритет при этом дается тем данным, которые получены от датчиков с минимальной оценкой погрешности.

Перечисленные обстоятельства и возможности определяют стратегию и постановку задачи построения бесконтактного человеко-машинного интерфейса и системы, предназначенной для обработки слабоструктурированной информации, к которой, несомненно, относятся графические образы и звуковые сигналы. Подобная концепция не противоречит общему подходу к организации систем обработки данных от многих датчиков, применяющемуся в авиационных и космических системах. Идеи представления и кластеризации слабоструктурированной информации изложены, например, в работе [10]. Составными частями интегрированного подхода должны стать методы обработки текстовой информации, изображений и лингвистической информации. Для их реализации необходимы алгоритмы извлечения ключевой информации из данных различной природы в виде классифицирующих признаков. Построение классификаторов может основываться на использовании возможностей ассоциативной памяти и искусственных нейронных сетей (ИНС).

Задача интегрированного интерфейса и интеллектуальной системы управления, таким образом, заключается в реализации комбинированных команд передачи информации от человека к системе. Необходимо вложить в искусственный интеллект способность распознавать вид передаваемой информации с оценкой

степени уверенности. При этом должен быть определен вид классификатора и способ его обучения. Применительно к управлению робототехническими системами целесообразно изначально определить базовый набор команд и произвести обучение классификатора с учетом наложенных ограничений. Рассмотрим основные этапы построения и распознавания команд различной природы.

2. Формирование и распознавание речевого сигнала

Рассмотрим особенности и основные этапы обработки в канале речевого управления робототехнической системой [6, 7].

2.1. Форматная обработка речевой команды

При форматном подходе к распознаванию команд установлена следующую последовательность операций.

1. Чтение команды с микрофона или из базы команд (формирование речевого сигнала, который может быть оцифрован и визуализирован).
2. Предварительная обработка. Вычисление дискретного вейвлет-преобразования сигнала. Разбиение сигнала на форманты (в преобразованном сигнале происходит выделение и определяется количество формант). На Рис. 3 рассмотрен пример формантного представления речевой команды. Здесь вертикальными полосами выделены участки сигнала, соответствующие отдельным буквам команды.
3. Сравнение выделенных формант с базой данных команд (каждой ставится в соответствие наиболее близкая эталонная форманта).
4. Распознавание команды на основе дискретной цепи Маркова (проводится анализ и подсчитывается степень отклонения распознаваемого сигнала от эталона, принимается решение о команде и степени уверенности ее распознавания).
5. Вывод распознанной команды (визуализируется сигнал с формантной разметкой, соответствующей команде, и степень доверия).

2.2. Обработка речевой команды без разбиения на форманты

В этом случае имеем следующую последовательность операций.

1. Чтение команды и ее преобразование.

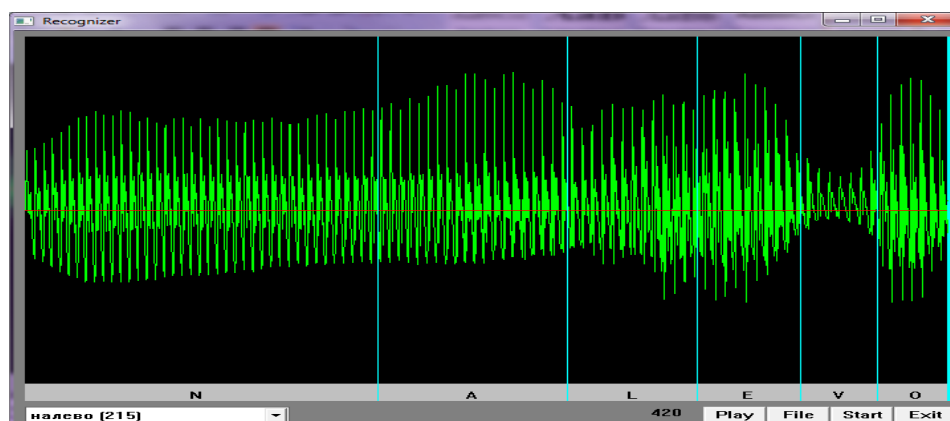


Рис. 3. Пример развертки речевой команды «налево»

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

3. Выделение информативных признаков (вычисление кепстральных коэффициентов).

4. Распознавание речевой команды (загрузка весовых коэффициентов нейронной сети и подача на вход выделенных признаков, на выходе сети – идентификатор распознанной команды).

5. Вывод распознанной команды (на экран выводится сигнал соответствующей команды и степень доверия).

3. Формирование и распознавание жестового сигнала

Особенности формирования и распознавания жестов были изучены в работах [8, 9].

3.1. Распознавание команды на основе РРА

Рассмотрим схему перевода жестового языка глухонемых в текстовый вид.

1. Подача на вход кадров видеоряда, представляющих дальностные изображения.

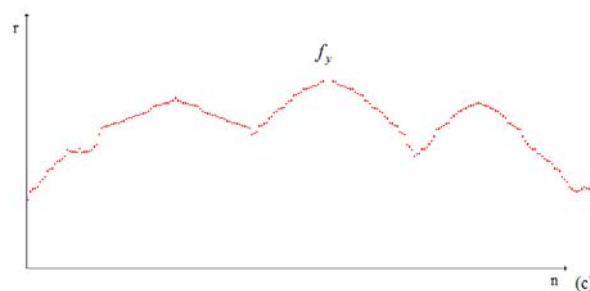
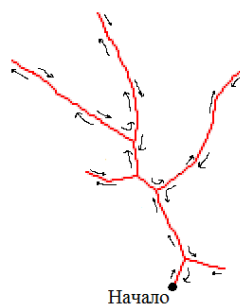
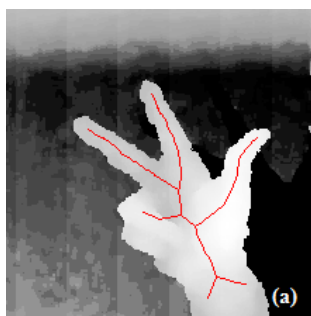


Рис. 4. Пример получения и развертки скелета руки по оси ординат (a) –дальностное изображение руки, (b) – скелет руки, (c) – развертка

вам Д, Ж, З, Й, Ц, Ч,Щ, Ъ, Ы, Ь, анализируются по траекториям движения ключевой точки руки.

5. Вывод новой буквы в текст, формирующий команду.

3.2. Распознавание специальных жестов управления

1. Формирование сигнала трехмерным сенсором.

2. Покадровая обработки видеоряда. Построение дальностного изображения, выделение изображения ладони, сглаживание.

3. Выделение информативных параметров в виде позиции ладони руки на основе обработки дальностного изображения и его преобразование в бинарное изображение. За центр ладони принимается центр окружности, вписанной в фигуру руки и имеющей максимальный радиус.

4. Распознавание жеста. Часть динамических жестов распознается путем анализа изменений позиции центра ладони по трем осям пространства. Для распознавания статических жестов вычисляются площади описанного и вписанного многоугольников бинарного изображения руки.

5. Вывод распознанной команды в систему управления.

4. Эксперименты по распознаванию речевых и жестовых команд

На основе анализа работ [6-9] можно составить результирующую таблицу качества распознавания речевых и жестовых команд (Табл. 1). В Табл. 2 представлены результаты распознавания жестов азбуки глухонемых РРА.

Здесь точность распознавания определяется как доля команд, действительно принадлежащих классу, относительно всех команд, которые были отнесены классификатором к этому классу. Формирование точного слова, передающего команду, только на основе распознавания букв (Табл. 2) не дает требуемой надежности из-за высокой вероятности ошибки распознавания

Табл. 1. Точность распознавания команд управления

	Способ подачи и обработки команды		
Команды:	Речь – форманты	Речь – слитно	Жест – команда
Старт	0.83	1.00	0.90
Прямо/вперед	0.92	0.98	0.99
Назад	0.94	0.96	1.00
Стоп	0.83	0.97	0.92
Налево	0.97	0.99	1.00
Направо	0.96	0.99	1.00
Поворот	-	-	0.98
Вверх/вниз	-	-	0.99

Примечание: прочерк означает отсутствие данных

отдельного символа. Проблема решается на основе сравнения полученных слов с командами, хранимыми в базе знаний. Текущей точности распознавания букв, в случае применения ограниченных словарей, вполне достаточно для уверенного распознавания команды.

В Табл. 3 для сравнения приведены результаты исследования качества распознавания подобных команд, полученные в зарубежной работе [2].

Как видно, результаты распознавания речевых и жестовых команд в экспериментах проведенных как в России, так и за рубежом, примерно одинаковы. Однако, в зависимости от наличия или отсутствия звуковых шумов или световых эффектов, могут оказаться эффективнее те или иные методы формирования команд или их комбинация.

Табл. 3. Точность распознавания команд управления

	Способ подачи и обработки команды	
Команды:	Речь – слитно	Жест – команда
Go	0.90	-
Forward	1.00	0.85
Backwards	0.95	0.90
Stop	0.75	0.80
Left	0.80	0.90
Right	1.00	1.00
Faster	0.80	0.95
Slower	0.70	0.95

Табл. 2. Точность распознавания жестов РРА

	Показатели качества распознавания																									
Буквы	А	Б	В	Г	Д	Е	Ж	З	И	К	Л	М	Н	О	П	Р	С	Т	У	Х	Ш	Ъ	Ы	Э	Ю	Я
Точность	1	1	.61	1	.61	.88	.9	.49	.61	0.8	.95	1	1	1	.91	.78	.78	.59	.74	.95	.43	.95	.71	1	1	1

5. Предложения по построению комбинированной системы управления и ее интерфейса

Общий анализ имеющейся информации показывает, что весь цикл интеллектуального управления, независимо от метода подачи команды, содержит пять основных последовательно выполняемых этапов. Процесс может быть унифицирован и представлен в виде конвейера, в котором происходят процессы передачи и преобразования информации. Общий вид системы интеллектуального управления представлен на Рис. 5.

Пункты 1 и 5 определяют способ формирования и передачи распознанной команды в робототехническое устройство. Передача, в зависимости от способа, может осуществляться как последовательно (поток), так и блоками данных. Поточковая передача – это непрерывный процесс продвижения и преобразования информации от датчика к приемнику. Сначала полученные от датчика данные записываются в буфер памяти ЦВМ, затем начинается их обработка. Размер буфера не должен быть слишком большим и не должен переполняться в случае значительного увеличения потока данных, поэтому скорость обработки данных должна быть достаточно высока. В блочном режиме информация передается порциями строго определенного размера. Датчик не начинает передачу следующего блока данных пока не получит подтверждения об успехе или запроса на передачу нового блока. Блочная передача данных характеризуется наличием буфера определенного размера и имеет менее строгие требования на скорость обработки, чем в первом случае.

Пункты 2-3, связанные с предварительной обработкой, включают методы выделения целевых объектов, устранения шумов, скелетизации, сегментации, вейвлет преобразования, вычисления инвариантов и коэффициентов, развертки скелетов, выделения других информативных признаков. Эти методы образуют базу методов предварительной обработки. В нашем случае единой моделью объекта распознавания служат развертка, называемая также «траекторией» в виде цифровой последовательности, физический смысл которой зависит от способа подачи команды. Траекторию формируют развертки речевого сигнала в виде последовательности фонем, коэффициентов ДПФ и кепстра, развертка изображения руки для специального жеста управления или последовательности букв РРА.

Пункт 4 служит непосредственно для распознавания. В него могут входить такие классификаторы, как нейронные сети, цепи Маркова, метрики для сравнения разверток и инвариантов, программно реализованные методы группового учета аргументов, опорных векторов и др. Есть смысл уменьшить число видов классификаторов в системе. Один из подходов заключается в замене всех классификаторов на универсальную нейронную сеть, в которой происходит только смена весовых коэффициентов в соответствии с типом решаемой задачи. В базу знаний, сопряженную с базой данных, следует поместить все проанализированные экспертами прецеденты, включая речевые команды, разложенные на фонемы, речевые корпуса команд, базу эталонных жестов РРА и их разверток, базу специальных жестов.

В настоящее время в свободном доступе имеются библиотеки и программные платфор-

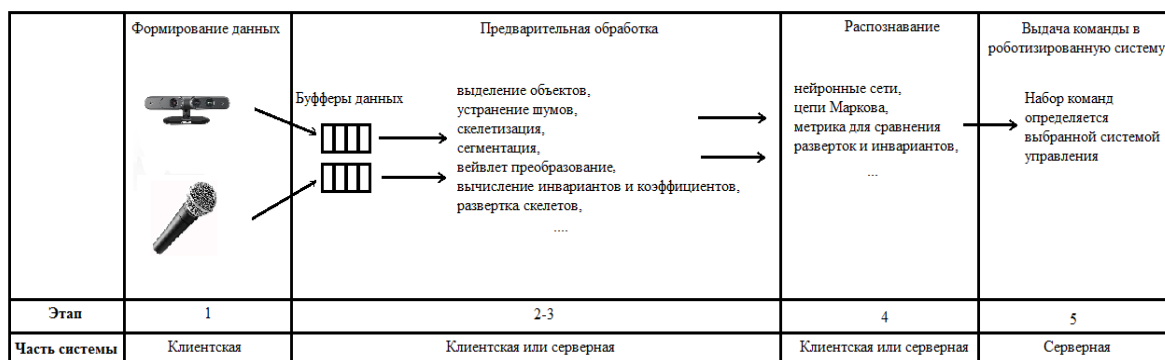


Рис. 5. Схема работы системы интеллектуального управления

мы, которые могут быть использованы для построения комбинированной системы управления. В то же время отдельные фазы обработки информации требуют применения новых подходов, еще не заложенных в общедоступное программное обеспечение. Для реализации комбинированной системы необходимы инструментальные средства, позволяющие упростить и ускорить процесс ее построения. В качестве таких инструментальных средств можно применить программный комплекс, описанный в работе [11]. Комплекс содержит развитый графический интерфейс для представления решаемой задачи в виде блочной схемы, базу данных и знаний с большим количеством модулей обработки информации (предобработчики, классификаторы), средства визуализации. Имеется вычислительное ядро, обеспечивающее выполнение схем алгоритмов в конвейерно-параллельном режиме в гетерогенной среде. На наш взгляд, целесообразно выделить часть технологической цепочки, которая работает непосредственно с датчиками, и реализовать ее в виде отдельного динамически подключаемого модуля.

Заключение

Настоящая работа содержит некоторые предложения по созданию комбинированной системы управления робототехническими системами, содержащей человеко-машинный интерфейс, основанный на использовании речевых и жестовых команд. В отличие от работы [3], где для распознавания жестов человека используется скелет тела человека, что требует расстояния от сенсора не менее 2-3 метров, в нашем случае используется только скелет руки. Это позволяет существенно уменьшить критическое расстояние между прибором и человеком. В работе [5] для распознавания жестов используются специальные сенсоры, прикрепленные к телу человека, что неудобно и не всегда допустимо. В нашем случае создается принципиальная возможность бесконтактного распознавания жестов руки в случаях, когда расстояние между сенсором и человеком составляет порядка 30 см. Проведенные экспериментальные исследования, в которых получены характеристики точности распознавания речевых и жестовых команд, показывают, что разные подходы обеспечивают примерно равное

качество распознавания команд. Увеличение точности распознавания и, соответственно, надежности системы может быть обеспечено при параллельном вводе команды разными способами и с применением комитетных методов принятия решений, в том числе построенных на основе нейронных сетей и специальных метрик. Применение готовых инструментальных средств для построения прикладных систем [11] существенно облегчает построение комбинированной системы бесконтактного управления роботизированными системами с нейросетевыми классификаторами.

Литература

1. Bhuiyan M., Picking R. Gesture-controlled user interfaces, what have we done and what's next? Centre for Applied Internet Research (CAIR), Glyndwr University, Wrexham, UK. – URL: http://www.glyndwr.ac.uk/computing/research/pubs/sein_bp.pdf (дата обращения 13.02.2015).
2. Tasevski J., Nikolić M., Mišković D. Integration of an Industrial Robot with the Systems for Image and Voice Recognition. – Serbian journal of electrical engineering, 2013, p.1-13. – URL: http://www.researchgate.net/publication/251898859_Integration_of_an_Industrial_Robot_with_the_Systems_for_Image_and_Voice_Recognition (дата обращения 13.02.2015).
3. Fardana A.R., Jain S., Jovancevic I., Suri Y., Morand C., Robertson N.M. Controlling a Mobile Robot with Natural Commands based on Voice and Gesture. – URL: <http://home.eps.hw.ac.uk/~cgb7/readinggroup/papers/RobotCommandingByVoiceAndGesture.pdf> (дата обращения 13.02.2015).
4. Shotton J., Fitzgibbon A., Cook M., Sharp T., Finocchio M., Moore R., Kipman A., Blake A. – Real-Time Human Pose Recognition in Parts from Single Depth Images Computer Vision and Pattern Recognition. – 2013 IEEE Conference on Computer Vision and Pattern Recognition (20-25 June 2011), ISBN: 978-1-4577-0394-2, pp: 1297-1304. – URL: <http://research.microsoft.com/pubs/145347/bodypartrecognition.pdf> (дата обращения 13.02.2015).
5. Urban M., Bajcsy P. Fusion of Voice, Gesture, and Human-Computer Interface Controls for Remotely Operated Robot. – National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign, Champaign, Illinois, USA. – URL: <http://isda.ncsa.illinois.edu/peter/publications/conferences/2005/Fusion05Robot.pdf> (дата обращения 13.02.2015).
6. Котомин А.В. Предобработка звукового сигнала в системе распознавания речевых команд. – Труды XV Молодежной научной конференции «Наукоемкие информационные технологии» SIT-2011 (20-24 апреля 2011г.). – Переславль-Залесский: Изд-во «Университет города Переславля», 2011, с.25–38.
7. Мокров Е.В., Хачумов В.М., Теплоухова Н.Е. Применение фоновых подходов к распознаванию речевых команд. – В сб. Ракетно-космическое приборостроение

- и информационные технологии. 2011. – Труды IV Всероссийской научно-технической конференции «Актуальные проблемы ракетно-космического приборостроения и информационных технологий» (15-17 июня 2011 г.) – М.: Радиотехника, 2012, с.418-424.
8. Nahapetyan V.E., Khachumov V.M. Automatic Transformation of Russian Manual Alphabet Gestures into Textual Form. – Scientific and Technical Information Processing, 2014, Vol.41, No.5, pp.302-308
9. Нагапетян В.Э., Хачумов В.М. Распознавание жестов руки в задаче бесконтактного управления беспилотным летательным аппаратом. – Автометрия, № 2, 2015.
10. Хачумов М.В. Модели представления и кластеризации слабоструктурированной информации. – Искусственный интеллект и принятие решений, № 4, 2013, с.62-71.
11. Талалаев А.А., Фраленко В.П. Комплекс инструментальных средств для проектирования нейросетевых прикладных систем. – Научно-технический вестник Поволжья, №4, 2013, с.237-243.

Нагапетян Ваагн Эдвардович. Старший разработчик "Aragast Ben Web technologies". Окончил Российского университета дружбы народов в 2014 году. Кандидат физико-математических наук. Автор 13 печатных работ. Область научных интересов: компьютерное зрение, человеко-машинные интерфейсы, интеллектуальное управление. E-mail: vahagnahapetyan@gmail.com

Хачумов Вячеслав Михайлович. Заведующий лабораторией Института системного анализа РАН. Окончил Ленинградский кораблестроительный институт в 1971 году. Доктор технических наук, профессор. Автор 190 печатных работ и двух монографий. Область научных интересов: искусственный интеллект, высокопроизводительные вычисления, обработка изображений и распознавание образов. E-mail: vmh48@mail.ru