

*На правах рукописи*



**КОТЮЖАНСКИЙ ЛЕОНИД АНАТОЛЬЕВИЧ**

**ИНТЕРФЕЙС БЕСКОНТАКТНОГО ЧЕЛОВЕКО-МАШИННОГО  
ВЗАИМОДЕЙСТВИЯ НА ОСНОВЕ ДАННЫХ  
СЕНСОРА-ДАЛЬНОМЕРА**

**Специальность 05.13.11 – Математическое и программное  
обеспечение вычислительных машин, комплексов  
и компьютерных сетей**

**Автореферат диссертации на соискание ученой степени кандидата  
технических наук**

**Екатеринбург – 2014**

Работа выполнена на кафедре информационных систем и технологий  
ФГАОУ ВПО «Уральский федеральный университет имени первого  
Президента России Б.Н.Ельцина»

|                       |  |
|-----------------------|--|
| Научный руководитель  | д-р физ.-мат. наук, доцент<br><b>ОСИПОВ Александр Владимирович</b><br>профессор кафедры «Информационные системы и технологии» ФГАОУ ВПО «Уральский федеральный университет имени первого Президента России Б.Н.Ельцина»  |
| Официальные оппоненты | д-р техн. наук, проф.<br><b>ЖУКОВ Дмитрий Олегович</b><br>профессор Института криптографии, связи и информатики ФГБОУ ВПО «Академия ФСБ РФ»<br><br>канд. техн. наук<br><b>НИКОНОВ Вячеслав Викторович</b><br>доцент кафедры «Автоматизированные системы управления и информационные технологии» ФГБОУ ВПО «Московский государственный университет приборостроения и информатики» |
| Ведущая организация   | ФГБОУ ВПО «Уральский государственный университет путей сообщения»  |

Защита диссертации состоится 21 мая 2014 г. в 10<sup>00</sup> часов на заседании диссертационного совета Д-212.288.07 на базе ФГБОУ ВПО «Уфимский государственный авиационный технический университет» по адресу:  
450000, г. Уфа, ул. К. Маркса, 12.

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВПО «Уфимский государственный авиационный технический университет» и на сайте <http://www.ugatu.ac.ru/>.

Автореферат разослан «\_\_\_»\_\_\_\_\_2014 г.

Ученый секретарь диссертационного совета

д-р техн. наук, доц.



И. Л. Виноградова

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность исследования.** В настоящее время все большее значение приобретают новые в концептуальном и в программно-аппаратном плане формы человеко-машинного взаимодействия, требующие создания новых интерфейсов, в основном меню которых не используются шаблоны или панели инструментов, а, напротив, при задании спецификаций операций и операндов акцент переносится на обучающие примеры, интуитивность, жесты и распознавание речи.

Концепция подобных интерфейсов получила название естественный пользовательский интерфейс. Основной целью естественного интерфейса является минимизация использования механизмов манипулирования преобладающих в сегодняшних интерфейсах и сокращение «когнитивной» дистанции между намерением и реализацией этого намерения. Иными словами, требуется дать пользователю возможность сосредоточиться на его задаче, а не на технологии, необходимой для специфицирования этой задачи.

Исследования в области естественных интерфейсов управления жестами стали появляться еще в конце восьмидесятых годов, в частности была представлена сенсорная перчатка, позволяющая отслеживать сгибание пальцев и положение ладони пользователя посредством ультразвуковых датчиков для манипулирования виртуальным содержимым (Thomas G. Zimmerman). В ряде работ (Laura Dipietro, Angelo M, Hoa Phan, Jeffrey Ellen, Mike Bruch) описывалось управление при помощи жестов посредством закрепленных на руке человека сенсоров – магнитных, инерционных, ультразвуковых, инфракрасных и прочих видов датчиков. Обладая, как правило, высокой точностью отслеживания, что имеет решающее значение в системах захвата движения, использующихся в киноиндустрии и компьютерных играх, этот подход показал целый ряд недостатков. В частности, механическая составляющая устройств подвержена физическому износу, имеют место определенные неудобства для пользователя при длительном применении, зависимость конструктива от конституции конкретного пользователя, высокая стоимость устройств.

Альтернативой данному решению является использование технологий машинного зрения и обработки изображения для отслеживания положения рук, пальцев и других частей тела при помощи различных оптических сенсоров, находящихся на некотором расстоянии от пользователя. С середины девяностых годов проявилась тенденция к использованию веб-камер как инструмента взаимодействия с компьютером (R. Cipolla, A. Pentland, Lars Bretzner, Ivan Laptev, Tony Lindeberg, Vincent John Vincent).

**Степень разработанности темы.** Рассматривая преимущества и недостатки использования оптической камеры в качестве сенсора жестов, следует отметить высокую зависимость качества распознавания от окружающего пространства, сложности фона и уровня освещенности. Отсутствие нормального, дневного освещения и наличие сложного, движущегося фона часто сводит на нет всю работу даже самых высококачественных и сложных алгоритмов машинного зрения. Кроме того сложные рекурсивные и иерархические алгоритмы

распознавания жестов часто являются ресурсоемкими, и выдают производительность на уровне нижнего предела (около 15-30 кадров в секунду) реального времени даже на высокопроизводительных компьютерах. В то же время «веб-камера» сегодня является стандартной опцией для большинства существующих моделей ноутбуков, планшетов, телефонов. Поэтому в сочетании с постоянным ростом вычислительной мощности элементной базы компьютеров веб-камера по-прежнему является привлекательным решением в качестве сенсора управления мультимедийным контентом.

Наибольший рост практических приложений технологий распознавания жестов в последнее время связан с разработкой компании Microsoft – сенсором Kinect (2011-2012 г.г.), позволяющим получить трехмерную видео информацию в виде карты глубины или трехмерного облака точек видимого сенсору пространства.

Именно трехмерность данных позволяет решить одну из основных и весьма сложных задач машинного зрения – отсечение заднего фона и сегментацию объектов для их последующего анализа.

Сегодня имеется ряд частных решений отдельных аспектов распознавания жестов, однако острой остается проблема повышенной ресурсоемкости реализованных алгоритмов и ограниченной точности наведения курсора на элементы интерфейса.

Настоящая работа сфокусирована на преодолении описанных выше трудностей, развивая идею разработки универсальной системы эффективного бесконтактного управления компьютером.

**Объект исследования** – процесс обработки данных сенсора-дальномера в задачах распознавания жестов для бесконтактного управления программным обеспечением.

**Предмет исследования** – возможность детектирования жестов в потоке данных сенсора-дальномера для реализации функций бесконтактного управления.

**Цель исследования** – создание эффективного интерфейса бесконтактного человеко-машинного взаимодействия.

**Задачи исследования:**

- 1) разработать *концепцию* системы управления жестами на основе данных сенсора-дальномера;
- 2) предложить *модель* интерфейса бесконтактного управления программным обеспечением на основе обработки данных сенсора-дальномера;
- 3) предложить *методы* детектирования физических объектов, участвующих в формировании жеста: пользователь, руки, пальцы и прикосновений к поверхности;
- 4) разработать *программные модули*, осуществляющие необходимые действия для эффективного управления жестами:
  - отслеживание пользователя;
  - распознавание вытянутых рук пользователя;
  - распознавание пальцев пользователя;

- детектирование жестов;

5) разработать программное обеспечение для реализации модели интерфейса бесконтактного управления и детектирования прикосновений к поверхности, интегрирующее программные модули распознавания.

### **Научная новизна**

*В целом* новизна результатов заключается в создании новых способов человеко-машинного взаимодействия на основе анализа данных сенсора-дальномера: бесконтактное управление жестами и управление с помощью прикосновений к любому объекту в поле зрения сенсора-дальномера.

*По существу* новизна результатов определяется следующими отличительными признаками:

1) *концепция системы управления жестами*: а) система использует сенсор дальномер, который устанавливается рядом с экраном; б) перед сенсором имеется некая интерактивная зона, в которой система детектирует пользователя, его видимые руки и пальцы; в) на основе полученных от сенсора-дальномера координат и вычисленных скоростей найденных объектов система детектирует производимые пользователем эталонные жесты; г) каждый эталонный жест транслируется в управляющую команду;

2) *модель интерфейса бесконтактного управления* программным обеспечением на основе обработки данных сенсора-дальномера для иерархического детектирования пользователя, его рук, пальцев и жестов; формирования управляющих команд с возможностями удаленного управления любым программным обеспечением, любыми устройствами на любых платформах и операционных системах, поддерживающих протокол UDP.

3) *методы* иерархического детектирования пользователя на основе Z-сегментации облака точек и определения его характеристических параметров; рук и пальцев на основе геометрических критериев; *метод* распознавания жестов по анализу координат, скоростей, форм выделенных подоблаков; *метод* детектирования прикосновений к любой поверхности в поле зрения сенсора по вычисляемым картам отклонений.

4) *программные модули* (техническая новизна), осуществляющие необходимые действия для эффективного управления жестами:

- отслеживание пользователя;
- распознавание вытянутых рук пользователя;
- распознавание пальцев пользователя;
- детектирование жестов;

5) *программное обеспечение* (техническая новизна) для реализации модели интерфейса бесконтактного управления и детектирования прикосновений к поверхности, интегрирующее программные модули распознавания.

### **Теоретическая и практическая значимость работы**

*Значение результатов для теории* (методологии) естественных человеко-машинных интерфейсов и машинного зрения выражается в том, что предложенный общий подход к бесконтактному управлению компьютером с помощью жестов, основанный на специальной аналитической обработке численных дан-

ных, расширяет возможности проектирования интерфейсов взаимодействия нового типа.

*Значение результатов для практики* использования человеко-машинных интерфейсов: предложенные методы позволяют создавать новые программные и аппаратные решения для управления бытовой техникой (жидкокристаллические дисплеи), информационными терминалами в местах массового скопления людей, конструировать новые типы интерактивных проекторов с сенсорным мультитач экраном. Предложенное бесконтактное управление эффективно из-за высокой интуитивности и отсутствия механических и электронных датчиков в прямом контакте с пользователями. Ряд разработанных методов допускает использование машинного зрения в системах безопасности, сенсорных интерактивных мультитач проекциях, в системах распознавания языка глухонемых, в управлении медицинским оборудованием, в системах медицинской диагностики и др.

### **Методология и методы исследования**

В работе использовались методы математического моделирования и численного эксперимента, математическая статистика и математический анализ, методы цифровой обработки изображений, элементы теории алгоритмов, объектно-ориентированный анализ, объектно-ориентированное программирование и др. Алгоритмизация распознавания изображений проводилась с использованием библиотек OpenCV, OpenGL, Openframeworks, протокола UDP и языков программирования C++ и GLSL.

### **Положения, выносимые на защиту:**

1) Общая концепция и архитектура системы бесконтактного управления компьютером с помощью жестов, *отличающиеся* ролевыми функциями пользователя, сенсора и экрана в интерактивной зоне, обеспеченными соответствующими программными модулями интерфейса бесконтактного человеко-машинного взаимодействия: детектирование пользователей, рук, пальцев, жестов, прикосновений и их трансляция в сигналы управления.

2) модель интерфейса бесконтактного управления программным обеспечением на основе обработки данных сенсора-дальномера, *отличающаяся* тем, что

а) архитектура программного обеспечения разделена на два связанных между собой модуля – передатчик – осуществляет анализ данных с сенсора и выполняет функции детектирования пользователя, его рук и пальцев; – приемник – проводит конечный анализ полученных от передатчика данных и формирует управляющие команды;

б) обеспечена возможность реализации универсального набора жестов, где каждый жест используется для определенной операции;

в) обеспечена возможность реализации любого удаленного управления любым программным обеспечением, любыми устройствами на любых платформах и операционных системах, поддерживающих протокол UDP.

3) Методы иерархического детектирования пользователя, его рук, пальцев, *отличающиеся* последовательным способом разбиения данных об ис-

ходном облаке трехмерных точек на отдельные области, вычислением их интегральных характеристик и детектированием определенного объекта на основе геометрических критериев; метод распознавания жестов, *отличающийся* использованием параметризованных критериев анализа координат, скоростей, форм соответствующих подоблаков; метод детектирования прикосновений к любой поверхности в поле зрения сенсора-дальномера, *отличающийся* иерархическим анализом карт глубины.

4) Программные модули для эффективного управления жестами на расстоянии, *отличающиеся* соответствием концепции (см. п. 1) и модели (см. п. 2) интерфейса бесконтактного управления и являющиеся практической реализацией предложенных методов (см. п. 3).

5) Программное обеспечение интерфейса бесконтактного управления, *отличающееся* соответствием концепции (см. п. 1), модели (см. п. 2) интерфейса бесконтактного управления и детектирования прикосновений к поверхности, интегрирующее программные модули распознавания (см. п. 4).

**Степень достоверности и апробация результатов.** Достоверность результатов подтверждается обоснованностью применения математического аппарата, оценкой адекватности математических и информационных моделей, результатами тестирования алгоритмов и программного обеспечения, а также результатами использования разработанных систем бесконтактного управления в практике информационной деятельности нескольких предприятий.

Проект Котюжанского Л.А. «Человеко-машинный интерфейс на базе устойчивого распознавания движений и жестов» стал победителем конкурса проектов по Уральскому федеральному округу по программе «СТАРТ-10», работа «Разработка сенсорных контактных и бесконтактных мультимедийных устройств» представлялась на II Всероссийском конкурсе «Умник на СТАРТ» (Самара 2011, 26 сентября – 1 октября), в итоге был выигран грант для проведения НИОКР.

**Публикации.** По результатам выполненных исследований опубликовано 14 работ, среди них 3 статьи в рецензируемых научных журналах из перечня ВАК, 7 публикаций без соавторов. Получено свидетельство о государственной регистрации программы для ЭВМ.

**Структура диссертации.** Диссертация включает введение, три главы, заключение, список литературы и приложения. Общий объем диссертации составляет 146 страниц машинописного текста, из них 129 страниц – основной текст, 75 рисунков, 3 таблицы. Библиографический список содержит 177 наименований.

## СОДЕРЖАНИЕ ДИССЕРТАЦИИ

**Во Введении** обоснована актуальность темы исследований, определены цель и задачи работы, сформулированы научная новизна и практическая значимость полученных результатов, приведены основные положения, выносимые на защиту.

В первой главе приведен обзор существующих методов отслеживания частей тела пользователя во время движения. В отдельном параграфе описываются функциональные особенности сенсора дальномера Kinect, используемого в работе, рассматривается возможная альтернатива Kinect – использование стереокамер. Приводится краткий обзор алгоритмов стереозрения, позволяющих получить *карту глубины изображения* с двух камер.

На данный момент известны активные и пассивные методы восстановления информации о глубине реальной сцены. Активные методы используют ультразвуковые преобразователи или лазерное освещение рабочего пространства, дающие на выходе быструю и точную информацию о глубине (Feiner S., MacIntyre B., Seligmann D., Iddan G. J., Yahav G.). Однако у этих методов есть ограничения по отношению к диапазону измерений и стоимости аппаратных компонентов.

Пассивные методы, основанные на компьютерном зрении, обычно реализуются более простыми и недорогими сенсорами, определяющими расстояние. Такие методы способны генерировать информацию о глубине по полученной паре изображений и параметрам двух камер (Fehn C., Cooke E., Schreer O., Kauff P., Kanade T., Yoshida A., Oda K., Kano H., Tanaka M.).

Одна из фундаментальных проблем стереозрения заключается в установлении точного соответствия между левым и правым изображением стереопары (рис. 1). Под соответствием понимается расстояние (диспаратет) между пикселями одного и того же объекта на левом и правом изображении. Существующие локальные методы по вычислению карты диспаратета основаны на принципе «скользящего окна». Особенностью локальных методов является то, что при увеличении размеров скользящего окна с одной стороны уменьшается вероятность ошибки определения диспаратета пикселя, а с другой стороны значительно возрастает объем вычислений, что делает данные методы малоприменимыми для расчетов на стереопарах высокого разрешения в реальном времени. Кроме того появляется эффект «раздувания» переднего плана.

В данной работе предложен метод вычисления карты глубины, основанный на использовании двух *одномерных разреженных окон* – вертикального и горизонтального. Каждое окно строится из центра рассматриваемого пикселя. Под разрежением понимается, что окно состоит не из ближайших друг к другу пикселей, а из пикселей находящихся на некотором заданном расстоянии друг от друга, например, два, четыре, шесть пикселей. Это позволило существенно увеличить скорость вычислений при незначительном увеличении ошибки.

В основу расчетов положено определение итоговой *стоимости* двух окон  $C_d$ , путем выбора минимального значения из стоимостей каждого окна  $C_d = \min\{C_h, C_v\}$ .

Стоимость горизонтального окна вычисляется по формулам

$$C_h(x, y, d) = \sum_{i \in W_h(x, y)} P_c(i, d, y) \cdot \frac{1}{|x - i|}, \quad (1)$$



$$P_c(x, d, y) = |p(x, y) - q(x - d, y)|, \quad (2)$$

$p$  – пиксель на левом изображении,  $q$  – пиксель на правом изображении,

$d$  – перебираемый диспаратет,  $W_h$  – горизонтальная окрестность пикселя с центром в точке  $(x, y)$ .

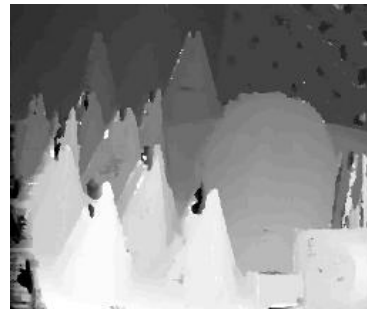
Аналогично проводятся вычисления стоимости вертикального окна. Подчеркнем, что операция суммирования выполняется по индексам суммирования, меняющимся с шагом неравным единице (эффект прореженности окна (окрестности)).



Левое изображение



Правое изображение



Карта глубины – чем ближе объект, тем ярче его изображение (60 уровней глубины)

Рисунок 1 – Формирование карты глубины по двум плоским изображениям

В настоящий момент в области GPU вычислений наиболее активно развиваются технологии CUDA и OpenCL, обеспечивающие высокую гибкость при взаимодействии между ЦПУ и GPU. В данной работе был выбран язык для программирования шейдеров GLSL, реализующий обработку графических данных, таких как текстуры, полигоны фрагментными и вершинными программами. Выбор в пользу GLSL обусловлен тем, что входными и выходными данными разработанного алгоритма являются изображения.

В процессе разработки данного метода обозначились пути возможного улучшения качества и оптимизации скорости вычислений. Так для практически полного удаления шумов и повышения четкости границ объектов на карте глубины предложено использование сегментации входного изображения. Это позволило находить *глубину* не для *каждого конкретного пикселя* изображения, а для группы пикселей – сегмента изображения. Дополняющие

решения – обработка предложенным алгоритмом только интересующих регионов изображения, либо вычисление значения глубины только для движущихся объектов в кадре стереокамеры.

Во **второй главе** рассматривается общая архитектура системы бесконтактного управления, основанная на многоступенчатом анализе данных с сенсора дальногомера для детектирования пользователя, определения координат его рук и пальцев, последующего анализа найденных координат для формирования управляющих команд. Управляющие команды подаются к исполнению при определенных жестах пользователей, простейшими из которых являются различные махи руками.

Концепция системы разделяет её архитектуру на два связанных между собой модуля – передатчик и приемник (рис. 2). Передатчиком считается модуль осуществляющий анализ данных с сенсора, и выполняющий функции детектирования пользователя, его рук и пальцев. Приемник проводит конечный анализ полученных от передатчика данных и формирует управляющие команды. К приемнику относится и непосредственно контролируемое жестами приложение.

Связь между приемником и передатчиком осуществляется по специально разработанному протоколу передачи данных GUIO (gesture user interface output), обеспечивающему передачу данных по локальной сети или сети интернет. Такое разделение архитектуры на два отдельных модуля дает системе ряд важных преимуществ:

- возможность реализации системы жестов, без привязки к конкретным жестам. Последнее объясняется тем, что сам передатчик, по сути, не осуществляет распознавание жестов и формирование управляющих команд. В задачу передатчика входит детектирование пользователей в облаке точек, полученном с сенсора дальногомера и распознавание у найденных пользователей вытянутых рук и пальцев. Полученные данные передатчик отправляет по протоколу GUIO в приемник, где и выполняется конечный анализ данных и формируются управляющие команды;

- возможность реализации любого удаленного управления любым программным обеспечением и устройствами на любых платформах и операционных системах, поддерживающих протокол UDP. Например, управление операционной системой, управление любыми бытовыми приборами, контроль робота-манипулятора, повторяющего движение руки – все это выполняется с одного компьютера, на котором запущена программа-передатчик, осуществляющая распознавание и передающая сведения конкретным удаленным приложениям или удаленному серверу;

- разделение системы на два отдельных модуля выгодно с точки зрения оптимального использования ресурсов ЦПУ и памяти.

Функциональная структура передатчика показана на рис. 3.

Передатчик является своеобразным фильтром трехмерного облака точек, который пошагово анализирует трехмерное облако точек видимых сенсору, отбрасывая ненужные точки.

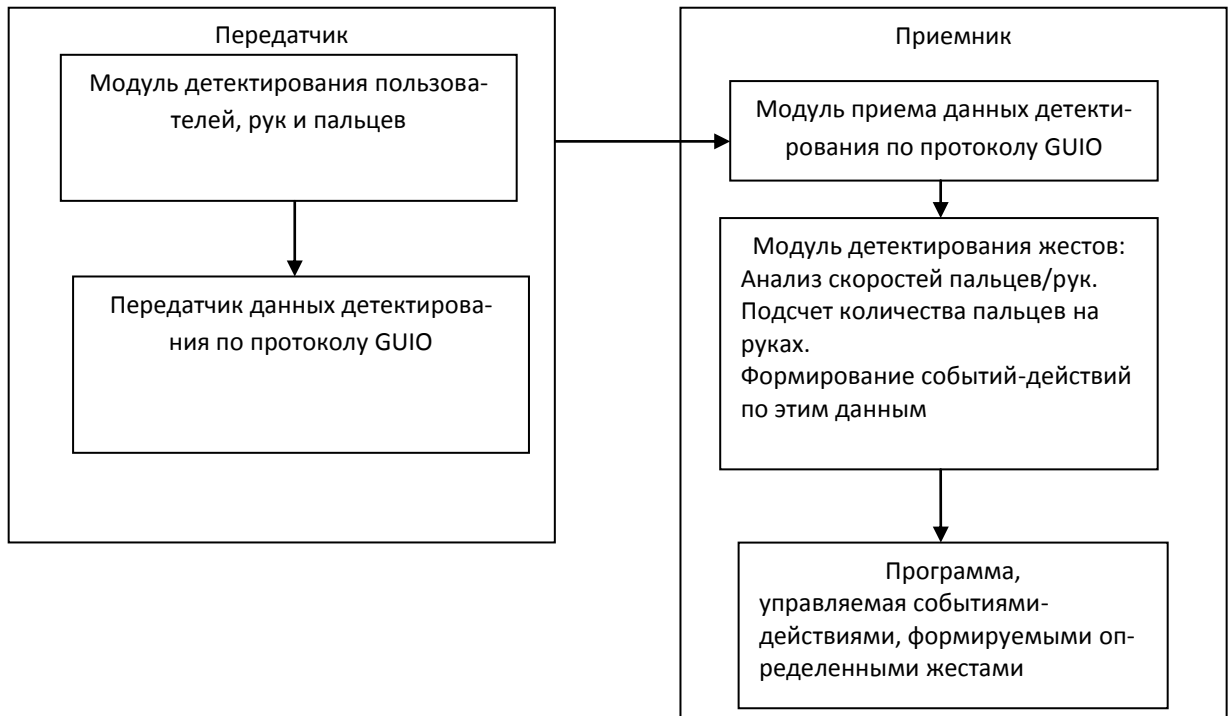


Рисунок 2 – Общая архитектура системы

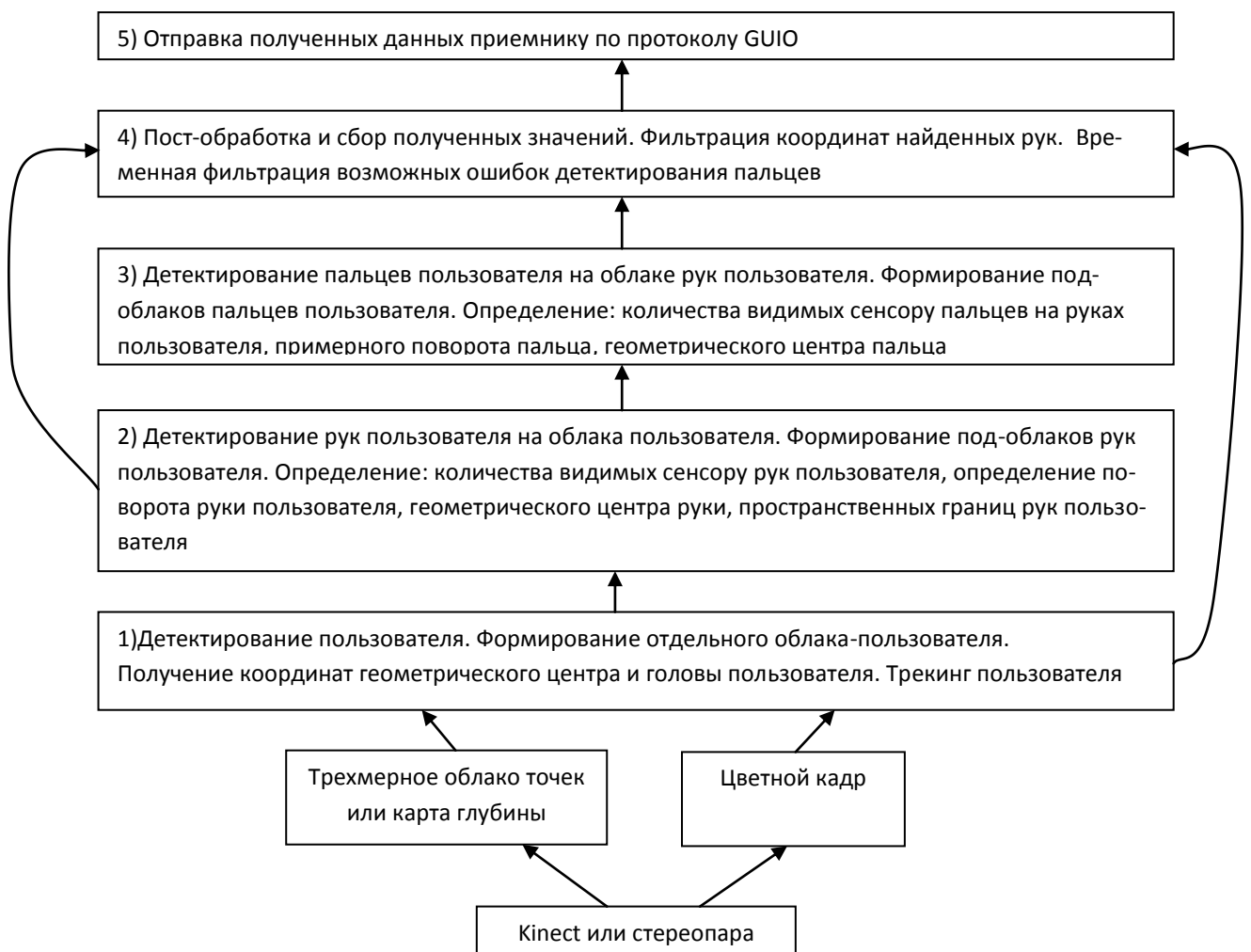


Рисунок 3 – Функциональная архитектура передатчика – модуля распознавания пользователя, его рук и пальцев

На вход модуля распознавания жестов (рис. 3) в каждый момент времени  $t$  с шагом  $\Delta t = \frac{1}{30}$  с от сенсора дальномера поступают данные о карте глубины  $dm$ , которая представляет собой двумерный массив значений глубины (удаленности от сенсора отдельных точек объекта воспринимаемого сенсором),  $dm = (dm_{ij})$ , где  $i, j$  – целочисленные координаты точки ( $i = 0, 1, \dots, width, j = 0, 1, \dots, height$ ).

Далее карта глубины сегментируется первичным  $Z$  – отсечением по правилу

$$bm_{ij} = \begin{cases} 0, & dm_{ij} \geq ft \\ 1, & dm_{ij} < ft \end{cases}, \quad (3)$$

где  $ft$  – константа (пороговое значение удаленности точки), являющаяся эмпирическим параметром моделируемой системы управления.

Следующая обработка бинаризованных данных заключается в «размытии» массива  $bm = (bm_{ij})$ , когда с помощью процедуры свертки строится массив  $\overline{bm} = (\overline{bm}_{ij})$ :

$$\overline{bm}_{ij} = \sum_{l,s \in N(i,j)} bm_{ls} \cdot m_{ls}. \quad (4)$$

Здесь  $N(i, j) - (3 \times 3)$  окрестность элемента  $bm_{ij}$ ,  $m_{ls}$  – элемент известной матрицы свертки

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \quad (5)$$

Далее множество  $\overline{bm}$  разделяется на подмножества – сегменты (блобы). Отдельный сегмент с номером  $S$  – соответствует множеству точек  $P_{S_k}(i, j), k = 1, 2, \dots, N_s$ , удовлетворяющих условиям:

$$\begin{aligned} \overline{bm}_{ij} &\neq 0, \\ \rho(P_{S_l}, P_{S_m}) &< R, \\ m_s &< N_s < M_s, \end{aligned} \quad (6)$$

где  $R, m_s, M_s$  – эмпирические параметры.

Для каждого сегмента формируется массив (облако) трехмерных точек  $C_{S_k}(x, y, z), k = 1, 2, \dots, N_s$ , (7)

определяются его центроид

$$Ctd_s = \frac{1}{N_s} \sum_{k=1}^{N_s} C_{S_k}(x, y, z) = (x_{Ctd_s}, y_{Ctd_s}, z_{Ctd_s}), \quad (8)$$

другие интегральные характеристики облака, значения которых используются для вычленения подмножества точек, возможно относящихся к голове пользователя, включая характеристические точки  $S_k \max$  и  $Ctd_H$  – точку вероятного

центра головы пользователя. Окончательный вывод об идентификации облака делается на основании анализа пикселей прямоугольной области цветного кадра в характеристической точке  $Ctd_h$  методом Виолы-Джонса (рис. 4).

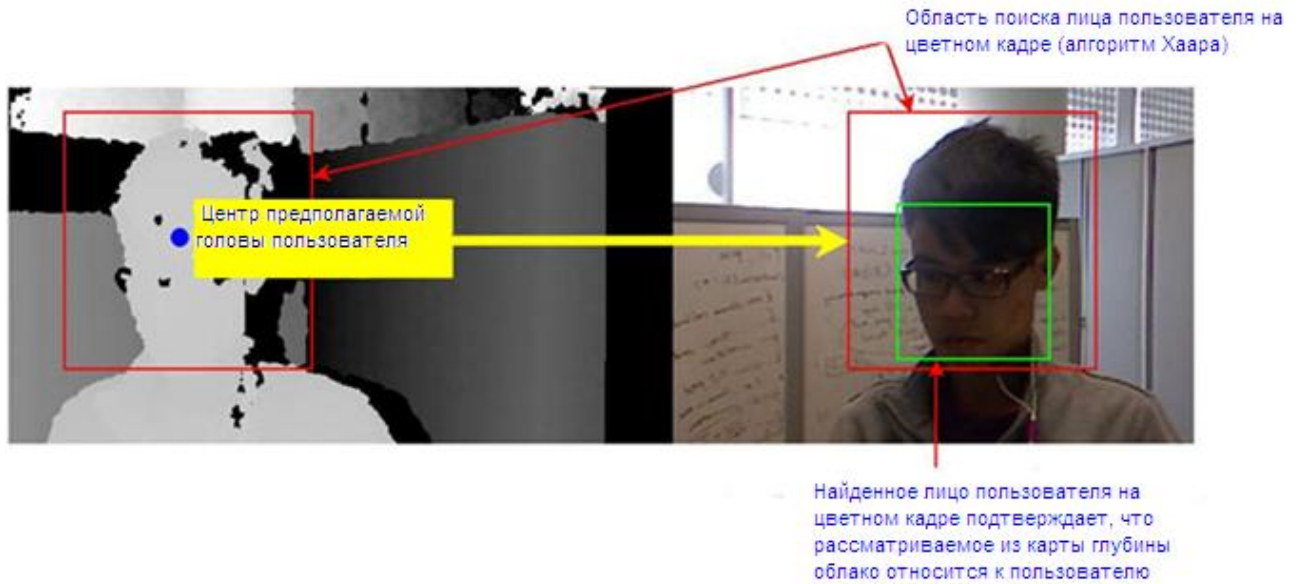


Рисунок 4 – Использование алгоритма Виолы-Джонса для определения пользователей

Необходимые вычисления выполняются при помощи GPU с использованием пиксельного шейдера, написанного на GLSL. В результате исключается возможность случайного избыточного сегментирования изображения пользователя.

Основная идея сегментирования рук заключается в  $Z$ -отсечении части пикселей карты глубины некоторой плоскостью

$$\pi : (z - z_0) = 0, \quad (9)$$

где

$$z_0 = \begin{cases} (Ctd_h)_z + Z_c, & (S_k \max)_z < (Ctd_h)_z + Z_c \\ \frac{((Ctd_h)_z + Z_c) + ((S_k \max)_z - Z_c) \cdot 0.85}{2}, & (S_k \max)_z \geq (Ctd_h)_z + Z_c \end{cases}, \quad (10)$$

$Z_c$  – эмпирический параметр.

Сегментированная карта глубины предполагаемых рук (рис. 5) для каждого конкретного пользователя строится по следующему правилу

$$\forall p \in D : p = \begin{cases} 0, & p \leq z_0 \\ p, & p > z_0 \end{cases}, \quad (11)$$

где  $p$  – значение пикселя карты глубины  $D$ .

Далее выполняется формирование облаков предполагаемых рук пользователя и окончательное детектирование рук на основе положений дифференциальной геометрии.

Одним из самых сложных и в тоже время самых значимых механизмов для бесконтактного управления является распознавание пальцев, позволяющее идентифицировать различные уникальные жесты. Если рассматривать руку как

некий манипулятор курсора, то пальцы могут играть роль модификатора курсора.

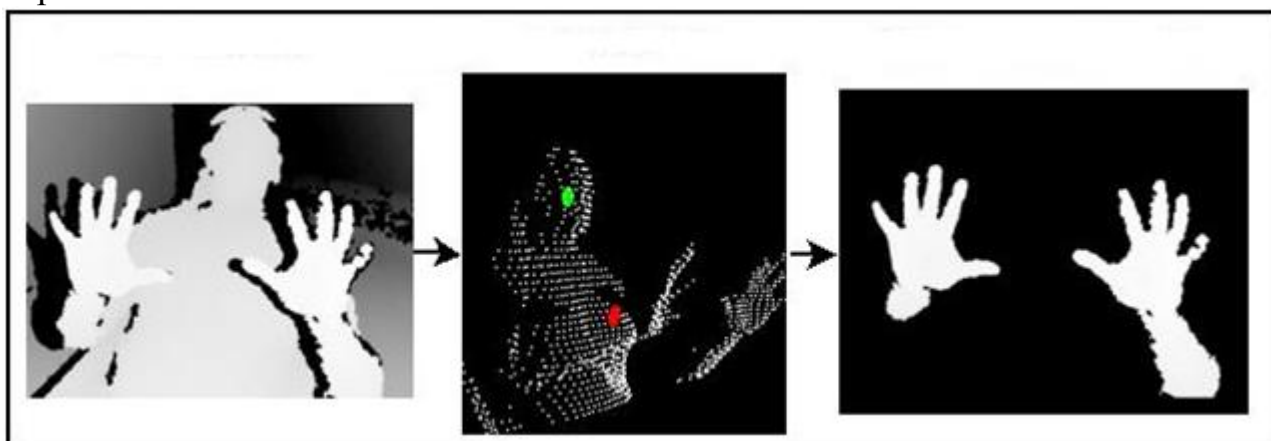


Рисунок 5 – Первичная сегментация рук по характеристическим точкам пользователя

Чтобы реализовать отмеченные возможности использовался комплексный подход, комбинирующий прямую работу над множествами трехмерных точек с их 2D изображениями (рис.6) по плану: нахождения контуров сегментированных изображений рук (каждый контур соответствует определенной руке пользователя), адаптивная аппроксимация контура руки в зависимости от занимаемой площади контура на изображении, поиск выпуклой оболочки на аппроксимированном контуре, уточнение аппроксимации выпуклого контура, детектирование пальца на основе анализа его формы.



Рисунок 6 – Получение бинаризованного изображения и его контура

В итоге из глобального облака формируется иерархическая система под-облаков с набором данных для последующего анализа с целью детектирования определенных жестов и формирования соответствующих команд управления.

**В третьей главе** описывается работа детектора жестов из предложенного конечного множества.

Процесс детектирования жестов начинается с обработки полученных от передатчика координат рук и пальцев, в результате которой определяется вектор-функция скорости ладони (пальца)  $\vec{v}(t)$ , вычисляемый через перемещения по координатным осям за единицу времени  $\Delta t$  (одна итерация).

В отдельном параграфе рассматривается предложенный алгоритм анализа  $\vec{v}(t)$  для распознавания того или иного жеста, состоящий в последовательной проверке выполнения системы некоторых условий.

Отдельно представлен специальный алгоритм фильтрации координат курсора руки.

В силу недостаточно высокого разрешения сенсоров-дальномеров (как правило, максимальное разрешение не превышает порог в  $640 \times 480$  пикселей) в системе используется сглаживание координат пальцев по времени. Это делается для того, чтобы минимизировать возможное колебание координат пальцев, при достаточно большой удаленности пользователя от сенсора. Сглаживание координат пальцев выполняется по формулам усреднения. Чем больше величина  $n$  (число усредняемых значений), тем более сглаженным выглядит отображаемое движение, но возрастает и эффект отставания сглаженных координат от реальных.

Тестирование показало, что оптимальное значение  $n$  попадает в промежуток от 5 до 10. В этом случае дрожание пальцев практически полностью отсутствует при сохранении достаточной реакции на движения пальцев пользователя.

Для наиболее точного наведения графического курсора при помощи движения руки необходимо минимизировать возможные ошибки определения центра руки. Под ошибками в данном случае понимаются, в частности, эффекты хаотического дрожания центра руки. Таковые возникают из-за недостаточно высокого разрешения сенсора или реального физического смещения руки пользователя при её фиксации в пространстве.

Для решения данной задачи необходимо сохранить баланс между чувствительностью системы к изменению положения руки и минимально допустимым дрожанием курсора.

При вычислении вектора скорости используется модифицированный фильтр Калмана для коррекции координат руки, дополнительные пороговые фильтры по абсолютным значениям координат рассматриваемых векторов и фильтр сглаживания координат по времени.

Предложенный подход детектирования рук и фильтрации курсора позволил добиться приемлемой точности *работы системы*, практически полностью устранив резкие ошибочные смещения координат заметные в OpenNi, и Kinect SDK.

Для реализации системы использовался язык C++, а также множество различных вспомогательных программных библиотек и технологий, в частности, OpenCV, OpenGL, GLSL и библиотеки работы с протоколом UDP. Связь с сенсором kinect осуществляется при помощи открытой библиотеки libfreenect.

В качестве основного каркаса приложения был выбран кроссплатформерный инструментарий Open Source Toolkit – OpenFrameworks.

Тестирование итогового программного комплекса проводилось для каждого модуля в отдельности и для всей системы в целом на записанном наборе данных из 1800 кадров. В результате были определены оптимальные значения

эмпирических параметров, что позволило поднять вероятность распознавания вытянутой руки до 95,7%, а вероятность корректного определения видимых пальцев до 94,8%.

**В заключении** сформулированы основные результаты работы.

**В приложениях** приведены: листинги модулей детектирования жестов и касания поверхности; представлены акты использования результатов диссертации.

## **ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ РАБОТЫ**

В работе представлены новые научно обоснованные информационно-технологические решения и разработки в области человеко-машинных интерфейсов; моделей, методов, алгоритмов и программных средств обработки изображений, имеющие значение для развития интерактивных технологий.

При этом получены результаты:

1) Общая концепция и архитектура системы бесконтактного управления компьютером с помощью жестов, *отличающиеся* ролевыми функциями пользователя, сенсора и экрана в интерактивной зоне, обеспеченными соответствующими программными модулями интерфейса бесконтактного человеко-машинного взаимодействия: детектирование пользователей, рук, пальцев, жестов, прикосновений и их трансляция в сигналы управления.

Это позволяет проектировать различные прикладные системы бесконтактного управления за счет интегрирования комбинаций соответствующих программных модулей при сохранении концептуальной общности проектируемых систем.

2) модель интерфейса бесконтактного управления программным обеспечением на основе обработки данных сенсора-дальномера, *отличающаяся* тем, что

а) архитектура программного обеспечения разделена на два связанных между собой модуля – передатчик – осуществляет анализ данных с сенсора и выполняет функции детектирования пользователя, его рук и пальцев; – приемник – проводит конечный анализ полученных от передатчика данных и формирует управляющие команды;

б) предусмотрена возможность реализации универсального набора жестов, где каждый жест используется для определенной операции;

в) предусмотрена возможность реализации удаленного управления любым программным обеспечением, любыми устройствами на любых платформах и операционных системах, поддерживающих протокол UDP.

Это позволяет *задавать инструментарий* проектирования гибких сценариев бесконтактного управления.

3) Методы иерархического детектирования пользователя, его рук, пальцев, *отличающиеся* последовательным способом разбиения данных об исходном облаке трехмерных точек на отдельные области, вычислением их интегральных характеристик и детектированием определенного объекта на основе геометрических критериев; метод распознавания жестов, *отличающийся* ис-



пользованием параметризованных критериев анализа координат, скоростей, форм соответствующих подоблаков; метод детектирования прикосновений к любой поверхности в поле зрения сенсора-дальномера, *отличающийся* иерархическим анализом карт глубины.

Это позволяет выполнять надежное, быстрое детектирование управляющих *элементов* на основе анализа данных сенсора дальномера.

4) Программные модули, написанные на языках C++ и GLSL для эффективного управления жестами на расстоянии, *отличающиеся* соответствием концепции (см. п. 1) и модели (см. п. 2) интерфейса бесконтактного управления и являющиеся практической реализацией предложенных методов (см. п. 3).

Это позволяет получить набор программных модулей для создания различных систем бесконтактного управления.

5) Программное обеспечение на языках C++ и GLSL интерфейса бесконтактного управления, *отличающееся* соответствием концепции (см. п. 1), модели (см. п. 2) интерфейса бесконтактного управления и детектирования прикосновений к поверхности, интегрирующее программные модули распознавания (см. п. 4).

Это позволяет:

а) реализовать возможность удаленного управления любым программным обеспечением, любыми устройствами на любых платформах и операционных системах, поддерживающих протокол UDP;

б) создать сенсорный экран на любой поверхности, находящейся в поле зрения сенсора дальномера;

в) обеспечивает оптимизационный баланс между скоростью обработки данных сенсора дальномера и надежностью результата детектирования, что подтверждает эффективность разработанных алгоритмов.

**Перспективы дальнейшей разработки темы.** В ходе дальнейших исследований планируется разработка методов универсального и надежного детектирования различных объектов в облаке точек на основе многослойной самообучающейся нейронной сети нового типа. Предполагается построить улучшенные алгоритмы детектирования касаний к поверхности с использованием пространственного фильтра облака трехмерных точек, на основе модели аналитической аппроксимации поверхности касания.

## ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

### *В рецензируемых журналах из списка ВАК*

1. Вычисление карты глубины стереоизображения на графическом процессоре в реальном времени / Л.А. Котюжанский // *Фундаментальные исследования*. 2012. №6 (2). С. 444–449.

2. Реализация симуляции флюидов в системе интерактивной проекции / Л.А. Котюжанский // *Вектор науки Тольяттинского государственного университета*. 2012. №3(21). С. 23–26.

3. Интерфейс бесконтактного управления / Л.А. Котюжанский // *Фундаментальные исследования*. 2013. 4 (1). С. 44–48.

### ***В других изданиях***

4. Нейросетевые технологии распознавания сложных образов / Л.А. Котюжанский // *Отчетная конференция молодых ученых: науч. труды XI отчетной конф. молодых ученых ГОУ ВПО УГТУ – УПИ: сб. статей*. Екатеринбург: ГОУ ВПО УГТУ-УПИ, 2007. Ч 3. С. 223–225.

5. Модель авторской системы универсального распознавания образов / Л.А. Котюжанский, Н.И. Потанин // *Отчетная конференция молодых ученых: науч. труды XII отчетной конф. молодых ученых ГОУ ВПО УГТУ-УПИ: сб. статей*. Екатеринбург: ГОУ ВПО УГТУ-УПИ, 2007 – С.21 – 26.

6. Свид. о гос. рег. программы для ЭВМ № 2006612201. SKYNET A.I. CORE / Л.А. Котюжанский. Зарег. 15.01.2007. М.: Фед. сл. по интел. собств. (Роспатент).

7. Универсальная система распознавания образов – структура, обучение и тестирование системы / Л.А. Котюжанский, В.А. Пухов // *Отчетная конференция молодых ученых: науч. труды XIV отчетной конф. молодых ученых УГТУ-УПИ: сб. статей*. Екатеринбург: УГТУ-УПИ, 2008. Ч. 3. С. 297-300.

8. Компактные нейронные сети, распознающие лица / Н.И. Потанин, Л.А. Котюжанский // *Интерфейс «Мозг-Компьютер»: мат. XV междунар. конф.* Ростов-на-Дону: Изд. ЮФУ, 2009. Т. 2. С. 146.

9. Физическая симуляция и визуализация поверхности воды для интерактивной проекции изображения на плоскость / Л.А. Котюжанский, М.Л. Щелкунов, В.М. Коренберг, Т.А. Матвеева // *Новые образовательные технологии в вузе: сб. мат. 7-й междунар. науч.-метод. конф.* Екатеринбург: ГОУ ВПО «УГТУ-УПИ имени первого Президента России Б.Н. Ельцина», 2010. Ч. 2. С. 293-296.

10. Интерактивные системы. Человеко-машинный интерфейс, основанный на распознавании жестов / В.М. Коренберг, М.Л. Щелкунов, Л.А. Котюжанский // *XVII международная конференция молодых ученых по приоритетным направлениям развития науки и техники: науч. труды XVII междунар. конф. молодых ученых по приоритетным направлениям развития науки и техники*. Екатеринбург: ГОУ ВПО «УГТУ-УПИ имени первого Президента России Б.Н. Ельцина», 2010. Ч. 3. С. 27- 34.

11. Дополненная реальность как новый интерфейс взаимодействия человека с компьютером / В.М. Коренберг, М.Л. Щелкунов, Л.А. Котюжанский // *XVII международная конференция молодых ученых по приоритетным направлениям развития науки и техники: науч. труды XVII междунар. конф. молодых ученых по приоритетным направлениям развития науки и техники*. Екатеринбург: ГОУ ВПО «УГТУ-УПИ имени первого Президента России Б.Н. Ельцина», 2010. Ч. 3. С. 41-46.

12. Реализация симуляции флюидов в системе интерактивной проекции / Л.А. Котюжанский // *Современные инструментальные системы, информационные технологии и инновации: мат. IX междунар. науч.-практ. конф.* Курск: Юго-Зап. гос. ун-т, 2012. С. 104-107.

13. Лабораторный мультитач стол / Л.А. Котюжанский, Н.В. Четверкин, Л.Е. Королев // *Новые образовательные технологии в вузе (НОТВ – 2012): сб. мат. IX междунар. науч.-метод. конф.* Екатеринбург: УрФУ, 2012. С.80-84.

14. Visiontouch – новая технология интерактивных проекций для учебного процесса / Л.А. Котюжанский, Л.Е. Королев, А.С. Курлов // *Новые образовательные технологии в вузе (НОТВ – 2013): сб. мат. X междунар. науч.-метод. конф.* Екатеринбург: УрФУ, 2013 [Электронный ресурс]. – Режим доступа: <http://notv.urfu.ru/notv/818/1839> (дата обращения 14.04.2013).

Диссертант

Л. А. Котюжанский

**КОТЮЖАНСКИЙ ЛЕОНИД АНАТОЛЬЕВИЧ**

**ИНТЕРФЕЙС БЕСКОНТАКТНОГО ЧЕЛОВЕКО-МАШИННОГО  
ВЗАИМОДЕЙСТВИЯ НА ОСНОВЕ ДАННЫХ  
СЕНСОРА-ДАЛЬНОМЕРА**

Специальность 05.13.11 – Математическое и программное  
обеспечение вычислительных машин, комплексов  
и компьютерных сетей

**А В Т О Р Е Ф Е Р А Т**  
диссертации на соискание ученой степени  
кандидата технических наук

ЛР № 020258 от 08.01.1998

Подписано в печать 00.00.2000. Формат 60x84 1/16  
Бумага офсетная. Печать плоская. Гарнитура Times New Roman.  
Усл. печ. л. 1,0. Усл. кр.-отт. 1,0. Уч.-изд. л. 0,9  
Тираж 100 экз. Заказ № \_\_\_\_

Уфимский государственный авиационный технический университет  
Центр оперативной полиграфии  
450000, Уфа-центр, ул. К. Маркса, 12