На правах рукописи

Алфимцев Александр Николаевич

РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДОВ ЗАХВАТА, ОТСЛЕЖИВАНИЯ И РАСПОЗНАВАНИЯ ДИНАМИЧЕСКИХ ЖЕСТОВ

Специальность 05.13.17 – Теоретические основы информатики

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Москва – 2008

Работа выполнена в
Московском Государственном Техническом Университете им. Н.Э. Баумана

Научный руководитель: доктор технических наук, профессор
Девятков В. В.

Официальные оппоненты: доктор технических наук, профессор
Артамонов Е. И.

кандидат технических наук
Бобков А. В.

Ведущая организация: Институт Проблем Передачи Информации РАН
имени А. А. Харкевича

Защита диссертации состоится « 26 » июня 2008 г. в 12 час. 00 мин. на заседании диссертационного совета Д 212.141.10 при Московском государственном техническом университете им. Н.Э. Баумана по адресу: 105005, Москва, 2-я Бауманская ул., д. 5.

С диссертацией можно ознакомиться в библиотеке МГТУ им. Н.Э. Баумана.

Ваш отзыв на автореферат в одном экземпляре, заверенный печатью организации, просьба направлять по адресу:
105005, Москва, 2-я Бауманская ул., д. 5, МГТУ им. Н.Э. Баумана,
ученому секретарю диссертационного совета Д 212.141.10.

Автореферат разослан «___ » _____2008 г.

Ученый секретарь
диссертационного совета,
кандидат технических наук, доцент С. Р. Иванов

1. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В настоящее время разработка и исследование человеко-машинных интерфейсов, основанных на распознавании образов и визуальном представлении мультимедийной информации, становится передним краем в развитии современного математического и программного обеспечения. Перед разработчиками подобных интерфейсов ставится задача использования естественных для человека способов общения с компьютерами с помощью жестов, голоса, мимики и других модальностей. Жесты являются особенно перспективными для построения интерфейсов управления программным и аппаратным обеспечением компьютеров, роботов, позволяют расширить возможности интерфейса для людей с дефектами слуха и речи.

В связи с этим, актуальность темы диссертации с теоретической точки зрения диктуется необходимостью разработки методов, моделей и алгоритмов захвата, отслеживания и распознавания жестов, совершаемых человеком в реальном времени, в частности руками, пригодных для создания интерфейса управления работой компьютера с их помощью.

Актуальность темы с прикладной точки зрения определяется необходимостью создания программных систем, способных обеспечить с помощью жестов интерфейс с персональным компьютером в реальном времени, используя только видеокамеры.

Объект исследования: методы, алгоритмы и программы захвата, отслеживания и распознавания жестов человека.

Предмет исследования: типы жестов, структура методов и алгоритмов захвата, отслеживания и распознавания динамических жестов, их взаимосвязь, сложность, надежность, устойчивость, позволяющие распознавать динамические жесты в реальном времени.

Цель работы и задачи исследований. Целью работы является разработка общей методологии захвата, отслеживания и распознавания динамических жестов человека, совершаемых руками, включая модели, методы и алгоритмы, теоретическое и экспериментальное обоснование работоспособности этой методологии в реальном времени с высоким уровнем надежности для создания работоспособных человеко-машинных интерфейсов.

Для реализации этой цели были поставлены следующие задачи:

1. Осуществить сравнительный аналитический обзор существующих методов захвата, отслеживания и распознавания динамических жестов человека.
2. Провести классификацию жестов выполняемых человеком и выбрать алфавит динамических жестов, пригодный для создания человеко-машинного интерфейса для управления компьютером.

3. Разработать вычислительно эффективный алгоритм захвата и отслеживания кисти человека на сложном фоне.
4. Разработать вычислительно эффективную модель и алгоритм распознавания динамических жестов человека.
5. Разработать методологию мультимодального распознавания сцен, определяемых динамическими жестами.
6. Провести эксперименты по оценке надежности и работоспособности системы в реальном времени, подтверждающие теоретические результаты.

Методы исследования. Основной задачей при планировании исследования было гармоничное сочетание теоретических проработок и экспериментальных проверок. Надежность, устойчивость и достоверность полученных алгоритмов и моделей проверялась на специально подготовленной доверительной выборке. Методы исследований базировалась на статистическом анализе и математическом моделировании, теории нечеткой логики и нечетких множеств, методах объектно-ориентированного программирования и разработки интеллектуальных систем, теории распознавания образов.

Научная новизна. Разработана новая комплексная методология захвата, отслеживания и распознавания динамических жестов в видеопотоке. В рамках этой комплексной методологии получены следующие новые результаты.

Разработан алгоритм захвата и отслеживания кисти человека в видеопотоке на сложном фоне, обладающий более высокой надежностью и устойчивостью работы по сравнению с известными из литературы аналогами.

Разработан алгоритм и вычислительно-эффективная модель для распознавания жестов, основанная на нечетких конечных автоматах, сложность распознавания с помощью которой составляет $O(mn)$, где m - количество нечетких автоматов используемых для распознавания, n - количество состояний нечеткого конечного автомата.

Разработана методология мультимодального распознавания сцен, определяемых жестами, с использованием нечетких операторов агрегирования. Методология позволяет повысить надежность распознавания жестов за счет использования дополнительных источников информации, учесть степень важности каждой модальности, непосредственно в процессе иерархического распознавания сцен.

В работе предложен алфавит жестов, позволяющий широко использовать его в различных приложениях для создания интерфейсов человек-компьютер.

Экспериментально показано, что предложенная архитектура системы распознавания динамических жестов позволяет с высокой степенью

надежности распознавать в реальном времени жесты независимо от индивидуума.

Практическая значимость и реализация. На основе разработанных алгоритмов создано программное обеспечение захвата и отслеживания и распознавания жестов, позволяющее использовать его в различных человеко-машинных интерфейсах на основе жестов. Материалы работы используются в учебном процессе кафедры информационных систем и телекоммуникаций МГТУ им. Н.Э. Баумана в курсе «Обработка изображений в информационных системах».

Программное обеспечение реализовано на персональном компьютере. Для захвата и отслеживания кадра используется Web-камера. Программное обеспечение системы написано на языке программирования C++ в объектно-ориентированной нотации. Документация программной реализации удовлетворяет требованиям, предъявляемым к программным продуктам ГОСТ 19.105–78.

Система имеет следующие характеристики:

1. Скорость работы в реальном времени (15 кадров в секунду).
2. Устойчивость к шуму, характерному для недорогих, «домашних» видеокамер (Web-камер).
3. Функционирование с кадрами низкого 320×240 пикселей/8 бит, и среднего 640×480 пикселей/8 бит качества.

Программный модуль распознавания жестов прошел экспериментальную проверку в системе, обеспечивающий интуитивный интерфейс между человеком и телевизором, разработанной в соответствии с генеральным соглашением между МГТУ им. Н.Э. Баумана и компанией «NXP Semiconductors founded by Philips».

Публикации. Основные результаты работы изложены в семи научных публикациях, из них в журналах по списку ВАК - 1.

Апробация результатов работы:

1. Презентация научных исследований в области интуитивного интерфейса для инженеров компании Philips, 10.07.2006, Россия, г. Москва.
2. Доклад по результатам научно-исследовательского проекта в соответствии с генеральным соглашением, 19.12.2006, Нидерланды, г. Эйндховен.
3. 2-й всероссийской конференции «Теория динамических систем в приоритетных направлениях науки и техники», 25.06.2007, Россия, г. Ижевск.
4. 23-й международной конференции робототехника и заводы будущего CARS & FOF 07, 17.08.2007, Колумбия, г. Богота.
5. 2-й международной конференции «Системный анализ информационные технологии» САИТ-2007,10.09.2007, Россия, г. Обнинск.

6. 30-й конференции молодых ученых и специалистов ИППИ РАН «Информационные технологии и системы» ИТиС-2007, 18.09.2007, Россия, г. Звенигород.
7. 16-й международной конференции в Центральной Европе по компьютерной графике, визуализации и компьютерному зрению WSCG08, 06.02.2008, Чехия, г. Пльзень.

Структура и объем работы. Диссертационная работа состоит из введения, пяти глав, заключения и списка литературы, занимающих 165 страниц текста, в том числе 41 рисунок на 35 страницах, 13 таблиц на 21 странице, список литературы на 11 страницах.

Научные положения, выносимые на защиту:

1. Итоги сравнительного анализа моделей и методов распознавания динамических жестов.
2. Алгоритм захвата и отслеживания кисти человека на сложном фоне.
3. Нечеткая модель для распознавания динамических жестов, основанная на нечетких конечных автоматах и алгоритм распознавания динамических жестов с использованием этой модели.
4. Методология мультимодального распознавания сцен, определяемых динамическими жестами, с использованием нечетких операторов агрегирования.
5. Архитектура системы распознавания жестов человека и экспериментальные результаты работы системы на доверительных выборках.

2. СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность работы, сформулированы цель, задачи исследования и научные положения, выносимые на защиту. Приведена структура диссертации, формы апробации и реализации результатов.

В первой главе проведен обзор существующих классификаций и алфавитов жестов, наиболее часто используемых при общении людей, как ограниченных, так и неограниченных по здоровью. Анализ классификаций и алфавитов показал, что они не подходят для использования в человеко-машинных интерфейсах на базе жестов, так как содержат много специфической информации о тональности общения, зонах человеческого тела, интенсивности переживаний.

В зависимости от способа выполнения жесты были разделены на статические и динамические жесты. Статические жесты выполняются заданием определенного положения кисти и пальцев в пространстве, вне зависимости от времени. Динамические жесты выполняются движением какой-либо части человеческого тела, чаще всего кистью, во времени и пространстве. Описан выбранный алфавит динамических жестов, созданный

с ориентацией на создание человеко-машинного интерфейса, включающий в себя все базовые жесты языка глухонемых, выполняющиеся в одно движение, имеющие форму геометрических фигур или букв латинского алфавита (рис. 1). Продемонстрировано, что с помощью данных жестов можно кодировать любой жест из международного дактильного алфавита для слепоглухих Кармела.

Кроме того, с помощью динамического жеста можно более естественно передать многие команды управления интерфейсом, такие как: вверх, вниз, левее, поворот. Динамические жесты легче распознать при неоднородной текстуре фона и различном освещении. И поскольку динамические жесты осуществляются во времени, то появляется возможность наблюдать за параметрами жеста как функциями времени, что дает дополнительную информацию, увеличивающую надежность распознавания.

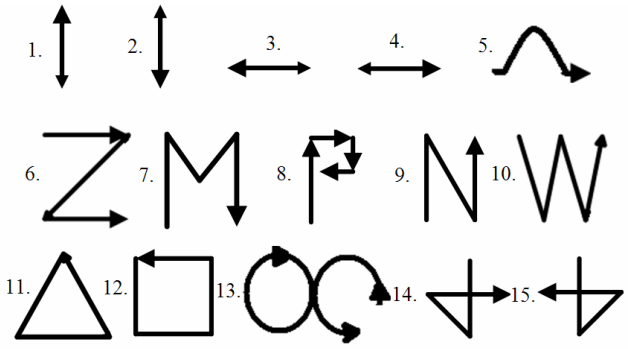

Рис. 1. Алфавит динамических жестов

Проведен анализ популярных математических моделей распознавания жестов. Показано, что оценки вычислительной сложности распознавания жестов с помощью этих моделей зависят от квадрата числа состояний (СММ), нейронов (нейросети) или вершин (Байесовы сети), используемых для распознавания, умноженных на число символов наблюдаемой последовательности. Вследствие этого, с ростом этих величин, практическое использование для распознавания жестов в реальном времени указанных моделей, из-за высоких затрат процессорного времени и памяти компьютера, становится невозможным. Тем самым обоснована необходимость создания новых вычислительно-эффективных методов.

Проведен обзор современных систем распознавания динамических жестов человека. Выделены три основные области применения подобных систем: интуитивный интерфейс человек-компьютер, автоматический перевод жестов глухонемых, приложения виртуальной реальности.

Во второй главе проведен анализ основных алгоритмов захвата и отслеживания области интересов: алгоритмов основанных на анализе перемещения, алгоритмов основанных на анализе цвета, алгоритмов основанных на анализе характерных признаков Хаара. Описаны два новых разработанных алгоритма: алгоритм захвата и отслеживания областей

интересов на сложном фоне, основанный на последовательном выделении объектов по перемещению, цвету и кластерам и алгоритм, основанный на параллельном каскадном детекторе, с использованием характерных признаков Хаара.

Основная идея первого алгоритма заключается в последовательной обработке и кластерном разбиении области интересов. Рассмотрим данный алгоритм по шагам. Кадр, получаемый видеокамерой в момент времени t и имеющий по горизонтали V, а по вертикали W пикселей, обозначим $I_t(V,W)$. Под областью интересов $Ob_t(X,Y)$ понимается множество пикселей кадра $I_t(V,W)$, очерчивающих искомый объект. *Захватом* области интересов называется выделение ее в кадре в момент времени t. *Отслеживанием* области интересов называется процесс последовательного захвата в кадрах $I_t(V,W)$, $I_{t+1}(V,W)$,..., $I_{t+k}(V,W)$ областей интереса $Ob_t(X_t,Y_t)$, $Ob_{t+1}(X_{t+1},Y_{t+1})$,..., $Ob_{t+k}(X_{t+k},Y_{t+k})$.

Шаг 1. На первом шаге алгоритма кадр $I_t(V,W)$ поступает от Web-камеры с разрешением 640×480/320×240, 8 бит.

Шаг 2. Второй шаг алгоритма – это фильтрация входного кадра $I_t(V,W)$. Для выполнения фильтрации было отдано предпочтение медианному фильтру перед фильтром Гаусса. Так как, при применении фильтра Гаусса в кадре получалось некоторое “размытие” областей изображения, шум которых описывался распределением с нулевым математическим ожиданием.

Шаг 3. На третьем шаге алгоритма, используя кадры $I_t(V,W)$ и $I_{t+1}(V,W)$ и применяя алгоритм, основанный на вычитании соседних кадров, находится перемещающийся объект и осуществляется захват области интересов $Ob_{t+1}(X_{t+1},Y_{t+1})$. Для выделения перемещающегося объекта используется фактор изменения яркости пикселей, относящихся к перемещающемуся объекту, в последовательности двух смежных $I_t(V,W)$ и $I_{t+1}(V,W)$ кадров. Если разность яркостей пикселей превышает заданный порог, то этот пиксель кадра $I_{t+1}(V,W)$ считается принадлежащим перемещающемуся объекту $Ob_{t+1}(X,Y)$. В данном случае порог найден экспериментально и равен 20. Для получения явных значений яркости пикселя, каждый кадр переводился из цветового пространства *RGB* в полутоновое цветовое пространство.

Шаг 4. Чтобы выделить в полученной области интересов $Ob_{t+1}(X,Y)$ только изображение кисти, на четвертом шаге ищутся пиксели, значение цвета которых совпадает со значением цвета кожи человека. В цветовом пространстве *HSV* (параметры *H*, *S*, *V* соответственно обозначают *Hue* (тон), *Saturation* (насыщенность), *Volume* (яркость)) для цвета кожи значения параметра *H* лежат в промежутке от 18 до 22, параметра *S* от 5 до 10.

Шаг 5. На пятом шаге алгоритма наложением четырехсвязной маски, отсекаются одиночные пиксели. Эти пиксели считаются шумом. Этих шагов достаточно, чтобы захватить и отследить только одну кисть человека

находящуюся в кадре. Но, чтобы найти правую, левую кисть и лицо пользователя управляющего системой издалека, необходимо применить алгоритм кластеризации *c*-средних.

Шаг 6. На шестом шаге, алгоритм кластеризации разбивает пиксели найденной области интересов $Ob_{t+1}(X_{t+1},Y_{t+1})$, принадлежащие перемещающемуся объекту и распределению цвета кожи, на три кластера, соответственно: правая кисть, левая кисть, лицо (рис. 2).

Шаг 7. На заключительном, седьмом шаге алгоритма для каждого кластера вычисляется центр тяжести. Путем отслеживания этих центров, строятся траектории перемещения кистей, с помощью которых распознается выбранный алфавит жестов для каждой руки.

Рис. 2. (а) Входной кадр, (б) Найдены пиксели области интересов $Ob_{t+1}(X_{t+1},Y_{t+1})$, принадлежащие перемещающемуся объекту и распределению цвета кожи человека,
(в) Найдены и маркированы три кластера: правая кисть, левая кисть, лицо

В конце главы описан второй разработанный алгоритм захвата и отслеживания областей интересов, основанный на параллельном каскадном детекторе, с использованием характерных признаков Хаара, который способен, как и рассмотренный алгоритм, отслеживать кисти и лицо человека в видеопотоке.

Параллельный каскадный детектор состоит их трех параллельных каскадов. Каждый каскад это цепь классификаторов, основанных на характерных признаках Хаара. Каскады, состоящие из пятнадцати классификаторов, предназначены для захвата и отслеживания правой и левой кистей человека, каскад из двадцати восьми классификаторов предназначен для захвата и отслеживания лица человека. Кроме своей структуры, оригинальной особенностью параллельного каскадного детектора, является обучение его на специально сформированной обучающей выборке. Выборка состояла из изображений кисти, захваченных при разных условиях освещения, что позволило в дальнейшем повысить устойчивость работы детектора в неконтролируемых условиях.

Найденная верхняя оценка вычислительной сложности разработанных алгоритмов составила $O(N)$, где N – количество пикселей кадра $I_t(V,W)$.

В третьей главе предлагается нечеткая модель для распознавания динамических жестов, основанная на нечетких конечных автоматах и алгоритм распознавания динамических жестов с использованием этой модели.

Для формирования нечеткой модели каждого динамического жеста, последний многократно повторяется и траектория каждого повторения фиксируется. Число повторений обычно равно 10-20. Так, например, для жеста, имеющего вид буквы «Z», траектории показаны на рис. 3а.

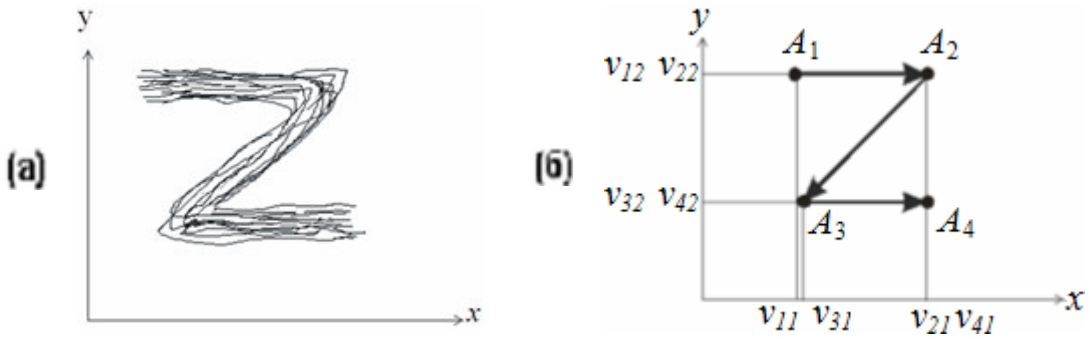

Рис. 3. (а) Траектории жеста, повторенного несколько раз,
(б) Граф жеста буквы «Z»

Обобщенно траекторию перемещения всех жестов, имеющих вид буквы «Z» можно представить в виде графа, показанного на рис. 3б. Вершина A_1 этого графа объединяет множество точек, принадлежащих началу траекторий, вершины A_2 и A_3 соответствуют множествам точек перегиба траекторий, вершина A_4 объединяет множество точек концов траекторий, дуги графа указывают на направление перемещения центра тяжести объекта по траекториям. Этот граф может служить основой для построения нечеткой модели жеста. Каждая вершина графа объединяет характерные точки с определенным сходством. Множество точек, относящихся к одной вершине, составляют кластер. Каждая точка в общем случае принадлежит m-мерному пространству и является набором значений характерных признаков y_1, y_2,…, y_m. Для нахождения кластеров использовался алгоритм кластеризации c-средних.

Для того чтобы можно было учесть время, вместо графа на рис. 3б, (для случая двумерного пространства) использовались два графа, полученные в результате проекции траекторий перемещения центров тяжести кисти руки на ось абсцисс и ось времени, а также на ось ординат и ось времени (рис. 4). В общем случае m-мерного пространства таких проекций $y_i(t)$ будет m: (i=1,…,m).

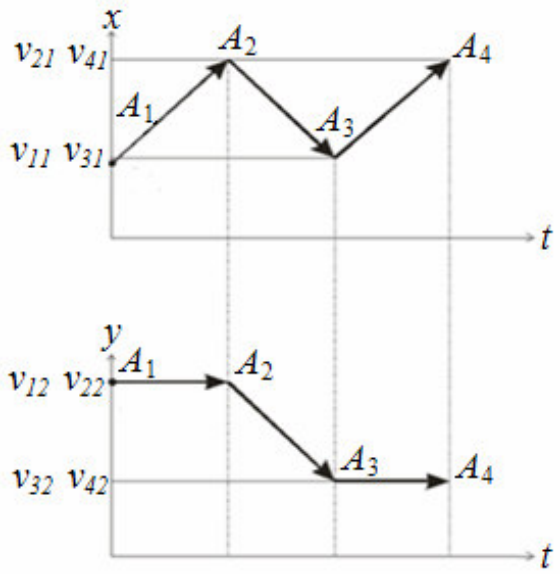

Рис. 4. Проекции графа жеста

Значение $y_i(t)$ в некоторый момент времени, будем называть *отсчетом* $y_i(t)$. Последовательность $n+1$ отсчетов $Y_i[t_0, t_n] = \{y_i(t_0), y_i(t_1), y_i(t_2), \ldots, y_i(t_n)\}$ i-ой проекции одного и того же жеста в течение нескольких подряд идущих моментов времени $t_0, t_1, t_2, \ldots, t_n$ (в течение временного интервала $[t_0, t_n]$) назовем *сигналом*. Сопоставим каждому отсчету $y_j(t_i)$ одного и того же сигнала состояние $b_j(t_i)$ конечного автомата M_j. Введем функцию выходов φ конечного автомата M_j: $\varphi(b_j(t_i)) = y_j(t_i)$ и функцию переходов f: $f(b_j(t_i), t_{i+1}) = b_j(t_{i+1})$.

Таким образом, каждый отсчет является значением функции выхода $y_j(t) = \varphi(b_j(t))$ автомата M_j, каждый сигнал является последовательностью значений функций выхода $\mathbf{y}_j(t) = (y_j(t_0), y_j(t_1), \ldots, y_j(t_n))$ одного и того же автомата M_j.

Представим автомат M_j, соответствующий какой-либо проекции некоторого жеста, его графом переходов (рис. 5а). Каждая вершина графа помечена символом b_i, $i = 0, 1, \ldots, 12$. (вершины обозначены кружками). Каждая пара соседних вершин b_i, b_{i+1}, i=0, 1, 2,…, 11 соединена дугой, направленной от вершины i к вершине i+1. Дуги, направленные от вершины i к вершине i+1 помечены символом t_i в алфавите $T = \{t_0, t_1, t_2, t_3, \ldots, t_{m-1}\}$. Если выписать обозначения всех дуг слева направо, то получим последовательность символов $t_1t_2t_3t_4t_5t_6t_7t_8t_9t_{10}t_{11}t_{12}\Lambda$ (здесь Λ - пустой символ, который может опускаться). Эта последовательность может рассматриваться как слово или предложение некоторого языка $L = L(G)$, порождаемого автоматной грамматикой $G_j = \{V, T, P, S = b_0\}$, $V = \{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10}, b_{11}\}$, $T = \{t_0, t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9, t_{10}, t_{11}, \Lambda\}$, $P = \{b_0 \to t_1b_1, b_1 \to t_2b_2, b_2 \to t_3b_3, b_3 \to t_4b_4, b_4 \to t_5b_5, b_5 \to t_6b_6, b_6 \to t_7b_7, b_7 \to t_8b_8, b_8 \to t_9b_9, b_9 \to t_{10}b_{10}, b_{10} \to t_{11}b_{11}, b_{11} \to t_{12}\, b_{12}, b_{12} \to \Lambda\}$.

Каждой дуге графа соответствуют две инцидентные вершины b_i и b_{i+1}. Координатой вершины b_i на оси абсцисс является t_i и $\varphi(b_i(t_i)) = y_i(t_i)$, а координата вершины b_{i+1} на оси абсцисс есть t_{i+1} и $\varphi(b_{i+1}(t_{i+1})) = y_{i+1}(t_{i+1})$.

Полагаем, что отсчеты $y_i(t_i)$ одного и того же кластера, соответствующие l различным траекториям одного и того же жеста, могут изменяться в пределах среднеквадратичного отклонения s_i от проекции центра кластера $v_i(t_i)$: $s_i = \sqrt{\dfrac{\sum_{l=1}^{N}[y_i^l(t_i) - v_i(t_i)]^2}{N}}$,

где N-число отсчетов принадлежащих кластеру, v_i - координата центра i-ого кластера, $y_i^l(t_i)$ отсчет, принадлежащий *i-ому* кластеру. Для простоты полагаем, что s_i одно и то же для всех i и равно s. Для каждого множества отсчетов $y_i^l(t_i)$ задаем треугольную функцию принадлежностей $\mu_i(y)$, определяемую точками, $y_i^- = v_i - s$, $y_i = v_i$, $y_i^+ = v_i + s$, причем $\mu_i(y_i^-) = 0$, $\mu_i(y_i) = 1$, $\mu_i(y_i^+) = 0$ (рис. 5б).

Вершину b_i с координатами t_i, y_i заменим множеством вершин $b_{ri} \in B(b_i)$ с координатами, изменяющимися по оси ординат в пределах области $y_i^- = y_i - s$, $y_i^+ = y_i + s$. Каждая вершина b_{ri} соответствует какому-либо пикселю, а множество $B(b_i)$ вершин (пикселей) вычисляется как множество всех пикселей, с координатой t_i по оси ординат. Вместо одной дуги (b_i, b_{i+1}) теперь будем иметь множество дуг $\{(b_{ri}, b_{r(i+1)}) \mid b_{ri} \in B(b_i),\ b_{r(i+1)} \in B(b_{i+1})\}$, соединяющих каждую вершину множества $B(b_i)$ с каждой вершиной множества $B(b_{i+1})$ и помеченных тем же символом t_{i+1}, что и дуга (b_i, b_{i+1}).

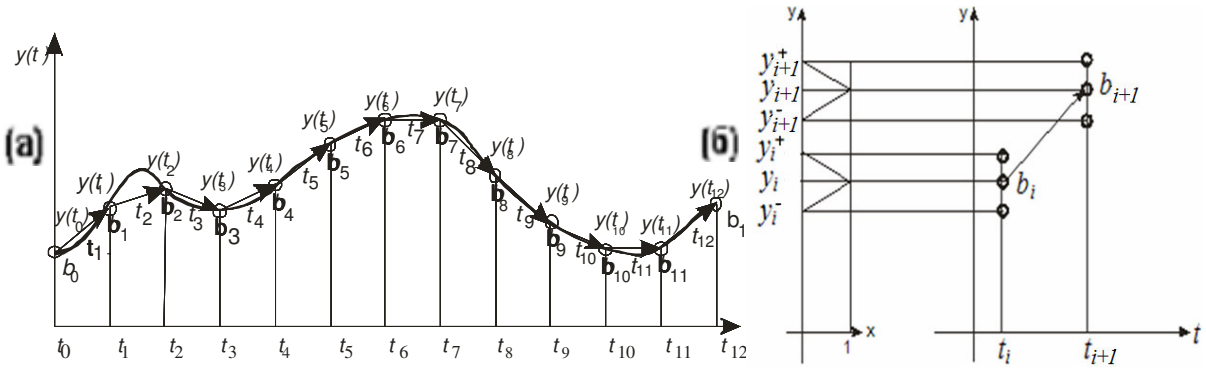

Рис. 5. (а) Граф переходов четкого автомата, (б) Функции принадлежности вершин, графа переходов нечеткого автомата

Будем полагать, что функция принадлежности каждой дуги $(b_{ri}, b_{r(i+1)})$, инцидентной вершинам $b_{ri} \in B(b_i)$ и $b_{r(i+1)} \in B(b_{i+1})$, для которых $\varphi(b_{ri}) = y_{ri}$, $\varphi(b_{r(i+1)}) = y_{r(i+1)}$ определяется как: $\mu_{(b_{ri}, b_{r(i+1)})}(t_{i+1}) = \min\{\mu_i(y_{ri}), \mu_{i+1}(y_{r(i+1)})\}$.

Нечеткая грамматика $G_F=\{V,\ T,\ P_F,\ S_F\}$ получается из четкой грамматики $G=\{V,\ T,\ P,\ S\}$ следующим образом. Единственный начальный нетерминальный символ четкой грамматики заменяется множеством начальных нетерминальных символов: $S_F = B(b_0)$. Множество правил P_F нечеткой грамматики G_F будет следующим:

$$P_F = \{b_{ri} \to t_{i+1}\, b_{r(i+1)},\ \mu(b_{ri} \to t_{i+1}\, b_{r(i+1)}) = \mu_{(b_{ri}, b_{r(i+1)})}(t_{i+1}),\ i=0,\ldots,\ n\text{-}1\ \}.$$

Для каждого жеста k=1,…, K строится множество нечетких эталонных грамматик $G^k_{F1},\ G^k_{F2},\ldots,\ G^k_{Fm}$, базируясь на изложенных принципах. Будем говорить, что четкая грамматика G', содержащая правила $\{b_i \to t_{i+1} b_{i+1},\ i=0,\ldots,\ n\text{-}1\}$, сравнима с нечеткой грамматикой G_F, если существует последовательность правил $\{b_{ri} \to t_{i+1} b_{r(i+1)},\ i=0,\ldots,\ n\text{-}1\}$ этой нечеткой грамматики, для которых имеет место $b_i=b_{ri}$ для всех $i=0,\ldots,n\text{-}1$. Алгоритм распознавания динамических жестов с использованием модели, основанной на нечетких конечных автоматах и соответствующем им множестве эталонных нечетких грамматик $G_{F1},\ G_{F2},\ldots,\ G_{Fm}$, будет следующим.

Шаг 1. Распознаваемый жест обрабатывается с теми же шагами дискретизации по временной оси, что и эталонные жесты, и для него строится множество четких грамматик $G_1\ G_2,\ \ldots,\ G_m$, ему соответствующих.

Шаг 2. Осуществляется сравнение четких грамматик $G_1\ G_2,\ \ldots,\ G_m$, соответствующих распознаваемому жесту, с каждой соответствующей нечеткой эталонной грамматикой $G^k_{F1},\ G^k_{F2},\ldots,\ G^k_{Fm}$. Здесь $k \in \{1,\ldots,\ K\}$, а K-число распознаваемых жестов.

Шаг 3. Для тех множеств нечетких эталонных грамматик $G^k_{F1},\ G^k_{F2},\ldots,\ G^k_{Fm}$ сравнение с которыми оказалось успешным, вычисляется соответствующее множество значений функций принадлежности $\mu_{G_1,G^k_{F1}},\ \mu_{G_2,G^k_{F2}},\ \ldots,\ \mu_{G_m,G^k_{Fm}}$ по формуле: $\mu_{G_j,G^k_{Fj}} = \min\limits_{j\in\{1\ldots m\}}\{\mu_{j(b_{ri}b_{r(i+1)})}(t_{i+1})\}$, а затем значение меры A_k, характеризующей близость распознаваемого жеста к эталонным жестам k по формуле:

$$A(G,G_k) = A_k = \max\{\mu_{G_1,G^k_{F1}},\ \mu_{G_2,G^k_{F2}},\ \ldots,\ \mu_{G_m,G^k_{Fm}}\}.$$

Шаг 4. Распознаваемый жест считается совпадающим с тем эталонным жестом k, для которого значение меры A_k оказалась максимальным.

Шаг 5. Если не было ни одного успешного сравнения грамматик, то распознавание этого жеста заканчивается неудачей (жест не был распознан).

Вычислительная сложность распознавания динамических жестов с помощью нечетких моделей равна $O(mn)$, где m - количество нечетких автоматов, n - максимальное количество состояний нечеткого конечного автомата используемого для распознавания.

В четвертой главе предлагается методология мультимодального распознавания сцен, определяемых жестами, с использованием операторов агрегирования (операторов Суджено или Шоке), использующих нечеткую меру. В общем случае каждый кадр $I_t(V,W)$ может содержать L объектов

$\theta_l, l = 1,..., L$, подлежащих распознаванию. Распознавание объекта θ_l состоит в вычислении множества значений операторов агрегирования $A_1, A_2,..., A_K$ по множеству значений функций принадлежности $\mu_1(y_1), \mu_2(y_2),..., \mu_m(y_m)$, вычисляемых для распознаваемого объекта θ_l, где $y_1 \in Y_1, y_2 \in Y_2,..., y_m \in Y_m$, а $Y_1, Y_2,..., Y_m$ - множество модальностей, характеризующих объект θ_l. Объекты различных множеств Θ_l могут находиться в определенных, в общем случае *r*-арных отношениях $\Xi \in \Theta_{l_1} \times \Theta_{l_2} \times ... \times \Theta_{l_r}$, $\{l_1, l_2,..., l_r\} \subseteq \{1,..., L\}$. Каждое такое отдельное отношение Ξ_1 будем называть *сценой* 1-го уровня. Сценами *s*-го уровня будем называть сцены $\Xi_s \in \Xi_{s-1} \times \Xi^1_{s-1} \times ... \times \Xi^v_{s-1}$, где $\Xi_{s-1} \times \Xi^1_{s-1} \times ... \times \Xi^v_{s-1}$ - сцены (*s*-1)-го уровня.

Таким образом, методология мультимодального распознавания сцен, определяемых жестами, состоит из следующих шагов.

Шаг 1. Каждый объект θ_{l_1}, θ_{l_2}, ..., θ_{l_r}, входящий хотя бы в одну сцену первого уровня Ξ_1 распознается отдельно сопоставлением соответственно с эталонными объектами $\Theta^{k_{l1}}{}_{l_1}, \Theta^{k_{l2}}{}_{l_2},..., \Theta^{k_{lr}}{}_{l_r}$, $k_{l_r} = 1,..., K_{l_r}$, $\{l_1, l_2,..., l_r\} \subseteq \{1,..., L\}$ с помощью операторов агрегирования $A_{l_1}, A_{l_2},..., A_{l_r}$.

Шаг 2. Каждая сцена 1-го уровня Ξ_1, для всех объектов которой найдены сходные с ними эталонные объекты, считается распознанной и для нее вычисляется значение оператора агрегирования A_1. После этого осуществляется переход к шагу 3. Если таких сцен не найдено, то распознанных сцен 1-го уровня и выше не существует и выполнение прекращается.

Шаг 3. Задается значение уровня *s*=2 и осуществляется переход к шагу 4.

Шаг 4. Если найдены сцены *s*-го уровня Ξ_s для всех сцен (*s*-1)-го уровня которых найдены ненулевые значения операторов агрегирования, то эти сцены Ξ_s считаются распознанными, для них вычисляются значения операторов агрегирования A_s. Если существуют сцены уровня *s*+1, то шаг 4 снова выполняется со значением *s*=*s*+1, в противном случае выполнение прекращается.

В пятой главе рассматривается архитектура системы распознавания динамических жестов (рис. 6). В *блоке формирования графов жестов (ФГЖ)* на основе рассмотренных алгоритмов захвата и отслеживания области интересов создается граф выполненного жеста. Сюда также включены алгоритмы захвата и отслеживания простых объектов (квадрат, прямоугольник, окружность) в кадре $I_t(V,W)$ и алгоритмы распознавания человека, использующиеся в блоке распознавания сцены.

На этапе обучения системы полученный граф жеста поступает на вход *блока обучения*, который формирует нечеткие конечные автоматы и

множества нечетких эталонных грамматик G^k_{F1}, G^k_{F2},…, G^k_{Fm}, k=1,…, K, в соответствии с рассмотренной методологией. Нечеткие конечные автоматы, множество нечетких эталонных грамматик и ряд настроечных параметров системы сохраняются в *базе знаний*.

На этапе распознавания, граф жеста, сформированный блоком ФГЖ, обрабатывается в *блоке распознавания жестов*. В этом блоке осуществляется распознавание жестов с помощью эталонных нечетких моделей из базы знаний. Если распознавание закончилось успешно, то *блок принятия решений* выдает управляющее воздействие, в зависимости от типа распознанного жеста.

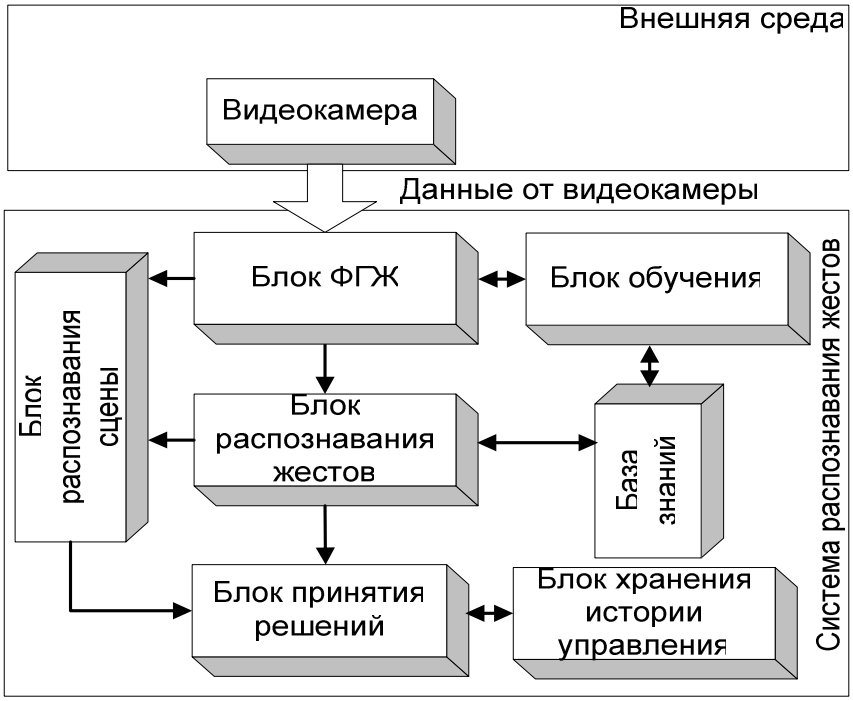

Рис. 6. Архитектура системы распознавания жестов

В *блоке распознавания сцены* на основе рассмотренной методологии мультимодального распознавания, базирующейся на нечетких операторах агрегирования, проводится распознавание сцен *s*-го уровня. Для того чтобы провести распознавание сцен первого уровня, блок получает результаты распознавания объектов сцены из блока распознавания жестов и блока ФГЖ. Результаты работы блока распознавания сцены могут влиять на принимаемое решение об управляющем воздействии.

В *блоке хранения истории управления* сохраняется последовательность распознанных жестов и соответствующих им управляющих воздействий за определенное время, в частности с целью интерпретации принятых решений по управлению. Все сцены и жесты, распознанные ранее этого периода, утрачиваются.

В пятой главе с системой распознавания жестов проводится серия экспериментов. В первых экспериментах были найдены параметры, при которых алгоритмы захвата и отслеживания области интересов достигали лучших результатов отслеживания кисти человека. В следующих экспериментах было проведено сравнение результатов работы обоих алгоритмов. Для этого были найдены значения надежности и устойчивости работы алгоритмов. В общем случае, под *надежностью* понимается процент успешных захватов из числа всех попыток. *Устойчивость* это процент равный разности ста процентов и процента коэффициента ложных захватов. *Коэффициент ложных захватов* это процент ложных захватов из числа всех попыток.

Выяснено, что надежность, устойчивость и время работы алгоритма, основанного на последовательном выделении объектов по перемещению, цвету и кластерам, в среднем равна 93%, 99.74%, 56 мс (для кадра разрешением 320×240 пикселей) соответственно. Данный алгоритм показал более высокую устойчивость к захвату и отслеживанию кисти под разными углами, по сравнению алгоритмом, основанном на параллельном каскадном детекторе. Поэтому этот алгоритм был использован, как основной алгоритм в блоке формирования графов жестов системы.

Для блока распознавания жестов системы были проведены следующие эксперименты. 1. Нахождение надежности распознавания жеста, выполняемого одной рукой одним человеком. 2. Нахождение надежности распознавания жестов, выполняемых двумя руками по очереди одним человеком (рис. 7). 3. Нахождение надежности распознавания жестов, выполняемых одной рукой различными людьми.

Основное отличие третьего эксперимента определения надежности распознавания в том, что система обучалась одним пользователем, а тестировалась группой других пользователей.

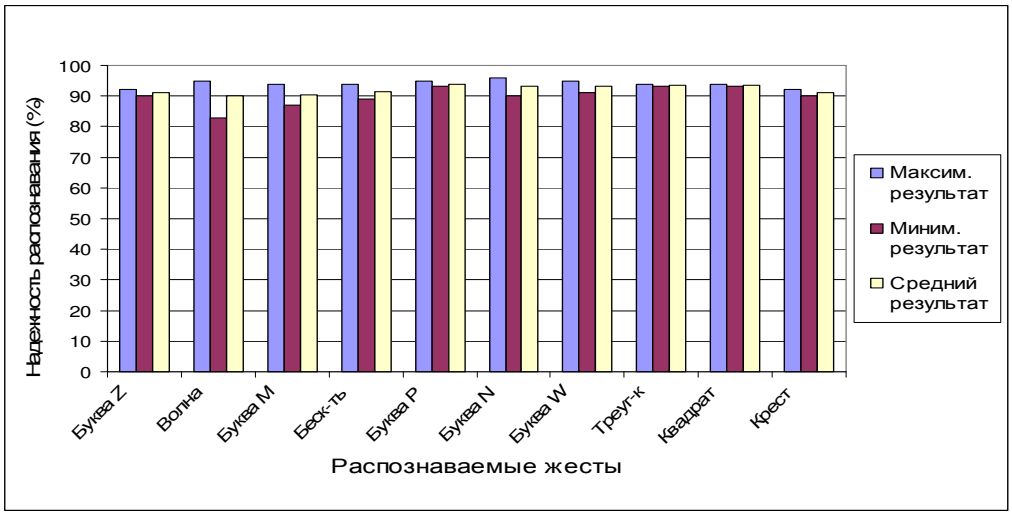

Рис. 7. Результаты экспериментов по распознаванию жестов, выполняемых двумя руками по очереди одним человеком

Надежность распознавания, в проведенных экспериментах, превышает девяносто процентов, что позволяет успешно использовать систему распознавании для реального интерфейса человек-компьютер. В таком интерфейсе данные о распознанных динамических жестах могут быть использованы как команды управления программным обеспечением компьютера, заменяя интерфейс, основанный на использовании клавиатуры и мыши.

В заключении сформулированы основные результаты, полученные в работе.

3. ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Проведен сравнительный анализ моделей и методов захвата, отслеживания и распознавания динамических жестов, пригодных для создания человеко-машинных интерфейсов, по критериям вычислительной сложности. На основе сравнительного анализа выявлены недостатки существующих методов, обоснована необходимость создания новых вычислительно-эффективных методов.

2. Выполнена классификация жестов с точки зрения удобства построения на их базе человеко-машинного интерфейса. Разработан алфавит динамических жестов, исключающий двусмысленность передаваемой информации за счет выбора жестов, не использующихся в обычном общении и состоящих из базовых жестов языка немых, интуитивно понятных пользователю.

3. Разработан новый алгоритм захвата и отслеживания кистей человека на сложном фоне. Алгоритм не требует дополнительных маркеров на теле человека выполняющего жест, захватывает кисти в помещении с различным фоном и освещением, в реальном времени, с вычислительной сложностью $O(N)$, высокой надежностью (93%) и устойчивостью (99.74%).

4. Разработан алгоритм и модель для распознавания динамических жестов, основанная на нечетких конечных автоматах. Главными преимуществами нечеткой модели является возможность строить распознаватель, имея всего несколько примеров в обучающей выборке, строить нечеткие автоматы разной длины, распознавать жесты с траекторией, содержащей пересечения, распознавать жесты с надежностью не менее 90%, и вычислительной сложностью $O(mn)$, где m - количество нечетких автоматов, n - максимальное количество состояний нечеткого конечного автомата используемого для распознавания.

5. Предложена новая методология мультимодального распознавания сцен, определяемых жестами, основанная на нечетких операторах агрегирования. Методология позволяет учитывать степень важности каждой модальности в процессе иерархического распознавания сцен, расширять интеллектуальность интерфейса системы, задавать сцены на основе

статических и динамических объектов, повышать надежность распознавания отдельных объектов (например, жестов) на заданной сцене за счет использования отношений между этими объектами и другими объектами сцены (фоновыми объектами).

6. Разработана архитектура программной системы распознавания динамических жестов независимо от индивидуума. Проведено экспериментальное апробирование системы, подтвердившее теоретические ожидания высокой надежности, устойчивости и скорости работы алгоритмов, пригодность их для создания реальных человеко-машинных интерфейсов на базе динамических жестов.

Основные результаты диссертации изложены в следующих работах:

1. Алфимцев А.Н. Логико-вероятностный подход к построению Экспертной системы на основе Нейронных и Байесовых сетей // Прогрессивные технологии, конструкции и системы в приборо- и машиностроении: Сб. трудов Всерос. конф.-М., 2004.-Т. 3.- С. 35-37.
2. Алфимцев А.Н. Современные тенденции принятия управляющих решений на основе распознавания жестов // Информационные технологии и системы: Сб. трудов Всерос. конф.- М., 2007.- С. 152-157.
3. Девятков В.В., Алфимцев А.Н. Распознавание динамических жестов // Применение теории динамических систем в приоритетных направлениях науки и техники: Сб. трудов Всерос. конф.- Ижевск, 2007.- С. 15-23.
4. Девятков В.В., Алфимцев А.Н. Распознавание манипулятивных жестов // Вестник МГТУ им. Н.Э.Баумана. Сер. Приборостроение.- 2007.- Т. 68, № 3.- С.56-75.
5. Девятков В.В., Алфимцев А.Н. Параллельный захват и отслеживание динамических жестов руки // Системный анализ и информационные технологии: Сб. трудов Межд. конф.- М., 2007.- С. 89-94.
6. Devyatkov V., Alfimtsev A. Gesture-based control of telerobots // Proc. of 23rd ISPE International Conference on CARS & FOF 07.- Bogota, 2007.- P. 59-67.
7. Devyatkov V., Alfimtsev A. Optimal Fuzzy Aggregation of Secondary Attributes in Recognition Problems // Proc. of 16-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision.- Plzen, 2008.- P. 33–41.