

*На правах рукописи*

Стародубцев Илья Сергеевич

**МОДЕЛИ, АЛГОРИТМЫ И ПРОГРАММНЫЙ КОМПЛЕКС  
ДЛЯ ПОСТРОЕНИЯ ЕСТЕСТВЕННОГО  
ЧЕЛОВЕКО-КОМПЬЮТЕРНОГО ВЗАИМОДЕЙСТВИЯ НА  
ОСНОВЕ ЖЕСТОВ**

Специальность 05.13.18 — Математическое моделирование,  
численные методы и комплексы программ

Автореферат диссертации на соискание учёной степени  
кандидата физико-математических наук

Екатеринбург — 2015

Работа выполнена в ФГАОУ ВПО “Уральский федеральный университет имени первого Президента России Б.Н. Ельцина” на кафедре информатики и процессов управления Института математики и компьютерных наук.

Научный руководитель: кандидат технических наук, доцент  
Авербух Владимир Лазаревич.

Официальные оппоненты: Горбачев Дмитрий Викторович,  
доктор физико-математических наук, ФБГОУ  
ВПО «Тульский государственный университет»,  
профессор кафедры прикладной математики и  
информатики института прикладной математики  
и компьютерных наук.

Клименко Станислав Владимирович,  
доктор физико-математических наук, профессор,  
ФГАОУ ВПО «Московский физико-технический институт  
(государственный университет)», профессор кафедры физико-  
технической информатики.

Ведущая организация: ФГБУН Институт систем обработки изображений  
РАН, г. Самара.

Защита состоится 23 сентября 2015 г. в 14:00 на заседании диссертационного  
совета Д 212.285.25 на базе ФГАОУ ВПО “Уральский федеральный университет  
имени первого Президента России Б.Н. Ельцина” по адресу: 620000, г. Екатеринбург,  
пр. Ленина 51, зал заседаний диссертационных советов, комн. 248.

С диссертацией можно ознакомиться в библиотеке и на сайте ФГАОУ ВПО  
“Уральский федеральный университет имени первого Президента России  
Б.Н. Ельцина”, <http://dissovet.science.urfu.ru/news2/>

Автореферат разослан «\_\_\_» \_\_\_\_\_ 20\_\_ г.

Ученый секретарь диссертационного  
совета, доктор физико-математических  
наук, профессор



Пименов В.Г.

# 1 Общая характеристика работы

**Актуальность работы.** В современном мире наличие большого количества компьютеризированных систем порождает задачу человеко-компьютерного взаимодействия. Один из подходов к её разрешению состоит в создании естественных методов человеко-компьютерного взаимодействия – разработке естественных интерфейсов.

Использование различных видов естественных интерфейсов ставится необходимо в случаях, когда использование классических человеко-компьютерных интерфейсов становится невозможным, или связанным с определёнными затруднениями. Наиболее яркими примерами являются задачи, связанные с управлением медицинским оборудованием, навигацией роботизированных летательных и колёсных платформ и навигацией и взаимодействием с объектами в средах виртуальной и расширенной реальности.

Так, расширение компьютеризации в медицинской сфере привело к тому, что многие этапы диагностических исследований и операций проводятся при помощи сложной аппаратуры, управляющейся компьютерами. Однако часто в таких системах управления используются традиционные человеко-компьютерные интерфейсы, рассчитанные на классические устройства ввода. При этом, как правило, практически не учитывается специфика работы специалиста-оператора, в данном случае врача. В частности, чаще всего игнорируются требования *«зоны стерильности»*. Классические устройства ввода не подразумевают возможности стерилизации без внесения конструктивных изменений.

В это случае интерфейс, основанный на жестах может быть полезен, так как позволяет реализовывать бесконтактные методы человеко-компьютерного взаимодействия. На основе таких методов могут быть разработаны интерфейсы, которые могут как эмулировать классические устройства ввода (например, *бесконтактный тачскрин («touchless screen»)*, *виртуальная бесконтактная клавиатура*, *виртуальный джойстик*), так и реализовывать новые.

В случае управления роботизированными платформами использование жестов может помочь в создании более высокоуровневых методов взаимодействия, позволяя при помощи такого же количества команд контролировать большее число параметров управления.

**Целью диссертационной работы** является разработка моделей объектов, необходимых для захвата движения человека и реализация методов захвата движения для построения естественного человеко-компьютерного взаимодействия на основе жестов.

**Методы исследования, достоверность и обоснованность результатов.** Результаты, представленные в диссертационной работе, получены с применением методов математического моделирования, методов компьютерного зрения, теории множеств, методов обработки изображений и методов вычислительной математики. Достоверность и обоснованность полученных результатов подтверждается соответствующими математическими доказательствами, использовани-

ем апробированных общепринятых математических методов и согласованностью результатов, полученных различными методами.

В работе получены и **выносятся на защиту** следующие результаты.

1. Математическая модель силуэта антропоморфного гибкого объекта, используемого в распознавании.
2. Метод динамического распознавания позы человека.
3. Численный метод оценки количества движения в сцене на основе накопительных разностных буферов.
4. Метод захвата движения точечного объекта на основе карт глубин сцены.
5. Программно-аппаратный комплекс «**VirtualHand**», обеспечивающий возможности по созданию [человеко-компьютерных] интерфейсов, основанных на жестах.

**Научная новизна.** Основные результаты являются новыми, они обобщают и дополняют результаты отечественных и зарубежных исследований по данной теме.

**Теоретическая и практическая значимость работы** состоит в том, что разработанные методы и подходы могут быть применены при решении задач захвата движения, технического зрения и распознавания образов и визуального анализа последовательности изображений. Предложенный программный комплекс и вычислительные алгоритмы может быть применён при разработке средств [человеко-]компьютерного взаимодействия для сред виртуальной и расширенной реальности.

**Апробация работы.** Основные результаты, представленные в диссертации, докладывались и обсуждались ранее на различных, в том числе международных, научных конференциях и семинарах:

- на XII международном семинаре «Супервычисления и математическое моделирование» (Саров, 2010 г.);
- XIII международном семинаре «Супервычисления и математическое моделирование» (Саров, 2011 г.);
- на международном Российско-Корейском семинаре (Екатеринбург, УрФУ, 2011 г.);
- на XIV международной конференции «Супервычисления и математическое моделирование» (Саров, 2012 г.);
- на расширенном семинаре кафедры Прикладной математики и информатики ТулГУ (Тула, 2014 г.);

- на объединенном семинаре лаборатории математических методов обработки изображений Института систем обработки изображений ИСОИ РАН и Кафедры геоинформатики и информационной безопасности СГАУ (Самара, 2014 г.)
- на семинарах кафедры информатики и процессов управления Института математики и компьютерных наук УрФУ (г. Екатеринбург) и на семинаре сектора компьютерной визуализации Института математики и механики УрО РАН (г. Екатеринбург).

**Публикации.** По результатам диссертации лично автором и в соавторстве опубликованы 14 работ:

- 3 работы в российских рецензируемых научных журналах, рекомендованных ВАК [1–3],
- 11 работ в других журналах и материалах международных конференций [6–16],
- 2 свидетельства о государственной регистрации программ для ЭВМ в Роспатенте [4, 5].

**Структура и объем диссертации.** Диссертация состоит из списка основных обозначений и соглашений, пяти глав, списка литературы из 108 наименований и списка публикаций автора по теме диссертации, содержит 24 рисунка и 3 алгоритмических вставки на псевдокоде.

## 2 Краткое содержание работы

**Во введении** обосновывается актуальность работы, формулируются цели, отмечается новизна, научная и практическая ценность работы. Также в этой части проводится обзор развития данной области исследований.

**В первой главе** описываются математические модели, методы и алгоритмы захвата движения и обработки видимых карт глубин. Глава разделена на две части. **В первом разделе первой главы** описывается подход, реализующий «*point-based*» модель взаимодействия. Такая модель подразумевает слежение за одной или несколькими точками интереса без использования дополнительной контекстной информации о строении человеческого тела.

Вводятся определения *точки интереса*  $POI(R)$ , соответствующей реальному объекту  $R$ , а также *центрально-проекционного* и *ортогонально-проекционного кадра*, которые в последствии используются как аналог понятия изображения(кадра) видео последовательности, используемого в видеоаналитике и техническом зрении.

Для захвата движения точки интереса производится анализ количества движения в сцене при помощи накопительных разностных буферов вида:

1. Абсолютный разностный буфер.

$$A_k(x, y) = \begin{cases} A_{k-1}(x, y) + 1, & |b(x, y) - f_{t_{n-k}}(x, y)| > T_{buf}(x, y), \\ A_{k-1}(x, y), & |b(x, y) - f_{t_{n-k}}(x, y)| \leq T_{buf}(x, y), \end{cases} \quad (1)$$

2. Позитивный разностный буфер.

$$P_k(x, y) = \begin{cases} P_{k-1}(x, y) + 1, & (b(x, y) - f_{t_{n-k}}(x, y)) > T_{buf}(x, y), \\ P_{k-1}(x, y), & (b(x, y) - f_{t_{n-k}}(x, y)) \leq T_{buf}(x, y), \end{cases} \quad (2)$$

3. Негативный разностный буфер.

$$N_k(x, y) = \begin{cases} N_{k-1}(x, y) + 1, & (b(x, y) - f_{t_{n-k}}(x, y)) < -T_{buf}(x, y), \\ N_{k-1}(x, y), & (b(x, y) - f_{t_{n-k}}(x, y)) \geq -T_{buf}(x, y), \end{cases} \quad (3)$$

где

- $\{f_{t_i}\}_{i=1}^n$  – последовательность кадров;
- $b(x, y)$  – опорный кадр;
- $A_1(x, y) = P_1(x, y) = N_1(x, y) = 0$  для  $\forall x, y \in \mathbb{R}$ ;
- $T_{buf}(x, y)$  – пороговая функция;

При использовании такого способа анализа движения в кадре становится возможным применение класса жестов, названных *повторными*, для инициализации точки интереса. В качестве примера таких жестов можно привести циклические жесты, вовремя которых объект преодолевает какое-то расстояние в пространстве, после чего проходит или возвращается в точку, из которого жест начался.

Использование таких жестов позволяет увеличить защиту

- от инициализации точки интереса «случайными» движениями, т.к. такой жест легко воспроизвести, но относительно сложно сделать это случайно.
- от шумового загрязнения точек фона создаваемого случайными объектами или ошибками распознавания. В этом случае наибольшее количество движения доставляет конец движущегося объекта.

Задачу слежения заточкой интереса можно разбить на две подзадачи: *Инициализация точки интереса* и *слежение за точкой интереса*.

Исходя из описанных выше соображений, **процедуру инициализации точки интереса** можно представить в виде следующего алгоритма:

1. По  $n$  кадрам пересчитать накопительные разностные буферы  $A$ ,  $N$  и  $P$  (1-3);
2. Построить множество  $A^\delta = \{(x, y) : A_k(x, y) \geq \delta\}$ , где  $\delta > 0$ .
3. Провести пороговую обработку с порогом  $T(x, y)$  множества  $A^\delta$ , тем самым получив сегментацию  $A^\delta = G_1 \cup G_2$  :

$$\begin{aligned} G_1 &= \{(x, y) : (x, y) \in A^\delta, f_t(x, y) \geq T(x, y)\}, \\ G_2 &= \{(x, y) : (x, y) \in A^\delta, f_t(x, y) < T(x, y)\}, \\ G_1 \cap G_2 &= \emptyset. \end{aligned} \quad (4)$$

4. Вычислить координаты  $(x_I, y_I)$  точки  $I$  по формулам

$$\begin{aligned} x_I &= \alpha \frac{\sum_x \sum_y P_k(x, y) S_P(x, y) x}{\sum_x \sum_y P_k(x, y) S_P(x, y)} + (1 - \alpha) \frac{\sum_x \sum_y N_k(x, y) S_N(x, y) x}{\sum_x \sum_y N_k(x, y) S_N(x, y)}, \\ y_I &= \alpha \frac{\sum_x \sum_y P_k(x, y) S_P(x, y) y}{\sum_x \sum_y P_k(x, y) S_P(x, y)} + (1 - \alpha) \frac{\sum_x \sum_y N_k(x, y) S_N(x, y) y}{\sum_x \sum_y N_k(x, y) S_N(x, y)}, \end{aligned} \quad (5)$$

где

- $\alpha \in [0, 1]$  – параметр, отвечающий за вес опорного кадра;
- $S_P, S_N$  – весовые функции, позволяющие более тонко манипулировать областью и параметрами поиска движения;
- суммирование ведётся по пространственным координатам области кадра.

5. Проверить принадлежность полученной точки  $I$  ко множеству  $G_2$  (4).

- (a) Если полученная точка  $I$  с координатами  $(x_I, y_I)$  удовлетворяет условию  $(x_I, y_I) \in G_2$ , то такая точка считается точкой интереса и за ней начинается слежение.
- (b) В противном случае инициализация считается не успешной и ожидается следующий кадр.

После успешной инициализации точки интереса за ней начинается слежение. Поиск нового положения точки интереса происходит в несколько этапов. Для проведения **процедуры поиска нового положения точки интереса** необходимо:

1. Оценить количество движения в пространственной  $r$  окрестности и малой временной окрестности старой точки интереса с координатами  $(x_I^{t-1}, y_I^{t-1})$ :

$$\begin{aligned}
 x_m &= \frac{\sum_{(x,y) \in O_r(x_I^{t-1}, y_I^{t-1})} A_2(x, y)x}{\sum_{(x,y) \in O_r(x_I^{t-1}, y_I^{t-1})} A_2(x, y)}, \\
 y_m &= \frac{\sum_{(x,y) \in O_r(x_I^{t-1}, y_I^{t-1})} A_2(x, y)y}{\sum_{(x,y) \in O_r(x_I^{t-1}, y_I^{t-1})} A_2(x, y)}.
 \end{aligned} \tag{6}$$

2. Уточнить центр региона поиска с учётом истории:

$$\begin{aligned}
 x_c &= \alpha x_m + (1 - \alpha)x_I^{t-1}, \\
 y_c &= \alpha y_m + (1 - \alpha)y_I^{t-1}.
 \end{aligned} \tag{7}$$

3. Вычислить координаты  $(x_I^t, y_I^t)$  точки  $I^t$  по формулам

$$\begin{aligned}
 x_I^t &= \frac{\sum_{(x,y) \in O_\rho(x_c, y_c)} S_P(x, y)(M - f(x, y))x}{\sum_{(x,y) \in O_\rho(x_c, y_c)} S_P(x, y)(M - f(x, y))}, \\
 y_I^t &= \frac{\sum_{(x,y) \in O_\rho(x_c, y_c)} S_P(x, y)(M - f(x, y))y}{\sum_{(x,y) \in O_\rho(x_c, y_c)} S_P(x, y)(M - f(x, y))}.
 \end{aligned} \tag{8}$$

где

- $\rho \in (0, r]$ ;  $M = \max_{(x,y) \in O_\rho(x_c, y_c)} \{f(x, y)\}$ ;
- $S_P(x, y)$  – весовая функция.

4. Проверить принадлежность полученной точки  $I^t$  ко множеству  $G_2$  (4). Если полученная точка  $I^t$  с координатами  $(x_I^t, y_I^t)$  удовлетворяет условию  $(x_I^t, y_I^t) \in G_2$ , то полученные координаты будут новыми координатами точки интереса в кадре  $f_t(x, y)$ . Иначе считается, что точка потеряна.

*Замечание.* При практической реализации процедуры поиска нового положения точки интереса окрестность рассматривалась в метрике, порожденной расстоянием Минковского<sup>1,2</sup> с коэффициентами  $p = 2$ ,  $p = 1$ ,  $p = \infty$ ; наиболее

<sup>1</sup>Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. Изд. четвертое, переработанное. Москва: Наука, 1976. С. 544.

<sup>2</sup>Вулих Б. З. Краткий курс теории функции вещественной переменной. 2-е изд., переработанное и дополненное. Москва: Наука, 1973. С. 352.

эффективные с точки зрения использования вычислительных ресурсов результаты были получены в случае  $p = \infty$ .

**Во втором разделе первой главы** описывается подход реализующий «*full-body-interaction*» модель взаимодействия. Такая модель подразумевает взаимодействие при помощи распознавания (или оценки) текущей позы тела (или его части) оператора в пространстве. Для проведения такой оценки необходимо последовательно решить следующие задачи:

1. *Сегментация*. Необходимо решить задачу сегментации сцены, в данном случае заключающейся в выделении оператора в видимой сцене.
2. *Скелетизация*. Для дальнейшей оценки позы оператора как древовидно-упорядоченного набора координат точек, соответствующих шарнирам в иерархической модели человека, необходимо решить задачу скелетизации множества точек, соответствующих оператору, полученному на этапе сегментации.
3. *Слежение*. Результирующая оценка заключается в создании гибкого модельного объекта, имеющего те же параметры, что и пользователь, и установка его в положение, максимально близкое к наблюдаемому. Таким образом, эту задачу можно, в свою очередь, разбить на две подзадачи:
  - (а) *Первоначальная оценка параметров пользователя*. Решается единожды при захвате (инициализации) объекта интереса,
  - (б) *Подгонка (слежение)*. Решается для каждого кадра в течение наблюдения.

Решение **задачи сегментации** основывается на анализе движения в сцене. Предполагается, что сцена (объекты фона) неподвижна, или мало подвижна, в сравнении с пользователем. Тогда можно воспользоваться механизмом *накопительных разностных буферов*. Для разметки будем пользоваться модифицированным алгоритмом кристаллизации<sup>3,4</sup>. В качестве начального центра кристаллизации возьмем точку с координатами  $(\bar{x}, \bar{y})$ :

$$\begin{aligned} \bar{x} &= \frac{\sum_x \sum_y P_k(x, y) S_p(x, y) x}{\sum_x \sum_y P_k(x, y) S_p(x, y)}, \\ \bar{y} &= \frac{\sum_x \sum_y P_k(x, y) S_p(x, y) y}{\sum_x \sum_y P_k(x, y) S_p(x, y)}, \end{aligned} \tag{9}$$

где  $S_p(x, y)$  – вспомогательная весовая функция;  $P_k(x, y)$  – позитивный разностный буфер.

<sup>3</sup>Gonzalez R.C., Woods R.E. Digital image processing. Second edition edition. Prentice-Hall, 2002. P. I–XVII, 794.

<sup>4</sup>Форсайт Д., Понс Ж. Компьютерное зрение. Современный подход. Вильямс, 2004. С. 928.

Отдельно рассматривается случай, когда какой-либо объект перекрывает видимый силуэт пользователя, разделяя его на несколько частей. Такая ситуация может возникнуть также в случае возникновения самоперекрытия силуэта. В этом случае для инициализации алгоритма строится многолистная фигура на основе расширенного циркулярного скелета. Необходимо заметить, что появление самоперекрытия возможно только после этапа инициализации скелета, которая проводится по силуэту с наложенными ограничениями (ограничения на позу). После построения составной поверхности для каждой простой поверхности, соответствующей сегментам, запускается процедура *расширенного выращивания* из нескольких точек – центров кристаллизации. При практической реализации центры кристаллизации выбираются для каждого сегмента на основе эвристики (близость по глубине + расстояние до расширенного циркуляра) или случайно (метод Монте-Карло<sup>5,6</sup>).

В результате действия алгоритма получается частично размеченное множество, в котором выделено замкнутое ограниченное подмножество, содержащее размеченную границу.

Далее следует описание **задачи скелетизации** как построения множества срединных осей для замкнутого ограниченного множества точек, соответствующих наблюдаемому силуэту. Для этого используется аппроксимация множества с помощью многоугольной фигуры, что в данном случае корректно<sup>7</sup>.

При определении позы возможны ситуации, когда при получении кадра из точки наблюдения становится не видна какая-либо часть фигуры. Возникают случаи самоперекрытия. При этом результаты скелетизации могут быть неинформативны. При восстановлении скрытой части возникает множество с самопересечениями<sup>8</sup>.

Для работы с такими случаями вводятся определения *простой поверхности* и *составной поверхности*. На их основе строго **строится система понятий и определений**, позволяющая, в некотором роде свести случаи поверхности, допускающей самоперекрытие или самопересечение, к случаю плоского множества.

После чего дается формулировка **задачи слежения**. Требуется оценить положение и позу оператора в пространстве по наблюдаемому силуэту с заданной точностью. Для решения предлагается представление силуэта пользователя в *иерархическом виде*, представленном на рисунке 1:

---

<sup>5</sup>Джонс М.Т. Программирование искусственного интеллекта в приложениях. Москва: ДМК Пресс, 2004. С. 312

<sup>6</sup>Metropolis N., Ulam S. The Monte Carlo Method. Journal of the American Statistical Association, Vol. 44, № 247, 1949. P. 335–341.

<sup>7</sup>Местецкий Л.М. Непрерывный скелет бинарного растрового изображения. Труды международной конференции Графикон-98. Москва: МГУ, 1998. С. 71–78.

<sup>8</sup>Мехедов И.С. Многолистная плоская фигура и её срединная ось. Известия вузов. Математика. 2011. №12. С. 42–53.

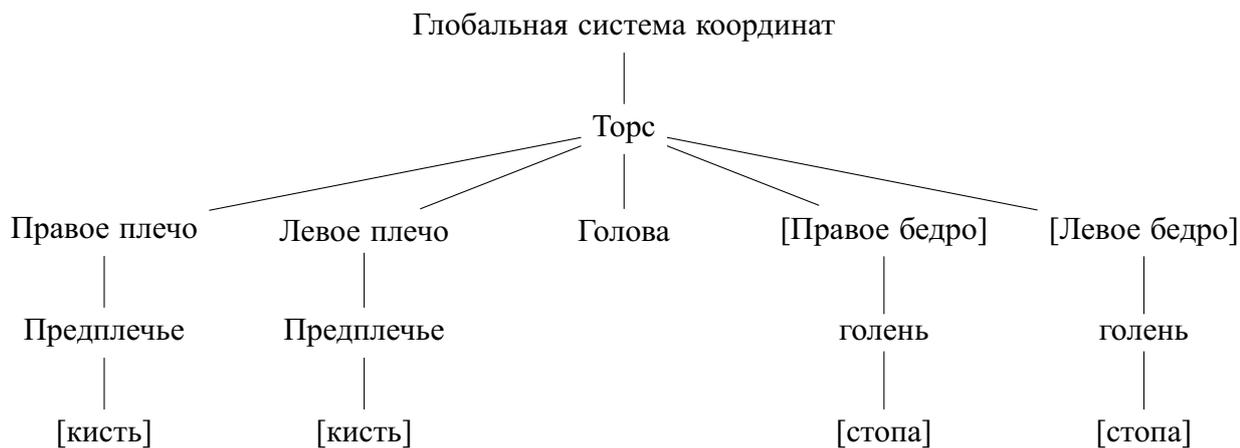


Рис. 1: Иерархическое представление силуэта пользователя

Все элементы в таком представлении являются жесткими, имеют «длину» и «толщину». Соединяются при помощи гибких шарниров  $p_i \in D_i$ , где  $D_i \subseteq \mathbb{R}^2$  – множество допустимых деформаций  $i$ -того шарнира.

Для **первоначальной оценки** параметров сегментов вводится понятие *расширенного циркулярного дерева*, которое является обобщением понятия *циркулярного дерева*<sup>9</sup> на случай силуэтов с самоперекрытиями/самопересечениями. Такое расширение становится возможно за счёт использования механизмов многолистных плоских множеств, описанных выше. Для самой оценки предлагается использовать дополнительный «искусственный» силуэт. Также предлагается использование специальной позы, с наложенными ограничениями, которую принимают пользователь и искусственный силуэт. При этом **алгоритм предварительной оценки параметров** пользователя принимает следующий вид:

1. Поиск скелета и скелетных циркуляров силуэта пользователя как плоского множества.
2. Выделение значимой части скелета, избавление от паразитных отростков, *стрижка* скелета. Одновременно с этим выделение изоморфной части скелетов силуэтов пользователя и искусственно<sup>10,11</sup>.
3. Оценка длин сегментов  $l_{seg}$  изоморфной части скелета. Оценки «ширины» («толщины») сегментов силуэта  $a_{seg}$  как среднего значения радиуса кругов в циркулярном представлении ребра, порождающего сегмент<sup>12,13</sup>, в расширенном циркулярном дереве,

<sup>9</sup>Местецкий Л.М. Сравнение изображений гибких объектов на основе нормализации. Труды 17 международной конференции ГРАФИКОН-2007. Москва, 2007. С. 203–210.

<sup>10</sup>Okhlopkov A., Domakhina L. Shape comparison based on skeleton isomorphism. VISAPP, 2009. P. 237–242.

<sup>11</sup>Yuille A. L., Zhu S. C. Forms: A flexible object recognition and modeling system. IJCV, 1996. Vol. 20. P. 187–212.

<sup>12</sup>Домахина Л. Г. Об одном методе сегментации растровых объектов для задач преобразования формы. Труды 13 всероссийской конференции Математические методы распознавания образов (ММРО-13). 2007. С. 311–314.

<sup>13</sup>Местецкий Л.М., Петрова Л. Г. Расчет гомеоморфизма многоугольников с изоморфными базовыми скелетами // Искусственный интеллект. 2006. № 2. С. 192–196.

$$a_{seg} = \frac{\sum_{c_r \in C_{seg}} r}{l_{seg}}, \quad (10)$$

где  $C_{seg}$  – множество циркуляров, центры которых лежат на собственном ребре сегмента.

При этом необходимо решить собственно задачу слежения: на текущем кадре оценить позу пользователя, получив соответствующий набор параметров, или, если это невозможно, сообщить о потере цели и перейти к процедуре повторной калибровки.

Для решения этой задачи предлагается *метод активного скелета*, являющийся некоторым подобием метода активного контура<sup>14</sup>.

Метод состоит в следующем: Представлен эталонный образец в виде гибкого объекта, включающего в себя описание расширенного циркулярного графа объекта с заданием множества допустимых деформаций (положений шарниров). В такой модели силуэт объекта будем рассматривать как замкнутую область, ограниченную огибающей семейства кругов циркуляра. С помощью применения трансформаций к элементам скелета происходит моделирование различных движений объекта и его частей, которые влекут за собой изменение силуэта объекта. Исходя из этого, задачу слежения можно записать так: необходимо найти такой элемент  $v_*^\alpha$ , что

$$v_*^\alpha = \operatorname{argmin}_{v^\alpha \in V} \mu(v^\alpha(E), F), \quad (11)$$

где:

- $\mu(E, F)$  – функция «непохожести» эталонного и предъявляемого тестовых образцов, является аналогом функции энергии в методе активных контуров.
- $E$  – эталон, заданный в виде гибкого объекта, деформацией которого будем добиваться «максимальной похожести» с тестовым образцом;
- $V = \{v^\alpha\}_{\alpha \in A}$  – совокупность всевозможных деформаций эталона,  $A \subset \mathbb{R}$  – множество индексов;
- $v^\alpha(E)$  – гибкий объект, получающийся в результате применения трансформации  $v^\alpha \in V$  к эталону  $E$ ;
- $F$  – силуэт, выделенный на тестовом кадре;  $T$  – осевой граф циркуляра эталона, а  $S$  – его силуэт.

Использование иерархического скелета даёт возможность описать позу в виде совокупности положений соответствующих шарниров  $\vec{p} = (p_1, \dots, p_k)$ ,  $p_i \in D_i$ ,

<sup>14</sup>Blake A., Isard M. Active contours - the application of techniques from graphics, vision, control theory and statistics to visual tracking of shapes in motion. Springer, 1998. P. I–XII, 1–352.

где  $D_i$  – множество допустимых деформаций  $i$ -того шарнира,  $D_i \subseteq \mathbb{R}^2$ . Тогда  $V = \{v(\vec{p})\}_{\vec{p} \in D}$ , где  $D = \bigcup_{i=1}^K D_i$ ,  $D \subseteq \mathbb{R}^{2k}$ . Таким образом, задачу (11) можно представить в следующем виде: необходимо найти такое  $\vec{p}_* \in D$ , что

$$\vec{p}_* : \mu(\vec{p}_*, F) = \mu(v_*^\alpha(E), F) = \min_{\vec{p} \in D} \mu(v(\vec{p}), F). \quad (12)$$

Использование иерархического представления скелета позволяет представить эталонный и распознаваемый объекты как объединение своих сегментов:  $E = \bigcup_{i=1}^K E_i$  (и соответствующий ему силуэт  $S = \bigcup_{i=1}^K S_i$ ),  $F = \bigcup_{i=1}^K F_i$ . Также, полученный изоморфизм  $\phi : F \leftrightarrow E$  будет обладать свойством  $\phi : F_i \leftrightarrow E_i$ . Всё это позволяет ввести понятие частичной позы  $s_i(\vec{p}) = p_i \in D_i$  и связанному с ним понятием частичной энергии

$$\mu_i(s_i(\vec{p}), F) = \text{Area}(F_i \setminus S_i) + \text{Area}(S_i \setminus F_i) \quad (13)$$

Полную энергию в это случае можно рассматривать как сумму частичных по всем сегментам  $\mu(\vec{p}, F) = \sum_{i=1}^K \mu_i(s_i(\vec{p}), F)$ . Это позволяет провести декомпозицию задачи подгонки (12) на  $K$  подзадач вида

$$\vec{p}_* : \mu_i(s_i(\vec{p}_*), F) = \min_{s_i(\vec{p}) \in D_i} \mu_i(s_i(\vec{p}), F), \quad i = 1, \dots, K \quad (14)$$

Последовательное решение этих задач даст искомый набор  $\vec{p}_*$ , являющийся решением (12).

Минимизация (14) для  $i$ -того шарнира выполняется за один шаг, при условии, что выполнена минимизация для всех его предков. Пусть  $T_i^F$  – серединная ось  $i$ -того сегмента тестового плоского силуэта  $F$ . Тогда, если отрезок  $L$ , получившийся в результате проекции на плоскость наблюдения записывается в радиальной системе координат как  $(\rho, \theta)$ , то, зная длину  $l$  образа  $T_i^F$  отрезка  $L$ , углы  $\alpha_i^*$  и  $\phi_i^*$ , которыми задаётся положение  $p_i = (\alpha_i, \phi_i)$   $i$ -того шарнира, можно получить из следующих простых геометрических соотношений:

$$\begin{aligned} \alpha_i^* &= \theta, \quad \alpha_i^* \in [0, 2\pi) \\ \cos(\phi_i^*) &= \frac{\rho}{l}, \quad \phi_i^* \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \\ (\alpha_i^*, \phi_i^*) &\in [0, 2\pi) \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \cap D_i \end{aligned} \quad (15)$$

Второе уравнение может иметь два решения, поэтому для выбора одного из них воспользуемся информацией о глубине сцены в соответствующей точке.

Получив значение  $p_i^* = (\alpha_i^*, \phi_i^*)$ , являющееся решением задачи (14) для  $i$ -того шарнира, можно решать эту же задачу для его потомка, если такой есть. Найдя все  $K$  таких параметров, мы получим  $\vec{p}_*$ , являющимся решением (12).

Во второй главе описываются функциональные особенности реализованного программно-аппаратного комплекса «*VirtualHand*». Программный комплекс имеет модульную архитектуру, представленную на рисунке 2.



Рис. 2: Общая структура комплекса

Такой подход позволяет заменять отдельные модули на их аналоги с наиболее подходящим к конкретной задаче функционалом. Это даёт возможность сделать комплекс легко расширяемым и имеющим возможность настройки для различных условий использования, задач и конкретных пользователей. Так же такой подход даёт возможность подключения различного оборудования, в том числе и различных типов датчиков.

Важной особенностью является реализация и использование *системы модификаторов*. Рассматривая жест как размеченную траекторию точки интереса, естественно будет введение логической функции-триггера, результат которой зависел бы от состояния системы. В качестве таких функций выступают модификаторы. В зависимости от значения модификатора положение и траектория точки интереса может интерпретироваться по-разному. В качестве модификатора может выступать положение точки интереса в пространстве, любой из уже

распознанных жестов, состояние аппаратных триггеров (кнопок, датчиков и т.д.) и/или их комбинации. В разделах 2.3.2 и 2.3.3 приводятся примеры видов модификаторов и их использования. Как один из примеров приводится разработка и реализация метафоры «бесконтактного тачскрина» и виртуальных объектов типа «виртуальная кнопка» и «виртуальная клавиатура». Такой подход позволяет создать легко расширяемые словари жестов, являющиеся основой для языков и основанных на них человеко-компьютерных интерфейсов.

Таким образом, становится возможным создать надёжный программно-аппаратный комплекс, который

- При необходимости гибко подстраивается для конкретной задачи и конкретного пользователя;
- Поддерживает различные виды датчиков, в том числе и пока еще не существующих;
- Позволяет использовать различные языки жестов, в том числе и построенные на пользовательских словарях. Также возможно построение новых лексических конструкций языка как комбинаций уже существующих;
- Позволяет использовать как методы взаимодействия, рассчитанные на классические устройства ввода, так создавать новые, в том числе и уникальные для конечного приложения;
- Может быть использован вместо классических устройств ввода, при этом сохраняет их функционал и позволяет осуществлять бесконтактный ввод информации.
- Является основой для создания [человеко-компьютерного] интерфейса для широкого круга задач;

**В третьей главе** дано описание каждого из модулей программного комплекса с точки зрения прикладного разработчика. Описывается положение модуля в структуре комплекса и форматы данных, ожидаемых на входе и на выходе работы модуля с подробным описанием полей и типов данных.

В заключении главы приводится используемый в разработке инструментарий: языки программирования и сторонние библиотеки. При разработке использовались библиотеки *OpenCV*, *Qt* (4.\*,5.\*), *OSCPack*. Большинство модулей реализовано на языках программирования *C* и *C++*. В некоторых местах применялись язык *C#* и скриптовые языки *Lua*, *Javascrit*, *Python*. Модули ядра, также как и некоторые места внешних модулей, реализованы на языках *C* и *C++*.

Описанные модули анализа карт глубин, представленные в третьей главе в разделе 2.1 и 2.2, «Иницируемый подход» и «Скелетный подход» реализованы в соответствующих программных комплексах «*Point module for VirtualHand*» и «*Skeleton module for VirtualHand*» [4, 5]

**В заключении** подводятся итоги, проводится анализ примененных подходов и методов и кратко описываются основные результаты диссертации. К основным результатам можно отнести:

1. Построение программно-аппаратного комплекса, являющегося основой для разработки интерфейсов человеко-компьютерного взаимодействия при помощи жестов. Модульная архитектура комплекса позволяет:
  - Использовать различные алгоритмы без изменения основного кода.
  - Подключать различные датчики глубины сцены, в том числе и еще не существующие, и единообразно работать с ними.
  - Заменять модули на другие, наиболее подходящие в конкретном случае.
2. Разработка и реализация алгоритмов захвата движения и слежения за объектами в пространстве, основанных на обработке карт глубин сцены и использующих
  - Инициализацию точки интереса с помощью специального ключевого жеста.
  - Полное или частичное распознавание позы человека.
3. Разработка подхода к созданию жестовых языков, основанного на системе модификаторов. Такой подход позволяет увеличивать выразительную способность языка, изменяя интерпретацию траектории точки интереса в зависимости от значения модификатора. Так же такой подход позволяет легко создавать индивидуальные словари для жестового языка, что даёт возможность оптимизации интерфейса под конкретную задачу и под конкретного пользователя.

## Список публикаций автора

**Статьи, опубликованные рецензируемых научных изданиях, определенных ВАК:**

1. *Стародубцев И.С.* Инициализируемый захват движения и слежение на основе карт глубин для жестового интерфейса // Научно-технический вестник Поволжья. Казань, 2013. № 3. С. 264–268.
2. *Стародубцев И.С.* Захват движения на основе иерархического гибкого скелета // Научно-технический вестник Поволжья. Казань, 2014. № 1. С. 159–162.
3. *I. Starodubtsev, V. Averbukh, N. Averbukh, D. Tobolin* Professional Natural Interfaces for Medicine Applications // Communications in Computer and Information Science / Ed. by C. Stephanidis. Springer International Publishing, 2014. Vol. 435. P. 435–439.

### Патенты и свидетельства о регистрации программ:

4. *Стародубцев И.С.* Свидетельство о государственной регистрации программы для ЭВМ № 2014613954 «Point module for VirtualHand». / Федеральная служба по интеллектуальной собственности (Роспатент). Зарегистрировано 14 апреля 2014 г.
5. *Стародубцев И.С.* Свидетельство о государственной регистрации программы для ЭВМ № 2014614169 «Skeleton module for VirtualHand». / Федеральная служба по интеллектуальной собственности (Роспатент). Зарегистрировано 17 апреля 2014 г.

### Другие публикации:

6. *Авербух В.Л., Дернов Г.С., Зырянов А.В., Стародубцев И.С., Флягина Т.А., Щербинин А.А.* Жестовые методики ввода информации в интерактивных системах компьютерной визуализации // XII Международный семинар «Супервычисления и математическое моделирование». Тезисы. Саров, Госкорпорация по атомной энергии «РОСАТОМ», 2010, С. 10–11.
7. *Starodubtsev I.S.* Fast Marker Tracking Based on OpenCV Library for Augmented Reality Applications // Международный Российско-Корейский семинар. 29-31 мая 2011. Abstract of reports. Екатеринбург, УрФУ, 2011. С. 10–11.
8. *Стародубцев И.С., Авербух В.Л.* Манипуляторы для систем научной визуализации // XIII международный семинар «Супервычисления и математическое моделирование». Тезисы. Саров. ФГУП «РФЯЦ ВНИИЭФ», 2011, С. 124–125.

9. *Зырянов А.В., Стародубцев И.С.* Интерфейсы на основе жестов для систем научной визуализации // SCVRT'2011 Труды международных научных конференций «Ситуационные центры и информационно-аналитические системы класса 4i (SC-IAS4i-2011)» и «Системы виртуального окружения для комплексной безопасности и антитеррористической защищенности зданий и сооружений (VRTerro2011)». Москва-Протвино. Институт физико-технической информатики. 2011. С. 138–139.
10. *Авербух В.Л., Авербух Н.В., Бахтерев М.О., Васёв П.А., Зырянов А.В., Манаков Д.В., Стародубцев И.С., Щербинин А.А.* Системные и визуализационные предпосылки создания виртуального испытательного стенда // Вопросы оборонной техники. Серия 14. 2012. Выпуск 2, С. 20–26.
11. *Стародубцев И.С., Авербух В.Л.* Манипуляторы для систем научной визуализации // Труды XIII Международного семинара «Супервычисления и математическое моделирование». Саров. ФГУП «РФЯЦ ВНИИЭФ», 2012, С. 506–509.
12. *Авербух В.Л., Бахтерев М.О., Васев П.А., Михайлов И.О., Стародубцев И.С.* Разработка средств визуального представления сеток // XIV Международная конференция «Супервычисления и Математическое Моделирование». Тезисы. ФГУП «РФЯЦ ВНИИЭФ». Саров. 2012, С. 14–15.
13. *Стародубцев И.С., Авербух В.Л.* Взаимодействие с виртуальными объектами // XIV Международная конференция «Супервычисления и Математическое Моделирование». Тезисы. ФГУП «РФЯЦ ВНИИЭФ». Саров. 2012, С. 144–145.
14. *Averbukh V.L., Starodubtsev I.S., Tobolin D.U.* The Gesture Interface For Control Of Angiographic Systems // Современные компьютерные и информационные технологии: сборник трудов международной научной Российско-Корейской конференции, Екатеринбург, УрФУ, 2012, С. 97–107.
15. *Стародубцев, И.С.* Проект «ВИРТУАЛЬНАЯ РУКА» // XIV Международная конференция «Супервычисления и Математическое Моделирование». Тезисы. Саров. ФГУП «РФЯЦ ВНИИЭФ». 2014. С. 124–125.
16. *Авербух В.Л., Авербух Н.В., Стародубцев И.С., Тоболин Д.Ю.* Использование жестовых интерфейсов при взаимодействии с объектами // Научная перспектива. 2014. №10 (56). С. 57–66.