

Вежневц Владимир Петрович

АЛГОРИТМЫ АНАЛИЗА ИЗОБРАЖЕНИЯ ЛИЦА ЧЕЛОВЕКА ДЛЯ
ПОСТРОЕНИЯ ИНТЕРФЕЙСА ЧЕЛОВЕК-КОМПЬЮТЕР

05.13.11 — математическое и программное обеспечение вычислительных
машин, комплексов и компьютерных сетей

Автореферат
диссертации на соискание учёной степени
кандидата физико-математических наук

Работа выполнена в Московском государственном университете им. М.В.Ломоносова.

Научный руководитель: кандидат физико-математических, доцент
Баяковский Юрий Матвеевич

Официальные оппоненты: доктор физико-математических наук,
профессор
Горбунов-Посадов Михаил Михайлович

кандидат физико-математических наук,
Переберин Антон Валерьевич

Ведущая организация: Государственный Научно - Исследовательский
Институт Авиационных Систем (ГосНИИАС)

Защита состоится “ ____ ” _____ 2004 в ____ ч. ____ мин на заседании Диссертационного совета Д 501.001.44 в Московском государственном университете им. М.В.Ломоносова по адресу: 119992, ГСП-2, г. Москва, Воробьевы Горы, МГУ им. М.В. Ломоносова, 2-ой учебный корпус, факультет вычислительной математики и кибернетики, аудитория 685

С диссертацией можно ознакомиться в библиотеке факультета вычислительной математики и кибернетики МГУ им М.В. Ломоносова.

Автореферат разослан “ ____ ” _____ 2004.

Учёный секретарь Диссертационного совета
доктор физико-математических наук, профессор

Трифонов Н.П.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Объект исследования и актуальность темы.

В настоящее время вычислительная техника используется во многих областях человеческой деятельности, являясь удобным и многофункциональным инструментом решения широкого круга задач. Однако, в настоящее время пользователи ЭВМ вынуждены использовать способы взаимодействия, слабо адаптированные к возможностям человеческого восприятия и ограничивающие способности человека к обмену информацией. Основная цель усовершенствования и развития интерфейса человек-компьютер заключается в организации обмена информацией с ЭВМ таким образом, чтобы:

- Снизить время освоения программных и аппаратных средств;
- Снизить уровень ошибок при передаче информации;
- Сделать работу с ЭВМ возможной для людей, не имеющих возможности пользоваться традиционными средствами интерфейса;
- Снизить утомляемость, увеличить субъективное удовлетворение пользователя от работы;

Для достижения поставленных целей необходимо применение средств взаимодействия, более полно использующих коммуникативные способности человека. Человек наделен большим количеством возможностей воспринимать и передавать информацию: зрение, слух (в т. ч. устная речь), жесты и движения, мимика, осязание и др. Во взаимодействии человека и компьютера существуют два информационных потока:

- управляющие команды и данные, передаваемые компьютеру для обработки;
- результаты вычислений и другая информация, представляемая компьютером пользователю.

Распространенный в настоящее время человеко-машинный интерфейс использует зрение, как основной канал представления информации пользователю, отображая данные в виде условных знаков на экране компьютера.

Воспринимать информацию естественными для человека способами (распознавать речь, жесты, мимику и т.д.) современные средства интерфейса практически не в состоянии.

Лицо человека является важным источником информации при общении между людьми. Выражение лица, мимика, артикуляция при разговоре, движения головой являются удобным, естественным и, что важно, необременительным способом передачи информации. Неспособность компьютера с одной стороны воспринять, а с другой стороны воспроизвести столь естественные для человека способы общения затрудняет передачу и восприятие информации при работе с ЭВМ.

Для того, чтобы “научить” компьютер распознавать и реагировать на движения головы, мимику, изменение выражения лица, направление взгляда, требуются устойчивые алгоритмы анализа и распознавания изображения лица человека.

Для представления вычислительной системы в качестве партнера, имитирующего аспекты человеческого общения, и превращения взаимодействия пользователя с вычислительной системой в естественное общение с виртуальным персонажем, используются так называемые агенты виртуальной реальности. Для их обозначения в литературе принят термин “аватары” (*avatars*). Агентами виртуальной реальности могут быть модели реально существующих людей, управляемые другими пользователями, или виртуальные актеры, управляемые компьютером. Использование таких агентов в системах виртуальной реальности и аудиовизуальных интерфейсах позволяет организовать общение пользователя с системой наиболее естественным образом и облегчить ему освоение. Это особенно актуально для пользователей, не являющихся профессионалами в области информационных технологий.

В рамках данной диссертационной работы будут рассмотрены следующие задачи:

1. Обнаружение и выделение лица на изображении и в видеопотоке.
2. Распознавание антропометрических точек лица на изображении и в видеопотоке.
3. Управление курсором “мыши” с помощью движений головой и мимики

пользователя.

Решение двух первых задач является необходимым условием для создания интеллектуальных интерфейсов, распознающих и реагирующих на движения головой и изменение выражения лица пользователя. Решение второй задачи также необходимо для автоматизированного построения фотореалистичных моделей лиц для интерфейсов, использующих агентов виртуальной реальности. Решение третьей задачи делает возможным использование компьютера людьми, неспособными использовать традиционные средства интерфейса по состоянию здоровья.

Использование в системах построения интерфейса накладывает дополнительные условия и требования на методы, которые могут использоваться для решения этих задач, а именно:

- Высокая скорость работы (в ряде случаев - реальное время);
- Функционирование на изображениях низкого (порядка 320x240 пикселей) и среднего (порядка 640x480 пикселей) разрешения;
- Устойчивость к уровню шума, характерному для недорогих бытовых видеокамер, подключаемых к компьютеру.

Цели работы

1. Разработать методы обнаружения и выделения лица на изображении низкого разрешения (порядка 320x240 пикселей) и высокого уровня шума (характерного для бытовых видеокамер), работающие в реальном времени;
2. Разработать алгоритм автоматического выделения антропометрических точек лица (контуров глаз, зрачков, бровей, носа, подбородка, рта) на фронтальной фотографии среднего разрешения (порядка 640x480 пикселей) для построения фотореалистичных моделей виртуальной реальности;
3. Разработать алгоритм управления курсором "мыши" с помощью движений головы и мимики пользователя;

4. Разработать алгоритм автоматического определения положения антропометрических точек лица в видеопотоке низкого разрешения в реальном времени.

Научная новизна работы

Разработанные алгоритмы обнаружения и локализации лица человека на основе цветовой сегментации кожи обладают значительно большей устойчивостью обнаружения, сохранив скорость существующих методов.

Предложенные алгоритмы выделения антропометрических точек лица на изображении и в видеопотоке обладают более высокой устойчивостью и точностью распознавания, нежели описанные в литературе аналоги.

Разработан новый алгоритм управления курсором мыши с помощью движений головы и мимики пользователя. Алгоритм предоставляет возможность гибкой настройки метода трансляции движений головы в перемещения курсора, что предоставляет возможность настройки интерфейса на индивидуальные особенности пользователя с нарушениями двигательного аппарата.

Практическая значимость и реализация

Разработаны и доведены до реализации методы и алгоритмы решения нескольких актуальных задач машинного зрения и обработки изображений. Программные реализации описываемых в диссертации методов удовлетворяют требованиям и ограничениям, сформулированным при постановке задач.

Реализована система управления компьютером для детей, страдающих ДЦП. Система успешно прошла экспериментальную проверку и используется в ряде детских учебных учреждений. По заказу Министерства Образования РФ на основе данной системы создан ряд тренажеров и вспомогательных программ для организации дистанционного обучения.

На основе разработанных алгоритмов обнаружения и локализации лица и его антропометрических точек на изображении построена система автоматизированного создания фотореалистичных 3D моделей головы по фотографиям. Система разрабатывалась в лаборатории Компьютерной Графики и Мультимедиа кафедры АСВК факультета ВМиК МГУ им. М.В. Ломоносова по заказу Samsung Advanced Institute of Technology.

Апробация работы и публикации

Результаты работы докладывались и обсуждались на:

- 6-ой международной конференции по машинному зрению Asian Conference on Computer Vision (ACCV) 2004, Jeju, Korea, Январь 2004;
- 13-й международной конференции по компьютерной графике и машинному зрению GraphiCon 2003, Россия, Москва, 2003;
- 12-й международной конференции по компьютерной графике и машинному зрению GraphiCon 2002, Россия, Нижний Новгород, 2002;
- 3-ей международной конференции “Цифровая обработка информации и управление в чрезвычайных (экстремальных) ситуациях” ЦОИУЧС 2002, Беларусь, Минск, 2002;
- 10-ой всероссийской конференции “Математические методы распознавания образов-10” (ММРО-10), Россия, Звенигород, 2001;
- семинаре по компьютерной графике и машинному зрению Ю.М. Баяковского (ф-т ВМиК МГУ);
- научно-исследовательском семинаре по автоматизации программирования под руководством проф. М. Р. Шура-Бура (ф-т ВМиК МГУ).

Основные результаты работы изложены в 7-и научных публикациях. Системы, использующие разработанные методы, защищены российскими и международным патентами.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность исследуемой проблемы, сформулирована цель и задачи диссертационной работы, научная новизна, практическая ценность; описана структура диссертации.

Первая глава посвящена задаче обнаружения лица на изображении и в видеопотоке.

Существующие алгоритмы обнаружения лица можно разбить на две категории: методы эмпирического распознавания и методы моделирования изображения лица. К первой категории относятся методы, отталкивающиеся от опыта человека в распознавании лиц и делающие попытку формализовать и алгоритмизовать этот опыт. Вторая категория опирается на инструментарий распознавания образов, рассматривая задачу обнаружения лица как частный

случай общей задачи распознавания. По набору тренировочных изображений строится модель изображения лица, и задача обнаружения сводится к проверке входного изображения на удовлетворение полученной модели.

Среди методов эмпирического обнаружения лица выделяется семейство методов, использующих цвет кожи как признак присутствия лица. Данные методы получили широкое распространение, поскольку сочетают в себе несколько важных преимуществ: малая вычислительная сложность, высокая скорость обработки, простота реализации; устойчивость к изменению ориентации и масштаба лица; устойчивость к изменению освещения (за исключением цветного); устойчивость к изменению выражения лица и частичного перекрытия лица другим объектом сцены.

Обнаружение лиц по цвету кожи производится обычно в два этапа:

1. выделение пикселей, близких по цвету к коже;
2. поиск лиц на изображении с использованием априорного знания о структуре человеческого лица.

Для выделения пикселей с цветом кожи используется модель распределения оттенка кожи в цветовом пространстве, получаемая на этапе предварительной тренировки. В литературе предложено большое количество различных типов моделей распределения оттенка кожи (подробный обзор приведен в работе /7/). На основании сравнения существующих методов цветовой сегментации кожи оптимальным представляется метод моделирования распределения с помощью Баейсовской карты вероятностей /7/.



Рис. 1. Результаты цветовой сегментации кожи.

Одной только цветовой информации недостаточно для устойчивого обнаружения лица, поэтому требуется расширить набор признаков дополнительными свойствами изображения лица. Наиболее часто используются ограничения на форму областей лиц (лицо имеет характерную овальную форму).

Большая часть существующих методов выделения лица по цвету сталкивается с проблемами в случае неточного выделения кожи по цвету и присутствия на изображении значительного количества объектов, близких по цвету к коже. Причина заключается в том, что объединение пикселей цвета кожи в области и проверка каждой из полученных областей на удовлетворение ограничениям по форме делаются независимо. У подобного подхода есть существенный недостаток - методы группировки пикселей кожи, не учитывающие ограничения на форму лиц, могут ошибочно сгруппировать в одну область пиксели, реально относящиеся к разным областям.

Перед автором стояла задача разработать метод выделения потенциальных областей лиц, обладающий следующими свойствами: устойчивость к неизбежным неточностям цветовой сегментации, точность выделения областей, высокая скорость работы. Важным требованием является сохранение высокой скорости работы методов наряду с увеличением устойчивости, чтобы сохранить важнейшее преимущество использования цвета кожи для обнаружения лица - быстродействие.

Проблему выделения эллиптических областей цвета кожи можно сформулировать, как задачу оптимизации в пространстве параметров эллипса (x_c, y_c, θ, a, b) . Однако получение формулы целевой функции в явном виде не является тривиальной задачей. Чтобы избежать трудностей, связанных с поиском явного задания целевой функции и разработки метода оптимизации, не подверженного проблеме локальных экстремумов, было разработано два алгоритма, использующих последовательные приближения с помощью деформируемой эллиптической модели лица, жестко фиксирующие возможную форму областей лиц.

Первый метод (эллиптической деформации) использует деформируемую контурную модель /5/, форма которой задается эллипсом. Вдоль границы модели расположены области (называемые "щупами"), в которых происходит оценка плотности пикселей, близких по цвету к коже. Модель инициализируется вблизи ожидаемого положения лица на изображении, деформация модели производится по шагам. Для каждого "щупа" производится расчет плотности пикселей кожи во внутренней и внешней окрестности границы модели. В зависимости от полученных плотностей "щуп" перемещается пер-



Рис. 2. Иллюстрации к предлагаемым методам обнаружения лица. (а) Метод эллиптической деформации. (б) Метод статистических моментов

пендикулярно границе модели. После того, как на каждом шаге все “щупы” занимают свое новое положение, набор точек центров “щупов” аппроксимируется эллипсом, после чего центры “щупов” проецируются на его границу. Столь простая схема алгоритма гарантирует возможность получения быстрой программной реализации. Несмотря на простоту, она обладает высокой устойчивостью к различным исходным данным и не требует высокой точности начального приближения. Кроме того, она обладает большой гибкостью и позволяет с легкостью встраивать дополнительные условия и ограничения на возможную форму выделяемых областей (пропорции, положение, ориентации эллипса).

Задача отслеживания перемещения лица в видеопотоке обладает определенными особенностями по сравнению с обнаружением лица на статическом изображении - во-первых, известно положение лица на предыдущем кадре, и во-вторых, требуется быстрая сходимость метода. Учитывая эти особенности, предлагается метод статистических моментов отслеживания перемещения лица. Метод основан на вычислении статистических моментов множества пикселей цвета кожи, попадающих в эллипс большего размера, чем область лица на предыдущем кадре. Для пересчета новых параметров эллипса используются координаты центра масс пикселей с высокой вероятностью кожи μ_x, μ_y и центральные моменты второго порядка $\mu_{11}, \mu_{20}, \mu_{02}$. По ним вычисляются координаты нового центра эллипса (μ_x, μ_y) , вектор направления его главной оси $(\mu_{11}, \mu_{20} - \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2})$ и длины его осей.

Сравнение предложенных методов с существующими показало, что предлагаемые автором методы обладают существенно более высокой устойчивостью к зашумленным исходным данным и ошибкам цветовой сегментации, сохраняя простоту программной реализации и высокую скорость работы, позволяющую использовать разработанные алгоритмы в системах реального времени.



Рис. 3. Примеры выделения областей лиц на изображениях с помощью предложенных методов

Вторая глава посвящена описанию системы автоматического распознавания антропометрических точек лица (контуров глаз, зрачков, бровей, носа, подбородка, рта) на фронтальных изображениях для построения трехмерной модели головы человека по набору фотографий.

Для автоматизированного создания 3D модели по фотографии необходимым является распознавание антропометрических точек лица для последующей адаптации трехмерной модели к индивидуальным особенностям лица. Большинство существующих систем автоматизированного построения модели головы по фотографиям используют ручное выделение антропометрических точек лица на изображениях. Алгоритмы, разработанные автором, и построенная на их основе система распознавания антропометрических точек лица позволяет в значительной степени автоматизировать этот процесс.

Для обнаружения области лица на изображении используются мето-

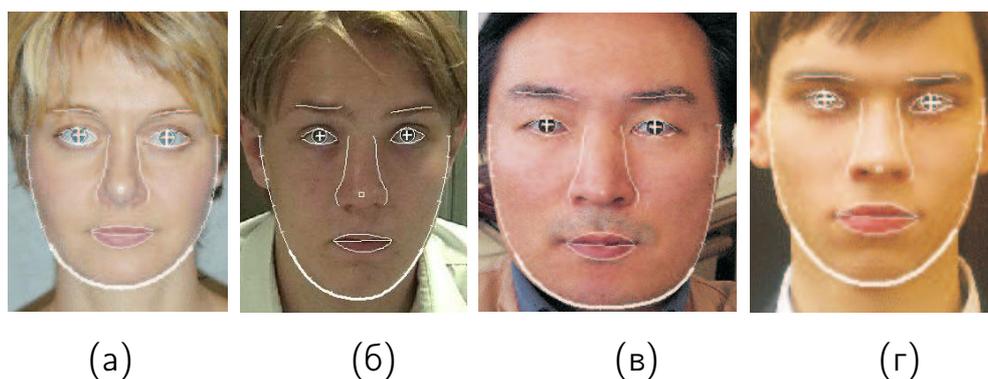


Рис. 4. Типичные примеры выделения антропометрических точек.

ды, описанные в первой главе. Следующим шагом в распознавании является обнаружение положения глаз на изображении. Обнаружение положения глаз производится с помощью поиска областей резкого изменения интенсивности красного канала путем фильтрации красного канала изображения функцией специального вида:

$$V_n(x, y) = \frac{\alpha}{|R_{n,x,y}|} \sum_{r \in R_{n,x,y}} \left\{ I(r) - \frac{1}{|P_{n,r}|} \sum_{p \in P_{n,r}} I(p) \right\}^2 \quad (0.1)$$

Здесь I - красный канал исходного изображения, p и r обозначают координаты пикселей, $R_{n,x,y}$ - прямоугольник размера $(n \times 7)$, с центром в точке (x, y) а $P_{n,r}$ - эллипс с размерами осей $(n \times n/3)$, ориентированный по осям изображения, с центром в точке r . Коэффициент α и масштаб фильтра n - параметры фильтрации. Использование красного канала сохраняет резкий контраст между интенсивностями пикселей радужной оболочки и ее окружения даже для светлых глаз (голубых, зеленых).

После обнаружения приблизительного положения глаз производится поиск координаты центра зрачка. В случае доминирующего фронтального освещения производится обнаружение пикселей, принадлежащих блику в центре зрачка. В случае освещения, близкого к равномерному, производится свертка изображения функцией, подчеркивающей круговые области темных пикселей.

$$W(x, y, c) = \frac{\sin((x^2 + y^2)/c)}{(x^2 + y^2)/c} \quad (0.2)$$

Параметр c контролирует радиус фильтра и выбирается исходя из соотношения размеров лица и радиуса радужной оболочки. Центр масс наиболее

темных пикселей изображения после свертки берется в качестве приближенного центра зрачка.

Вычисление радиуса радужной оболочки r и точных координат центра зрачка (x_c, y_c) производится с помощью следующей функции:

$$f_{\Theta}(x_c, y_c, r) = \int_{\theta \in \Theta} R(x_c + r \cos \theta, y_c + r \sin \theta) d\theta \quad (0.3)$$

где $R(x, y)$ - красная компонента пикселя (x, y) , $\Theta = [-\pi/4, \pi/6] \cup [5\pi/6, 5\pi/4]$. Для заданных координат центра (x_c, y_c) наиболее вероятным значением радиуса радужной оболочки r будет величина, соответствующая максимальному значению $\frac{d}{dr} f_{\Theta}$. После обнаружения точного положения центров глаз производится нормализация изображения (поворот и масштабирование). Отрезок, соединяющий центры глаз, делается горизонтальным, его длина приравнивается к фиксированной величине.

Обнаружение точного контура глаза производится путем поиска точек, являющихся локальными минимумами яркости в области границы глаза. Эксперименты показали, что в отличие от часто используемых точек резкого изменения яркости, точки локальных минимумов являются более устойчивыми признаками для распознавания контуров глаз. Как видно из рисунка 5, области значительных минимумов с высокой вероятностью соответствуют границе глаза и кожи. Обнаружение точек минимума производится отдельно для каждой строки изображения, рассматривая яркость изображения вдоль строки как одномерную функцию $f_y(x)$. Для того, чтобы снизить случайные флюктуации $f_y(x)$, перед поиском минимумов производится низкочастотная фильтрация функции.

Для обнаружения прямоугольника, ограничивающего область губ, применяется анализ статистики цветов области лица. На основе результатов обнаружения кожи строится гистограмма цветов кожи, попавших внутрь уточненного прямоугольника лица. Используя полученную гистограмму, внутри прямоугольника лица в районе ожидаемого положения губ выделяются пиксели нехарактерных для кожи данного человека цветов. Для вычисления прямоугольника используется итерационный алгоритм, основанный на вычислении центра масс и вторых центральных моментов области пикселей с низкой вероятностью.

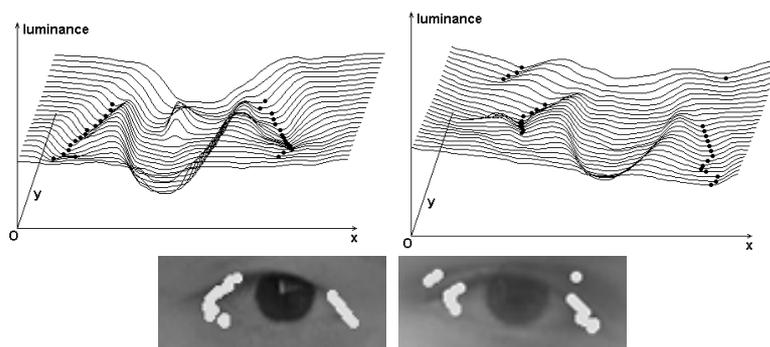


Рис. 5. Точки значительных локальных минимумов яркости для обнаружения контуров век

Для обнаружения контуров носа используется векторное поле градиента яркости изображения. Для обнаружения границ ноздрей используется метод контурных шаблонов, представляющих собой типичную форму крыльев носа. Критерий согласия шаблона и изображения вычисляется следующим образом:

$$GoF(q) = \sum_{p \in S(q)} Fit(p) \quad (0.4)$$

$$Fit(p) = \begin{cases} 1, \text{ если } \exists r \in \Omega(p) : \frac{\nabla I(\vec{r}) \cdot \vec{a}(p)}{|\nabla I(\vec{r})| |\vec{a}(p)|} \geq 0.9, |\nabla I(\vec{r})| > T_1; \\ 0, \text{ иначе;} \end{cases} \quad (0.5)$$

Где p , q и r - координаты пикселей изображения, $S(q)$ - множество точек шаблона, $\Omega(p)$ - окрестность точки p , $\nabla I(\vec{r})$ - градиент яркости изображения в точке r , $\vec{a}(p)$ - вектор касательный к кривой шаблона в точке p . T_1 задает минимальную абсолютную величину градиента. $Fit(p)$ является индикаторной функцией, проверяющей наличие точек с достаточной абсолютной величиной градиента и направлением, близким к направлению касательной к кривой шаблона в окрестности точки шаблона p . Для обнаружения наиболее вероятной пары анализируются пары координат шаблона с высокими значениями критерия согласия в левой и правой части лица. Вычисление кривых формы носа производится, исходя из полученных координат крыльев носа, контуров глаз и априорных знаний о пропорциях черт лица.

Обнаружение контура лица (линии щек и подбородка) осуществляется с помощью деформируемой контурной модели. Как и в случае с контуром носа, обычно признаками обнаружения границы лица служили точки резкого

изменения яркости, вычисляемые на основе абсолютной величины градиента яркости изображения. Как показали тесты описанных в литературе методов, использование такой информации в большом количестве случаев недостаточно для устойчивого распознавания. Во-первых, фон изображения, или его отдельные фрагменты, могут слабо отличаться по яркости от области лица. Во-вторых, для устойчивого обнаружения границы лица требуется рассматривать также и направление наибольшего изменения яркости и цвета. Для обнаружения контура лица была разработана методика обнаружения величины и направления резкого изменения цвета - "цветовых краев" изображения. Деформируемая модель, применяемая для обнаружения контура подбородка, инициализируется внутри области лица и расширяется в сторону границ лица до тех пор, пока не встретит достаточное количество пикселей с резким изменением цвета или яркости.

Реализация разработанных алгоритмов используется в системе автоматического построения моделей головы по двум фотографиям (фронтальной и профильной). Результаты работы алгоритмов представлены на рисунке .

Третья глава описывает систему, реализующую управление передвижением курсора и срабатыванием кнопки "мыши" при помощи движений головы и мимики пользователя, регистрируемых видеокамерой. Разработка системы велась совместно с факультетом психологии МГУ им. М.В. Ломоносова. Сотрудник факультета психологии, к.ф.-м.н., д.ф.н. А.Н. Кричевец занимался разработкой общей концепции системы, необходимых требований к работе системы и характеру перемещения курсора в зависимости от действий пользователя. Перед автором стояла задача разработки и реализации алгоритмов определения движений головы и мимики пользователя в видеопотоке, а также трансляции движений головы в сигналы перемещения курсора.

Система предназначена для инвалидов, которые не могут справиться с клавиатурой и мышью - в первую очередь для детей, страдающих церебральным параличом. Этот недуг затрагивает прежде всего двигательный аппарат, однако подвижность головы в значительной степени сохраняется, что позволяет использовать движения головы для управления компьютером. Существует и другой тип управления, использующий распознавание направления взгляда для указания положения курсора на экране. Однако, как показывают исследо-

вания, проведенные в ряде лабораторий по разработке человеко-машинного интерфейса, взгляд используется человеком для обозначения области его фокуса внимания, а не для совершения действий. Движения глаз часто происходят неосознанно, и использование направления взгляда, как инструмента для управления курсором, ведет к заметному дискомфорту. Движения головой значительно лучше поддаются сознательному контролю со стороны пользователя.

Существующие коммерческие системы опираются на использование подсветки в инфракрасном диапазоне для облегчения обнаружения маркеров, располагаемых на лице или голове пациента. Цена систем, по функциональности аналогичных разработанной, колеблется от 900 до 15,000 долларов и требует инфракрасных источников света и видеокамер, чувствительных в ИК диапазоне. Разработанная система использует в качестве дополнительного оборудования бытовую видеокамеру стоимостью около двадцати долларов (дешевизна оборудования является необходимым условием для внедрения системы в отечественных медицинских центрах). Изображения, получаемые с нее, характеризуются низким разрешением (около 320x240 пикселей), высоким уровнем шума и низкой четкостью. Система должна функционировать на компьютере невысокой вычислительной мощности, оставляя достаточно вычислительных ресурсов для работы пользовательских приложений.

Для выполнения строгих требований к устойчивости распознавания в условиях низкого качества изображения, а также использования небольших вычислительных ресурсов, в системе используется реперная конструкция (шлем) с контрастными маркерами для облегчения задачи распознавания движения головы.

Положение маркеров на изображении, получаемом с видеокамеры, транслируется в сигналы перемещения курсора мыши. В системе используются три маркера - два расположены с левой и правой стороны головы, примерно на одинаковом удалении от камеры. Третий расположен примерно по центру между ними, на меньшем удалении от камеры. После обнаружения положений маркеров в текущем кадре производится вычисление координат центрального маркера в системе координат, связанной с двумя боковыми маркерами. Положение точки начала координат O вычисляется как среднее арифметическое

координат боковых маркеров. Длины проекций вектора, соединяющего точку O с центральным маркером, на прямую, проходящую через центры боковых маркеров и на перпендикуляр, проходящий через точку O , дают координаты центрального маркера u и v в системе координат, связанной с боковыми маркерами. Такой выбор координатной системы делает алгоритм нечувствительным к повороту головы в плоскости изображения и позволяет отслеживать только наклон и повороты головы вне плоскости изображения.

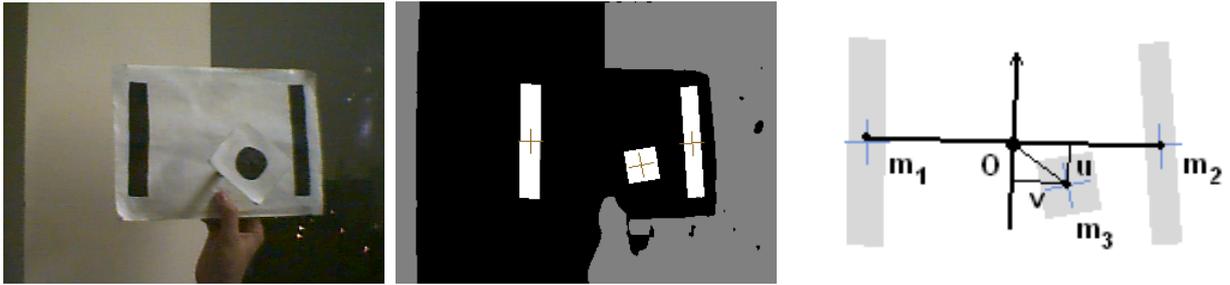


Рис. 6. Маркеры, используемые для распознавания движений головы.

В соответствии с рекомендациями психологов, в системе предусмотрено два различных способа трансляции поворота головы в движения курсора - по скорости (аналогично манипулятору типа “джойстик”) и по положению (аналогично манипулятору типа “мышь”). Первый тип управления означает, что поворот головы на определенный угол влечет перемещение курсора в соответствующем направлении со скоростью, пропорциональной углу поворота. Во втором типе управления положение курсора напрямую зависит от угла поворота головы. Система позволяет комбинировать два вида управления в зависимости от задания, стоящего перед пользователем, и его персональными предпочтениями и возможностями. На основе вычисленных координат u и v перемещение курсора мыши вычисляется по следующим формулам:

$$\begin{aligned} x_t &= x_{t-1} + \alpha_x \cdot (f_d(u_t - u_0) + \int_{u_{t-1}-u_0}^{u_t-u_0} f_m(u) du) \\ y_t &= y_{t-1} + \alpha_y \cdot (f_d(v_t - v_0) + \int_{v_{t-1}-v_0}^{v_t-v_0} f_m(v) dv) \end{aligned} \quad (0.6)$$

В формуле (0.6) (x_{t-1}, y_{t-1}) и (x_t, y_t) есть координаты курсора мыши в предыдущий и текущий момент времени, (u_{t-1}, v_{t-1}) и (u_t, v_t) - координаты центрального маркера относительно боковых на предыдущем и текущем кадре. Точка с координатами (u_0, v_0) соответствует “нейтральному” положе-

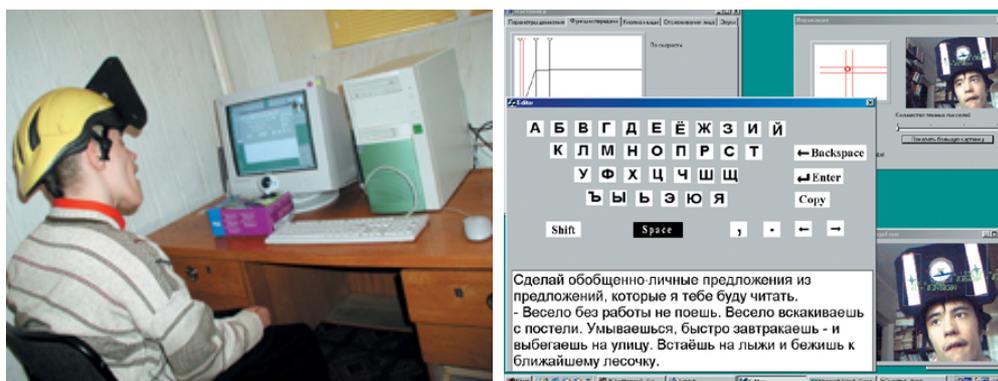
нию маркеров, когда перемещения курсора не происходит. Кусочно-линейные функции $f_d()$ и $f_m()$ регулируют скорость и характер перемещения курсора в зависимости от положения головы. Коэффициенты α_x и α_y регулируют интенсивность движения по каждой из осей.

Срабатывание кнопки мыши должно быть связано с изменением мимики пользователя, поскольку необходимо полностью исключить использование рук. Наиболее предпочтительным оказалось открывание рта, поскольку это действие могут успешно совершать большинство пациентов - потенциальных пользователей системы и, в отличие от моргания, это действие производится человеком осознанно. Система следит за мимикой пользователя, и, если он держит рот открытым достаточно долгое время, происходит срабатывание кнопки мыши. В системе предусмотрено несколько вариантов функционирования кнопки мыши: залипающая кнопка и срабатывание кнопки только при отсутствии перемещения курсора мыши (для части пациентов оказывается достаточно сложным одновременно следить за движением курсора и контролировать мимику).

Для выделения областей маркеров на изображении используется яркостная сегментация (маркеры представляют собой темные области на светом фоне). К полученному после сегментации изображению применяется морфологическая фильтрация для снижения уровня шума. Следующим шагом является выделение связных компонент, после чего производится поиск областей, соответствующих маркерам. Распознавание осуществляется по характеристикам областей (площадь, периметр, удлиненность, компактность и ориентация главной оси) и положению маркеров на предыдущем кадре.

Для увеличения устойчивости распознавания используются маркеры характерной формы, достаточно протяженные по площади: боковые маркеры - вертикально ориентированные вытянутые прямоугольники, центральный маркер - круговая область. Это позволяет минимизировать вероятность случайного совпадения характеристик связных областей, относящихся к маркерам, и фоновых областей на изображении. Вытянутая форма прямоугольников позволяет использовать ориентацию главной оси инерции в качестве дополнительного признака областей при начальном обнаружении и отслеживании маркеров.

Для упрощения использования системы были разработаны алгоритмы автоматического обнаружения маркеров на первом кадре и последующего определения параметров яркостной сегментации, а также адаптация этих параметров по ходу сеанса. Наличие подобных алгоритмов позволяет полностью исключить необходимость ручной настройки системы и делает ее устойчивой к изменению условий освещения.



(a)

(б)

Рис. 7. Пример сеанса работы с системой

В четвертой главе описываются алгоритмы распознавания и отслеживания положения антропометрических точек лица в видеопотоке. Цель разработки подобных алгоритмов - освободить пользователя системы, описанной в третьей главе, от необходимости использовать дополнительные маркеры для работы.

В качестве характерных антропометрических точек были выбраны две точки в районе глаз, одна в нижней части лица (рот) и одна в области носа (не лежащая в плоскости глаз и рта). Такой выбор точек связывает их с характерными чертами лица (что облегчает их обнаружение), и дает возможность вычислить ориентацию головы в трехмерном пространстве по проекциям точек на изображение.

Для определения положения черт лица не требуется столь же высокой точности и подробности распознавания, как в случае построения трехмерной модели головы. Учитывая это, а также невысокое качество изображения и низкое разрешение кадров видеопотока, генерируемого недорогими видеокameraми, методы, разработанные для решения этой задачи, нацелены на устойчивое распознавание положения черт лица, без дополнительных деталей.

Выделение области лица на текущем кадре производится с помощью метода статистических моментов (см. главу 1). Результатом применения этого метода является выделенная эллиптическая область лица. Дальнейшие действия построены в предположении, что область лица уже найдена.

Обычно системы отслеживания черт лица в видеопотоке фокусируются на определении положения центров глаз на изображении. Однако, если пользователь носит очки (что является скорее правилом для пользователей ЭВМ, чем исключением), определение положения центров глаз значительно усложняется из-за бликов стекол. По результатам испытаний линия бровей была выбрана как более устойчиво обнаруживаемый признак, нежели центры глаз.

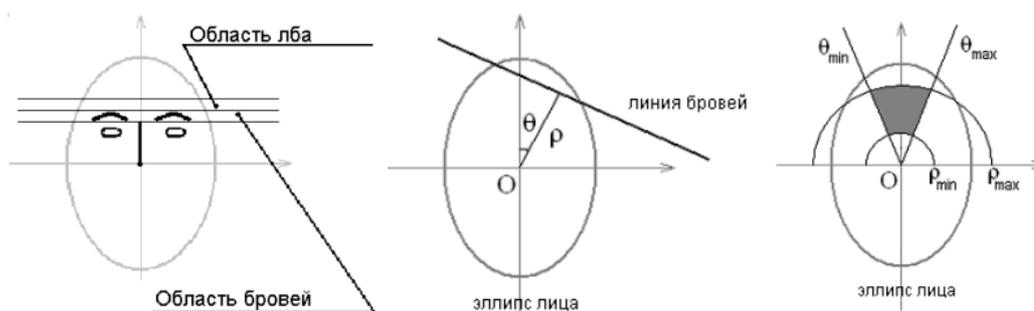


Рис. 8. Иллюстрация к методу обнаружения бровей

Поиск линии бровей происходит путем обнаружения прямоугольной области, внутри которой содержится значительное количество пикселей с абсолютной величиной градиента яркости $\nabla I(x, y)$, превышающей заданный порог. Дополнительно требуется отсутствие пикселей с большими абсолютными значениями градиента непосредственно над областью бровей - в области лба (рис. 8). Обнаружение производится с помощью модифицированного преобразования Хафа. Используется система координат с началом в центре эллипса лица и осями, направленными вдоль осей эллипса. Положение линии бровей задается двумя параметрами ρ и θ (рис. 8). Границы отрезков допустимых значений параметров $[\rho_{min}, \rho_{max}]$ и $[\theta_{min}, \theta_{max}]$ задаются, исходя из возможной неточности в определении эллипса лица. Каждой ячейке фазового пространства $[\rho_{min}, \rho_{max}] \times [\theta_{min}, \theta_{max}]$, дискретизированного с шагами $\Delta\rho$ и $\Delta\theta$, ставится в соответствие счетчик, изначально полагаемый равным нулю. Для компенсации дискретности разбиения используется следующий алгоритм обнаружения линии бровей: для каждого пикселя (x, y) , такого что

$$|\nabla \overline{I(x, y)}| > \theta:$$

1. Перевести координаты (x, y) пикселя в систему, связанную с эллипсом (x', y') ;
2. Для каждого $\theta \in [\theta_{min}, \theta_{max}]$ подсчитать $\rho = x' \cdot \cos \theta + y' \cdot \sin \theta$;
3. Если $\rho \in [\rho_{min}, \rho_{max}]$, то увеличиваются счетчики ячеек фазового пространства со следующими координатами: $(\lfloor \rho / \Delta \rho \rfloor, \lfloor \theta / \Delta \theta \rfloor)$, $(\lfloor \rho / \Delta \rho \rfloor + 0.5, \lfloor \theta / \Delta \theta \rfloor)$, $(\lfloor \rho / \Delta \rho \rfloor - 0.5, \lfloor \theta / \Delta \theta \rfloor)$;
4. Если $(\rho - \Delta \rho) \in [\rho_{min}, \rho_{max}]$, то уменьшаются счетчики ячеек фазового пространства со следующими координатами: $(\lfloor (\rho - \Delta \rho) / \Delta \rho \rfloor, \lfloor \theta / \Delta \theta \rfloor)$, $(\lfloor (\rho - \Delta \rho) / \Delta \rho \rfloor + 0.5, \lfloor \theta / \Delta \theta \rfloor)$, $(\lfloor (\rho - \Delta \rho) / \Delta \rho \rfloor - 0.5, \lfloor \theta / \Delta \theta \rfloor)$.

После рассмотрения всех пикселей изображения лица выбирается линия бровей с параметрами (ρ', θ') , соответствующими наибольшему значению счетчика. Использование варианта преобразования Хафа, компенсирующего дискретность разбиения фазового пространства, дает повышенную по сравнению с традиционным алгоритмом устойчивость распознавания. Еще одним важным отличием приведенного алгоритма от традиционного преобразование Хафа является встроенный в алгоритм механизм “штрафования” (шаг 4) нежелательных конфигураций обнаруживаемого объекта. Обнаружение линии бровей дает две из требуемых четырех точек лица.

Предлагаемый алгоритм отслеживания положения губ с использованием карты вероятности цвета губ (0.7) отличается от алгоритмов, использующих только яркостную информацию или заранее обученный цветовой классификатор, высокой надежностью распознавания и устойчивостью к различным условиям освещения.

$$Lip(x, y) = \frac{u^*(RGB)}{v^*(RGB)} + (1 - Skin(RGB)) \quad (0.7)$$

В (0.7) RGB - цветовые координаты пикселя с координатами (x, y) в пространстве RGB, $Skin(RGB)$ - вероятность принадлежности цвета к оттенку кожи, u^*, v^* - координаты соответствующего цвета в пространстве CIEL*u*v*. На основе получаемой карты вероятностей производится поиск



Рис. 9. Примеры карты вероятности цвета губ

областей пикселей, с высокой вероятностью имеющих форму, близкую к эллиптической. Для увеличения надежности обнаружения производится шумоподавляющая фильтрация карты вероятностей. Обнаружение координат губ происходит путем нахождения координат с наибольшим значением критерия согласия, вычисляемого по следующей формуле:

$$E(x_c, y_c) = \alpha \cdot \sum_{(x,y) \in In(x_c, y_c)} Lip(x, y) - \beta \cdot \sum_{(x,y) \in Out(x_c, y_c)} Lip(x, y) \quad (0.8)$$

Области $In(x_c, y_c)$ и $Out(x_c, y_c)$ задаются ожидаемой формой и размером губ на изображении. Результаты отслеживания представлены на рисунке 10(б). Положение губ предоставляет информацию о положении третьей точки.

Для определения ориентации головы по положению черт лица на изображении необходимо знать координаты еще одной точки, не лежащей в плоскости рта и бровей. Идеально для этого подходит кончик носа, однако устойчивого отслеживания кончика носа в неизвестных условиях освещения достичь весьма проблематично. С другой стороны, при соответствующем положении камеры, ноздри представляют собой четко видимые, характерные особенности изображения. Обнаружение координат ноздрей производится путем поиска контрастных областей определенной формы с помощью критерия согласия, задаваемого следующей формулой:

$$E(x_c, y_c) = \alpha \cdot \sum_{(x,y) \in In(x_c, y_c)} I(x, y) - \beta \cdot \sum_{(x,y) \in Out(x_c, y_c)} I(x, y) + \gamma \cdot \sum_{(x,y) \in In(x_c, y_c)} |\nabla I(x, y)|; \quad (0.9)$$

Области $In(x_c, y_c)$ и $Out(x_c, y_c)$ задаются шаблоном (см. рисунок 10(а)). Результаты отслеживания представлены на рисунке 10(б).

Предложенные алгоритмы функционируют в реальном времени на ЭВМ с процессором Pentium II (тактовая частота 800 МГц) со скоростью 8 кадров в секунду.

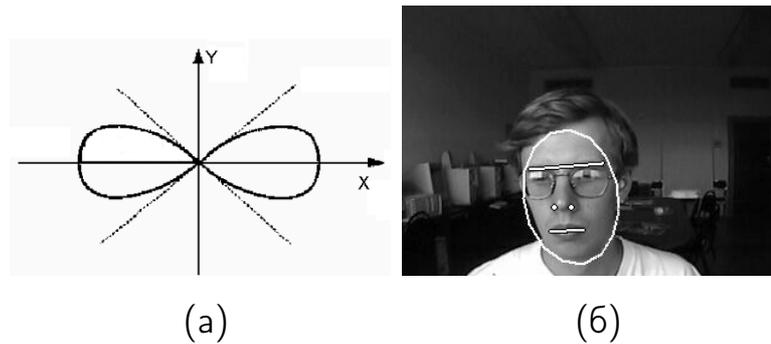


Рис. 10. (а) Шаблон для обнаружения ноздрей. (б) Пример обнаружения полного набора антропометрических точек

В заключении сформулированы основные результаты работы.

Основные результаты работы состоят в следующем:

1. Предложены новые алгоритмы выделения лица человека на изображении, работающие в реальном времени и удовлетворяющие следующим ограничениям: работа с изображениями низкого разрешения (порядка 320x240 пикселей) и высокого уровня шума, характерного для бытовых видеокамер.
2. Разработан алгоритм автоматического выделения антропометрических точек лица на фронтальной фотографии среднего разрешения (порядка 640x480 пикселей) для построения фотореалистичных моделей виртуальной реальности.
3. Разработан алгоритм управления курсором “мыши” с помощью движений головы и мимики пользователя.
4. Разработан алгоритм для автоматического определения положения антропометрических точек лица в видеопотоке низкого разрешения в реальном времени.

Основное содержание диссертационной работы изложено в следующих публикациях:

- [1] *Вежневек В. П.* . Локализация человеческого лица на цветном растровом изображении // Труды конференции “Математические методы распознавания образов” (ММРО-10). — 2001.
- [2] *Вежневек В. П.* . Использование контурных моделей для выделения черт лица на фронтальном изображении // Труды конференции “Математические методы распознавания образов” (ММРО-10). — 2001.
- [3] Automatic extraction of frontal facial features for 3-d face modeling / *V. Vezhnevets, S. Soldatov, A. Degtiareva, I.-K. Park* // Proc. Sixth Asian Conference on Computer Vision. — 2004.
- [4] *Vezhnevets V.* Face and facial feature tracking for natural human-computer interface // Труды конференции ГрафиКон'2002. — 2002.
- [5] *Vezhnevets V.* Method for localization of human faces in color-based face detectors and trackers // Труды конференции Digital Information Processing And Control In Extreme Situations'2003. — 2003.
- [6] *Vezhnevets V., Degtiareva A.* Robust and accurate eye contour extraction // Труды конференции ГрафиКон'2003. — 2003.
- [7] *Vezhnevets V., Sazonov V., Andreeva A.* A survey on pixel-based skin color detection techniques // Труды конференции ГрафиКон'2003. — 2003.