УДК 004.93

МНОГОМОДАЛЬНАЯ СИСТЕМА ДЛЯ БЕСКОНТАКТНОЙ РАБОТЫ С ПЕРСОНАЛЬНЫМ КОМПЬЮТЕРОМ

А. Л. Ронжин,
канд. техн. наук, старший научный сотрудник
А. А. Карпов,
аспирант
Санкт-Петербургский институт информатики и автоматизации РАН

Рассмотрена многомодальная система ICANDO, разработанная для помощи людям без рук или с проблемами двигательных функций рук при работе с персональным компьютером. В этой системе объединяются модули автоматического распознавания русской речи и отслеживания движений головы. Разработанная система была применена для бесконтактного (без использования рук) управления графическим пользовательским интерфейсом для задач Интернет-коммуникации и работы с текстовыми документами.

The paper describes multimodal system ICANDO developed for assistance to persons without hands or with disabilities of their hands or arms. This system combines the modules for Russian speech recognition and head tracking in one multimodal system. The developed system was applied for hands-free control of Graphical User Interface for such tasks as Internet communication and work with text documents.

Введение

Многомодальные интерфейсы способны обрабатывать несколько естественных для человека способов ввода информации: речь, письменный ввод, жесты руками, направление взгляда, движения головы и тела совместно с мультимедийной системой вывода информации. Многомодальные системы представляют новое направление в информатике и концептуально изменяют традиционные интерфейсы, вплоть до отказа от клавиатуры и различных устройств-манипуляторов.

Группой речевой информатики СПИИРАН разработана одна из первых российских многомодальных систем ICANDO (Intellectual Computer AssistaNt for Disabled Operators), предназначенная для бесконтактного управления персональным компьютером (без использования клавиатуры и мыши). Такая система необходима в основном для помощи людям, имеющим проблемы с двигательными функциями рук, или же вообще без рук. Вместо клавиатуры и мыши для управления графическим интерфейсом компьютера здесь используются голосовые команды и движения головой. Вернее, происходит отслеживание не всей головы, а только позиции кончика носа, поскольку кончик носа человека является центром лица, и, когда пользователь двигает головой (поворачивает влево, вправо, наклоняет или поднимает голову), кончик носа синхронно двигается в эту сторону, что позволяет использовать его для управления курсором мыши на экране монитора. Такая система бесконтактного управления компьютером также может успешно применяться для игровых приложений, в системах виртуальной реальности и в ряде других робототехнических приложений.

В системе ICANDO используются голосовые команды на русском языке. Для распознавания русской речи применяется разработанная в СПИИРАН оригинальная система автоматического распознавания речи SIRIUS (SPIIRAS Interface for Recognition and Integral Understanding of Speech) [1].

Обработка видеоинформации

Для отслеживания движений головы могут применяться как аппаратно-ориентированные подходы, так и программно-ориентированные. С точки зрения программной реализации более простой способ – когда пользователь надевает на голову специальные устройства (шлем, очки виртуальной реальности или специальные конструкции с отражающими метками), но такие устройства дороги, требуют длительной предварительной на-

стройки и вносят дополнительный дискомфорт при работе. Поэтому разрабатываются автоматические способы обнаружения лица, его характерных черт и отслеживания перемещения лица в видеопотоке без использования искусственных маркеров. Такой подход более сложен в программной реализации, но не накладывает дополнительных ограничений на пользователя и обеспечивает максимальную естественность при работе с компьютером.

Для многомодальной системы ICANDO был разработан программный метод отслеживания движений головы пользователя на основе метода Лукаса–Канаде для оптического потока [2]. Определение первоначального положения лица пользователя на изображении с видеокамеры реализуется программным путем с помощью детектора объектов Хаара, который определяет прямоугольные графические области, которые с высокой степенью вероятности содержат изображение лица человека [3]. Размер этой области должен быть не менее 250×250 точек для того, чтобы захватывать только одно лицо, достаточно близко расположенное по отношению к камере, это ускоряет процесс обработки видеопотока.

Для управления курсором мыши в реальном времени был разработан специальный алгоритм, включающий в себя два режима: калибровка (или настройка) и отслеживание. При калибровке производится привязка координат курсора к положению кончика носа. На рисунке показано, как пользователь производит настройку курсора мыши. В окне, расположенном по центру дисплея, отображается изображение, поступающее с видеокамеры. Учитывая стандартные пропорции лица, предположительное положение кончика носа автоматически отмечается синей точкой на экране. В течение нескольких секунд настройки пользователь должен совместить реальное изображение своего носа с этой точкой. По истечении времени калибровки курсор мыши выставляется по центру рабочего стола и «привязывается» к положению кончика носа. При управлении курсором в режиме отслеживания алгоритм иногда «теряет» позицию кончика носа пользователя по причине недостатка света, очень быстрых перемещений головы или выхода из зоны видеозахвата. Для решения этой проблемы введена специальная голосовая команда «калибровка», которая запускает процесс калибровки снова.

Механизм многомодального объединения информации

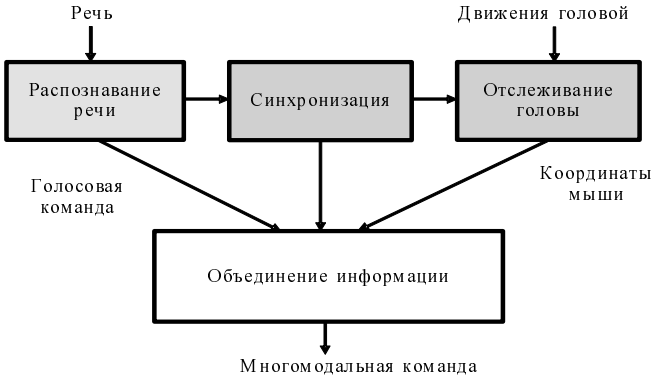
В системе ICANDO используются две естественные входные модальности: речь и движения головы. Так как обе модальности являются активными, то они должны непрерывно отслеживаться компьютером [4]. Каждая из модальностей передает свою семантическую информацию: положение головы (носа) определяет положение курсора мыши в данный момент времени, а речевой сигнал передает информацию о действии, которое должно быть выполнено с некоторым объектом рабочего стола компьютера. Для распознавания речи используется система распознавания речи SIRIUS, основанная на скрытом марковском моделировании и морфемном представлении языка и речи, что обеспечило инвариантность к грамматическим отклонениям и высокую скорость обработки по сравнению с общепринятыми целословными моделями языка. Перечень голосовых команд системы ICANDO состоит из 41 команды (например, «печать» или «сохранить») [5].

Перемещение курсора зависит только от положения кончика носа и отрабатывается непрерывно по мере обработки видеопотока. В том случае, когда система распознавания речи зафиксировала и распознала некоторую голосовую команду, ее необходимо выполнить с учетом информации о положении курсора на экране. Синхронизация модальностей производится следующим образом: текущее положение курсора вычисляется в момент определения начальной границы речи. Это связано с той проблемой, что во время произнесения фразы пользователь может непреднамеренно незначительно перемещать голову и тем самым менять положение курсора, в результате чего он будет указывать на другой графический объект. Кроме того, речевое намерение формируется в сознании в соответствии с целью и ситуацией до того, как произносится голосовая команда. Для объединения информации, поступающей от двух модальностей, используется фреймовый метод позднего объединения, когда поля некоторой структуры заполняются данными по мере их поступления, а по окончании процесса распознавания речи выполняется многомодальная команда.

Результаты экспериментов по использованию системы

В качестве аппаратного обеспечения используется миниатюрная USB веб-камера Logitech QuickCam for Notebooks Pro, обеспечивающая разрешение 640×480 точек при 25 кадрах в секунду. Также камера записывает аудиосигнал с частотой 16 кГц при помощи встроенного в камеру микрофона. Использование профессиональной цифровой видеокамеры (например, Sony DCR-PC1000E) позволило достичь большей точности распознавания графических объектов и речи, но, учитывая, что система должна быть доступна для большинства пользователей, мы применяем камеру стоимостью до 50 ˆ.

Тестирование системы было произведено пятью пользователями, которые имели незначительный опыт работы с персональным компьютером, а также одним пользователем с ограниченными возможностями (без рук). В ходе экспериментов пользо-



■ *Работа с компьютером посредством системы ICANDO*

ватели работали с приложениями операционной системы Microsoft Windows. Задача включала в себя управление текстовым редактором NotePad, а также доступ в Интернет посредством MS Internet Explorer.

В ходе тестирования было проведено сравнение скорости работы при помощи стандартного управления компьютером (клавиатура+мышь) и при использовании многомодальной системы ICANDO. Экспериментально было установлено, что многомодальный способ ввода оказался примерно в два раза медленнее, чем стандартный клавиатурно-ориентированный способ. Однако такое замедление вполне приемлемо, так как система разрабатывается для помощи людям со специфическими потребностями. Точность распознавания голосовых команд составила свыше 97 % для каждого из пользователей.

Заключение

В статье представлена многомодальная система ICANDO для бесконтактной работы с персональным компьютером. Описаны общая архитектура системы, процесс видеообработки, а также механизм объединения модальностей. Результаты тестирования системы позволяют заключить, что разработанная многомодальная система может успешно применяться для бесконтактного управления компьютером пользователями-инвалидами.

В ноябре 2005 года работа системы демонстрировалась в программе «Время» на «Первом канале» телевидения, и реальный пользователь без рук успешно работал с персональным компьютером посредством разработанного многомодального интерфейса. Работа выполняется при поддержке гранта EC SIMILAR NoE FP6: IST-2002-507609.

Литература

1. **Ронжин А. Л., Карпов А. А., Ли И. В.** Система автоматического распознавания русской речи SIRIUS // Искусственный интеллект. № 3. 2005. С. 590–601.
2. **Bouguet J.-Y.** Pyramidal implementation of the lucas kanade feature tracker // Intel Corporation. Microprocessor Research Labs: Tech. Rep. 2000.
3. **Lienhart R., Maydt J.** An Extended Set of Haar-like Features for Rapid Object Detection // IEEE International Conference on Image Processing ICIP'2002: Proc. 2002. P. 900–903.
4. **Карпов А. А., Ронжин А. Л.** Многомодальные интерфейсы в автоматизированных системах управления // Изв. вузов. Сер. Приборостроение. 2005. Т. 48. № 7. С. 9–14.
5. **Ronzhin A. L., Karpov A. A.** Assistive multimodal system based on speech recognition and head tracking // 13-th European Signal Processing Conference EUSIPCO-2005: Proc. Turkey, 2005.