

УДК 004.5

Распознавание жестов

С.А. Кладов, студент

*Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Программное обеспечение ЭВМ и информационные технологии»*

Научный руководитель: О.В. Rogozin, к.т.н., доцент

*Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Программное обеспечение ЭВМ и информационные технологии»*

irudakov@bmstu.ru

Введение

Проблема взаимодействия человека с компьютером встает перед разработчиками с самого зарождения информационных технологий и появления первых ЭВМ. Одним из самых популярных средств взаимодействия является всем известная «мышь». За десятилетия прогресса она сильно поменялась: лазеры сменили механические компоненты, манипулятор стал беспроводным. Многие усовершенствовались, но концепция осталась прежней, так же как и недостатки, например, сложности при работе в трех измерениях и обязательное наличие поверхности, на которой работает мышь.

В последнее время появилась новая технология реализации интерфейсов – распознавание жестов. Основная идея этой технологии заключается в том, чтобы максимально приблизить взаимодействие человека и компьютера к естественному взаимодействию между людьми.

Что такое жест?

Жест – это форма невербального общения, при которой некоторое действие или движение человеческого тела или его части передает информацию. Жест может содержать движение как одной, так и нескольких частей тела и нести эмоциональную нагрузку, что делает его очень информативным средством общения и взаимодействия.

Классификация жестов

Рассмотрим классификацию жестов:

- Тип жеста

- Простые жесты
- Комплексные жесты
- Способ обработки
- Обработка после ввода
- Обработка в процессе ввода
- Длительность
- Статические жесты
- Динамические жесты

Основной идеей классификации типа жестов является тот факт, что жест может быть представлен различными способами. Среди простых часто используются пальцевые жесты (устройства с сенсорными экранами), жесты руками (Kinect, MYO). Примером комплексных жестов являются движения, которые задействуют несколько частей тела или состоят из нескольких простых (Wii, Kinect).

Классификация жестов по способу обработки разделяет их на 2 категории: жесты, обрабатываемые после ввода и жесты, обрабатываемые в процессе ввода. Пример жеста второй категории – увеличение/уменьшение фотографии в устройствах фирмы “Apple”.

Длительность жеста показывает, изменяется ли он со временем. Следовательно, жесты могут быть разделены на статические и динамические. Статические жесты представляются в виде одного изображения, в то время как динамические жесты – это последовательность изображений либо видеофрагмент.

Система распознавания жестов.

Упрощенная структура системы распознавания жестов приведена на схеме 1. Она состоит из 3 блоков:

- Устройство ввода
- Блок распознавания
- Реагирующий блок

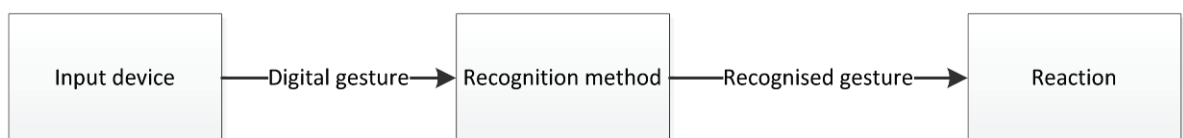


Схема 1. Структура системы распознавания жестов

Устройство ввода является точкой входа в систему. Оно преобразовывает жест в цифровую форму и передает его в блок распознавания, который сравнивает поступившую информацию с шаблонами, хранящимися в базе данных, и в качестве результата

возвращает распознанный жест. В итоге реагирующий блок выполняет некоторые функции в соответствии с распознанными жестами.

Популярные методы

По мере развития технологии распознавания жестов определились наиболее популярные методы:

- Скелетный метод
- 3D-распознавание
- 2D-распознавание

Скелетный метод использует математическую модель человеческой руки. Преимуществом этого метода является скорость работы, так как он использует высоко детализированную модель, которая описана с помощью длин сегментов и углов между ними. Примеры скелета человеческой ладони приведены на рисунках 1 и 2с.

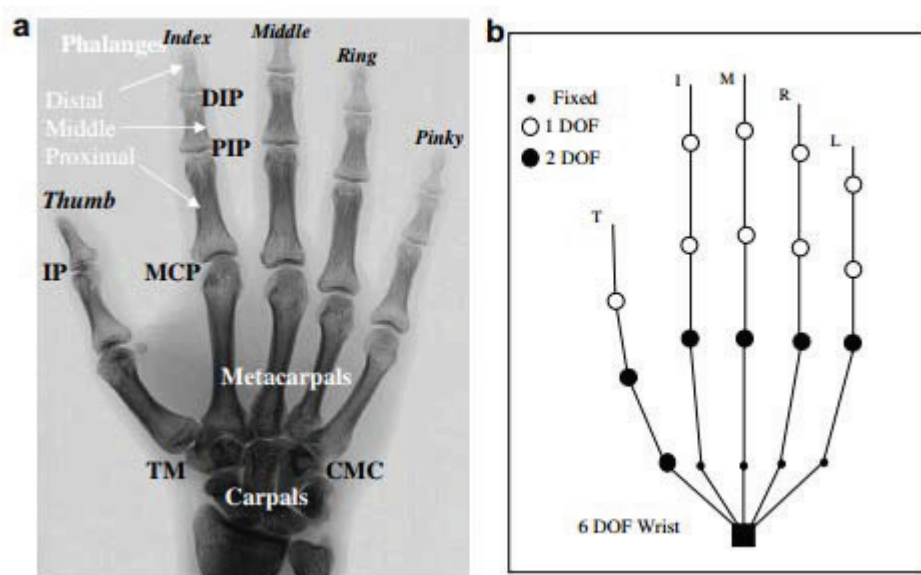


Рис. 1. Скелет ладони [3]

Следующий метод использует 3D-модель для идентификации жеста (рис. 2а, 2б). Такие модели могут быть представлены в виде сложных трехмерных поверхностей и классифицироваться с помощью нейронных сетей. Недостатком этого метода является его ресурсоемкость. Построение модели, обучение нейронной сети и ее использование могут потребовать значительных ресурсов.

Метод 2D-распознавания схож с предыдущим методом, но оперирует двумерными изображениями вместо объемных моделей (рис. 2д, 2е). Как следствие, снижается вычислительная сложность и отпадает необходимость в специальном оборудовании, так как для получения изображений может быть использована обычная веб-камера. Основным недостатком этого метода является низкая точность.

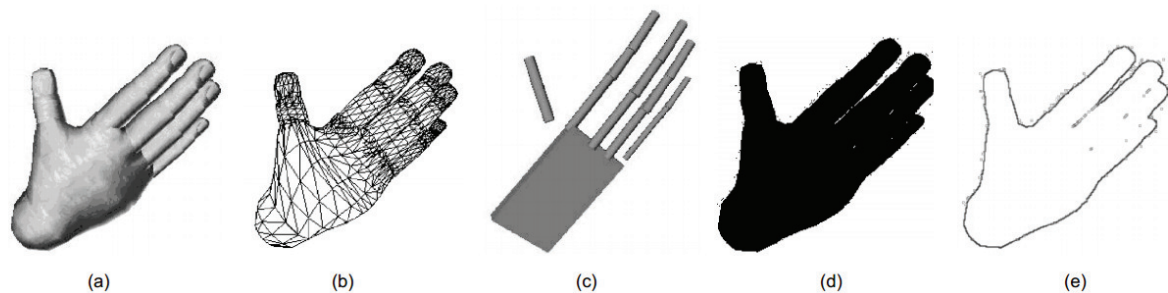


Рис. 2. Варианты представления ладони [2]

Стоит отметить, что некоторые источники делят алгоритмы на 2 категории:

- Использующие 3D-модель
- Использующие внешнее подобие

Алгоритмы, основанные на 3D-моделировании жестов, используют шарнирную модель человеческой руки для определения ключевых параметров. На основе этих параметров затем производится классификация жестов. Алгоритмы, использующие внешнее подобие, напрямую соотносят образ руки, ладони и их движения с определенными жестами [2]. Основываясь на этих определениях, можно сделать вывод, что первая категория алгоритмов соответствует скелетному методу, а вторая категория объединяет методы 2D- и 3D-распознавания из первой классификации.

Устройства ввода

Все вышеописанные методы используют различные устройства ввода. Они могут быть разделены на 2 группы. Первая группа устройств (взаимодействующие) включает:

- Перчатки с датчиками;
- Специальные контроллеры;

Отличительной особенностью этих устройств заключается в том, что пользователь напрямую взаимодействует с ними, и жесты выполняются пользователем и устройством совместно. Основным недостатком этой группы является некоторая неуклюжесть таких жестов. Это объясняется тем, что человеку неудобно использовать устройства-посредники для жестового общения. Эту неловкость можно преодолеть с помощью неконтактных подходов, основанных на видео. Они подразумевают использование нескольких видеокамер и алгоритмов компьютерного зрения для интерпретации жестов [2]. Поэтому вторая группа (наблюдающие устройства) - это:

- Камеры с возможностью определения глубины изображения;
- Стереокамеры;
- Обычная камера;

Также допустимы комбинации этих групп. Например, система, разработанная Робертом Вангом и Йованом Поповичем (Robert Y. Wang and Jovan Popović), использует цветную перчатку и видеокамеру [1] (см. рисунок 3).



Рис. 3. Комбинация перчатки и камеры

Применение технологии

Системы, основанные на распознавании жестов, могут быть использованы во многих областях деятельности, таких как: компьютерные игры, андроиды, управление с помощью выражения лица, альтернативные варианты интерфейсов и т.д.

В настоящее время многие системы уже используют преимущества технологии распознавания жестов. Наиболее известными примерами являются Kinect, Nintendo Wii и MYO. Kinect и Wii используют распознавание жестов для создания новых игровых ощущений, в то время как MYO демонстрирует новый уровень взаимодействия с различными устройствами от телевизора до квадрокоптера.

Заключение

Распознавание жестов является очень мощным инструментом для создания интерфейсов и выводит человеко-машинное взаимодействие на новый уровень. Поэтому данная технология представляется очень перспективной в плане развития.

Список литературы

1. Robert Y. Wang and Jovan Popović, (2009), "Real-Time Hand-Tracking with a Color Glove", ACM Transactions on Graphics, Vol. 28(3).
2. Pavlovic, V., Sharma, R. & Huang, T. (1997), "Visual interpretation of hand gestures for human-computer interaction: A review", IEEE Trans. Pattern Analysis and Machine Intelligence., July, 1997. Vol. 19(7), pp. 677 -695.
3. A Erol, G Bebis, M Nicolescu, RD Boyle, X Twombly, "Vision-based hand pose estimation: A review", Computer Vision and Image Understanding Volume 108, Issues 1-

2, October–November 2007, Pages 52-73 Special Issue on Vision for Human-Computer Interaction, [doi:10.1016/j.cviu.2006.10.012](https://doi.org/10.1016/j.cviu.2006.10.012).