

Язык жестов для взаимодействия с компьютером

Иванов С.Ю., Шихов А.И., Константинов Ю.А.

НИУ ВШЭ МИЭМ. Департамент компьютерной инженерии. isu@itas.miem.edu.ru

Аннотация. Статья посвящена разработке программных средств анализа визуальных образов, полученных с web-камер. Рассматриваются различные варианты ввода информации и их особенности. Предлагается использование языка жестов для средств визуального ввода информации, и описываются средства для реализации такого языка. В статье рассматриваются алгоритмы определения направления движения объекта и определения силуэта объекта любой формы. Алгоритмы реализованы на *VB*. Рассматривается система распознавания жестов. Неотъемлемой частью системы является подсистема формирования эталонов визуальных объектов, которая позволяет добавлять новые визуальные объекты в базу эталонов и получать упрощенное изображение в виде силуэта. Реализована специальная процедура, позволяющая выделять объект независимо от фона. Основное внимание уделяется надежному и безусловному распознаванию жестов и скорости распознавания. Система является открытой и расширяемой. Первый вариант рассчитан на распознавание 16 жестов. Система разрабатывалась для учебных курсов, где рассматривается распознавание визуальных объектов.

Ключевые слова: визуальные объекты, распознавание, жесты

В статье рассматриваются различные способы взаимодействия человека и компьютера, способы ввода информации и их определяющие влияние на проектирование интерфейсов.

В настоящее время в вычислительной технике пользователи в основном используют четыре способа ввода информации: ввод символов с клавиатуры, ввод координат объектов, речевой ввод и ввод информации, полученной в результате анализа отображения визуальных объектов.

Первый способ традиционный, предполагает использования клавиатуры. Нужен экран для отображения вводимой информации. Данный способ предполагает использование некоторого языка. Каждому символу присваивался код, по которому происходило его распознавание. Группы символов составили слова и фразы, которые позволили задавать множество команд и запросов. Мощность такого множества чрезвычайно велика. Однако *скорость ввода* здесь невелика, а *вероятность ошибки* высока.

Второй способ, определение объекта по его координатам. Выбор действия осуществляется за счет использования кликов, двойных кликов и указания пункта меню. Такой подход позволил резко сократить *скорость ввода* и уменьшить *вероятность ошибки*. Кроме того, пользователь работает с допустимыми командами в нужном формате. Возникает возможность обучения пользователя во время работы с системой.

Количество команд здесь мало («указать», «выполнить», «меню»). Мощность этого способа взаимодействия повышается за счет множества объектов на экране и меню (список допустимых команд) для выбранного объекта. Такой подход требует обязательного наличия экрана и размещенных на нем объектов с заданными в меню командами. Сейчас для такого способа ввода кроме «мыши» широко используют сенсорная панель и сенсорный экран. Возникли новые возможности, но количество команд увеличилось ненамного.

Третий способ, речевой ввод, оказался очень удобным для пользователя. Основной трудностью стало распознавание фраз пользователя, их интерпретация. Экран здесь необязателен, нужен микрофон. Команды можно задавать на некотором расстоянии от микрофона, а сам микрофон может быть на значительном расстоянии от вычислительной системы. Сейчас

речевой ввод успешно используют сотовые компании в поисковых системах. Мощность этого способа велика, но ограничивается словарем распознаваемых слов. Однако *скорость ввода* меньше, чем в первом случае, *вероятность ошибки* высока.

Четвертый способ – анализ изображения, полученного с видеокамеры. Здесь наиболее сложная задача – распознавание визуальных объектов. Основные параметры объектов: форма и цвет. Необходима настройка получаемого изображения. Очень важна освещенность.

При анализе и интерпретации предъявляемых объектов важное значение имеет точность отображения этих объектов и выделение некоторой части объекта, которая будет служить некоторой упрощенной моделью жеста. Распознавание жеста будет производиться по этой модели. Такой подход позволяет определять жест у различных объектов и решать проблему удаленности объекта.

Многие производители web-камер вводят специальную процедуру настройки. Процедура требует предъявления объекта для его распознавания и создания эталона. Как правило, эталон один, а программа распознавания определяет перемещение этого объекта. Оказалось, что перемещение объекта достаточно просто отследить [Девятков В.В. и др., 2007]. Некоторые компании упростили задачу распознавания, они используют безусловные цветовые метки. Для этого они используют цветной скотч или раскрашенные перчатки. Такой подход позволяет значительно упростить программное обеспечение и повысить надежность и скорость распознавания. Требования надежности и скорости распознавания очень важны для геймеров, которые составляют основную массу пользователей, применяющих такой способ ввода управляющей компьютером информации для больших экранов [Фомин Я. А., 2012]. Здесь используется довольно ограниченное количество образов, которые соответствуют командам «мыши». Мощность такого языка невысока.

Часто разработчики программного обеспечения для Web камер, используют наиболее простые способы распознавания, основанные на распознавании движения объекта, и взаимодействуют с компьютером ограниченным набором команд, часто совпадающими с командами «мыши». Качество изображения, получаемого с Web камер, сейчас достаточно для использования более сложных методов распознавания. Можно распознавать контуры различных объектов, получать их многократное изображение и уточнять образ визуального объекта, сохранять его в памяти и тем самым увеличить надежность распознавания демонстрированного объекта [Малашин Р.О. и др., 2013]. Все это позволяет распознавать различные жесты пользователя.

При распознавании визуальных объектов наиболее просто решается задача распознавания при точном совпадении изображения, предъявляемого объекта, с эталонным изображением. Небольшие расхождения с эталоном могут не позволить распознать объект. Возникает проблема о критериях оценки соответствия объекта эталону. Наиболее простое решение – это попиксельное сравнение двух изображений и определение процента соответствия. Но любое несоответствие понижает этот процент. Например, при различной освещенности, часть руки может иметь другой цвет, то есть пиксели другого цвета будут отображаться по-другому. Данную проблему часто решают за счет частой настройки и частого создания эталона. Такой подход плохо подходит для случая, когда объектов некоторое число. Настройка будет занимать значительное время и раздражать пользователя. Небольшие изменения в окружающей среде могут свести на нет проведенную настройку. Более того, изменение расстояния до объекта так же может понизить качество распознавания. Возникает проблема масштабирования.

Для реализации языка жестов предлагается на базе полученного изображения формировать простую модель объекта, которую будем рассматривать как жест. Жест сравнивается с базой эталонов и определяется его код, который затем преобразуется в определенную команду для компьютера. Формирование такой модели направлено на надежное и безусловное распознавание жеста. Были проанализированы различные свойства визуальных объектов и были исключены свойства, которые могли дать существенную ошибку при распознавании. В первую очередь – это цветовые оттенки. Настройка на конкретный цвет значительно упрощает распознавание. Однако такой подход неудобен для пользователя. В качестве модели объекта берем его силуэт. Силуэт хранится в матрице, и она сравнивается с матрицей эталона.

Язык жестов развивается с незапамятных времен. В середине XVIII глухой французский учитель Лоран Клерк создал первую школу для глухих в США. На VII Всемирном Конгрессе по проблемам глухоты в Вашингтоне в 1975 г. была принята и утверждена Международная

жестовая речь. Согласно 14 статье Федерального закона «О социальной защите инвалидов в Российской Федерации», язык жестов признается как средство межличностного общения. Существуют стандарты для отображения букв (дактилизирование) и фонем (хирем). На рисунке 1 представлена таблица соответствия жестов буквам русского алфавита.

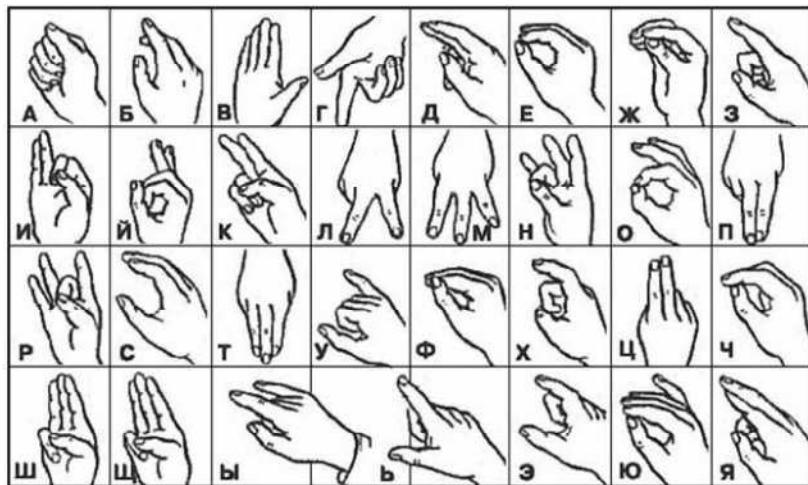


Рис.1 Таблица соответствия жестов и букв русского алфавита



Рис.2 Пример эталонов жестов для автоматизированного распознавания

Проблема разработки и развития языка жестов является сложной и актуальной. Этой проблемой занимаются тысячи людей. В данной работе предлагается решение лишь небольшой задачи автоматизированного распознавания жестов с помощью компьютера. С целью надежного распознавания жестов предлагается набор эталонных жестов, которые содержатся в базе жестов и используются для распознавания демонстрируемого жеста. Это явилось основным критерием отбора. В полученных эталонах старались исключить расположение объектов под углами некратными 90 градусам. На рисунке 2 приведены несколько эталонов для одной руки.

При разработке эталонов жестов сознательно стремились не учитывать демонстрацию четвертого (безымянного) пальца, так как пользоваться им затруднительно. Значения для жестов еще не выбраны, им соответствует только номер. При необходимости номеру может быть поставлена буква или команда. Решается задача определения наиболее удобных для демонстрации жестов и присваивание им наиболее часто встречающихся букв. Частоту использования букв можно взять в соответствии со стандартом для клавиатур (QWERTY), а наиболее удобные жесты определить опытным путем.

Для анализа жеста необходимо только два кадра, а не весь видеопоток. Это позволяет значительно сократить количество операций по преобразованию данных и сократить объем памяти для хранения изображения. Фиксируется изображение, содержащее объект, и изображение, на котором объект уже отсутствует. Разработан алгоритм, который определяет объект любой формы по двум изображениям. Каждому изображению соответствует числовая матрица. После преобразования получаем контур объекта, который закрашиваем и получаем силуэт. Затем силуэт используем для распознавания жеста, сравнивая его с эталоном. Так же реализован алгоритм, с помощью которого можно определить направление движения объекта.

Список литературы

[Девятков В.В. и др., 2007] Девятков В.В., Алфимцев А.Н. Параллельный захват и отслеживание динамических жестов руки // Системный анализ и информационные технологии: Сб. трудов Межд. конф.- М., 2007.

[Фомин Я. А., 2012] Фомин Я. А. Распознавание образов: теория и применения. — 2-е изд. — М.: ФАЗИС, 2012.

[Малашин Р.О. и др., 2013] Малашин Р.О., Луцив В.Р. “Восстановление силуэта руки в задаче распознавания жестов с помощью адаптивной морфологической фильтрации бинарного изображения”. Санкт-Петербург, 2013.