

УДК 007.61, 681.3.01

М.Ю. Катаев, Л.В. Широков

Методика определения жестов руки, наблюдаемых с помощью видеокамеры

Предложена методика определения жестов руки человека из изображений, получаемых с помощью видеокамеры. Приведены обзоры устройств и алгоритмов, где реализовано выделение жестов на изображении. Поясняется особенность предлагаемого подхода на фоне известных и приводятся первичные результаты обработки изображений и выделение жестов.

Ключевые слова: изображение, методы обработки изображений, жесты руки.

В настоящее время разработка и исследование взаимодействия человек–компьютер (человеко-машинные интерфейсы) приобретают особое внимание благодаря развитию математического и программного обеспечения. Одной из центральных задач этого взаимодействия является разработка и применение естественного для человека способа общения с компьютером, чтобы ускорить общение, сделать его более надежным и расширить круг решаемых компьютером задач и круг пользователей. К естественным способам общения можно отнести жесты (рук, головы, тела), голоса, мимики и др.

Одним из возможных вариантов управления комплексом оборудования являются статические и динамические жесты рук. Существуют алгоритмы, позволяющие выделять на изображении руку человека и распознавать жесты. К прямым аналогам можно отнести продукт компании Leap Motion, систему управления с помощью жестов – Leap 3D. Другой вариант – контроллер движений Kinect, разработанный в компании Microsoft. Приложение NPointer позволяет управлять компьютером без использования мыши. Программная система Flutter позволяет управлять музыкой и видео на Mac при помощи жестов. Фирма Pioneer разработала программно-аппаратный комплекс «Air Gesture» для управления аудиовидеооборудованием фирмы с помощью жестов. Компания «Атанор» разработала систему распознавания жестов для проведения презентаций. Система распознает жесты человека с помощью видеокамеры и специальных сенсоров глубины и позволяет ведущему управлять показом слайдов руками. Израильская компания eyeSight Mobile Technologies совместно с корейским производителем Innodigital разработала первую в мире телевизионную приставку под управлением операционной системы Android, которая оснащена интегрированной технологией распознавания жестов. Инновационный продукт получил имя nTobeBox. Японская компания NEC разработала систему управления жестами, которая использует детектор движения и проектор для отображения виртуального «пульта управления».

Помимо указанных выше разработок в сфере управления устройствами через жесты рук, которые считываются при помощи видеоустройств, показывают, что существует множество разработок, которые являются экспериментальными. Проведенный обзор [1–12] по данному направлению показывает, что данный вид взаимодействия человек–компьютер является перспективным как в научном плане, так и в коммерческом. Отметим, что указанные выше устройства имеют весьма малое распространение за счет привязанности к соответствующим техническим устройствам.

Развитие подходов детектирования и распознавания жестов особенно важно для таких устройств, как, например, STB (Set-Top-Box) и другие устройства, когда на первый план встает ограниченность в памяти и производительности вычислительного процессора. В данной статье приводится описание простой и легко реализуемой методики выделения статических жестов руки на изображении.

Типы жестов руки. На рис. 1 приведены некоторые из статических жестов, которые возможно диагностировать из анализа видео. Видно, что для диагностики жестов необходимо фиксировать положение руки и пальцев, при этом возникает множество однотипных ситуаций, которые могут быть представлены как жест, однако не будучи таковым.

При анализе жестов необходимо понять, какие может принимать положения и формы кисть, относительно стоящего/сидящего человека. Кисть практически не гнется в запястье в сторону большого пальца, в сторону мизинца гнется примерно на 70–90°. В сторону тыльной и внутренней стороны ладони кисть гнется (может гнуться, и эта величина зависит чаще всего от возраста) на 90°.

Локтевой сустав позволяет повернуть ладонь примерно на 180° в сторону большого пальца. Вращение кисти относительно линии, проходящей через центр ладони и перпендикулярной ей, производится на 270° . Если повернуть тыльной стороной к себе, то поворот осуществляется от 0 до 270° против часовой стрелки. Если повернуть внутренней стороной к себе, то поворот осуществляется от 90° до -180° по часовой стрелке. Кисть имеет пять пальцев (это определяет набор жестов и соответствующих алгоритмов распознавания).



Рис. 1. Положения руки и пальцев во время отображения жестов

Рассмотрим возможные комбинации с пальцами (см. рис. 1). Если ладонь сжать в кулак и по отдельности распрямлять каждый палец, то получится шесть вариантов: кулак, открыт большой, указательный, средний, безымянный с мизинцем, мизинец. Далее можно собирать различные комбинации, при некоторых комбинациях распрямляется третья фаланга сжатых пальцев (если это позволяют кистевые суставы). Конечно, растяжение связок и вращение в суставах индивидуально, так что здесь приведены усредненные данные. Для пальцев существует 2 позиции: сжат, разогнут. Так же углы поворота выбираются кратными 90° (это делается для облегчения обработки данных).

Итог: жест определяется положением ладони относительно видео устройства (вращением в различных плоскостях) и сгибанием/сжатием пальцев.

При детектировании жестов возникают вопросы, которые должен решать разработчик.

1. *Продолжительность времени, когда определенное положение руки и пальцев считается жестом.* Имеется в виду время фиксации руки в статичном положении, для определения жеста. Человек во время работы с устройством может выполнять произвольные или непроизвольные жесты-кулачки. Они могут возникать при общении с другим человеком выражением эмоций либо при совершении других действий, не связанных со взаимодействием с устройством, что выражается в непреднамеренных изменениях положения руки и пальцев (например, дрожание).

2. *Выбор видеоустройств.* На продолжительность времени сказывается и несовершенство воспринимающих датчиков, в нашем случае видеокамеры. Такие камеры имеют порог скорости восприятия информации (типично 25–30 кадров/с), что не позволяет уследить за всеми движениями руки (отдельные части руки могут двигаться с большей скоростью, чем другие (ситуация, когда локоть практически статичен, а ладонь движется быстро)). В такой ситуации при съемке получаем размазанный жест, который достаточно сложно интерпретировать. Конечно, можно было бы использовать дорогие камеры скоростной съемки, но система задумывалась изначально для работы с типичными камерами и их аналогами.

3. *Вычислительные алгоритмы.* Сложность алгоритмов и низкое быстродействие элементов вычислительной системы не всегда позволяют выполнить действие за доли секунды и требуют определенного времени, которое зависит от разрешения изображения, от качества съемки и других параметров.

Рассмотрим период времени, в котором человек должен зафиксировать определенный жест. Это время, по нашему мнению, лежит в интервале от 0,5 до 2 с или в среднем 1–1,5 с на жест (около 30–45 кадров при скорости съемки 30 кадров/с). В течение 1–1,5 с устройство должно провести не только съемку движения руки, но и принять решение: человек выполнил просто движение рукой или хочет выполнить команду с помощью определенного жеста.

Почему время детектирования жеста не больше 2 с? Ответ связан с физиологией человека. В теле человека существует движение крови с частотой примерно 1,2–1,5 Гц, что сказывается на статичности жестов, т.е. как бы человек не хотел удержать руку в одном положении, пульс будет создавать колебания, которые будут смещать руку. Также чем больше человек будет удерживать руку в одном положении, т.е. постоянно напрягать одну группу мышц, то в этих мышцах будет создаваться тремор, интенсивность и амплитуда которого пропорциональны времени. Свой вклад вносит и дыхание человека – при вдохе грудная клетка и брюшная полость расширяются, в момент выдоха наоборот сжимаются, подобные колебания передаются во все части тела.

Алгоритмы выделения руки человека. Метод гистограмм. Метод заключается в построении двух гистограмм изображения и их анализа. Отсекается фоновое изображение. Изображение переводится в полутоновое изображение (Gray Scale). Строится две гистограммы (по оси X и Y) оставшегося изображения руки. Принцип построения гистограммы: изображение разбивается на N частей (по горизонтали или вертикали, в зависимости от необходимой гистограммы), в каждой части получаем среднее значение оттенков серого по всем пикселям области. Нормализуем значения всех областей по максимальному значению из всех областей, далее проводим анализ гистограммы и, если надо, выводим ее на экран для наглядности.

Метод зависит от освещения, методов фильтрации, следовательно, система обработки должна иметь хорошую приспособляемость к изменению внешней среды. В некоторых случаях достаточно сложно отследить логику изменения гистограммы при изменении исследуемых частей изображения (положения руки в пространстве, относительно камеры), следовательно, анализ гистограммы предоставляется слишком трудоемкой задачей.

Метод вычисления площади. В этом подходе строится функция площади выделенного контура руки с различным количеством пальцев. Контур должен быть замкнутым. Данная функция позволяет достаточно эффективно отделять жесты один от другого, но существенно зависит от мешающих факторов (освещенность, контраст, элементы не удаленного фона). Единственное ограничение метода связано с тем, что с его помощью можно определить лишь шесть жестов.

Контурный метод. Поиск контура сводится к поиску изменения яркости полутонового изображения, также поиск контура возможен по поиску цветовых перепадов на изображении. При контурном анализе изображения можно выделить два аспекта:

1. Удаление лишних объектов (фона) с изображения, чтобы они не вносили искажение при поиске контуров (фильтрация, удаление шумов и пр.).
2. Настройка фильтра поиска контура так, чтобы незначительные детали (текстура и внутренние особенности объекта, если они нам не нужны) не выделялись фильтром поиска контуров.

При контурном анализе могут возникать проблемы: разрывность контуров, необходимо дополнительно соединять разделенные части контура, чтобы упростить сам анализ; создается проблема выделения ненужных нам контуров (остатки фона или ненужных деталей искомого объекта). Также затруднительно проводить сложный анализ изображения (определение жестов) по контуру объекта (руки). Контурный анализ чаще всего проводится для определения относительных размеров объекта, формы, сложности или проводится прямое сопоставление с шаблоном.

Анализ контура относительно центра тяжести объекта. Данный метод включает в себя предыдущие два метода (метод гистограмм и контурный метод). Анализируя гистограммы изображения, находим центр тяжести объекта. Получаем контур объекта. Строим функцию расстояния точки контура относительно центра тяжести объекта. Этот метод представляется менее трудоемким, чем представленные выше методы. При визуализации функции расстояния довольно просто увидеть логику изменения минимумов и максимумов при изменении исследуемых частей изображения (положения руки и пальцев в пространстве, относительно камеры).

Результаты работы алгоритма. Нами применялись различные численные алгоритмы для устойчивого выделения руки на изображении: фиксация центра тяжести бинарного изображения руки, выделение контура руки, достижение условия его замкнутости (при помощи морфологических операций). Также нами был организован поиск алгоритмов устойчивого детектирования положения пальцев на руке, для различных пространственных положений руки. Среди всех возможных алгоритмов нами был отобран простой, но эффективный алгоритм, который устойчив в работе, требует минимального объема памяти и быстр в работе. Идея связана с вычислением функции расстояния от контура фигуры до центра тяжести. Тогда распознавание отдельных жестов, как и отличие одного жеста от другого, производится путем анализа этой функции расстояния. Анализируя экстремумов функции расстояния, возможны распознавание типа жеста и сравнение жестов.

$$d(i) = \sqrt{(x_c - x_i)^2 + (y_c - y_i)^2},$$

где i – текущий номер точки контура; $\{x_c, y_c\}$ – координаты центра тяжести изображения руки и $\{x, y\}$ – координаты контура, начиная с нижней части контура, перпендикулярной центру тяжести.

На рис. 2 показаны фотографии руки с изображением определенного числа пальцев (рис. 2, а), контур выделенной на фоне руки (рис. 2, б) и график функции расстояния d (рис. 2, в). На рис. 2, в хорошо видно, что количество пиков, которые выделяются при помощи нашего подхода однозначно

определяют число пальцев на руке и тем самым отделяют один жест от другого. Отметим, что первый пик на изображении рис. 2, *в* соответствует левому нижнему краю руки и всегда детектируется мешающий и не учитывается при дальнейших расчетах. Поиск числа пальцев на изображении сводится к простейшей задаче определения числа максимумов кривой. Таким образом, анализ кривой расстояния d позволяет не только провести выделение типа жеста, но и легко построить алгоритм распознавания жеста и определения той или иной команды для устройства.

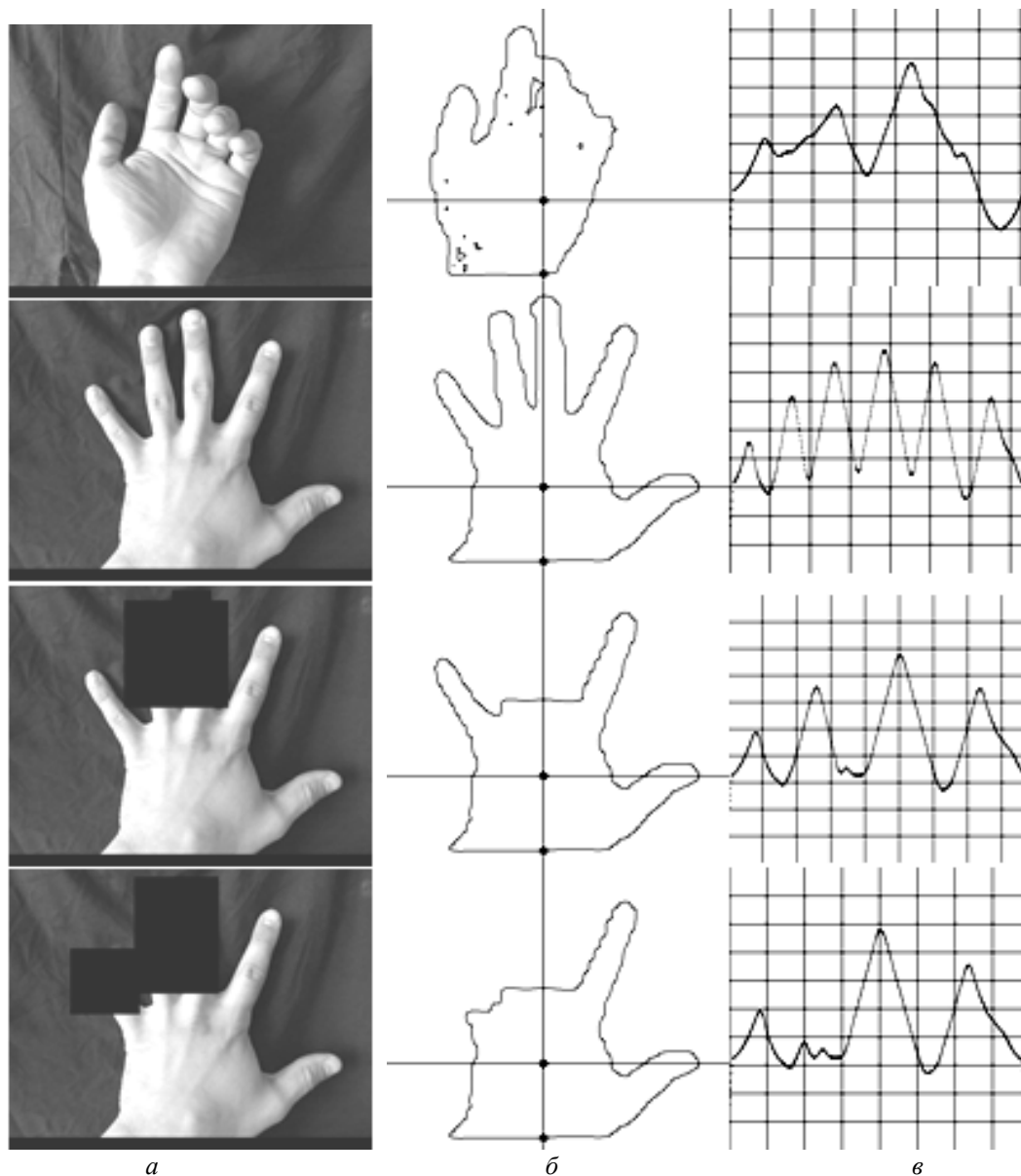


Рис. 2. Анализ жестов. *а* – изображение руки; *б* – бинарный контур руки с определенным центром тяжести и *в* – рассчитанная функция расстояния d

В настоящее время нами получены результаты, позволяющие выделять жесты руками в статическом режиме. Динамический, т.е. произвольный, режим движения руки и распознавания жестов является пока в состоянии разработки.

Заключение. Разработан и рассмотрен на примере распознавания статических жестов метод определения числа раскрытых пальцев на руке. Подобраны алгоритмы обработки изображений, позволяющие устойчиво проводить детектирование положения руки и кисти на изображении. Обработка реальных снимков показала высокую надежность обнаружения числа пальцев на руке и их последовательность открытия. В дальнейшем этот метод планируется использовать для обнаружения и распознавания динамических жестов.

Литература

1. Системы компьютерного зрения [Электронный ресурс]. – Режим доступа: <http://beta.iitvision.ru/?p=173>, свободный (дата обращения: 04.04.2013).
2. Гонсалес Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М.: Техносфера, 2006. – 1072 с.
3. Zhou H. Tracking articulated hand motion with Eigen dynamics analysis / Y. Zhou, T.S. Huang // Proc. of the Intern. conf. on computer vision, Nice (France), 14–16 Oct. 2003. – Washington DC: IEEE Computer Soc. – 2003. – Vol. 2. – P. 1102–1109.
4. Wu Y. Non-stationary color tracking for vision-based human computer interaction / Y. Wu, T.S. Huang // IEEE Trans. Neural Networks. – 2002. – Vol. 13, №4. – P. 948–960.
5. Manresa C. Hand tracking and gesture recognition for human computer interaction / C. Manresa, J. Varona, R. Mas, F.J. Perales // Electron. Lett. Comput. Vision Image Anal. – 2005. – Vol. 5, №3. – P. 96–104.
6. Mckenna S. A comparison of skin history and trajectory-based representation schemes for the recognition of user-specific gestures / S. Mckenna, K. Morrison // Pattern Recognition. – 2004. – Vol. 37. – P. 999–1009.
7. Chen F. Hand gesture recognition using a real-time tracking method and hidden Markov models / F. Chen, C. Fu, C. Huang // Image Vision Comput. – 2003. – Vol. 21, №8. – P. 745–758.
8. Ng C.W. Gesture recognition via pose classification / C.W. Ng, S. Ranganath // Proc. of the 15th Intern. conf. on pattern recognition, Barcelona (Spain). – 2000. – Vol. 3. – P. 699–704.
9. Oka K. Real-time fingertip tracking and gesture recognition / K. Oka, Y. Sato, H. Koike // Proc. IEEE. Comput. Graphics Appl. – 2002. – Vol. 22, №6. – P. 64–71.
10. Huang C. A model-based hand gesture recognition system / C. Huang, S. Jeng // Machine Vision Appl. – 2001. – Vol. 12, №5. – P. 243–258.
11. Cutler R. View-based interpretation of real-time optical flow for gesture recognition / R. Cutler, M. Turk // Proc. of the 3rd IEEE conf. on face and gesture recognition, Nara (Japan), 14–16 Apr., 1998. – P. 416–421.
12. Lu S. Using multiple cues for hand tracking and model refinement / S. Lu, D. Metaxas, D. Samaras, J. Oliensis // Proc. of the IEEE conf. on computer vision and pattern recognition, Madison (USA). – 16–22 June, 2003. – P. 443–450.

Катаев Михаил Юрьевич

Д-р техн. наук, профессор каф. автоматизированных систем управления (АСУ) ТУСУРа,
профессор Юргинского технологического института (филиала НИ ТПУ)
Тел.: 8-960-975-27-85, (382-2) 70-15-36
Эл. почта: kataev.m@sibmail.com

Широков Леонид Владимирович

Аспирант каф. АСУ ТУСУРа
Тел.: 8-962-778-33-67, (382-2) 70-15-36
Эл. почта: ngtrx89@gmail.com

Kataev M.Yu., Shirokov L.V.

The methodology of determining the hand gestures the observed with the camcorder

A procedure is proposed for determination of the hand gestures from the data processing of the images received with the help of camcorder. Provides reviews of devices and algorithms, which implements the allocation of gestures in the image. Explains the features of the proposed approach, on the background of well-known and are the primary results of processing of images and the selection of gestures.

Keywords: image, image processing methods, the gestures of the hand.