

УДК 004.93

## РАСПОЗНАВАНИЕ ЖЕСТОВ РУК В РЕАЛЬНОМ ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ СЕГМЕНТАЦИИ

Кучер М.Ю., Белов Ю.С.

ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана»,  
филиал, Калуга, e-mail: fn1-kf@mail.ru

Для взаимодействия человека и компьютера очень важно внедрять использование распознавания жестов рук. В данной статье будет рассмотрен один из новейших и быстрых методов распознавания жестов рук в реальном времени. Известно, что распознавание жестов на основе компьютерного зрения – важная часть взаимодействия человека с техникой. И хотя клавиатура и мышь являются на сегодняшний день самыми популярными устройствами для взаимодействия с компьютером, программное обеспечение быстро развивается и требуются новые типы передачи информации от пользователя к технике, особенно бесконтактные, в числе которых голос и жесты. В данной статье рассматривается один из самых новых и быстрых механизмов взаимодействия человека и компьютерных технологий при помощи движений рук и фигур, изображенных пальцами (жестов). Рассмотрим сначала коротко, что необходимо для этого сделать. Чтобы извлечь область руки из фона, потребуется использовать метод вычитания фона. Затем, чтобы распознать жест, нужно сегментировать пальцы (отличить их от всей руки на изображении). Последним шагом должен быть применен классификатор правил для прогнозирования и распознавания жеста. Метод сегментации пальцев был испытан на 1300 изображениях, результаты показали его высокую эффективность, которая превышает показатели уже существующих методов по времени и точности распознавания.

**Ключевые слова:** жест, распознавание, управление, сегментация, классификация изображений

## REAL-TIME HAND GESTURE RECOGNITION USING FINGER SEGMENTATION

Kucher M.Yu., Belov Yu.S.

Bauman Moscow State Technical University, Kaluga branch, Kaluga, e-mail: fn1-kf@mail.ru

It is very important for human-computer interaction to implement the use of hand gesture recognition. This article will discuss one of the newest and fastest methods for real-time hand gesture recognition. It is known that computer vision-based gesture recognition is an important part of human interaction with technology. And although the keyboard and mouse are by far the most popular devices for interacting with a computer, software is rapidly evolving and new types of information transfer from user to technology are required, especially contactless ones, including voice and gestures. This article examines one of the newest and fastest mechanisms of interaction between humans and computer technologies using the movements of hands and figures depicted by fingers (gestures). Let's first consider briefly what needs to be done to do this. To extract the area of the hand from the background, you will need to use the background subtraction method. Then, to recognize the gesture, you need to segment your fingers (distinguish them from the entire hand in the image). The final step is to apply a rule classifier to predict and recognize gesture. The finger segmentation method was tested on 1300 images, which showed its high efficiency, which exceeds the indicators of existing methods in terms of recognition time and accuracy.

**Keywords:** gesture, recognition, control, segmentation, image classification

Жест – это символ, сделанный при помощи направленного движения рукой или/и создания фигуры пальцами, который несет в себе некий индивидуальный смысл. Каждому жесту соответствует значение, которое превращает его в способ передачи информации. Примеры статических жестов изображены на рис. 1. Жесты можно использовать и для передачи команд от человека к технике, что отличается от наиболее привычных методов управления на сегодняшний день. В последнее десятилетие разработчики систем искусственного интеллекта стремились создать систему для распознавания жестов, которая работала бы быстро и эффективно. Для управления устройствами такая система обязательно должна работать в режиме реального времени, так же быстро

или даже быстрее, чем более привычные технологии взаимодействия с техникой (компьютерная мышь, сенсорное управление, управление при помощи клавиатуры, управление голосом) [1]. Развитие данной технологии управления было важно не только из-за цели изобрести новую быструю систему управления устройствами, но и для комфортного использования технологий людям с инвалидностью, для которых управление мышью или клавиатурой, или голосом является невозможным [2].

Коротко процесс распознавания руки на изображении можно описать так: сначала определяется область руки, и затем отслеживаются изменения в ее движении (или движении пальцами) [3].

Первым этапом берется исходное изображение с устройства ввода.



Рис. 1. Примеры жестов

Затем применяются функции распознавания жестов рук и выделяются признаки.

Признаки сравниваются с имеющимися в базе знаний и делается вывод.

Для отделения фона используются разные типы камер – обычная, стерео или ToF. Если обычная камера является сейчас очень популярной, то про другие две необходимо дать пояснения [4]. Они нужны для получения информации о глубине, то есть для получения карты глубины изображения и таким образом более быстрого отделения области руки от области фона.

В мире уже не первый год существуют подобные системы управления, но все они предполагают использование специальной перчатки либо других специальных устройств для обнаружения главных точек руки. Данный метод предполагает управление без устройств посредников. Также у данного метода в отличие от других подобных существующих методов распознавание жеста происходит не целиком, а путем выделения пути движения отдельных ее фрагментов, что упрощает и ускоряет один из наиболее важных этапов – процесс классификации [5].

Цель исследования – разработать систему обнаружения и распознавания жестов в реальном времени с использованием сегментации пальцев.

*Описание модели.* Для решения задачи распознавания жеста становится целесообразным комбинировать несколько алгоритмов. Работа системы обнаружения и распознавания жестов происходит в три этапа: обнаружение руки в кадре, сегментация пальцев, распознавание жеста.

*Этапы распознавания жестов.* Рассмотрим распознавание жеста пошагово. Первым делом рука обнаруживается с использованием метода вычитания фона, и результат преобразуется в двоичное изображение. Затем пальцы и ладонь сегментируются, чтобы облегчить распознавание пальцев. Наконец, жесты рук распознаются с помощью простого классификатора правил.

Прежде чем подробно разобрать все шаги, рассмотрим, что такое сегментация. В компьютерном зрении сегментация – это процесс разделения цифрового изображения на несколько сегментов (множество пикселей, также называемых суперпикселями). Цель сегментации заключается в упрощении и/или изменении представления изображения, чтобы его было проще и легче анализировать. Сегментация изображений обычно используется для того, чтобы выделить объекты и границы (линии, кривые, и т.д.) на изображениях. Более точно, сегментация изображений – это процесс присвоения таких меток каждому пикселю изображения, что пиксели с одинаковыми метками имеют общие визуальные характеристики [6].

Результатом сегментации изображения является множество сегментов, которые вместе покрывают всё изображение, или множество контуров, выделенных из изображения. Все пиксели в сегменте похожи по некоторой характеристике или вычисленному свойству, например по цвету, яркости или текстуре. Соседние сегменты значительно отличаются по этой характеристике.

В контексте данной статьи сегментация будет использоваться для разделения руки

на сегменты, каждый из которых содержит в себе по одному пальцу, что позволяет нам отслеживать изменения в каждом из фрагментов и распознавать не только статические, но и динамические жесты.

Наиболее легко и эффективно определять область руки на исходном изображении с помощью метода вычитания фона, но в некоторых случаях в результате вычитания фона включаются другие движущиеся объекты. Цвет кожи пользователя можно использовать, чтобы отличить область руки от других движущихся объектов, заранее добавив его в базу знаний. Размер изображения обнаруженной руки изменяется таким образом, чтобы распознавание жеста было инвариантным к масштабу изображения.

Чтобы сегментировать пальцы, первым делом необходимо найти палм-пойнт ладони. Он определяется как центральная точка ладони, которая находится методом преобразования расстояния. То есть в изображении каждому пикселю приписывается расстояние от него до ближайшего граничного пикселя. Расстояние до блоков используется для измерения расстояний между пикселями и ближайшими граничными пикселями [7]. Таким образом, в изображении с преобразованием расстояния двоичного изображения руки пиксель с наибольшим расстоянием выбирается в качестве ладони. Найденная точка ладони отмечена точкой зеленого цвета на рис. 2.

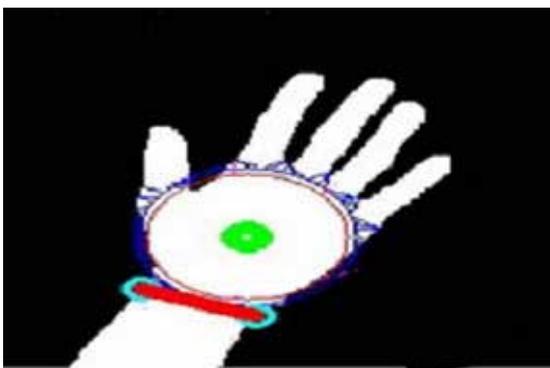


Рис. 2. Двоичный результат распознавания

Когда центр ладони найден, то найденная следом крайняя точка может описать окружность с палм-пойнтом ладони в качестве центральной точки. Получившийся круг называется внутренним кругом, потому что он находится внутри ладони. Радиус круга постепенно увеличивается, пока не достигнет края ладони. То есть радиус круга перестает увеличиваться, когда в круг входят черные пиксели. В итоге, когда ра-

диус максимального внутреннего круга получен, создается круг большего размера, радиус которого в 1,2 раза больше радиуса максимального внутреннего круга. На рис. 2 граница найденного круга обозначена синей окружностью.

Для каждой точки выборки на окружности находят ближайшую граничную точку и проводят к ней линию. Граничная точка оценивается просто. Если 8 соседей пикселя состоят из белых и черных пикселей, он помечается как граничная точка. Все найденные ближайшие граничные точки связаны, чтобы получить маску ладони, которую можно использовать для сегментирования пальцев и ладони. Используется больший круг вместо максимального внутреннего круга, чтобы получить более точную маску ладони для следующей сегментации.

Две точки запястья – это две конечные точки линии запястья в нижней части руки. Точки запястья – важные точки для распознавания жестов рук. Их можно искать следующим образом: если расстояние между двумя последовательными точками маски велико, эти две точки маски оцениваются как точки запястья.

С помощью маски ладони пальцы можно легко сегментировать. Та часть руки, которая закрыта маской ладони, – это и есть ладонь, а другие части руки – пальцы. Результат сегментации пальцев показан на рис. 3.



Рис. 3. Результат сегментации пальцев

*Определение вида обнаруженного пальца.* В сегментированном изображении пальцев алгоритм маркировки применяется для обозначения областей пальцев. В результате применения метода маркировки обнаруженные области, в которых количество пикселей слишком мало, считаются зашумленными и отбрасываются. Только области достаточного размера считаются пальцами и остаются.

Для каждой оставшейся области, то есть пальца, находят минимальную ограничивающую палец рамку. Затем центр минимального ограничивающего прямоугольника используется для представления центральной точки пальца. После чего центры пальцев соединяются с палм-поинтом ладони и вычисляется градус угла между этими линиями и линией запястья. Если степень меньше, это означает, что большой палец присутствует на изображении руки. Соответствующий центр – это центральная точка большого пальца. Если все градусы угла больше, большой палец не существует на изображении.

Чтобы обнаружить и распознать другие пальцы, сначала исследуется линия ладони. Чтобы найти линию ладони, необходимо начинать от ряда линии запястья. В каждом ряду руку пересекает линия, параллельная линии запястья. Если есть только один связанный набор белых пикселей на пересечении линии и руки, линия смещается вверх. Если на пересечении линии и руки имеется более одного связанного набора белых пикселей, линия считается кандидатом на линию ладони [8].

В случае если большой палец не обнаружен, линия, пересекающая руку с более чем одним соединенным набором белых пикселей на их пересечении, выбирается в качестве линии ладони. В случае наличия большого пальца линия продолжает двигаться вверх с краевыми точками ладони. Теперь, когда большой палец убран, остается только один связанный набор пикселей на пересечении линии и руки.

После того как линия ладони получена, ее делят на 4 части. По горизонтальной координате центральной точки пальца он распадается на определенные части. Если палец попадает в первую часть, это указательный палец. Если палец принадлежит второй части, это средний палец. Третья часть соответствует безымянному пальцу. Четвертая часть – мизинец. Результат распознавания пальцев показан на рис. 4. На этом рисунке желтая линия – это линия ладони, а красная линия параллельна линии запястья.

В некоторых случаях два или более пальца остаются близко, и между пальцами нет промежутка. Чтобы отличить данный случай от случая одного пальца, ширина минимального ограничивающего прямоугольника используется в качестве показателя различия.

Если ширина минимального ограничивающего прямоугольника равна обычному значению, обнаруженная область представляет собой один палец. Если ширина минимального ограничивающего прямоуголь-

ника в несколько раз превышает обычное значение, обнаруженная область соответствует нескольким пальцам, которые плотно прилегают друг к другу. Для надежности распознавания пальцев также учитываются расстояния и углы между пальцами, чтобы различать различные жесты.

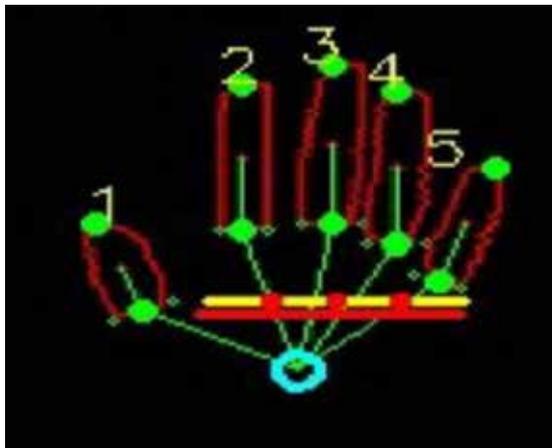


Рис. 4. Линия ладони

Когда пальцы обнаружены и распознаны, жест руки можно распознать с помощью простого классификатора правил. В классификаторе правил жест руки прогнозируется в соответствии с количеством и содержимым обнаруженных пальцев. Содержание пальцев означает, какие пальцы обнаруживаются.

Эффективность предлагаемого метода сильно зависит от результата обнаружения руки. Так как движущиеся объекты с цветом, подобным цвету кожи, могут попасть в результат обнаружения руки и затем ухудшат производительность распознавания жестов. Однако алгоритмы машинного обучения могут отличить руку от фона. А камеры ToF предоставляют информацию о глубине, которая может повысить эффективность обнаружения рук. Таким образом, в будущих работах методы машинного обучения и камеры ToF могут быть использованы для решения сложной фоновой проблемы и повышения надежности обнаружения рук.

#### Список литературы

1. Weiss K., Khoshgoftaar T.M., Wang D.D. A survey of transfer learning. *Journal of Big Data*. 2016. P. 1–40. URL: <https://journalofbigdata.springeropen.com/articles/s40537-016-0043-6> (дата обращения: 11.02.2022).
2. Басанько А.С., Белов Ю.С. Использование нейросетевого подхода для сегментации слов в рамках задачи офлайн-распознавания рукописного текста // Программные продукты, системы и алгоритмы. 2018. № 4. С. 71–73.
3. Kensert A., Harrison P.J., Spjuth O. Transfer learning with deep convolutional neural network for classifying cellu-

lar morphological changes. PubMed.gov. 2018. P. 1–11. URL: <https://www.biorxiv.org/content/10.1101/345728v1.full.pdf> (дата обращения: 11.02.2022).

4. Демин И.С., Белов Ю.С., Чухраев И.В. Обучение сверточной нейронной сети на базе архитектуры U-NET с использованием минимальных ресурсов // Электромагнитные волны и электронные системы. 2019. Т. 24. № 7. С. 24–29

5. Liu Yu Han. Feature Extraction and Image Recognition with Convolutional Neural Networks // Journal of Physics Conference Series. 2018. P. 1–7. URL: <https://iopscience.iop.org/article/10.1088/1742-6596/1087/6/062032/pdf> (дата обращения: 11.02.2022).

6. Sahu M., Saxena A., Manoria M. Application of Feature Extraction Technique: A Review // International Journal of Computer Science and Information Technologies. 2015. vol. 6 (3). P. 3014–3016. URL: <http://ijcsit.com/docs/Volume%206/vol6issue03/ijcsit20150603227.pdf> (дата обращения: 11.02.2022).

7. Kaoutar B.A., Jelodar A. Fine-Tuning VGG Neural Network For Fine-grained State Recognition of Food Images. 2018. URL: <https://arxiv.org/ftp/arxiv/papers/1809/1809.09529.pdf> (дата обращения: 11.02.2022).

8. Dakwale P., Monz Ch. Fine-Tuning for Neural Machine Translation with Limited Degradation across In- and Out-of-Domain Data. Proceedings of MT Summit XVI, vol. 1. P. 156–169. Sep. 18–22. 2017.