

УДК 004.932

М. С. Тухбатуллин, А. П. Кирпичников, С. А. Ляшева,  
М. П. Шлеймович

## РАСПОЗНАВАНИЕ ДИНАМИЧЕСКИХ ЖЕСТОВ НА ОСНОВЕ ВЫЧИТАНИЯ ФОНА

*Ключевые слова: интеллектуальный человеко-машинный интерфейс, обнаружение жестов, распознавание жестов, вычитание фона, анализ движения.*

*Рассмотрены подходы к обнаружению и распознаванию жестов при реализации технологии интеллектуального человеко-машинного интерфейса. Приведены основные методы обнаружения жестов в видеопотоке на основе анализа движения. Описана система распознавания жестов на основе вычитания фона. Приведены результаты экспериментального исследования системы.*

*Keywords: intelligent human-machine interface, gesture detection, gesture recognition, background subtraction, motion analysis.*

*Discusses the approaches to the detection and recognition of gestures in the realization of the technology of intelligent human-machine interface. Given the basic methods for the detection of gestures in a video stream based on the motion analysis. Describes a gesture recognition system based on background subtraction. Given the results of experimental studies of the system.*

В настоящее время активно развиваются технологии интеллектуального человеко-машинного интерфейса. На их основе строятся, например, вспомогательные средства управления смартфонами и планшетами. Концепция подобных интерфейсов получила название естественный пользовательский интерфейс. Основной целью ее реализации является сокращение времени между намерением пользователя ввести команду в систему и выполнением этого намерения. Тем самым сокращается время на взаимодействия между человеком и системой. Человеко-машинное взаимодействие в этом случае приближается к естественному взаимодействию между людьми. В числе других технологий интеллектуального интерфейса активно развиваются методы и средства распознавания жестов.

Рассмотрим основные подходы к распознаванию динамических жестов руки на видеоизображениях. В этом случае на вход подается последовательность кадров видеопотока, а на выходе формируется описание жеста. При распознавании выделяются области изображения, соответствующие руке, и выполняется анализ вида и поведения точек (пикселей) этих областей.

Одним из самых популярных подходов к обнаружению руки на изображении является цветовая сегментация. Этот подход заключается в выделение элементов изображения, имеющих определенный цвет. Результатом цветовой сегментации кожи является множество областей изображения, близких по цвету к человеческой коже. Эффективность данного подхода в значительной степени зависит от выбора цветового пространства (RGB, HSV, YCrCb, YUV и др.). Цветовые пространства, эффективно отделяющие цветность от компонентов яркости цвета, считаются более предпочтительными, поскольку они позволяют достичь некоторой независимости от освещенности. В [1] произведен обзор различных моделей цветности кожи и оценена их производительность. В результате для решения задачи распознавания жестов наиболее оптимальным оказывается применение цветового пространства HSV.

Одной из главных характеристик объекта на изображении является его форма. Например, кисть руки имеет характерный вид. Для ее обнаружения можно воспользоваться подходами на основе вы-

числения и анализа инвариантных моментов, дескрипторов Фурье и др.

При решении рассматриваемой задачи применяются также обучающие методы [2], которые условно можно разделить на обобщающие и различающие методы. Обобщающие методы формируют абстрактное представление о структуре объекта (обобщенная модель) на основе положительных примеров обучающей выборки. В процессе анализа нового изображения оно сопоставляется с обобщенной моделью для получения степени соответствия. Самыми популярными представителями данного подхода являются методы построения модели случайного поля, модели неявной форм, модели констелляции.

В методе случайного поля исходное изображение разбивается на множество фрагментов и каждый фрагмент относится к определенному классу, который описывает его смысл. Помимо этого, строится матрица вероятностей смежного расположения двух видов фрагментов. Множество классов и матрица вероятности составляют случайное поле. При этом большее значение имеет способ выбора фрагментов. Одним из реализаций данного принципа является метод Капура–Винна [3].

Метод обобщения неявной формы использует преобразование Хафа [4]. Во время обучения на изображении выделяются характерные точки. Далее для каждой характерной точки определяется вектор расстояния от ключевой точки до центра объекта. При исследовании изображения производится поиск ключевых точек и вектор, определенный при обучении указывает на пиксель, который предположительно является центром объекта. Данный пиксель добавляется в список предполагаемых центров объекта. Так же каждый элемент списка предполагаемого центра имеет количество голосов. После того как перебор всех ключевых точек завершен центром объекта считается наиболее вероятный претендент. Одним из реализаций данного принципа является метод Феррари [5].

При использовании модели констелляции объект рассматривается в виде совокупности составных элементов. На новом изображении каждый характерный участок относят к объекту или к фону. Далее при помощи перебора производится поиск наиболее

вероятного соответствии. Одним из реализаций данного принципа является метод Фергюса-Пероны [6].

Некоторые методы достаточно эффективно способны находить кисть по текстуре и форме с предварительным обучением. Здесь исходят из предположения, что различие текстур разных жестов у одного человека выше, чем различие текстур одинаковых жестов у разных людей. Основная проблема этого подхода заключается в автоматическом выборе признаков [7, 8].

Различающие методы характеризуются специальной функцией, которая классифицирует изображение и позволяет выделить объект. При обучении эти методы выделяют отличия между позитивными и негативными изображениями. На основе данных отличий выбирается параметр, благодаря которому можно отличить позитивное изображение от негативного и используют их как параметры классификатора. В результате появляется возможность разделить исходное изображение на два новых изображения, одно из которых содержит интересующие объекты, а другое – не содержит. Самыми популярными из различающих методов являются методы Виолы-Джонса, Лекуна и Папагеоигу. Отметим, что различающие методы являются более ресурсоемкими по сравнению с обобщающими методами.

Описанные подходы можно использовать как для распознавания статических, так и динамических жестов.

При обнаружении и распознавании динамических жестов можно применить методы анализа движения. Одним из наиболее простых и эффективных подходов является вычитание фона [9]. При его реализации осуществляется попиксельное сравнение текущего кадра с моделью фона, под которой понимается изображение сцены без движущихся объектов. В простейшем случае модель фона – это некоторый заданный опорный кадр. Для повышения надежности модель фона должна периодически обновляться. На основе результатов сравнения, которые часто выполняются путем попиксельного вычитания одного изображения из другого, производится обнаружение контуров и вычисление их признаков. В работе [10] приведен подход, основанный на предположении, что жестикация обычно является плавным движением. Исходя из этого, выделяются участки с характеристиками жеста.

В общем случае, процедура обнаружения и распознавания динамического жеста руки на основе анализа движения имеет следующие шаги:

1. Регистрация кадров видеопотока;
2. Сегментация кадров по движению;
3. Обнаружение изображения руки на кадрах видеопотока;
4. Определение соответствия между ключевыми точками изображения руки на последовательных кадрах;
5. Вычисление характеристик движения ключевых точек изображения руки;
6. Анализ характеристик движения ключевых точек изображения руки.

В случае выполнения условия медленного изменения положения руки на соседних кадрах для ее обнаружения можно воспользоваться следующим

подходом. Сначала необходимо найти область кисти руки на первом кадре, а затем на следующих кадрах искать в окрестности центра этой области.

Для отслеживания траектории движения заданных объектов, в том числе кисти руки при распознавании жеста, можно также воспользоваться методами на основе фильтра Калмана или вычисления оптического потока [11, 12].

Один из эффективных подходов основан на отслеживании кисти руки с учетом кинематической модели, под которой понимают информацию о форме, степени свободы подвижных элементов, взаимосвязи между ними и параметрах движения. Кисть можно представить в виде кинематической цепи, где ладонь представляется сильно вытянутым цилиндром с эллиптическим сечением, большой палец как два цилиндра с одним шарниром, а все остальные пальцы как три цилиндра с двумя шарнирами. Можно определять положение частей кисти, принимая во внимание ограничения на повороты элементов этой кинематической цепи [13].

Для исследования описанных подходов была разработана система управления компьютером с помощью жестов. Для формирования моделей жестов в составе программного обеспечения системы предусмотрен обучающий модуль. Еще один модуль предназначен для распознавания жестов пользователя по сформированным и сохраненным моделям. В нем осуществляется также преобразование жестов в последовательность команд.

Система функционирует следующим образом. На первом этапе выполняется обнаружение руки на видеопотоке с помощью метода вычитания фона. При этом в первые несколько секунд после запуска производится создание динамической модели фона. Поскольку предполагается, что программное обеспечение должно эффективно функционировать с недорогими web-камерами, то для улучшения качества изображений используется усреднение по трем кадрам. После вычитания фона для удаления шума применяются морфологические операции. На втором этапе выполняется получение контура кисти. На данном этапе осуществляется удаление шума, возникающего от произвольных движений пользователя, и отсеивание объектов, которые не относятся к кисти руки. На третьем этапе выполняется собственно распознавание жеста. На этом этапе на основе анализа контура определяется семантика жеста.

Созданная система была обучена распознавать семь жестов оператора:

1. Все пальцы выпрямлены и растопырены;
2. Все пальцы, кроме указательного, согнуты в кулак (данный жест удобен для манипуляции мелкими объектами);
3. Все пальцы, кроме указательного и большого, согнуты в кулак;
4. Все пальцы согнуты в кулак, кроме указательного и большого, которые полностью выпрямлены и находятся друг к другу под углом  $90^\circ$ ;
5. Все пальцы согнуты в кулак, кроме указательного и мизинца, которые полностью выпрямлены и находятся друг к другу под углом  $90^\circ$ ;

6. Все пальцы согнуты в кулак, кроме указательного и мизинца. Мизинец и указательный палец полностью выпрямлены по направлению вверх;

7. Все пальцы, кроме большого, согнуты в кулак, а большой палец полностью выпрямлен и находится под углом 90°.

Примеры анализа изображения жеста руки показаны на рис. 1 и 2.

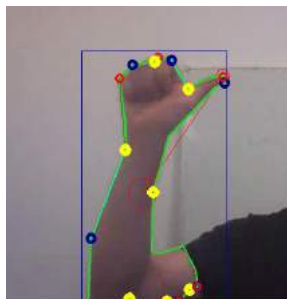


Рис. 1 - Пример изображения жеста руки



Рис. 2 - Пример выделения жеста руки

Таблица 1

Тип жеста	Число жестов	Распознано жестов	Не распознано жестов	Правильное распознавание	Неправильное распознавание
1	12	12	0	100%	0%
2	12	10	2	80%	20%
3	8	7	1	80%	20%
4	9	8	1	82%	18%
5	11	11	0	100%	0%
6	10	8	2	80%	20%
7	9	8	1	82%	18%

Для анализа эффективности распознавания жестов на видеопотоке были проведены экспериментальные исследования разработанной системы. При

проведении экспериментов было обработано 15000 кадров разрешением 480 × 640 пикселей. По результатам данного исследования были получены результаты, приведенные в таблице 1. Время распознавания жеста составляет не более 400 мс.

### Литература

1. Terrillon J., Shirazi M., Fukamachi H., Akamatsu S. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. // International Conference on Automatic Face and Gesture Recognition (FG), pages 54-61, 2000.
2. Wu Y., Huang T. S. View-independent recognition of hand postures. // IEEE Computer Vision and Pattern Recognition (CVPR). Hilton Head Island 2000. P. 84-94.
3. Kapoor, A. Located Hidden Random Fields: Learning Discriminative Parts for Object Detection / A. Kapoor, J. Winn // Proceedings of the European Conference on Computer Vision. - 2006. - Vol. 3. - P. 302-315.
4. Greenspan, M. Efficient Tracking with the Bounded Hough Transform / M. Greenspan, L. Shang, P. Jasiobedzki // Proceedings of Computer Vision and Pattern Recognition. - 2004. - Vol. 1. - P. 520-527.
5. Ferrari, V. From Images to Shape Models for Object Detection / V. Ferrari, F. Jurie, C. Schmid // Intl. J. Computer Vision. - 2010. - Vol. 87(3). - P. 284-303.
6. Fergus, R. Object Class Recognition by Unsupervised Scale-Invariant Learning / R. Fergus, P. Perona, A. Zisserman // Proceedings of Computer Vision and Pattern Recognition. - 2003. - Vol. 2, Sec. II. - P. 264-271.
7. Cui Y., Swets D., Weng J.. Learning-based hand sign recognition using shoslf-m. // Workshop on Automatic Face and Gesture Recognition. IEEE Int 1995. P. 201-206.
8. Triesch J., Von der Malsburg C. A gesture interface for human-robot-interaction. // International Conference on Automatic Face and Gesture Recognition (FG). IEEE Nara, Japan, April 1998. P.546-551.
9. Шлеймович, М.П. Методы вычитания фона в системе управления технологическим процессом / М.П. Шлеймович, М.В. Медведев, С.А. Ляшева, А.П. Кирпичников // Вестник технологического университета. 2015. Т.18, №13. С.166-170.
10. Yuan Q., Sclaroff S., and V. Athitsos. Automatic 2D hand tracking in video sequences. // IEEE Workshop on Applications of Computer Vision. 1995. P. 250-256.
11. Kalman R. E. A new approach to linear filtering and prediction problems. //Transactions of the ASME Journal of Basic Engineering 1960. С. 35-42.
12. Кирпичников, А.П. Обнаружение и сопровождение людей в интеллектуальных детекторах внештатных ситуаций /А.П. Кирпичников, С.А. Ляшева, М.П. Шлеймович. // Вестник Казанского технологического университета. 2014. Т. 17. № 21. С. 351-356.
13. Oikonomidis I., Kyriazis N., Paderleris P., Argyros A. Efficient model-based 3D tracking of hand articulations using Kinect // In Proceedings of the 22nd British Machine Vision Conference, BMVC'2011, University of Dundee, UK, Aug. 29-Sep. 1, 2011.

© М. С. Тухбатуллин – магистрант КНИТУ-КАИ, e-mail: tumarsal@gmail.com; А. П. Кирпичников – д.ф.-м.н., зав. кафедрой интеллектуальных систем и управления информационными ресурсами КНИТУ, e-mail: kirpichnikov@kstu.ru; С. А. Ляшева – к.т.н., доцент кафедры прикладной математики и информатики КНИТУ-КАИ, e-mail: stellyash@mail.ru; М. П. Шлеймович – к.т.н., доцент кафедры автоматизированных систем обработки информации и управления КНИТУ-КАИ, e-mail: shlch@mail.ru.

© М. С. Tukhbatullin – Master Student, KNRTU-KAI, e-mail: tumarsal@gmail.com; А. П. Kirpichnikov - Doctor of sciences in physics and mathematics, professor, Head of the Department of Intelligent Systems and Information Systems Control, Kazan National Research Technological University, kirpichnikov@kstu.ru; S. A. Lyasheva – PhD, Associate Professor of the Department of Applied Mathematics & Informatics, KNRTU-KAI, stellyash@mail.ru; M. P. Shleymovich – PhD, Associate Professor of the Department of Automated Information Processing Systems & Control, KNRTU-KAI, e-mail: shlch@mail.ru.