

УДК 004.413.2

UDC 004.413.2

**АЛГОРИТМ И МЕТОДЫ ОБНАРУЖЕНИЯ И
РАСПОЗНАВАНИЯ ЖЕСТОВ РУКИ НА
ВИДЕО В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ**

**ALGORITHMS AND METHODS FOR
DETECTION AND RECOGNITION OF HAND
GESTURES ON VIDEO IN REAL TIME**

Мурлин Алексей Георгиевич
к.т.н., доцент
*Кубанский государственный аграрный
университет, Кубанский государственный
технологический университет, г. Краснодар,
Россия*

Murlin Aleksey Georgievich
Cand.Tech.Sci, associate professor
*Kuban State Technological University, Krasnodar,
Russia*

Пиотровский Дмитрий Леонидович
д.т.н., профессор
*Кубанский государственный технологический
университет, г. Краснодар, Россия*

Piotrovskiy Dmitriy Leonidovich
professor
*Kuban State Technological University, Krasnodar,
Russia*

Руденко Елизавета Алексеевна
студент
*Кубанский государственный технологический
университет, г. Краснодар, Россия*

Rudenko Elizaveta Alekseevna
student
*Kuban State Technological University, Krasnodar,
Russia*

Янаева Марина Викторовна
к.т.н., доцент
*Кубанский государственный аграрный
университет, Кубанский государственный
технологический университет, г. Краснодар,
Россия*

Yanaeva Marina Victorovna
Cand.Tech.Sci, associate professor
*Kuban State Agrarian University, Kuban State
Technological University, Krasnodar, Russia*

Статья посвящена исследованию методов
обнаружения и распознавания жестов руки в
режиме реального времени и написанию
модифицированного алгоритма на их основе

The article deals with methods for detection and
recognition of hand gestures on video in real time and
writing a modified algorithm on their basis

Ключевые слова: МЕТОДЫ ОБНАРУЖЕНИЯ
ОБЪЕКТА, МЕТОДЫ РАСПОЗНАВАНИЯ,
АЛГОРИТМ

Keywords: METHODS OF OBJECT DETECTION,
RECOGNITION METHODS, ALGORITHM

Современный уровень развитие информационных технологий, алгоритмов и методов способствует появлению новых интерфейсов взаимодействия пользователей и устройств. Одним из перспективных вариантов из таковых являются разработка и исследование человеко-машинных интерфейсов, основанных на распознавании объектов. Перед разработчиками подобных интерфейсов ставится задача использования естественных для человека способов общения с компьютерами с помощью жестов, голоса, мимики и других модальностей.

Одной из основных задач расширения взаимодействия человека с компьютером является разработка общей методологии обнаружения и

распознавания динамических жестов человека, совершаемых руками. Для решения этой задачи необходимо на основе анализа существующих методов обнаружения и распознавания жестов человека, разработать алгоритм анализа потокового видео и распознавания жестов.

В настоящее время существует множество различных методов обнаружения объекта на изображении. Эти методы можно разделить на три основные группы:

- скелетные методы;
- методы на основе 3D модели объекта;
- методы на основе 2D модели объекта.

В скелетных методах исследуется контур силуэта: обычно отыскиваются углы, выступы, впадины и другие точки с высокими значениями кривизны. Для получения информации о форме контура применяются различные представления границы объекта.

В методах на основе 3D модели руки представляют в виде сложных трехмерных поверхностей и классифицируются с помощью нейронных сетей

Метод 2D распознавания схож с предыдущим, но использует двумерное изображение вместо объемных моделей.

Каждый из методов имеет преимущества и недостатки. Поскольку при разработке алгоритма объектом обнаружения на изображении является рука, то первый метод можно исключить из-за неудобства использования контура для определения конкретного жеста. Недостатком метода на основе 3D модели является его ресурсоемкость. Построение 3D модели, обучение нейронной сети и ее использование могут потребовать значительных ресурсов, так же не стоит забывать, что для использования данного метода требует камеры с возможностью определения глубины изображения. Поэтому чаще всего используются 2D методы, так как в них снижается вычислительная сложность и отпадает необходимость в

специальном оборудовании, поскольку для получения изображения может быть использована обычная веб-камера. Поскольку 2D методы являются менее точными, возникает необходимость в их усовершенствовании для получения не ресурсоемкого и эффективного алгоритма распознавания объектов.

Одним из перспективных методов распознавания образов считается метод Виолы-Джонса — алгоритм, позволяющий обнаруживать объекты на изображениях в реальном времени. Этот метод предложен в 2001 году Полом Виолой и Майклом Джонсом. Данный метод является основополагающим для поиска объектов на изображении в реальном времени в большинстве существующих алгоритмов распознавания и идентификации. Так же он является одним из лучших по соотношению эффективности распознавания и скорости работы. Алгоритм показывает отличные результаты и распознает объекты под небольшим углом, примерно до 30 градусов, и при различных условиях освещенности.

Основные принципы, на которых основан метод:

1. Возможность представления изображения в интегральном виде, что позволяет вычислять быстро необходимые объекты;
2. Использование признаков Хаара, т.е. признаков цифрового изображения, используемых в распознавании образов, с помощью которых происходит поиск нужного объекта;
3. Применение алгоритма бустинга (от англ. boost- улучшение, усиление) для выбора наиболее подходящих признаков для искомого объекта на данной части изображения, процедуры последовательного построения композиции алгоритмов машинного обучения, когда каждый следующий алгоритм стремится компенсировать недостатки композиции всех предыдущих алгоритмов;
4. Использование каскадов признаков для быстрого отбрасывания окон, где не найден объект.

Метод Виолы-Джонса с использованием признаков Хаара является одним из лучших алгоритмов для решения задач обнаружения объектов на изображениях в реальном времени, используя двухмерные изображения. Особенностью признаков Хаара, является наибольшая, по сравнению с остальными признаками, скорость, поэтому, для решения задачи построения алгоритма обнаружения руки в реальном времени, возьмём за основу метод Джонса – Виолы.

Первостепенной задачей во многих методах распознавания объектов в реальном времени является интегральное представление изображения. Оно используется, и в вейвлет-преобразованиях – интегральных преобразованиях, которые представляет собой свертку функции с сигналом, и в фильтрах Хаара – позволяющих анализировать различные частотные компоненты данных, в SURF – методе поиска особых точек изображения и создания их дескрипторов, инвариантных к масштабу и вращению, а также многих других алгоритмах. Интегральное представление позволяет быстро рассчитывать суммарную яркость произвольного прямоугольника, причем время расчета не зависит от площади прямоугольника.

Интегральное представление изображения представляет собой матрицу, размерность которой совпадает с размерностью исходного изображения. Элементы матрицы рассчитываются по следующей формуле:

$$II(x, y) = \sum_{i=0, j=0}^{i \leq x, j \leq y} I(i, j).$$

где $I(i, j)$ — яркость пиксела исходного изображения.

Каждый элемент матрицы $II[x, y]$ представляет собой сумму пикселей в прямоугольнике от (0,0) до (x,y). Расчёт матрицы занимает линейное время, пропорциональное числу пикселей в изображении.

Расчет матрицы можно производить по рекуррентной формуле:

$$II(x, y) = I(x, y) - II(x - 1, y - 1) + II(x, y - 1) + II(x - 1, y)$$

По такой интегральной матрице можно очень быстро вычислить сумму пикселей произвольного прямоугольника, произвольной площади. Пусть в прямоугольнике ABCD есть интересующий нас объект D (рис. 1).

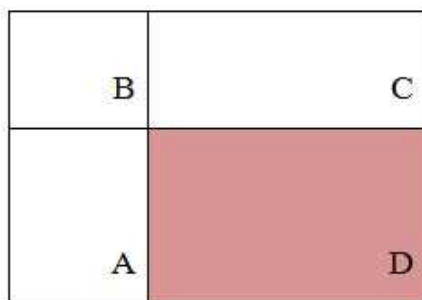


Рисунок 1 – Прямоугольник ABCD

Из рисунка 1 понятно, что сумму внутри прямоугольника можно выразить через суммы и разности смежных прямоугольников по следующей формуле:

$$S(ABCD) = L(A) + L(C) - L(B) - L(D)$$

Для сравнения изображений используем каскад Хаара. Каскад Хаара — это набор признаков, для которых считается их свёртка с изображением. Признак Хаара — отображение $f: X \Rightarrow D_f$, где D_f — множество допустимых значений признака. Если заданы признаки f_1, \dots, f_n , то вектор признаков $x = (f_1(x), \dots, f_n(x))$ называется признаковым описанием объекта $x \in X$. Признаковые описания допустимо отождествлять с самими объектами. При этом множество $X = D_{f_1} * \dots * D_{f_n}$ называют признаковым пространством.

Чаще всего используются прямоугольные признаки, изображенные на рисунке 2 и называемые примитивами Хаара:

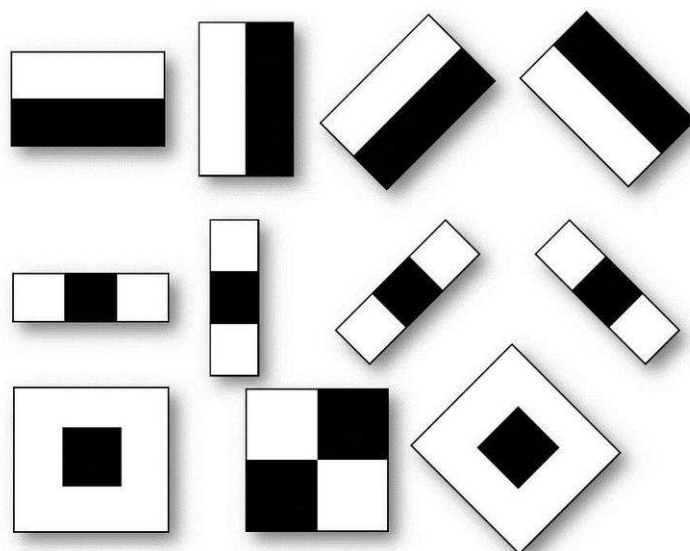


Рисунок 2 – Примитивы Хаара

В более расширенных методах используются дополнительные признаки рис. 3.

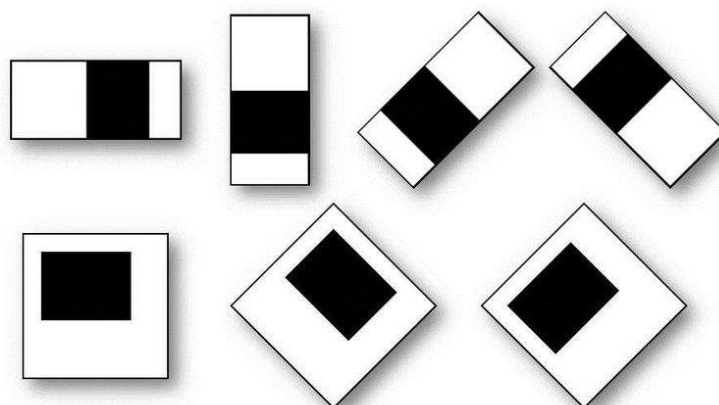


Рисунок 3 – Дополнительные признаки

Использование каскада Хаара имеет следующие плюсы:

1. Описывают те знания о классе объектов, которые трудно выделить на конкретном числе обучаемых данных.
2. Устойчивость к смене освещения, даже если это локальная смена освещения, устойчивость к шумам (примитивы представляют собой простейший полосовой фильтр).

3. Если примитивы были не очень маленькие, то сильно устойчивее корреляции при изменении масштаба (размер примитивов при этом не будет влиять на точность, если обход с маленьким шагом).

4. Если признаки на большом изображении рассчитать заранее и при сдвиге окна поиска брать уже посчитанные и актуальные для него — поиск будет значительно быстрее корреляции (нужно сравнить меньшее количество элементов).

5. Такие системы работают гораздо быстрее, чем системы работающие напрямую с пикселями.

Для выбора наиболее подходящих признаков для искомого объекта на части изображения используют алгоритм бустинга. Данный метод позволяет усилить простые классификаторы путём комбинирования примитивных «слабых» классификаторов в один «сильный». Под «силой» классификатора в данном случае подразумевается эффективность (качество) решения задачи классификации.

Развитием данного подхода явилась разработка более совершенного семейства алгоритмов бустинга — AdaBoost (adaptive boosting — адаптированное улучшение), предложенная Йоавом Фройндом (Freund) и Робертом Шапиром (Schapire) в 1999 году, который может использовать произвольное число классификаторов и производить обучение на одном наборе примеров, поочередно применяя их на различных шагах.

Видеопоток, получаемый с помощью видеокамеры, представляет собой последовательность кадров. Для каждого кадра вычисляется его интегральное изображение. Затем кадр сканируется окном малого размера (субокном), содержащим признаки Хаара.

Для каждого i -го признака соответствующий классификатор определяется формулой:

$$h_i(z) = \begin{cases} 1, & p_i f_i(z) < \theta_i \\ 0, & \text{иначе} \end{cases}$$

где z — субокно; θ_i — пороговое значение; p_i — направление знака неравенства; f_i — признак Хаара.

Таким образом, на основе проведённого анализа выбора оптимальных алгоритмов и методов обнаружения и распознавания жестов руки на видео в режиме реального времени разработан усовершенствованный алгоритм, который позволяет уменьшить вероятность ложных обнаружений и повысить быстродействие работы системы. Данный алгоритм основан на методе Виолы-Джонса с использованием признаков Хаара и состоит из следующих этапов:

1. Предварительная подготовка изображения:
 - а) Изменение размера изображения;
 - б) Перевод изображения в полутоновое;
 - в) Бинаризация изображения;
 - г) Пороговое преобразование;
 - д) Операции морфологической обработки изображения.
2. Обнаружение руки и параметров жеста:
 - а) Интегральное представление изображения;
 - б) Использование метода Виолы-Джонса;
 - в) Если на изображении найдена рука, то определить параметры жеста (количество не зажатых пальцев, положение руки и др.).
3. Распознавание жеста:
 - а) Для каждого типа жестов определены свои каскады классификаторов Хаара;
 - б) Если определенные классификаторы Хаара сработали и найдены параметры к какому-либо типу, то определен жест.

Схематически алгоритм представлен на рисунке 4.

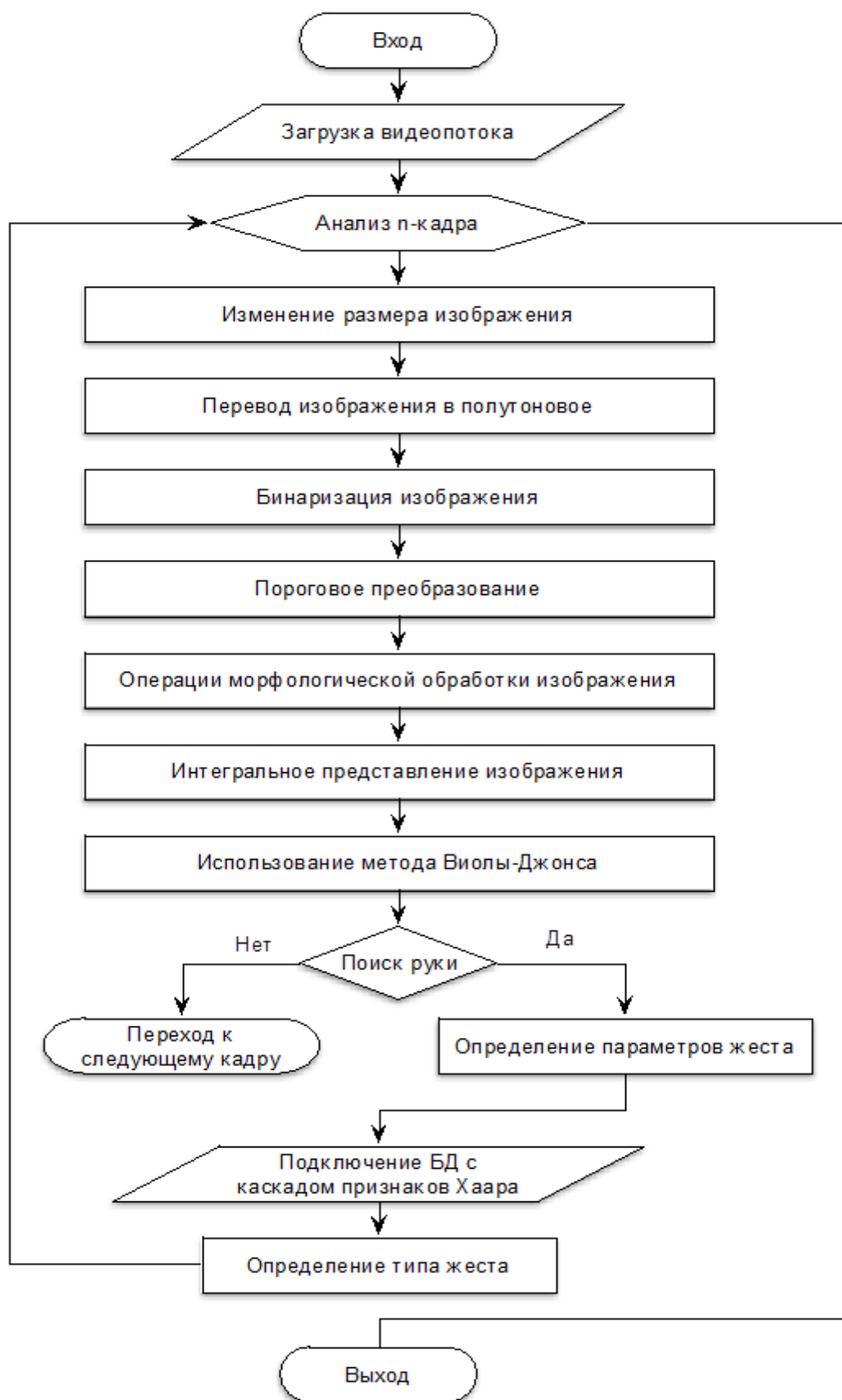


Рисунок 4 – Алгоритм распознавания жестов

В результате разработанный алгоритм позволяет уменьшить вероятность ложных обнаружений и повысить быстродействие работы системы, помогает избавиться от ограничений в виде помех на изображении и неразличимого объекта на фоне с помощью предварительной обработки изображения, так как данный метод работает гораздо лучше с полутонными более четкими изображениями, также избавиться от проблем, связанных с углом наклона руки путем тренировки новых каскадов.

Данный подход является универсальным, расширяемым и если использовать достаточное количество каскадов классификаторов, то метод нахождения работает очень быстро и практически безошибочно.

Список литературы

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М.: Техносфера, 2005. 1072 с.
2. Зайцева Г.Л. Жестовая речь. Дактилология: Учеб. для студ. Высш. Учеб. Заведений.- М.: Гуманит. Изд. центр ВЛАДОС, 2004.-192 с.
3. Форсайт Д., Понс Ж. Компьютерное зрение. Современный подход / Пер. с англ. А.В. Назаренко и И.Ю. Дорошенко.- М.: Издат. дом "Вильямс", 2004.- 928 с.
4. Хуанг Т.С. и др. Быстрые алгоритмы в цифровой обработке изображений. – М.: Радио и связь, 1984. – 224 с.
5. Шапиро Л., Стокман Д. Компьютерное зрение / Пер. с англ.-М.:БИНОМ. Лаборатория знаний, 2006.-752 с.

List of references

1. Gonsales P., Vuds R. Cifrovaja obrabotka izobrazhenij. M.: Tehnosfera, 2005. 1072 s.
2. Zajceva G.L. Zhestovaja rech'. Daktilologija: Ucheb. dlja stud. Vyssh. Ucheb. Zavedenij.- M.: Gumanit. Izd. centr VLADOS, 2004.-192 s.
3. Forsajt D., Pons Zh. Komp'juternoe zrenie. Sovremennyj podhod / Per. s angl. A.V. Nazarenko i I.Ju. Doroshenko.- M.: Izdat. dom "Vil'jame", 2004.- 928 s.
4. Huang T.S. i dr. Bystrye algoritmy v cifrovoj obrabotke izobrazhenij. – M.: Radio i svjaz', 1984. – 224 s.
5. Shapiro L., Stokman D. Komp'juternoe zrenie / Per. s angl.-M.:BINOM. Laboratorija znaniy, 2006.-752 s.