

МЕТОД ВИОЛЫ-ДЖОНСА ДЛЯ РАСПОЗНАВАНИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИИ

THE VIOLA-JONES METHOD FOR OBJECT RECOGNITION IN IMAGES

A. Tymchuk

Summary. This article describes the algorithm of Viola-Jones object recognition on the static image. Outlines the key points of the algorithm and used methods, determined the strengths and weaknesses of the algorithm. Conclusions are made regarding the relevance of the algorithm and the areas of its application.

Keywords: Viola-Jones, computer vision, Haar-like feature, machine learning, AdaBoost.

Тымчук Андрей Игоревич

ФГБОУ ВО «Кубанский Государственный Университет»,
timchuck.andrey2011@yandex.ru

Аннотация. В статье проводится разбор алгоритма Виолы-Джонса для решения задачи обнаружения объекта на статичном изображении. Обозначены ключевые моменты алгоритма и используемые в нём методы, определены сильные и слабые стороны алгоритма. Кроме того, сделаны выводы относительно актуальности алгоритма и сферах его применения.

Ключевые слова: Виола-Джонс, компьютерное зрение, признак Хаара, машинное обучение, AdaBoost.

В настоящее время в определённых областях человеческой деятельности стоит задача распознавания объекта на статичном изображении или в видеопотоке. К таким областям относятся:

- 1) Автоматический контроль состояния работника на рабочем месте (например, пилотов, машинистов, авиадиспетчеров и т.д.), там, где необходимо, чтобы работник имел должную концентрацию внимания;
- 2) Информационная помощь при обеспечении безопасности на объектах с большим скоплением людей;
- 3) Помощь в медицине при необходимости точного обнаружения определённого объекта;
- 4) Повышение качества оказания услуг за счёт оценки удовлетворённости потребителя.

Задача распознавания образов не имеет точного аналитического решения, что привносит сложности в разработку универсального алгоритма. Тем не менее, для того, чтобы наделять компьютерные системы возможностью своеобразного «зрения» создано и предложено большое число методов и алгоритмов. Среди них особо выделяются подходы, основанные на нейронных сетях, методе главных компонент, вейвлет-преобразованиях. Одним из самых известных и популярных алгоритмов в области компьютерного зрения является алгоритм Виолы-Джонса (Viola-Jones). Он был разработан Полом Виолой и Майклом Джонсом в 2001 году, и получил широкое применение в силу своей скорости и эффективности.

Метод состоит из двух этапов: алгоритм обучения и алгоритм распознавания. На практике скорость работы алгоритма обучения не важна. Однако крайне важна скорость работы алгоритма распознавания объектов. Метод имеет следующие ярко выраженные преимущества:

- 1) Возможность обнаружения большого количества объектов на исходном изображении;
- 2) Использование простых классификаторов в процессе работы алгоритма обнаружения показывает хорошую скорость и позволяет использовать этот метод в анализе видео на присутствие определённых объектов в кадре;
- 3) Возможность обучения классификатора нахождение любого объекта;
- 4) Проверка любого признака (в методе используются признаки Хаара, о которых будет рассказано далее) на определённой позиции в изображении занимает линейное время ввиду использования интегрального представления изображения;
- 5) Большое количество реализаций в различных open-source библиотеках и на различных языках программирования;
- 6) Малое количество ложных срабатываний при поиске объекта на изображении.

Недостатки метода:

- 1) Большое время обучения классификатора, так как требуется проанализировать большое количество тестовых изображений;
- 2) Зависимость точности распознавания от правильности подбора обучающей выборки при обучении классификатора;

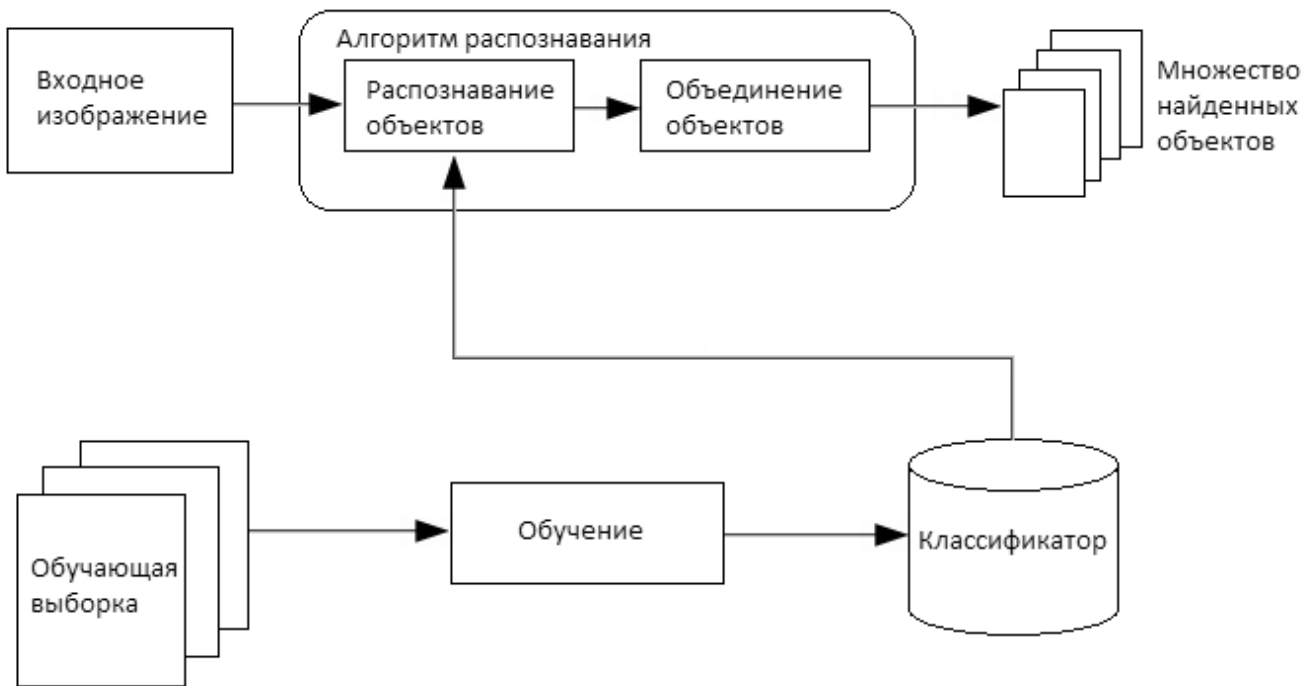


Рис. 1. Обобщённая схема алгоритма Виолы-Джонса

- 3) Сильное снижение эффективности распознавания объекта при увеличении угла поворота объекта;
- 4) Может наблюдаться большое количество близко расположенных друг к другу результатов ввиду применения различных масштабов для сканирующего.

Упрощённая схема алгоритма выглядит следующим образом. Перед началом распознавания алгоритм обучения на основе тестовых изображений обучает классификатор, состоящий из значений определённых признаков. Далее алгоритм распознавания ищет объекты на разных масштабах изображения, основываясь на обученном классификаторе. На выходе алгоритма выдаётся множество найденных объектов на разных масштабах. Схема алгоритма в графическом виде представлена на рисунке 1.

Основополагающей идеей алгоритма Виолы-Джонса для распознавания объектов является выделение локальных особенностей (признаков) изображения и последующего обучения алгоритма на них. Поиск на изображении выполняется по принципу сканирующего окна — изображение сканируется окном поиска, и на каждом положении окна применяется классификатор [2].

Система обучения и выбора наиболее значимых признаков полностью автоматизирована и не требует вмешательства человека.

Все операции с изображениями в методе выполняются с использованием интегральных представлений изображений. Интегральное представление позволяет быстро рассчитывать суммарную яркость произвольного прямоугольника на данном изображении, причем время расчета не зависит от размеров прямоугольника. Интегральное представление изображения — это матрица, совпадающая по размерам с исходным изображением. В каждом элементе ее хранится сумма интенсивностей всех пикселей, находящихся левее и выше данного элемента [2]. Элементы матрицы рассчитываются по формуле (1).

$$M(x, y) = \sum_{i=0, j=0}^{i \leq x, j \leq y} I(i, j) \tag{1}$$

В вышеуказанной формуле $I(i, j)$ представляет собой яркость пикселя с координатами i и j в исходном изображении. В итоге, каждый элемент интегрального изображения $M[x, y]$ содержит в себе сумму пикселей изображения в прямоугольнике от $(0,0)$ до (x, y) . Расчёт интегрального изображения выполняется за линейное время, которое прямо пропорционально числу пикселей исходного изображения. Соответственно интенсивность пикселей на любом прямоугольном участке изображения, который имеет в координаты $A(x1, y1), B(x2, y2), C(x3, y3), D(x4, y4)$ (соответствующий прямоугольник ABCD изображен на рисунке 2), можно рассчитать по формуле (2) [1].

$$RSum = M(A) + M(C) - M(B) - M(D) \quad (2)$$

В формуле всего три математические операции вне зависимости от размера прямоугольника. Также, вне зависимости от размера будет выполнено четыре обращения к памяти. Расчёт всей матрицы можно выполнить по рекуррентной формуле (3).

$$M(x, y) = I(x, y) - M(x-1, y-1) + M(x, y-1) + M(x-1, y) \quad (3)$$

Как уже говорилось ранее, основная идея алгоритма — распознавание особенностей на изображении. Особенности, которые использовали Пол Виола и Майкл Джонс, базируются на каскадах признаков Хаара. Основной причиной, почему в основу метода Виолы-Джонса легли примитивы Хаара, стала попытка уйти от пиксельного представления с сохранением скорости вычисления признака. Такое название они получили из-за интуитивной схожести с вейвлетами Хаара. Признаки Хаара представляют собой прямоугольные области, которые составлены из нескольких соседних прямоугольных областей, отмеченных как светлая или темная. Под признаком будем понимать трёхмерный вектор вида:

$$j = \{ \text{маска, положение, размер} \} \quad (4)$$

Из значений пары пикселей сложно вынести какую-либо осмысленную информацию при выполнении классификации, в то время как из нескольких признаков Хаара можно построить первый каскад системы по распознаванию объекта, который имеет вполне осмысленную интерпретацию.

В процессе движения сканирующего окна признак позиционируется на изображении, далее суммируются интенсивности пикселей в светлых и тёмных областях признака, после чего вычисляется разность между суммами. Эта разность представляет собой значение определенного признака, определенного размера, определенным образом позиционированного на изображении [1]. Соответственно значение каждого признака рассчитывается по формуле (5).

$$F = L - D \quad (5)$$

В данной формуле L — сумма значений яркостей точек, закрываемых светлой частью признака, а D — сумма значений яркостей точек, закрываемых темной частью признака. Если значение признака Хаара превышает определённый порог, который вычисляется в процессе обучения классификатора, то считается, что признак имеет место на изображении.



Рис. 2. Пример расположения прямоугольника в формуле (2)

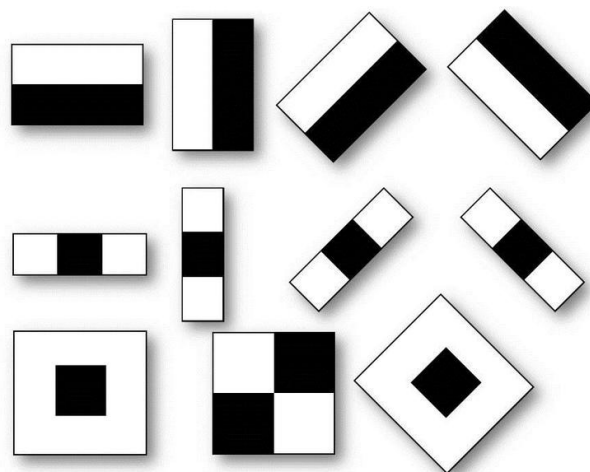


Рис. 3. Признаки Хаара, использующиеся в стандартном методе Виолы-Джонса

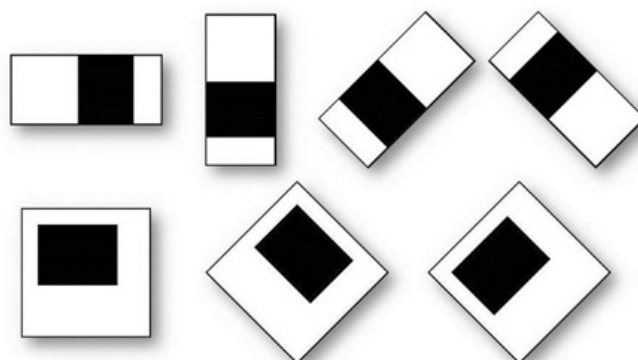


Рис. 4. Дополнительные признаки Хаара

Признаки (иногда их ещё называют примитивами) Хаара, которые используются в стандартном алгоритме Виолы-Джонса, изображены на рисунке 3.

Для каждой прикладной задачи могут использоваться различные признаки и совершенно разные методы. В стандартном методе Виолы-Джонса используются прямоугольные признаки. Такие признаки представляют собой прямоугольники, которые различаются между собой размерами сторон, углом наклона и расположением тёмных и ярких частей. Стоит отметить, что существует расширенная версия метода Виолы-Джонса (она, в частности, используется в библиотеке компьютерного зрения OpenCV), в которой применяются некоторые дополнительные прямоугольные признаки [2]. Данные признаки представлены на рисунке 4.

Операция вычисления суммы значений яркости пикселей области признака будет многократно повторяться в процессе сканирования изображения. Если данная операция будет требовать значительных вычислительных ресурсов, то это повлечёт увеличение общего времени работы алгоритма на маломощных системах. Поэтому в алгоритме Виолы-Джонса при выполнении вычислений авторы предлагают использовать интегральное представление изображения, которое было рассмотрено ранее. Благодаря этому, время расчёта каждого признака неизменно.

Таким образом, признаки Хаара дают точечное значение изменения яркости по осям X и Y . Рассмотрим пример — базу данных с изображениями человеческих лиц анфас. Общим для всех изображений будет то, что область в районе глаз темнее, чем область в районе щёк. Из этого следует, что общим признаком Хаара для лиц является два смежных прямоугольных региона, лежащих на глазах и щеках.

Поиск признаков Хаара на изображении выполняется по методу сканирующего окна. В общем виде работа данного метода состоит из следующих шагов [2]:

- 1) Окно сканирования постепенно перемещается по рассматриваемому изображению с шагом, равным одной ячейке изображения (исходный размер ячейки устанавливается заранее);
- 2) Для каждого положения окна вычисляются варианты расположения признаков посредством изменения расположения и масштаба признаков;
- 3) Обнаруженные признаки поступают на вход классификатору;
- 4) Изменяется масштаб сканирующего окна за счёт изменения размера ячейки изображения и повторяются шаги 1–3;
- 5) После вычисления признаков для всех масштабов классификатор определяет, присутствует ли

на изображении соответствующий объект или нет (в зависимости от того, как натренирован классификатор).

Скорость сканирования изображений сильно зависит от размера шага сканирующего окна в виде количества пикселей, на которое сдвигается сканирующее окно. В большинстве имеющихся реализаций метода Виолы-Джонса шаг постоянный и определяется на этапе компиляции, что означает, что окно смещается на постоянную величину и не зависит от содержания изображения. Однако, стоит отметить, что существуют реализации, в которых величина шага является адаптивной, то есть подстраивается под каждое изображение.

Так как в процессе «сканирования» нужно вычислить большое количество признаков, встаёт вопрос об оптимизации метода. Как правило, в данном случае приходят к тому, что классификатор должен реагировать только на определённое множество признаков и игнорировать признаки, которые не подходят для определяемого объекта. При обучении классификатора для метода Виолы-Джонса можно использовать машинное обучение.

Существуют самые разные реализации алгоритмов обучения классификатора. От качества реализации, а также качества и длительности обучения зависит точность, с которой классификатор сможет распознавать объекты на изображениях. В общем случае в процессе обучения используется множество изображений, в котором присутствует так называемая «обучающая выборка». В процессе обучения должен сформироваться алгоритм, который сможет классифицировать объект из исходного множества.

В контексте алгоритма обучения имеется множество объектов (в алгоритме обучения классификатора для метода Виолы-Джонса объектами являются изображения), разделённых некоторым образом на классы. Задаётся конечное множество объектов, для которых известно, к какому классу они относятся (к примеру, это может быть класс «номер машины», «профиль человека», «мотоцикл» и т.д.). Данное множество и носит название обучающей выборки. Принадлежность остальных объектов к оставшимся классам неизвестна. Соответственно, в процессе обучения классификатора нужно построить алгоритм, который будет способен классифицировать произвольный объект из исходного множества.

Задача классификации в машинном обучении относится к типу «Обучение с учителем» (Supervised learning), в ходе которой классификатор принудительно обучается с помощью примеров «стимул-реакция». В методе Виолы-Джонса классификация является двух классовою (объект либо есть на изображении, либо нет). На осно-

ве обучающей выборки восстанавливается зависимость (строится модель отношений стимул-реакция, которые пригодны для прогнозирования), то есть создаётся алгоритм, который способен для любого объекта определять ответ с определённой долей точности.

С математической точки зрения постановка задачи классификации может быть описана следующим образом. Имеется множество D , в котором хранится описание объектов, N — конечное множество номеров, принадлежащих классам. Между ними есть зависимость в виде отображения:

$$M: D \Rightarrow N \quad (5)$$

Обучающая выборка представляется в виде множества:

$$D_i = \{(d_1, n_1), \dots, (d_b, n_b)\} \quad (6)$$

В ходе решения задачи классификации строится функция F от множества D , которая способна классифицировать любой объект $d \in D$. Одним из обязательных условий для функции является то, что она должна хорошо работать на любых новых данных.

В виду сложности обучения, в алгоритме Виолы-Джонса в процессе обучения используется технология бустинга (AdaBoost). Бустинг представляет собой алгоритм усиления классификатора. Эффективный классификатор, дающий малое количество ошибок, называется «сильным», а противоположность такого классификатора — «слабым». Под слабым классификатором можно понимать классификатор, который даёт правильный ответ не чаще, чем при обычном угадывании. В процессе бустинга выполняется построение набора алгоритмов машинного обучения, в котором каждый последующий алгоритм пытается компенсировать недостатки предыдущих алгоритмов. AdaBoost был предложен Йоавом Фройндом (Yoav Freund) и Робертом Шапиром (Robert Schapire) в конце прошлого века в качестве решения вопроса «Как из множества плохих алгоритмов построить один хороший» [4].

Суть алгоритма бустинга заключается в том, что если имеется набор эталонных объектов (обучающая выборка) и множество простых классификаторов, то из простых классификаторов можно построить один лучший классификатор. При этом AdaBoost является адаптивным алгоритмом, поскольку в процессе составления и обучения финального классификатора акцент делается на эталоны, которые распознаются труднее остальных. Таким образом, алгоритм, составляемый в процессе обучения классификатора, подстраивается под самые труднораспознаваемые объекты.

Метод AdaBoost имеет следующие преимущества [4]:

- 1) Хорошая обобщающая способность, которая может улучшаться по мере увеличения числа базовых алгоритмов;
- 2) Простота реализации метода;
- 3) Метод не сильно влияет на время обучения.

Недостатки метода [4]:

- 1) Наличие возможности переобучения, когда алгоритм начинает положительно реагировать на сильно-зашумлённые объекты;
- 2) Требуются большие обучающие выборки;
- 3) Возможно построение слишком большого набора алгоритмов, которые будут требовать существенных ресурсов на вычисление классификации объекта.

В методе Виолы-Джонса AdaBoost позволяет получить серии классификаторов как последовательность фильтров, что особенно эффективно для классификации областей изображения. Каждый фильтр является отдельным классификатором AdaBoost с достаточно малым числом слабых классификаторов. Расположение фильтров в каскаде базируется на весовых значениях, которые присваивает алгоритм бустинга. Более тяжёлые взвешенные фильтры располагаются в самом начале цепочки каскада, для более быстрого отбрасывания областей изображения, не содержащих искомого объекта изображения.

В начале процесса обучения на каждом уровне фильтров в каскаде устанавливается низкий порог, чтобы фильтры пропустили большинство образов из тренировочного набора. Если во время работы какой-то из фильтров не пропускает область изображения, то тогда область сразу же классифицируется как область без искомого объекта. Когда фильтр пропускает область изображения, в дело вступает следующий фильтр в последовательности, который также выполняет проверку области на наличие искомого объекта. Области изображения, прошедшие через все фильтры, классифицируются как области, содержащие искомого объект.

На основании рассмотренного в данной статье материала можно сделать следующие выводы:

- 1) Алгоритм Виолы-Джонса отлично подходит для использования в системах реального времени ввиду малозатратности операции обнаружения объекта;
- 2) Качество обнаружения объекта сильно зависит от процесса и качества обучения классификатора, соответственно для правильной классификации важно построить правильный классификатор;
- 3) Алгоритм плохо подходит для описания изображений, так как основан на нахождении конкретных

признаков и не способен работать с содержанием изображения;

- 4) Алгоритм можно применять для нахождения любых объектов (нужно только обучить соответствующий классификатор, желательнее также использовать соответствующие признаки), несмотря на то, что изначально алгоритм разрабатывался для определения человеческого лица.

ющий классификатор, желательнее также использовать соответствующие признаки), несмотря на то, что изначально алгоритм разрабатывался для определения человеческого лица.

ЛИТЕРАТУРА

1. Метод распознавания лиц Виолы-Джонса: [Электронный ресурс]. URL: <http://oxozle.com/2015/04/11/metod-raspoznavaniya-lic-violy-dzhonsa-viola-jones>
2. Метод Виолы-Джонса как основа для распознавания лиц: [Электронный ресурс]. URL: <https://habrahabr.ru/post/133826>
3. Face Detection using Haar Cascades: [Электронный ресурс]. URL: http://docs.opencv.org/trunk/d7/d8b/tutorial_py_face_detection.html
4. Алгоритм AdaBoost: [Электронный ресурс]. URL: <http://www.machinelearning.ru/wiki/index.php?title=AdaBoost>

© Тымчук Андрей Игоревич (timchuck.andrey2011@yandex.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»

