

УДК 004.93'11

*Азаров Д. М.*

Магистрант инфокоммуникационных технологий и систем связи  
Санкт-Петербургский государственный университет  
Аэрокосмического приборостроения

*Гильмутдинов М. Р.*

доцент, к.т.н.

Санкт-Петербургский государственный университет  
Аэрокосмического приборостроения

## РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДА РАСПОЗНАВАНИЯ ОБЪЕКТОВ НА БАЗЕ АЛГОРИТМА ВИОЛЫ-ДЖОНСА

В статье описан предложенный метод распознавания объектов, использующий значение мягкого выхода и кластеризацию полученных результатов.

**Ключевые слова:** распознавание образов, алгоритм Виолы-Джонса, компьютерное зрение.

*Azarov D. M.*

Master student Information and Communication Technologies and Communication Systems

St. Petersburg State University of Aerospace Instrumentation

*Gilmudinov M. R.*

Associate professor, Ph.D.

St. Petersburg State University of Aerospace Instrumentation

## DEVELOPMENT AND RESEARCH OF OBJECT RECOGNITION METHOD BASED ON VIOLA-JONES ALGORITHM

The paper presents a method of object recognition using a soft output and clustering.

**Keywords:** object recognition, algorithm Viola-Jones, computer vision.

### Введение

В настоящее время существует множество задач, в которых требуется принять некоторое решение в зависимости от присутствия на изображении объекта или классифицировать его. Способность «распознавать» считается

основным свойством биологических существ, в то время как компьютерные системы этим свойством в полной мере не обладают.

Основополагающей идеей при создании алгоритма Виолы-Джонса для распознавания лиц является выделение локальных особенностей (признаков) изображения и последующего обучения алгоритма на них. Для определения локальных особенностей было предложено использовать *признаки Хаара* [1]. Под признаком  $j$  будем понимать трехмерный вектор вида

$$j = \{\text{маска, положение, размер}\} \quad (1)$$

Каждая *маска* характеризуется размером светлой и темной областей, пропорциями, а также минимальным размером. Пример часто используемых масок изображен на рис. 1.

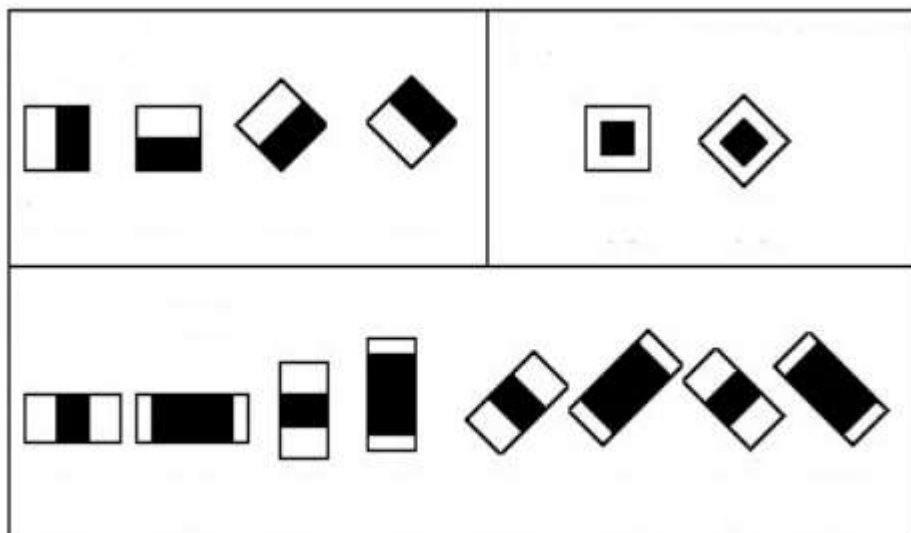


Рисунок 1 Пример набора масок из метода Виолы-Джонса

Обработка изображения выполняется с помощью *метода скользящего окна*. Так как искомый объект на изображении может иметь различный масштаб необходимо произвести поиск этого объекта с различным масштабом признаков. Эта процедура позволяет производить однотипные вычисления на разных областях исходного изображения и на разных масштабах признаков. Скользящее окно  $U$  описывается начальным размером  $U_x$  и  $U_y$  и величиной сдвига по осям  $\Delta U$ . В предложенном алгоритме минимальный размер скользящего окна был предложен 24x24 пикселя.

Базовый алгоритм Виолы-Джонса (далее *базовый* алгоритм) имеет ряд недостатков:

- длительное время работы алгоритма обучения. В ходе обучения алгоритму необходимо проанализировать большое количество тестовых изображений;
- большое количество близко расположенных друг к другу результатов из-за применения различных масштабов и скользящего окна.

Длительное время работы алгоритма обучения выходит за рамки данной статьи. В рамках данной статьи речь будет идти об исследовании эффективности распознавания базового алгоритма, разработке метода распознавания, использующий модификацию алгоритма распознавания Виолы-Джонса и постобработку в виде кластеризации. В конце статьи приведен сравнительный анализ базового алгоритма и предложенных улучшений.

### **Постановка задачи**

Описание базового алгоритма представлено в работе [1]. Алгоритм распознавания анализирует каждый регион отдельно и принимает жесткое решение о нахождении искомого объекта внутри рассматриваемого региона. *Отклик* признака  $f_j(x)$  вычисляется как разность интенсивностей пикселей в светлой и темной областях. Базовый алгоритм оперирует понятием *слабого классификатора*  $h_j(x)$ , который вычисляется следующим образом:

$$h_j(x) = \begin{cases} 1, & p_j f_j(x) < p_j \theta_j \\ 0, & \text{иначе} \end{cases} \quad (2)$$

где  $p_j$  (паритет),  $\theta_j$  (граница), параметры подбираемые в процессе процедуры обучения.

Окончательное решение принимается на основе значения сильного классификатора, значение которого вычисляется по следующей формуле:

$$H(x) = \begin{cases} 1, & \sum_{t=1}^T a_t h_{j(t)}(x) \geq \frac{1}{2} \sum_{t=1}^T a_t \\ 0, & \text{иначе} \end{cases} \quad (3)$$

где  $a_t = \log \frac{1}{\beta_t}$ ,  $\beta_t = \frac{e_{J^{(t)}}}{1 - e_{J^{(t)}}$ ,  $e_{J^{(t)}}$  - ошибка наилучшего слабого классификатора  $h_J$ , на итерации  $t$ .

Описанный алгоритм имеет существенный недостаток. После работы алгоритма на выходе имеется большое количество данных, обусловленных применением скользящего окна. Целью работы является разработка эффективного метода объединения большого количества разрозненных данных на базе алгоритма Виолы-Джонса и анализ полученных результатов. Объединение осуществляется с помощью алгоритма кластеризации.

### **Критерии оценки эффективности распознавания**

В данной работе рассматриваются известные критерии оценки [2], а также предлагается новый критерий на базе идеи Ground Truth. К известным критериям можно отнести:

- *истинно положительные примеры (TP)* – верно классифицированные примеры;
- *истинно отрицательные примеры (TN)* – верно классифицированные отрицательные случаи;
- при пропуске классификатором присутствующего на изображении объекта *возникает ошибка I рода* – ложный пропуск (*FN*);
- если классификатор определяет искомый объект на изображении, которого в действительности там нет, то *возникает ошибка II рода* – ложное обнаружение (*FP*).

Ошибка классификации происходит, когда классификатор относит входной объект к классу  $C_i$ , в то время как верный класс  $C_j$ , причем

$$i \neq j \text{ и } C_i \neq C_j$$

К типовым оценкам эффективности распознавания следует отнести:

*Доля истинно положительных примеров (True Positives Rate):*

$$TPR = \frac{TP}{TP + FN} \cdot 100\% \quad (4)$$

*Доля ложно положительных примеров (False Positives Rate):*

$$FPR = \frac{FP}{TN + FP} \cdot 100\% \quad (5)$$

Дополнительный критерий оценки эффективности работы алгоритма может быть сформулирован на основе Ground Truth. За основу взята работа [4] и предложен алгоритм вычисления численного критерия оценки  $er$ , как отношения непересекающихся частей к общей площади регионов (рис 2).

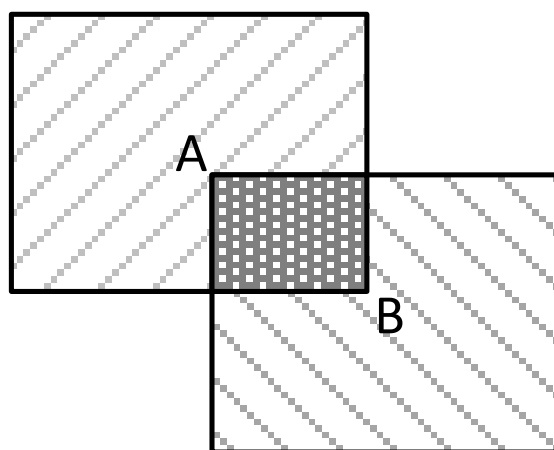


Рисунок 2 Пример вычисления критерия Ground Truth Base

Критерий  $er$  вычисляется по следующей формуле:

$$er(A, B) = \frac{|A \setminus B| + |B \setminus A|}{|A| + |B|} \quad (6)$$

Введенный критерий показывает ошибку несовпадения двух рассматриваемых регионов, при этом он учитывает взаимное положение объектов между собой. Численное значение критерия ограничено интервалом  $[0,1]$ , при достижении значения 1 можно сделать вывод, что объекты не пересекаются и получен результат  $FP$ . При значении  $er$  равным 0 получаем полное совпадение с указанным вручную регионом, что означает  $TP$ .

### **Модификация сильного классификатора в форме мягкого решения**

В оригинальном алгоритме 2001 года решение о том есть ли искомый объект в рассматриваемом скользящем окне принимается однозначное (жесткое) решение. Объект либо есть, либо его нет. Решение принимается на основе строго результата вычисления формулы (3). Идея мягкого выхода за-

ключается в нечетком результате, на основе анализа которого можно принять другое решение.

Мягкие вычисление – термин, введенный Лотфи Заде в 1994 году, обозначающий совокупность неточных, приближенных методов решения задач, зачастую не имеющих решение за полиномиальное время [5].

В проведенных опытах было обнаружено, в некоторых условиях алгоритмом принималось решение о том, что искомого объекта нет на текущем скользящем окне, в то время как в действительности объект существовал. Идея мягкого выхода заключается в неточной оценке нахождения объекта в рассматриваемом окне.

Напомним, в оригинальном алгоритме для того, чтобы классификатор вынес решение о том, что в рассматриваемом окне присутствует искомым шаблон, необходимо, чтобы выполнялось условие сильного классификатора (3). Модифицируем формулу (3) к следующему виду:

$$H(x) = \begin{cases} 1, & \text{если } \frac{2 \sum_{t=1}^T a^{(t)} h_j^{(t)}}{\sum_{t=1}^T a^{(t)}} \geq 1 \\ 0, & \text{иначе} \end{cases} \quad (7)$$

Далее избавимся от принятия решения по порогу. Будем рассматривать отношение в формуле (7), как самостоятельный числовой критерий, который будем называть мягкий выход:

$$Hs(x) = \frac{2 \sum_{t=1}^T a^{(t)} h_j^{(t)}}{\sum_{t=1}^T a^{(t)}} \quad (8)$$

Одним из возможных применений мягкого выхода является принятие однозначного решения о найденном объекте, исходя из значения  $Hs(x)$ . Однако применение такого критерия напрямую имеет ряд недостатков. На рис. 3 показан пример работы оригинального алгоритма Виолы Джонса с дополнительными значениями. Белым отмечено истинное лицо (Ground Truth). Красным выделены регионы, найденные алгоритмом Виола Джонса. В каждом найденном регионе числом записано значение мягкого выхода. Также есть отметка “GT”, как регион с наименьшей ошибкой Ground

Truth, по отношению к вручную проставленному региону (белым цветом). Видно, что найденный алгоритмом объект в реальности не имеет наибольшее значение мягкого выхода.



Рисунок 3 Пример работы оригинального алгоритма Виолы-Джонса с дополнительными значениями

На рис. 4 видно, что регион с максимальным значением мягкого выхода 1,0548 наоборот имеет ложно положительное срабатывание и никак не может быть классифицирован как искомый объект.



Рисунок 4 Пример ложного результата при максимальном значении мягкого выхода

На рис. 5 видно, что единственный найденный алгоритмом регион является ложно положительным срабатыванием.



Рисунок 5 Пример изображение с единственным найденным объектом с максимальным значением мягкого выхода

Приведенные выше примеры демонстрируют: *нельзя достоверно утверждать, что регион с максимальным значением мягкого выхода определенно является искомым объектом.*



Однако дальше будет показано, что введенный по формуле (8) мягкий выход можно использовать при кластеризации решений базового алгоритма Виолы-Джонса.

### **Кластеризация данных с учетом мягкого выхода**

Процесс кластеризации занимается поиском структуры в коллекции немаркированных данных, это процесс организации объектов в группы, члены которой в некотором смысле похожи. Кластер – множество похожих объектов, между тем как непохожие принадлежат другим кластерам.

Основные требования к алгоритму кластеризации:

- Масштабируемость (однотипная работа на разных объемах данных)
- Минимальные требования начальных знаний о природе объектов
- Способность отбрасывать шум и выбросы

Под указанные требования подходит алгоритм DBSCAN [6]. Алгоритм прост в реализации. Оперирует понятием сосед, расстояние до соседей и количество соседей. Параметры на базе этих понятий могут быть использованы для повышения эффективности решаемой задачи.

В задаче кластеризации данных можно учитывать дополнительную информацию, которой является значение мягкого выхода. Кроме того, используя эту информацию можно искусственно снизить порог распознавания и учитывать полученные регионы на этапе кластеризации данных.

При снижении порога значительно увеличивается количество найденных регионов. На рис. 6 показано среднее количество найденных объектов, в зависимости от значения мягкого выхода.

Идея предложенного метода заключается в проверке гипотезы: *при уменьшении порога срабатывания алгоритма распознавания и применяя алгоритм кластеризации объектов возможно улучшить эффективность распознавания.* При этом следует заметить, что параметры в алгоритме кластеризации подбираются эмпирически.



Рисунок 6 Зависимость количества найденных объектов от порога срабатывания

Проведя серию экспериментов был сформирован график зависимости (рис. 7), по которому была определена оптимальная пара значений для алгоритма кластеризации (количество соседей, минимальное расстояние между соседями).



Рисунок 7 Зависимость Ground Truth от порога алгоритма распознавания и количества ближайших соседей

Учитывая вероятность ложного срабатывания (таблица 1), выбираем подходящие параметры.

Таблица 1 Вероятность ложного срабатывания в зависимости от порога алгоритма распознавания

Порог	0,82	0,84	0,86	0,88	0,9	0,92	0,94	0,96	0,98	1	1,02
Вероятность ложного срабатывания	98%	78%	18%	15%	15%	14%	8%	2%	1%	0%	0%

### Анализ полученных результатов

Проводимые эксперименты анализируют изображения из базы данных YouTube Faces DB [3]. Изображения из этой базы данных представляют из себя отдельные кадры из видеозаписей, снятых на фиксированную камеру людей в различных позах и действиях. К каждому изображению вручную приписана информация о положении и размере лиц в кадре (Ground Truth). Сравнение производилось с использованием 300 тестовых фотографий разных людей. При сравнении использовались следующие критерии оценки: ложный пропуск ( $FN$ ) и ложное обнаружение ( $FP$ ). Оценка эффективности распознавания в данной работе ведется с двумя альтернативными реализациями алгоритмами Виолы-Джонса: Open CV [7] и открытой реализация стандартного алгоритма Viola Jones [8].

На рис. 8 показан результат по обоим критериям сравнения. Видно, что предложенный алгоритм эффективнее обрабатывает ошибки I рода чем стандартный алгоритм и реализация Open CV. В это же время ошибки II рода обрабатываются хуже, чем в Open CV, но заметно лучше, чем в оригинальном алгоритме. Это позволяет сделать вывод, что предложенный метод распознавания объектов эффективнее оригинального алгоритма.

При различных значениях параметров алгоритма кластеризации изменяется процентное соотношение ошибок I и II родов. Подобранные параметры

алгоритма кластеризации и пороговое значение мягкого выхода показывают одинаковые результаты на различных тестовых выборках. Это позволяет сделать вывод, что предложенный метод инвариантен к рассматриваемым исходным данным.

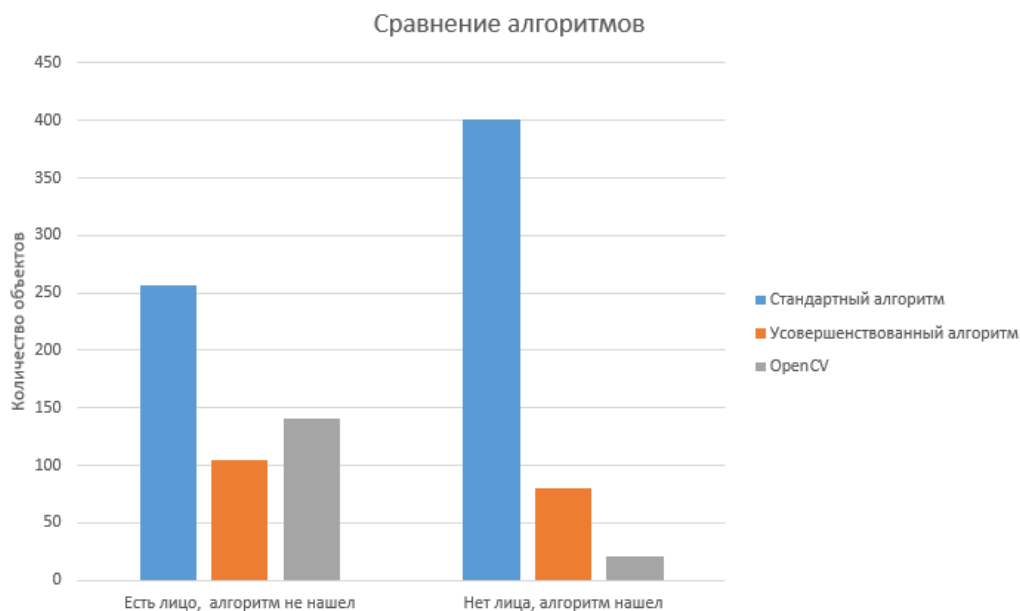


Рисунок 8 Сравнение алгоритмов

## Выводы

В рамках статьи был рассмотрен алгоритм Виолы-Джонса. Определены его основные недостатки: большое количество выходных данных алгоритма распознавания и длительное время работы алгоритма обучения. Было установлено, что использование мягкого выхода для однозначного ответа о принадлежности рассматриваемого региона к искомому объекту невозможно.

Предложен метод распознавания объектов, использующий измененное значение мягкого выхода и дополнительную постобработку в виде кластеризации полученных регионов. Приведенный метод показал хорошие результаты распознавания, что подтверждается большим количеством тестов.

Полученные результаты позволяют сделать вывод, что предложенный алгоритм на 59% эффективнее базового алгоритма Виолы-Джонса справляется с ошибками I рода и на 81% эффективнее против ошибок II рода. В

сравнении с известной реализацией алгоритма в библиотеки OpenCV предложенный алгоритм на 31% эффективнее справляется с ошибками I рода и на 64% хуже с ошибками II рода. В дальнейшей работе планируется улучшить эффективность работы предложенного алгоритма в борьбе с ошибками II рода.

### **Список литературы**

1. *P. Viola, M. Jones* Rapid Object Detection using a Boosted Cascade of Simple Features, 2001
2. Линда Шапиро, Джордж Стокман, Компьютерное зрение, Бином. Лаборатория знаний, 2006 г. 752 стр.
3. <http://www.cs.tau.ac.il/~wolf/ytfaces/> 07.04.2014
4. *R Grosse, Micah K. Johnson, Edward H. Adelson William T. Freeman* Ground truth dataset and baseline evaluations for intrinsic image algorithms, 2009
5. *Lotfi A. Zadeh*, Fuzzy Logic, Neural Networks, and Soft Computing, 1994
6. *H. Bäcklund, A. Hedblom, N. Neijman* A Density-Based Spatial Clustering of Application with Noise, 2011
7. <http://opencv.org> 07.04.2014
8. <https://github.com/RaphaelJ/Viola-Jones-Mono> 07.04.2014