

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное бюджетное образовательное учреждение  
высшего образования

«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий  
Кафедра «Прикладная математика и информатика»

01.03.02 ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА

СИСТЕМНОЕ ПРОГРАММИРОВАНИЕ И КОМПЬЮТЕРНЫЕ  
ТЕХНОЛОГИИ

### **БАКАЛАВРСКАЯ РАБОТА**

на тему: **Детектирование людей на изображениях с помощью метода  
Виолы-Джонса**

Студент(ка)	И. И. Ялаев	_____
Руководитель	В. С. Климов	_____
Консультант по аннотации	Н. В. Яценко	_____

**Допустить к защите**  
Заведующий кафедрой к.тех.н., доцент, А. В. Очеповский \_\_\_\_\_

« \_\_\_\_\_ » \_\_\_\_\_ 2017 г.

Тольятти 2017

## АННОТАЦИЯ

Тема: «Детектирование людей на изображениях с помощью метода Виолы-Джонса».

В данной выпускной квалификационной работе исследуются пути совершенствования методов распознавания образов. В исследовании рассматривается задача локализации человека на изображении с использованием метода Виола-Джонса. При этом предложено использовать несколько каскадов классификаторов, что позволило увеличить точность работы алгоритма. В работе представлены результаты апробации алгоритма на практике.

Структура ВКР представлена введением, тремя главами, заключением, списком литературы.

Во введении описывается актуальность проводимого исследования, формулируется цель и ставятся задачи, которые необходимо решить.

В первой главе проводится анализ методов детектирования людей на изображениях и сравниваются результаты их работы.

Во второй главе описывается проектирование программного обеспечения, в основе которого будет лежать алгоритм, позволяющий оптимизировать детектирование людей на изображениях.

В третьей главе будет проведена реализация и тестирование разработанного программного обеспечения, а также представлены результаты его работы.

В заключении представлены выводы по проделанной работе.

В работе использовано 3 таблицы, 21 рисунок, список литературы содержит 25 литературных источников. Общий объем выпускной квалификационной работы составляет 44 страниц.

## **ABSTRACT**

The title of the graduation work is “Detecting of People on Images by Means of Viola-Jones's Method”.

In this graduation work we study the ways of improvement of methods of recognition of images investigate. We research the problem of localization of the person on the image with use of a method Viola-Jones. At the same time it is offered to use several cascades of qualifiers that have allowed increasing the accuracy of work of an algorithm. In work results of approbation of an algorithm in practice are presented.

The structure of the graduation project is represented by an introduction, three parts a conclusion and a list of references.

In the first part the analysis of methods of detecting of people on images is carried out and results of their work are compared.

The second part describes the design of software, which is based on an algorithm that allows to detect people on images.

In the third part, the software was implemented and tested. Also, the results of the software were provided.

The graduation work consists of 3 tables, 21 figures and the list of references contains 25 sources. The total amount of final qualifying work is 44 pages.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	3
1 АНАЛИЗ СОСТОЯНИЯ ВОПРОСА О ПРОБЛЕМЕ ЛОКАЛИЗАЦИИ ЧЕЛОВЕКА В КАДРЕ .....	4
1.1 Постановка задачи локализации человека в кадре видеопотока .....	4
1.2 Локализация человека как задача бинарной классификации .....	7
1.3 Комбинирование технологий в задачах локализации человека.....	10
1.4 Сравнение результатов работы методов HOG SVM, Viola-Jones и GNG- FIS .....	22
2 РАЗРАБОТКА АЛГОРИТМА ПОИСКА ЛЮДЕЙ НА ИЗОБРАЖЕНИИ...	25
2.1 Метод Viola-Jones.....	25
2.2 Модификация алгоритма.....	33
3 ПРОГРАММНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА .....	34
3.1 Разработка программной части.....	34
3.2 Тестирование алгоритма и обсуждение результатов .....	36
ЗАКЛЮЧЕНИЕ .....	41
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	<b>Ошибка! Закладка не определена.</b>

## ВВЕДЕНИЕ

От интеллектуальных систем в большинстве случаев требуется выработка управляющих сигналов в зависимости от окружающей ситуации. Информация об окружающей ситуации может представлять собой потоковое видео, в том случае если она считывается с помощью цифровой камеры. При этом разработчикам интеллектуальных систем требуется обеспечить решение задачи распознавания образов в зависимости от требуемого функционала системы.

Очень часто разработчикам таких систем приходится сталкиваться с решением подзадачи локализации людей в кадре видеоряда.

Для ее решения используются такие методы локализации объектов на изображении как HOG SVM, Viola-Jones, GNG-FIS.

Известно, что метод Viola-Jones отлично справляется с локализацией на изображении статических объектов. Человек к таким объектам не относится, так как в процессе движения он может вращать головой, двигать руками, ногами и т.д.

По этой причине, при использовании метода Viola-Jones в классическом виде для локализации человека дает слабые результаты из низкой точности, вызванной большим (по сравнению с HOG SVM и GNG-FIS) количеством ошибок первого рода.

Исследования в данной ВКР направлены на поиск путей по уменьшению ошибок первого рода при использовании метода Viola-Jones в задачах локализации людей на изображениях.

# 1 АНАЛИЗ СОСТОЯНИЯ ВОПРОСА О ПРОБЛЕМЕ ЛОКАЛИЗАЦИИ ЧЕЛОВЕКА В КАДРЕ

## 1.1 Постановка задачи локализации человека в кадре видеопотока

От интеллектуальных систем в большинстве случаев требуется выработка управляющих сигналов в зависимости от окружающей ситуации.

Если в задачи интеллектуальной системы входит обеспечение безопасности, то она должна уметь распознавать ситуации связанные с опасным поведением людей (хулиганство, драки, бег в неположенных местах, наличие предметов, представляющих опасность).

Для того чтобы обеспечить данные системы необходимым функционалом требуется решение задач распознавания образов, что достигается путем применения специальных интеллектуальных алгоритмов.

Уже существуют решения таких технических задач как детектирование распознавание автомобильных номеров [1], обнаружение движения объектов [2], идентификация людей по лицу [3], контроль ситуации в общественном транспорте [4].

Однако при решении таких задач разработчикам систем часто приходится сталкиваться с решением подзадачи локализации людей в кадре видеоряда.

Один из методов, который можно использовать для решения данной подзадачи - метод Виолы–Джонса (Viola-Jones) [3]. Данный метод основан на построении каскада слабых классификаторов, которые при совместном использовании позволяют относительно точно определять присутствует ли в данной области изображения нужный объект или нет. Работа слабых классификаторов основана на признаках Хаара.

Для данного метода в библиотеке OpenCV существуют уже обученные каскады для поиска различных частей человеческого тела – лица, верхних и нижних частей тела, всего тела целиком. В OpenCV данные каскады носят названия – `frontalface`, `upperbody`, `lowerbody`, `fullbody` [5].

Вопрос области применения и эффективности метода Виолы–Джонса по-прежнему остается открытым и активно изучается в настоящее время [6].

С целью повышения точности локализации объектов учеными Далалом (Navneet Dalal) и Триггсом (Bill Triggs) было предложено в качестве признакового пространства использовать гистограммы направленных градиентов, которые характеризуют отличия в контурах объектов (данный метод носит название Histogram of Oriented Gradients) [7].

При этом для построения классификатора решающего задачу поиска ответа на вопрос присутствует объект в данной области изображения или нет применяется метод опорных векторов (Support Vector Machine). Эффективность работы такой связки «Histogram of Oriented Gradients + Support Vector Machine» (HOG SVM) в настоящее время изучается различными учеными [8]. Однако уже известны, слабые стороны такого подхода.

Со стороны HOG недостатком является необходимость приведения изображения к определенному стандарту с очисткой от шумов.

Со стороны SVM недостатком является отсутствие алгоритма выбора ядра при отсутствии возможности линейного разделения классов [9].

Существуют исследования по преодолению недостатков «Histogram of Oriented Gradients + Support Vector Machine» путем замены метода опорных векторов на деревья принятия решений с заданными параметрами [10].

Существуют исследования направленные на решение задач поиска людей на изображении с применением других подходов. Например, с использованием деформируемых моделей (Elastically Deformable Models). Данный метод основан на анализе ограничений геометрического характера [11].

В другой работе применяется анализ одновременно пространственных и временных признаков, чтобы обнаруживать людей на изображении в относительно сложных ситуациях [12].

Общей проблемой для алгоритма HOG SVM и его модификаций является гиперчувствительность к помехам и шумам на изображениях. Это приводит к тому, что при построении классификатора (определяющего наличие или отсутствие искомого объекта на изображении) он начинает ориентироваться на шумы, а не на действительно значимые признаки [13]. Что в конечном итоге приводит к снижению точности работы [14]. При этом необходимо учитывать, что шумы присутствуют на всех цифровых изображениях и связаны они, во-первых, с условиями съемки, во-вторых с техническими характеристиками оборудования для съемки, в-третьих с особенностями формата (сжатого) хранения изображения.

Существуют исследования по применению нейронных сетей в задачах распознавания образов. Так в одной из статей приводятся сравнения различных методов при решении задачи распознавания текста [15-18].

При распознавании более сложных объектов (машин, пешеходов, детей) применяются нейронные сети с большим количеством вычислительных слоев нейронов. Такие сети носят название «глубокие» (Deep Neural Network). Наиболее эффективные с точки зрения локализации людей на изображении являются сверточные нейронные сети. Так компанией NVIDIA такой тип нейронных сетей был успешно применен при реализации системы поддержки водителей, обеспечивающих локализацию автомобилей в транспортном потоке, а также пешеходов и дорожных знаков с помощью камеры. Однако недостатками сверточных сетей является долгое время обучения и необходимость составления большой обучающей выборки [19-20].

Существуют и другие типы нейронных сетей, которые с точки зрения машинного обучения относятся к «обучению без учителя». Благодаря используемому в них математическому аппарату, они позволяют обнаруживать закономерности в параметрах анализируемых объектов. Это позволяет в задачах анализа изображений автоматически делить образы на множество классов и отсеивать зашумленные образы [21-23].

К таким самообучающимся нейронным сетям можно отнести модель нейронного газа (Growing Neural Gas, GNG), предложенная Мартинесом (Т. Martinetz) в своей работе *Neural Gas Network learns topologies*. Достоинством предложенного подхода является малое время обучения [24].

## 1.2 Локализация человека как задача бинарной классификации

Рассмотрим проблему локализации человека как задачу бинарной классификации.

Предположим, что необходимо анализировать поток  $S$ , который состоит из последовательности отдельных изображений (кадров). Под кадром мы будем понимать анализируемое изображение. Каждое изображение можно представить в виде двухмерной матрицей  $I_t$ , где  $t$  – индекс кадра. Размер матрицы будет совпадать с размером изображения ( $width \times height$ ). Каждый элемент матрицы, в данном случае, будет равен положительному целому числу. Если изображение представлено в оттенках серого, то это значение будет лежать в диапазоне от 0 до 255. Если же изображение представлено в трёхкомпонентной цветовой модели RGB, то это значение будет лежать в диапазоне от 0 до  $255^3$ .

Для каждого изображения требуется определить, есть ли на нем искомый объект, и в какой части кадра он присутствует. Расположение объекта можно описать прямоугольником:

$$\text{rectangle}_\eta = \{x, y, w, h\} \quad (1)$$

где  $x$  – координата по оси  $x$  центра прямоугольника,  $y$  – координата по оси  $y$  центра прямоугольника,  $w$  – ширина прямоугольника,  $h$  – высота прямоугольника,  $\eta$  – номер найденного объекта.

Задача усложнена тем, что искомым объектам на изображении может быть несколько, а может и не быть вовсе.

Осуществление поиска объекта обычно выполняется техникой сканирующего окна, которое на каждой итерации вычисления смещается по изображению с фиксированным шагом (один или больше пикселей). После

прохода по всему изображению сканирующего окна перемещается в начало кадра, его размер увеличивается и начинается новый проход по изображению. Обозначим окно как  $C_r$ , где  $r$  – номер итерации сканирования. При этом размер сканирующего окна можно выразить через его высоту  $h_c$  и, соответственно, ширину  $w_c$ .

Для каждого значения  $r$  необходимо определить присутствует ли объект (человек) в данном окне. Фактически это означает, что необходимо классифицировать  $C_r$  при всех значениях  $r$  к одному из двух классов: «человек присутствует», «человек отсутствует». Если  $C_r$  относится к первому классу, то вектор  $rectangle_r$  пополняется объектом  $rectangle = \{x_r, y_r, w_r, h_r\}$ , где  $x_r$  – координата по оси  $x$  центра окна,  $y_r$  – координата по оси  $y$  центра окна,  $w_r$  – ширина сканирующего окна,  $h_r$  – высота сканирующего окна.

Таким образом, ясно, что каждое положение сканирующего окна необходимо классифицировать к одному из классов: {«нет людей»; «есть человек»} =  $-1; +1$ .

Задачу бинарной классификации можно представить в следующем виде.

Пусть дано множество  $\mathbf{X}$ , в котором содержится описание некоторых объектов  $o$ . При этом во множестве  $\mathbf{Y}$  пусть хранятся множества классов (в нашем случае их два). Задачей построения классификатора является нахождение такой решающей функции  $\mathbf{F}$ . При этом  $\mathbf{F}$  должен являться отображением множества  $\mathbf{X}$  в множество  $\mathbf{Y}$  ( $\mathbf{F}: \mathbf{X} \rightarrow \mathbf{Y}$ ).

Продолжая вышесказанное, признак  $f$  объекта  $o$  выражается как  $f: o \rightarrow \mathbf{D}_f$ , где  $\mathbf{D}_f$  – множество допустимых значений данного признака  $f$ . Когда известен набор признаков  $f_1, \dots, f_m$  для объекта  $o$ , то набор признаков  $x$  объекта  $o \in \mathbf{X}$  определяется следующим образом:  $x = (f_1(o), \dots, f_m(o))$ .

От найденного классификатора  $F$  требуется обеспечение возможности отнесение произвольного объекта  $o \in X$  к одному из классов  $Y$ . Таким образом, задачу локализация человека можно представить в виде следующей схемы (рисунок 1.1).

Наилучшим будет являться тот классификатор, который обеспечит минимальную вероятность ошибки  $P(x)$  для всех возможных допустимых значений  $x$ . Такой критерий можно записать как:  $P(x) \rightarrow \min_{x \in X}$ .

При расчете вероятности ошибки  $P(x)$  учитываются ошибки двух видов: ошибки первого рода (ложные срабатывания), ошибки второго рода (отсутствие срабатывания при наличии искомого объекта).

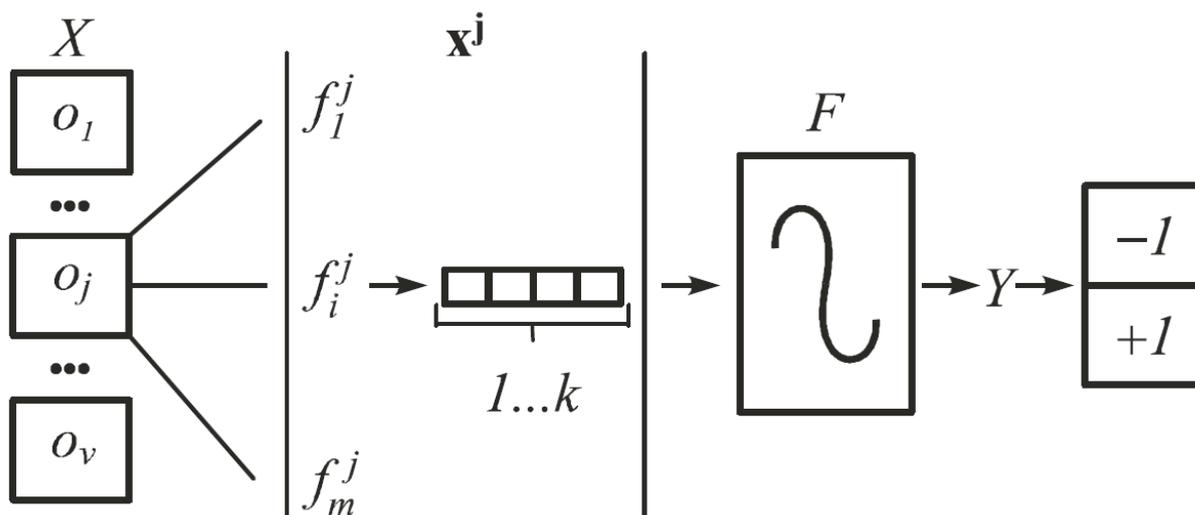


Рисунок 1.1 – Локализация человека как задача бинарной классификации

Для того чтобы оценить качество работы классификатора проводятся натурные эксперименты, по результатам которых строится ROC-кривая (receiver operating characteristic). Одним из наиболее важных показателей ROC-кривой является площадь под графиком, которая обозначается как AUC (area under ROC curve).

Для построения классификатора  $F$  используется набор данных, представленных в виде обучающей выборки  $D$ :  $D = \{(x_1, y_1), \dots, (x_v, y_v)\}$ , где  $y_j \in Y = \{-1, +1\}, j = 1 \dots v$ .

### 1.3 Комбинирование технологий в задачах локализации человека

Как видно из анализа литературных источников при решении задачи поиска людей, на изображении часто используются технологии:

- самообучаемые искусственные нейронных сетей, так как их математический аппарат наиболее устойчив к шумам на изображении;
- признаковое пространства Histogram of Oriented Gradients.

Наиболее современные исследования в области локализации людей на изображении направлены на комбинировании данных технологий с целью улучшения конечного результата.

Например, существуют исследования по разработке алгоритма на основе нейро-нечеткой модели растущего нейронного газа для обучения с подкреплением [25]. В данной работе используется двухпроходное обучение с последующей перемаркировкой классов, осуществляемой блоком нечеткого вывода. А в качестве признакового пространства применяется метод Histogram of Oriented Gradients.

Алгоритм перевода изображения в новое пространство Histogram of Oriented Gradients осуществляется так:

$$\Phi: o_j \rightarrow x^j \quad (2)$$

Пример преобразования исходного изображения в признаковое пространство Histogram of Oriented Gradients представлен на рисунке 1.2.

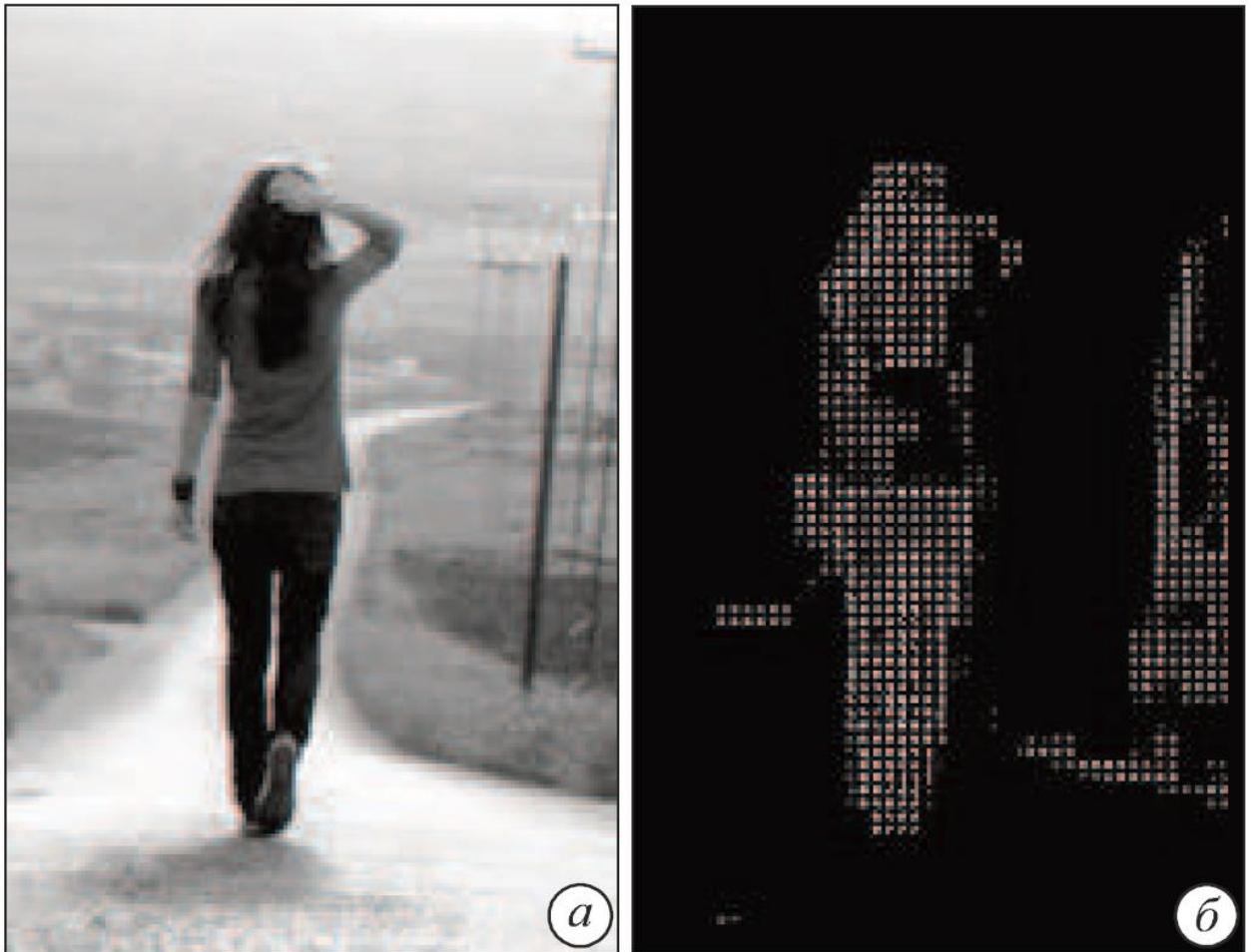


Рисунок 1.2 – Пример преобразования: а – оригинальный кадр, б – визуализация Histogram of Oriented Gradients

Использование данного метода обладает следующими достоинствами:

- высокая устойчивость к размытию вызванному движением объекта (что достигается путем сильного разбиения пространства малыми ядрами);
- хорошая устойчивость к геометрическим преобразованиям кадра.

Решение задачи поиска человека на изображении в рассматриваемом исследовании [25] решается следующим образом:

- на первом этапе исходное изображение переводится в оттенки серого и затем обрабатывается с помощью метода Histogram of Oriented Gradients;

- затем осуществляется классификация фрагмента изображения, выделенного сканирующим окном, алгоритмом растущего нейронного газа;
- после этого осуществляется переход к бинарной классификации путем маркировки с использованием нейронечеткой модели растущего нейронного газа;
- затем выполняется нечеткая перемаркировка и построение тепловой карты, что обеспечивает значительно уменьшение ошибки второго рода.

Предложенный авторами алгоритм проиллюстрирован на рисунке 1.3, где  $F_{GNG}$  – нейронный газ,  $l$  – обозначение номера класса,  $\mathbf{D}$  – обучающая выборка.

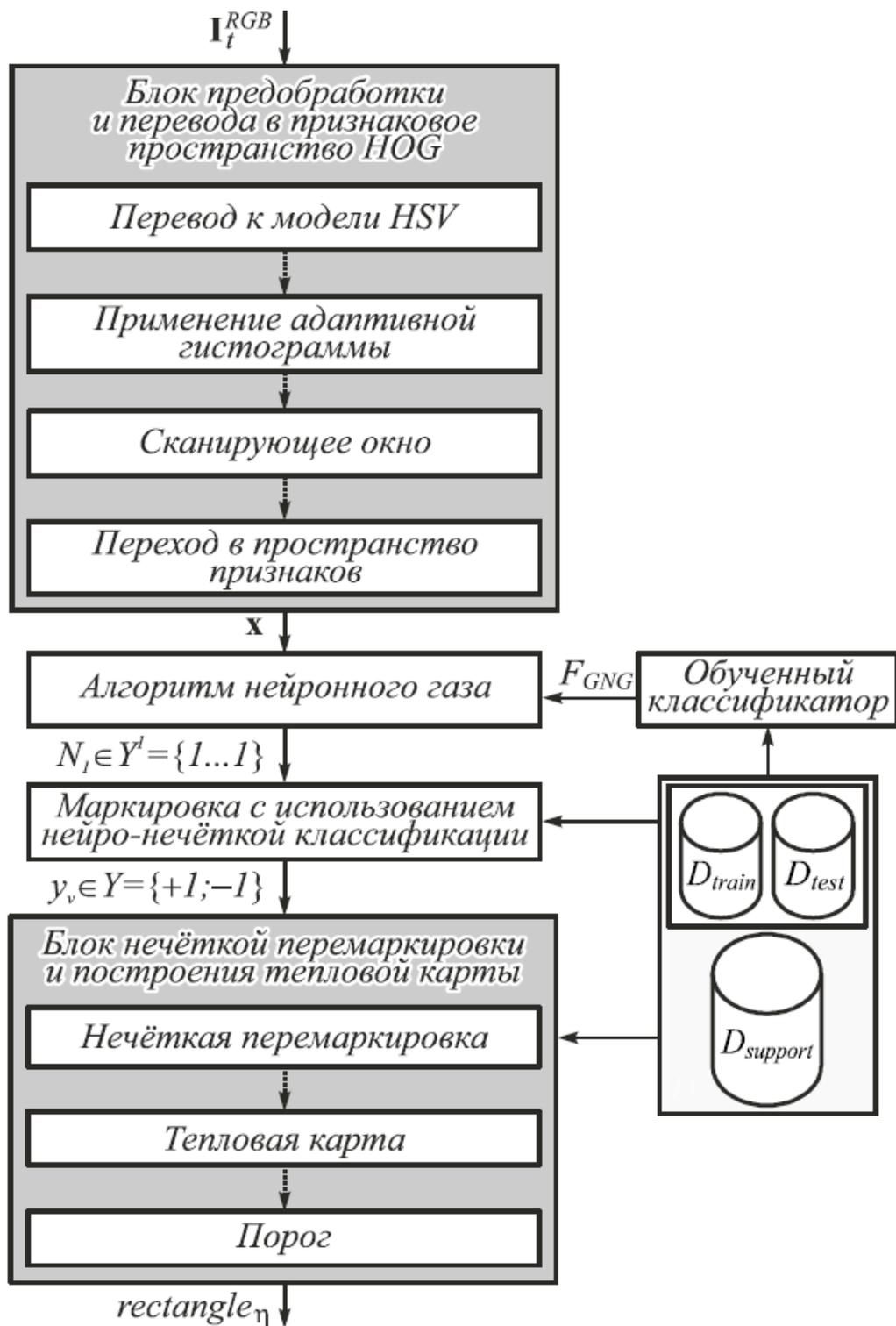


Рисунок 1.3 – Решение задачи локализации человека [25]

Авторы описывают необходимость для каждого кадра видеопотока  $I$ , проводить изменение стандартной цветовой модели RGB.

В работе указывается, что цветовая модель HSV (которая состоит из таких компонентов, как H – тон, S – насыщенность, V – значение) обеспечивает более высокую точность работы алгоритмов в задачах для локализации людей.

На рисунке 1.4 проиллюстрировано одно и тоже изображения, но в разных цветовых моделях.

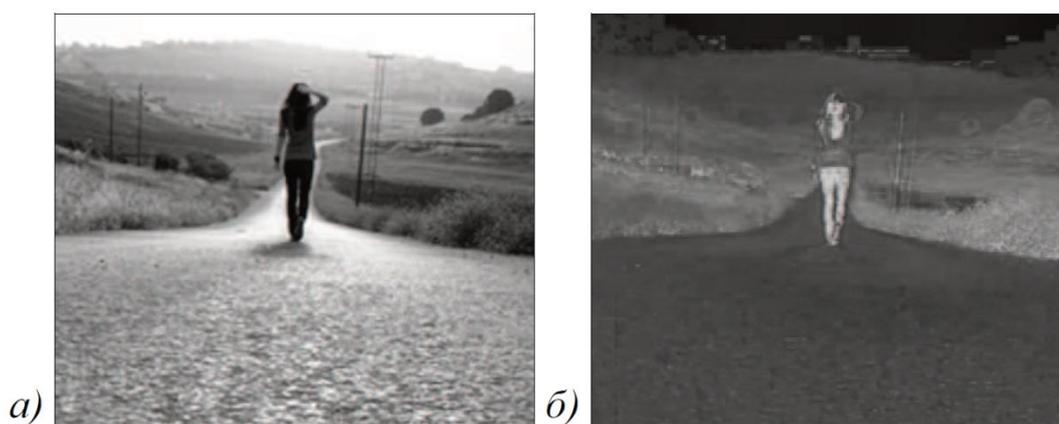


Рисунок 1.4 – Изображение в различных цветовых моделях: а) RGB, б) HSV

[25]

Авторы данной работы сообщают о необходимости повышения контрастности изображения путем вычисления гистограмм, относящихся к разным частям изображения. Такой подход носит название адаптивной эквализации.

Программная реализация такого подхода есть в открытой библиотеке машинного зрения OpenCV. Для этого в библиотеке реализована функция *adapthisted*.

Как только исходное изображение будет приведено к цветовой модели HSV и обработано методом адаптивной эквализации, становится возможным получение с помощью сканирующего окна набора объектов  $o$  классификации (прямоугольных фрагментов изображения).

Каждый объект  $o$ , представленный вектором признакового описания  $x$ . Пусть обучающая выборка, заданная конечным множеством  $D$ . Для

настройки алгоритма на поиск людей необходимо построить оптимальный классификатор  $\mathbf{F}$  (критерий оптимальности - наименьшая ошибка распознавания при всех значениях  $x$ ).

Для построения классификатора авторами работы предложено использовать алгоритм машинного обучения - растущий нейронный газ.

Алгоритм обучения нейронного газа можно представить в виде следующих последовательно выполняемых шагов:

1. Установка первоначальных параметров нейронного газа:

- $age_{\max}$  – максимально допустимое значение возраста ребра;
- $\varepsilon_{\xi}, \varepsilon_{\zeta}$  – коэффициенты обучения нейрона-победителя и связанных с ним соседних нейронов;
- $\lambda$  – количество итераций для вставки нового нейрона;
- $\tau$  – количество итераций, как критерий остановки обучения;
- $\alpha, \beta$  – коэффициенты для адаптации ошибок.

2. Задаются два начальных узла  $u$  и  $v$  в случайных точках  $\omega_u$  и  $\omega_v$ ,  $u$  и  $v$  соединяются ребром, возраст  $age$  которого равен нулю. Ошибка  $err$  в узлах  $u$  и  $v$  принимается  $= 0$ .

3. Выбирается следующий объект  $o$ , описанный вектором  $x$  из обучающей выборки  $\mathbf{D}$ .

4. Определяются узлы  $\psi_1$  и  $\psi_2$  – первый и второй по критерию близости к  $o$ . Мерой близости является расстояние Евклида:

$$E = \|\omega_{\psi} - x\|, \psi = \psi_1, \psi_2$$

5. Увеличивается на единицу возраст дуг, исходящих от  $\psi_1$ .

6. Счетчик ошибки нейрона  $\psi_1$  увеличивается на величину квадрата расстояния до  $o$ :  $E(\psi_1) = \|\omega_{\psi_1} - x\|^2$ .

7.  $\psi_1$  и узлы, соединённые с ним, смещаются по направлению к  $o$  на расстояния  $\Delta\omega_{\psi_1} = \varepsilon_{\xi}(x - \omega_{\psi_1})$ ,  $\Delta\omega_{\psi_n} = \varepsilon_{\xi}(x - \omega_{\psi_n})$ , при  $0 < \varepsilon_{\xi} \ll 1$ ,  $0 < \varepsilon_{\xi} \ll \varepsilon_{\xi}$

8. Если  $\psi_1$  и  $\psi_2$  соединены ребром, то возраст ребра обнуляется; иначе между  $\psi_1$  и  $\psi_2$  создается новое ребро с возрастом = 0.

9. Если имеются ребра, возраст которых превышает предельное значение  $age_{\max}$ , то они удаляются. Если имеются узлы, которые не связаны с другими узлами ребрами, то они убираются.

10. Если номер текущей итерации кратен величине  $\lambda$ , осуществляется вставка нового узла в точке  $\omega_{\rho} = 0,5(\omega_{\mathcal{G}} - \omega_k)$ , где  $\mathcal{G}$  – узел с наибольшей накопленной ошибкой;  $k$  – узел, соединенный ребром  $\mathcal{G}$  с наибольшей накопленной ошибкой. Ребро между  $k$  и  $\mathcal{G}$  удаляется, вместо него добавляется новый узел –  $\rho$  и добавляются рёбра между  $k$  и  $\rho$ , а также между  $\rho$  и  $\mathcal{G}$ . Ошибка  $err$  в узлах  $k$  и  $\mathcal{G}$  уменьшается умножением на константу  $\alpha < 1$ ; значение ошибки нового нейрона  $\rho$  инициализируется равным значению ошибки  $\mathcal{G}$ .

11. Ошибка  $err$  для всех узлов уменьшается путем умножения на  $\beta$  при  $\beta < 1$ .

12. Проверка условия остановки алгоритма. Если оно не выполнено все шаги, начиная со второго повторяются. Условием остановки в данном случае является выполнение фиксированного количества итераций обучения  $\tau$ . Примеры структур получаемые, с помощью алгоритма растущего нейронного газа (рисунке 1.5).

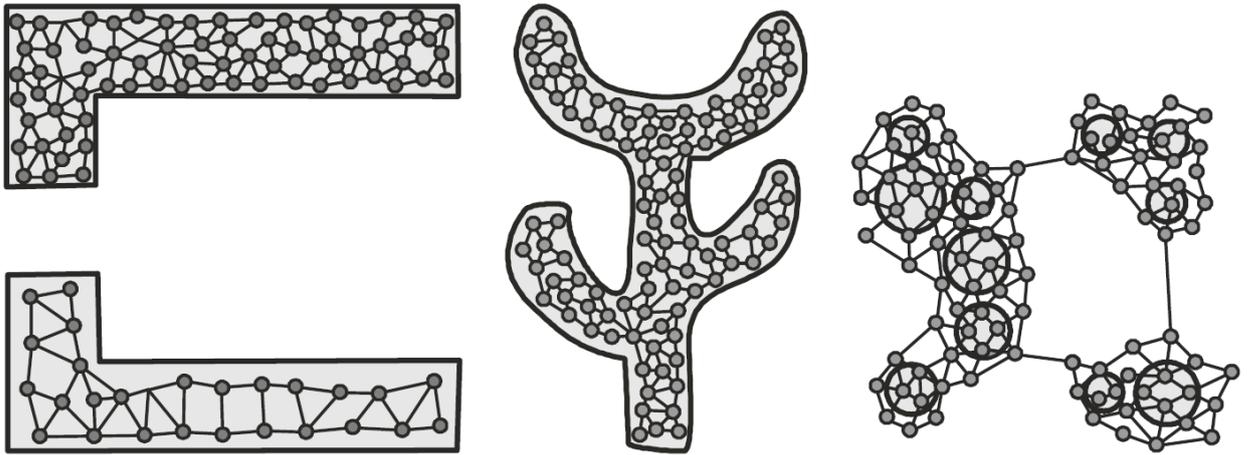


Рисунок 1.5 – Пример структуры *GNG*

Обучающая выборка, на которой производилась настройка и тестирование комбинированной системы локализации людей на изображении, бралась авторами из открытых источников. Обучающая выборка состоит из негативных изображений, которые не содержат изображения людей и позитивных, на которых присутствуют люди.

При обучении системы для получения лучших результатов авторы рекомендуют чередовать подачу негативных и позитивных изображений на вход системы.

Для проверки точности работы настроенной системы применялась тестирующая выборка, которая была получена путем разбиения первоначальной выборки на две равные части – обучающая и тестирующая.

Таким образом, обучающая выборка  $D_{train}$  состоит из 1500 позитивных и столько же негативных изображений. Тестирующая выборка  $D_{test}$  состоит из такого же числа позитивных и негативных изображений. Примеры изображений из данных выбор представлены на рисунке 1.6.

Размер всех изображений нормировался к значениям 128 пикселей по высоте и 64 по ширине.

Результатом обучения растущего нейронного газа является разбиение обучающего множества на  $N_i \geq 2$  классов. Как говорят авторы работы – с помощью нейронного газа фактически решается задача кластеризации.

Примечательно, что количество кластеров нейронный газ определяет самостоятельно.

При получении на вход GNG нового объекта  $o$ , отсутствующего в обучающей выборке  $D_{train}$ , алгоритм будет относить его к одному из известных классов. Таким образом, решается задача многоклассовой классификации, нам же необходимо решить задачу бинарной классификации, т.е. отнесения объекта  $o$  к одному из двух классов: «есть человек»/«люди отсутствуют».



Рисунок 1.6 – Некоторые примеры из обучающей выборки: а) негативные; б) позитивные

Так как в результате самоорганизации нейронного газа может получиться несколько кластеров, то получившийся классификатор будет мульти-классовым. На практике же требуется получения бинарного

классификатора - «есть человек»/«люди отсутствуют». Для решения данной проблемы авторы работы добавляют в систему блок нечеткого вывода, который осуществляет маркировку классов из большого количества меток классов к двум меткам класса.

Авторы комбинированного подхода объединяют нейронные сети и нечеткую логику в одну систему. Ими предложена структура нейро-нечеткой модели (GNG-FIS), схема функционирования которой представлена на рисунке 1.7. В предложенном алгоритме входные данные обработанные методом HOG подаются на нейронную сеть GNG и затем обрабатываются при помощи нечеткой системы FIS.

Получается, что в данной работе нейронная сеть действует как препроцессор нечетких выходов, основным назначением которого является приведения классификатора к двум классам.

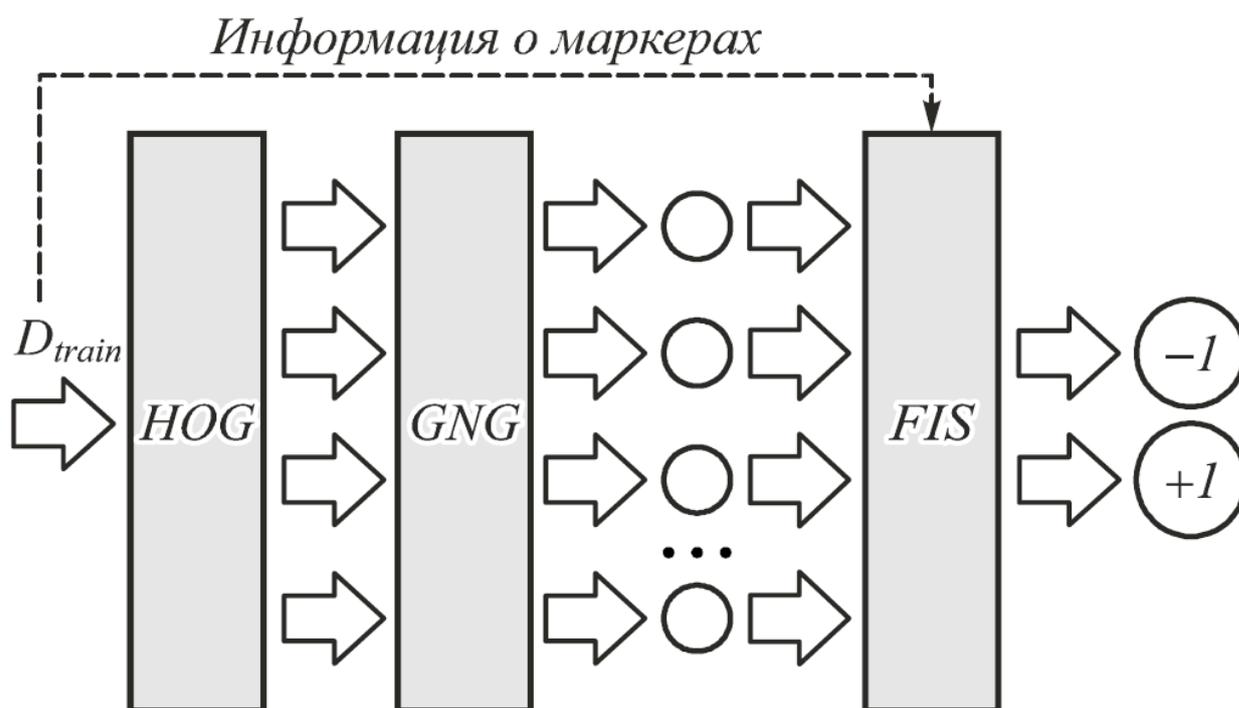


Рисунок 1.7 – Схема функционирования алгоритма GNG-FIS [25]

Маркировка выполняется следующим образом: сначала используется классический алгоритм, затем используются нечеткие правила.

На первом этапе маркировки выполняются следующие шаги:

- Создается множество объектов  $R$ , размер которого равен количеству классов  $N_l$ . Каждый объект  $R$  содержит набор параметров, описывающий состав кластера (класса):

- 1)  $n_{pos}$  – количество позитивных образов;

- 2)  $n_{neg}$  – количество негативных образов.

- Объект  $o$  из обучающего множества  $D_{train}$  подвергается классификации с помощью обученной сети GNG.

- Выбирается объект  $R_l \in R$ , где  $l$  – полученный номер класса из сети GNG. Используя заранее установленную маркировку из  $D_{train}$ , требуется увеличить один из параметров  $R_l$  следующим образом:

if образ=негативный then  $n_{neg} := n_{neg} + 1$ ;

else  $n_{pos} := n_{pos} + 1$ .

- По окончании первого этапа маркировки следующие классы маркируются как негативные, если:

- количество  $n_{neg}$  равно 1 ( $n_{pos} = 1$ );

- количество  $n_{neg}$  и  $n_{pos}$  равно ( $n_{neg} = n_{pos}$ );

- количество  $n_{neg}$  больше, чем  $n_{pos}$  ( $n_{neg} > n_{pos}$ ).

Затем на втором этапе для всех кластеров, соответствующих условию  $0 < n_{neg} < n_{pos}$ , вычисляется:

доля позитивных образов:

$$h_l(n_{pos}) = [n_{pos} / (n_{pos} + n_{neg})] \cdot 100\% \quad (3)$$

доля негативных образов:

$$h_l(n_{neg}) = [n_{neg} / (n_{pos} + n_{neg})] \cdot 100\% \quad (4)$$

среднее евклидово расстояние для позитивных образов до центра кластера:

$$m_l(n_{pos}) = (\sum E_{pos}^l) / n_{pos} \quad (5)$$

аналогичные вычисления для негативных образов до центра кластера:

$$m_l(n_{neg}) = (\sum E_{neg}^l) / n_{neg} \quad (6)$$

Переход к нечеткой модели доверия к кластерам выглядит следующим образом.

Пусть имеются входные переменные  $h_l(n_{pos})$ ,  $h_l(n_{neg})$ ,  $m_l(n_{pos})$ ,  $m_l(n_{neg})$  и выход  $q$ , характеризующий степень доверия к кластерам. Вводятся следующие терм-множества:

- $T_h(h_l) =$  маленькая, средняя, большая для доли позитивных/негативных образов;
- $T_m(m_l) =$  маленькая, средняя, большая для среднего евклидового расстояния образов до центра кластеров;
- $T_w =$  положительный, сомнительный, отрицательный для выхода  $q$ .

Нечеткий вывод осуществляется с использованием алгоритма Мамдани. Нечеткая база знаний состоит из 16 правил. Функции принадлежности реализована с использованием *gaussmf* симметричной гауссовской кривой.

Нечеткий вывод можно представить в виде следующего выражения

$$\begin{aligned} & \text{if } h_l(n_{pos}) = T_h(h_l) \& m_l(n_{pos}) = T_m(m_l) \& h_l(n_{neg}) = \\ & T_h(h_l) \& m_l(n_{neg}) = T_m(m_l) \\ & \text{then } q_l = T_q \end{aligned} \quad (7)$$

В результате нечеткой маркировки является то, что "положительные" и "сомнительные" кластеры маркируются как "+1", а "отрицательные" – как "–1". Т.е. результат работы алгоритма GNG-FIS – бинарная классификация объектов  $o$  к одному из двух классов  $y_i \in Y = -1; +1$ .

## **1.4 Сравнение результатов работы методов HOG SVM, Viola-Jones и GNG-FIS**

Результаты тестирования методов HOG SVM, Viola-Jones и GNG-FIS в задачи локализации людей представлены в таблице 1.1. Примеры работы данных методов представлены в таблице 1.2.

Таблица 1.1 – Результаты тестирования различных методов на тестовой выборке  $D_{test}$  состоящей из 2100 изображений

Название метода	Количество не распознанных (ошибка 2-го рода)	Количество ложных Распознаваний (ошибка 1-го рода)	Общая ошибка	Время обработки	Время обучения
HOG SVM	358	42	19,05%	40 мс	3-5 часов
Viola-Jones	801	45	40,29%	10 мс	180-300 часов
GNG-FIS	76	112	8,95%	20 мс	5 часов
GNG-FIS с использованием двухпроходной перемаркировки	120	52	8,19%	30 мс	6 часов

Таблица 1.2 – Примеры работы методов



Как видно из таблицы 1.1 метод Viola-Jones имеет наихудшую точность локализации людей на изображении. Причем наибольший вклад в общую ошибку вносят ошибки второго рода.

Однако в тестировании данного метода использовался каскад слабых классификаторов `fullbody`, идущий в комплекте вместе с библиотекой `OpenCV`.

Известно, что метод Viola-Jones отлично справляется с локализацией на изображении статических объектов. Человек к таким объектам не относится, так в процессе движения он может вращать головой, двигать руками и ногами и т.д. Но наиболее статичным элементом человеческого тела является лицо.

Поэтому в исследовании, проводимом в рамках ВКР выносятся гипотеза о возможности повышения точности работы системы (локализации человека на изображении) за счет использования в системе дополнительного каскада нацеленного на поиск относительно статичного элемента человеческого тела (например, лица).

Цель работы – повышение точности работы системы локализации человека на изображении с использованием метода Viola-Jones за счет совместного использования нескольких каскадов классификаторов.

Для достижения поставленной цели в ВКР необходимо решить следующие задачи:

1. Провести анализ состояния вопроса.
2. Разработать алгоритм одновременного использования нескольких каскадов слабых классификаторов.
3. Спроектировать, разработать и протестировать программную реализацию предложенных решений.

## 2 РАЗРАБОТКА АЛГОРИТМА ПОИСКА ЛЮДЕЙ НА ИЗОБРАЖЕНИИ

### 2.1 Метод Viola-Jones

Обобщенная схема распознавания в алгоритме Виолы-Джонса показана на рисунке ниже.

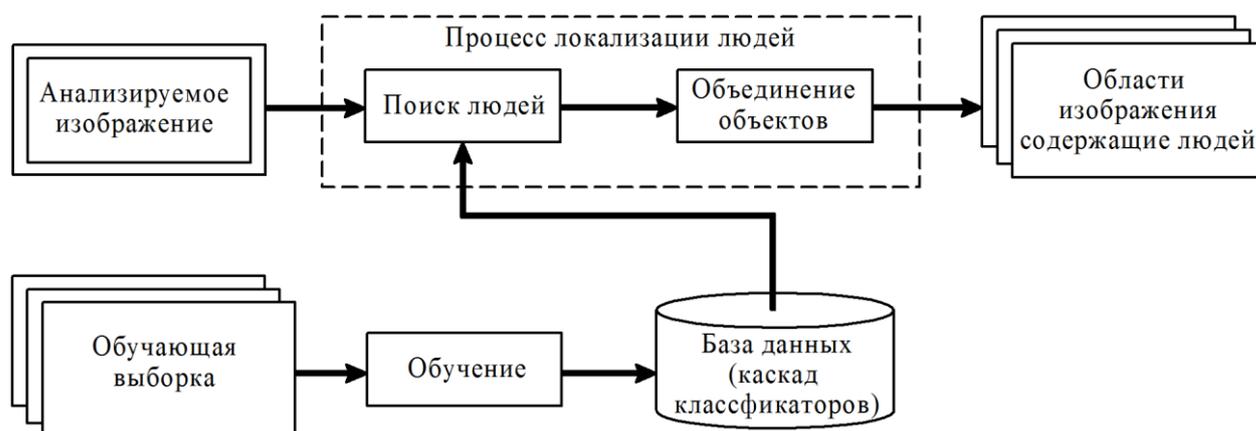


Рисунок 2.1 – Схема распознавания образов

Обобщенная схема алгоритма выглядит следующим образом: перед началом распознавания алгоритм обучения на основе тестовых изображений обучает базу данных, состоящую из признаков, их паритета и границы. Подробнее о паритете, признаке и границе будет рассказано в следующих пунктах. Далее алгоритм распознавания ищет объекты на разных масштабах изображения, используя созданную базу данных. Алгоритм Виолы-Джонса на выходе дает всё множество найденных необъединенных объектов на разных масштабах. Следующая задача – принять решение о том, какие из найденных объектов действительно присутствуют в кадре, а какие – дубли.

Рассмотрим признаки класса. В качестве признаков для алгоритма распознавания авторами были предложены признаки Хаара, на основе вейвлетов Хаара. Они были предложены венгерским математиком Альфредом Хааром в 1909 году.

В задаче распознавания лиц, общее наблюдение, что среди всех лиц области глаз темнее области щек. Рассмотрим маски, состоящие из светлых и темных областей.

Каждая маска характеризуется размером светлых и темных областей, пропорциями, а также минимальным размером. Совместно с другими наблюдениями были предложены следующие признаки Хаара, как пространство признаков в задаче распознавания для класса лиц.

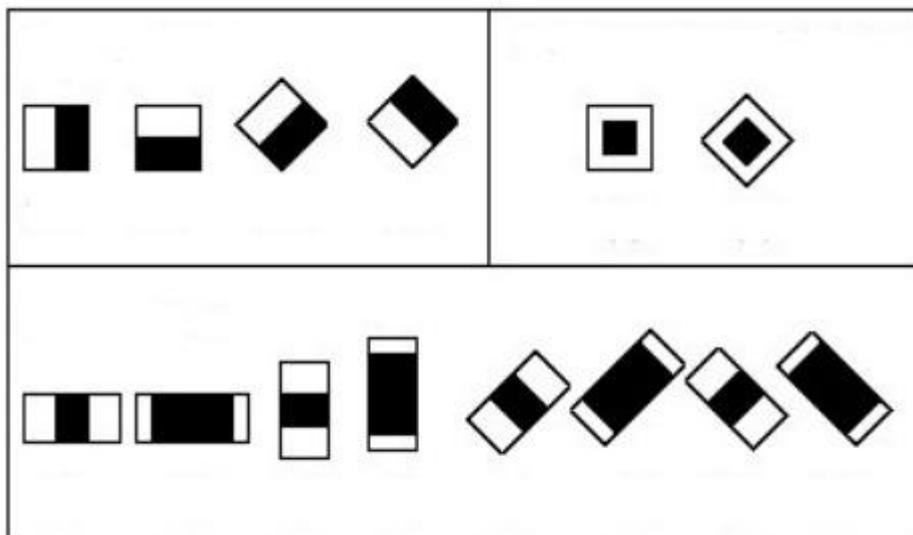


Рисунок 2.2 – Вариации признаков Хаара

Общий признак Хаара для распознавания лиц представляет набор двух смежных прямоугольников, которые лежат выше глаз и на щеках. Значение признака вычисляется как (8):

$$F = X - Y, \quad (8)$$

где  $X$  – сумма значений яркостей пикселей, соответствующих области признака, а  $Y$  – соответственно, черной области.

Видно, что если считать суммы значений интенсивностей для каждого признака это потребует значительных вычислительных ресурсов. Виолой и Джонсом было предложено использовать *интегральное представление изображения*, подробнее о нем будет далее. Такое представление стало

довольно удобным способом вычисления признаков и применяется также и в других алгоритмах компьютерного зрения, например *SURF*.

Обобщенная схема алгоритма обучения выглядит следующим образом. Имеется тестовая выборка изображений. Размер тестовой выборки приблизительно 10 000 изображений. На рисунке показан пример обучающих изображений лиц. Алгоритм обучения работает с изображениями в оттенках серого.



Рисунок 2.3 – Обучающая выборка из изображений лиц

При размере тестового изображения 24 на 24 пикселя количество конфигураций одного признака около 40 000 (зависит от минимального размера маски). Современная реализация алгоритма использует порядка 20 масок. Для каждой маски, каждой конфигурации тренируется такой слабый классификатор, который дает наименьшую ошибку на всей тренировочной базе. Он добавляется в базу данных. Таким образом алгоритм обучается. И на выходе алгоритма получается база данных из  $T$  слабых классификаторов. Обобщенная схема алгоритма обучения показана на рисунке.

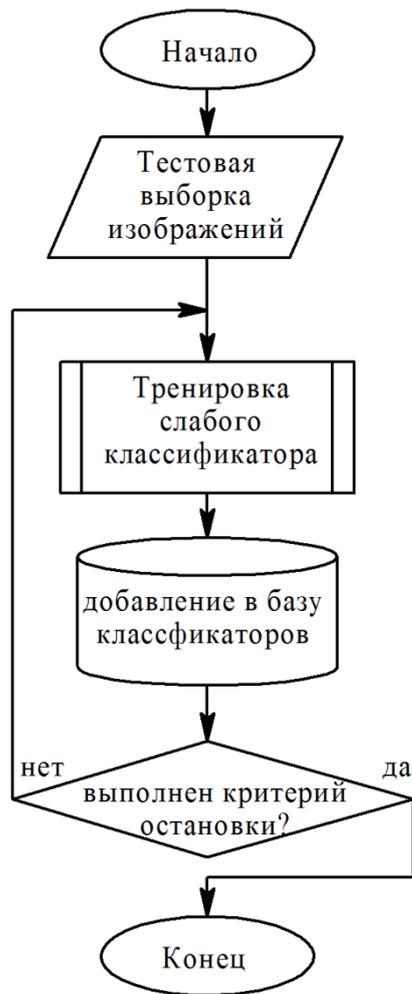


Рисунок 2.4 – Блок схема алгоритма

Обучение алгоритма Виола-Джонса – это обучение алгоритма с учителем. Для него возможно такая проблема как переобучение. Показано, что AdaBoost может использоваться для различных проблем, в том числе к теории игр, прогнозировании. В данной работе условие остановки является достижение заранее заданного количества слабых классификаторов в базе.

Для алгоритма необходимо заранее подготовить тестовую выборку из  $l$  изображений, содержащих искомый объект и  $m$  не содержащих. Тогда количество всех тестовых изображений будет (9):

$$\begin{aligned}
 n &= l + m \\
 \mathbf{X} &= \{x_1, x_2, \dots, x_n\}
 \end{aligned}
 \tag{9}$$

где  $X$  – множество всех тестовых изображений, где для каждого заранее известно присутствует ли искомый объект или нет и отражено во множестве  $Y$  (10).

$$Y = \{x_1, x_2, \dots, x_n\} \quad (10)$$

где:

$$y_i = \begin{cases} 1, \text{если объект присутствует на изображении } x_i \\ 0, \text{иначе} \end{cases} \quad (11)$$

Под признаком  $j$  будем понимать структуру вида (12):

$$j = \{\text{маска, положение, размер}\} \quad (12)$$

Тогда откликом признака будет  $f_j(x)$ , который вычисляется как разность интенсивностей пикселей в светлой и темной областях. Слабый классификатор имеет вид:

$$h_j = \begin{cases} 1, \text{если } p_j f_j(x) < p_j \theta_j \\ 0, \text{иначе} \end{cases} \quad (13)$$

Задача слабого классификатора – угадывать присутствие объекта в больше чем 50% случаев. Используя процедуру обучения AdaBoost создается сильный классификатор состоящий из  $T$  слабых классификаторов и имеющий вид:

$$H(x) = \begin{cases} 1, \text{если } \sum_{t=1}^T a_t h_j^{(t)}(x) \geq \frac{1}{2} \sum_{t=1}^T a_t \\ 0, \text{иначе} \end{cases} \quad (14)$$

Целевая функция обучения имеет следующий вид:

$$f_{\text{ц}}(h_j^{(1)}, h_j^{(2)}, \dots, h_j^{(T)}, a_1, a_2, \dots, a_T) = \arg \min \sum_{i=1}^n |H(x_i, T, h_j^{(1)}, h_j^{(2)}, \dots, h_j^{(T)}, a_1, a_2, \dots, a_T)|$$

Рассмотрим понятие интегрального представления изображения. Интегральное представление можно представить в виде матрицы, размеры которой совпадают с размерами исходного изображения  $I$ , где каждый элемент рассчитывается так (15):

$$I(x, y) = \sum_{i=0, j=0}^{i \leq x, j \leq y} I(r, c) \quad (15)$$

где  $I(r, c)$  — яркость пиксела рассматриваемого кадра.

Значения матрицы  $I(x, y)$  в соответствии с формулой, представленной выше - сумма яркостей в прямоугольной области заданной точками  $(0,0)$  и  $(x, y)$ . Расчет такой матрицы занимает линейное время. Для того, чтобы вычислить сумму прямоугольной области в интегральном представлении изображения требуется всего 4 операции обращения к массиву и 3 арифметические операции. Это позволяет быстро рассчитывать признаки Хаара для изображения в обучении и распознавании.

Например, рассмотрим прямоугольник выделенный штриховкой.

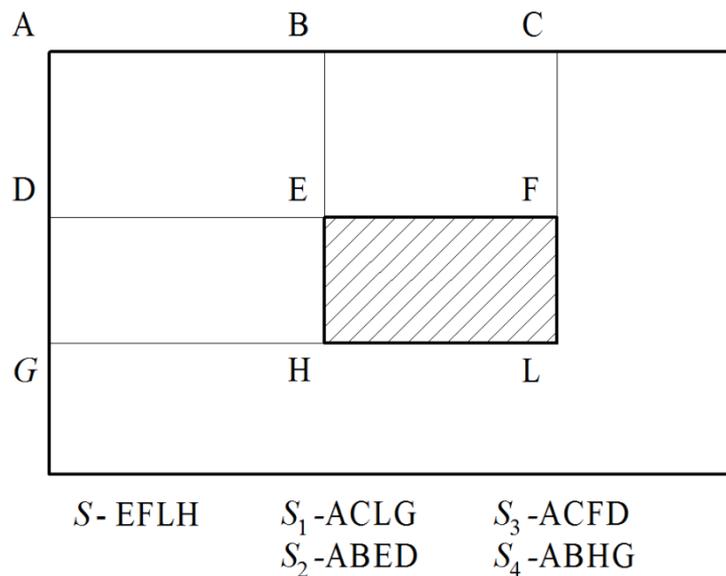


Рисунок 2.5 – Расчет суммы яркостей пикселей на интегральном представлении изображения

Сумму яркостей пикселей в  $S$  через интегральное представлении изображения можно выразить через  $S_1, S_2, S_3, S_4$  так (16):

$$I(S) = I(S_1) + I(S_2) - I(S_3) - I(S_4) \quad (16)$$

Рассмотрим алгоритм обучения. Перед началом обучения инициализируются веса  $w_{q,i}$ , где  $q$  – номер итерации,  $i$  – номер изображения (17).

$$w_{1,i} = \begin{cases} \frac{1}{2l}, & \text{если } y_i = 1 \\ \frac{1}{2m}, & \text{если } y_i = 0 \end{cases} \quad (17)$$

После процедуры обучения получится  $T$  слабых классификаторов и  $T$  значений,  $a, p_j = \{1, -1\}, \theta_j = [\theta_{\min}, \theta_{\max}]$  (рисунок 2.7):

$$\left. \begin{array}{l} t = 1 \dots T \\ w_{t,i} = \frac{w_{t,i}}{\sum_{k=1}^n w_{t,k}} \\ \text{foreach } j \\ e_j = \min_{p, \theta} \sum_{i=1}^n w_{t,i} |h_j(x_i) - y_i| \\ J^{(t)} = \arg \min_j (e_j) \\ \beta_t = \frac{e_j^{(t)}}{1 - e_j^{(t)}} \\ w_{t+1,i} = \begin{cases} w_{t,i} \beta_t, & \text{если } h_j^{(t)}(x_i) = y_i \\ w_{t,i}, & \text{иначе} \end{cases} \\ a_t = \log \frac{1}{\beta_t} \end{array} \right\}$$

Рисунок 2.6 – Алгоритм обучения

На каждой итерации цикла происходит обновление весов так, что их сумма будет равна 1. Далее для всех возможных признаков происходит подбор таких значений  $p, j, \theta$ , что значение ошибки  $e_j$  будет минимально на этой итерации. Полученный признак  $J(t)$  (на шаге  $t$ ) сохраняется в базу слабых классификаторов, обновляются веса и вычисляется коэффициент  $a_t$ .

В предложенном в 2001 году оригинальном алгоритме не была описана процедура получения оптимального признака на каждой итерации. Предполагается использование алгоритма AdaBoost и полный перебор возможных параметров границы и паритета.

Рассмотрим алгоритм локализации объектов на изображении. После обучения на тестовой выборке имеется обученная база знаний из  $T$  слабых классификаторов. Для каждого классификатора известны: признак Хаара, использующийся в этом классификаторе, его положение внутри окна размером  $24 \times 24$  пикселя и значение порога  $E$ .

На вход алгоритму поступает изображение  $I(r,c)$  размером  $w \times h$ , где  $I(r,c)$  – яркостная составляющая изображения. Результатом работы алгоритма служит множество прямоугольников  $R(x,y,w,h)$ , определяющих положение лиц в исходном изображении  $I$ .

Алгоритм сканирует изображение  $I$  на нескольких масштабах, начиная с базовой шкалы: размер окна  $24 \times 24$  пикселя и 11 масштабов. Разработчики метода Виола-Джонса рекомендуют, чтобы масштабы друг от друга отличаются в 1,25 раза. Алгоритм распознавания выглядит следующим образом.

Для каждого значения масштаба  $scale$  проводятся следующие действия:

Для текущего значения  $scale$  масштабируется размер скользящего окна и шаг.

Вычисляется  $T$  значений слабых классификаторов из баз определенной на этапе обучения.

Для того, чтобы классификатор вынес решение о том, что в рассматриваемом окне присутствует искомый шаблон – необходимо чтобы выполнилось условие сильного классификатора:

$$H(x) = \begin{cases} 1, \text{ если } \sum_{t=1}^T a_t h_j^{(t)}(x) \geq \frac{1}{2} \sum_{t=1}^T a_t \\ 0, \text{ иначе} \end{cases} \quad (18)$$

## 2.2 Модификация алгоритма

Для улучшения работы точности алгоритма предлагается одновременно использовать два независимых каскада классификаторов обученных под разные цели.

Первый каскад классификаторов будет локализовать человека путем анализа формы тела во весь рост (каскад fullbody).

Второй каскад классификаторов будет осуществлять поиск лиц на изображении (в анфас и профиль) и если лицо будет найдено, то с учетом его масштаба и положения будет выделяться область расположения всего человека (каскад face).

Результаты независимой работы двух классификаторов будут объединяться для получения более точного решения задачи локализации человека.



Рисунок 2.7 – Модификация метода Viola-Jones с целью уменьшения ошибок первого рода

## 3 ПРОГРАММНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА

### 3.1 Разработка программной части

В ходе выполнения исследований по выпускной квалификационной работе в среде разработки Visual Studio Community на языке программирования C++ была разработана и спроектирована программа для анализа потокового изображения, поступающего с веб-камеры с целью локализации людей.

Функциональные возможности программы:

- получение видеоряда с выбранного источника (например с веб-камеры);
- поиск и выделение людей на изображении с использованием метода Viola-Jones;
- при формировании результатов анализа изображения реализована возможность одновременного использования сразу нескольких каскадов классификаторов;
- реализована возможность подключения каскадов идущих в комплекте с библиотекой OpenCV;
- возможность задания своего цвета рамок для выделения людей для каждого используемого каскада.

Проект разработанного приложения в среде Visual Studio Community представлен на рисунке 3.1.

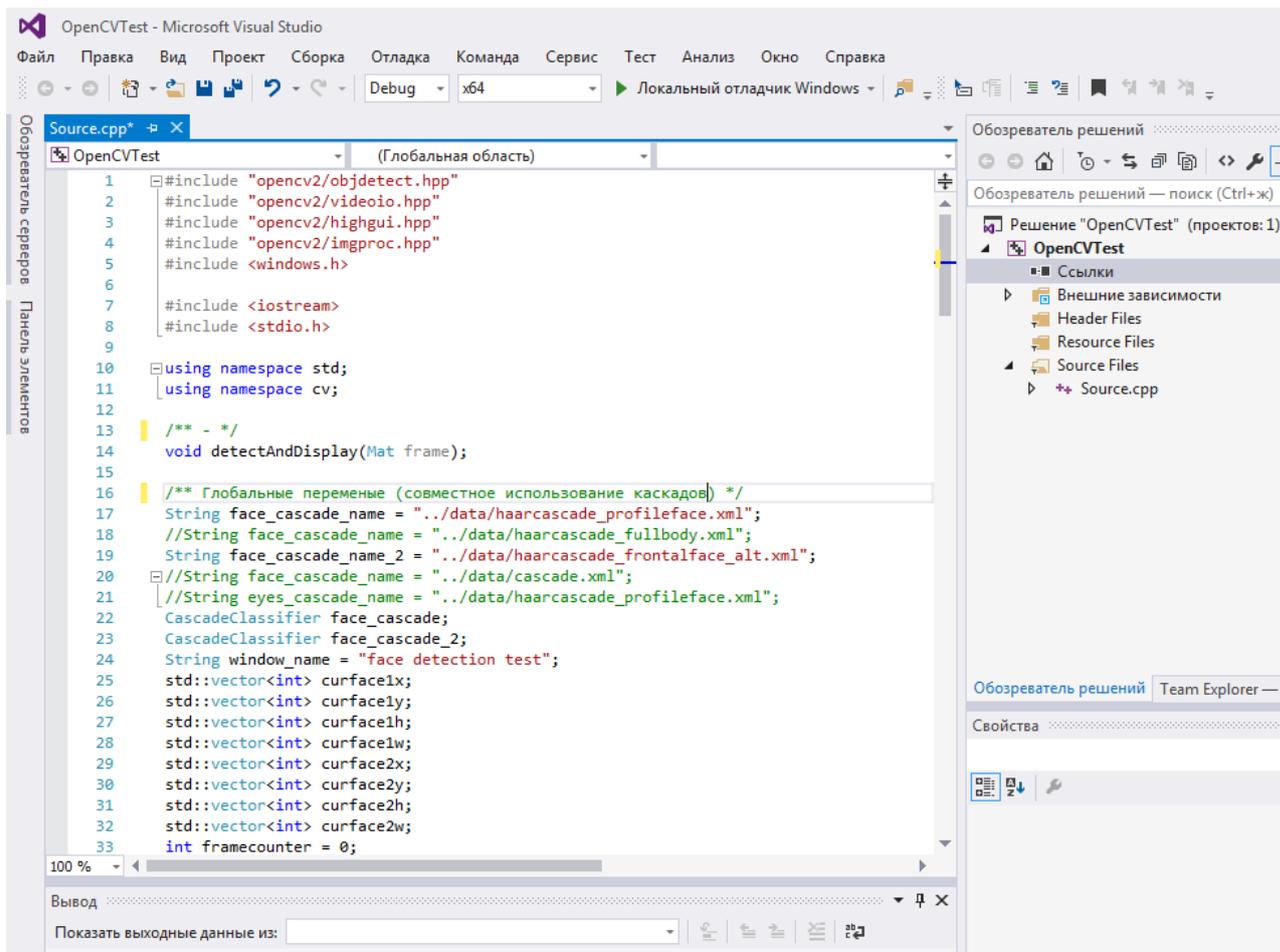


Рисунок 3.1 – Проект приложения в среде Visual Studio Community

После запуска приложение подключится к указанному в программном коде источнику изображений, в нашем случае – это веб-камера ноутбука и в режиме реального времени приступит к решению задачи локализации людей на видео.

В качестве результата программа будет выводить на экран пользователю анализируемый видео ряд, где в режиме реального времени с помощью рамок будет выделять найденных людей (рисунок 3.2).

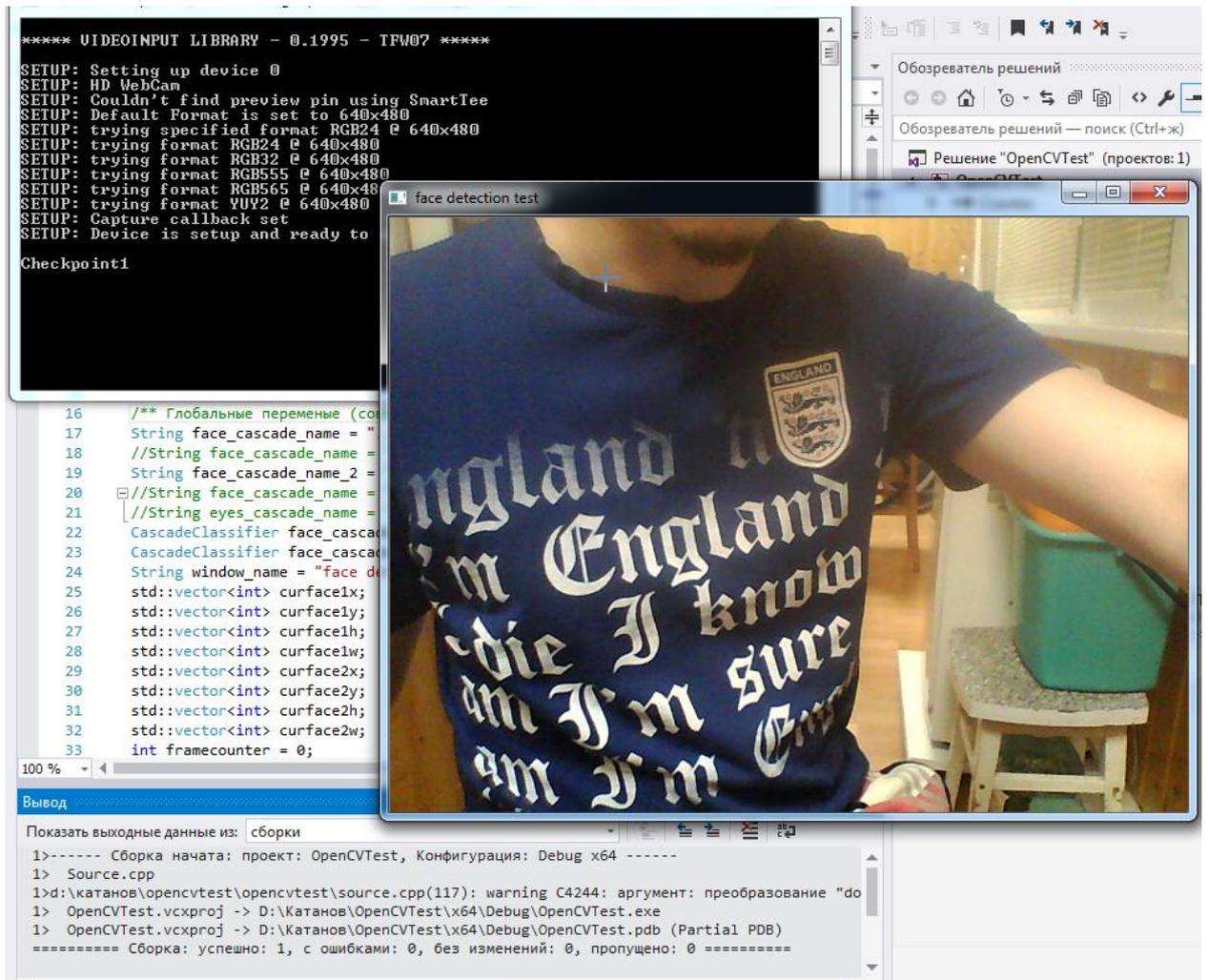


Рисунок 3.2 – Запуск приложения в среде Visual Studio Community

### 3.2 Тестирование алгоритма и обсуждение результатов

Примеры тестирования предложенного подхода в режиме реального времени с помощью веб-камеры с разрешением 0,3 МП (640×480) приведены на рисунках 3.3-3.7.

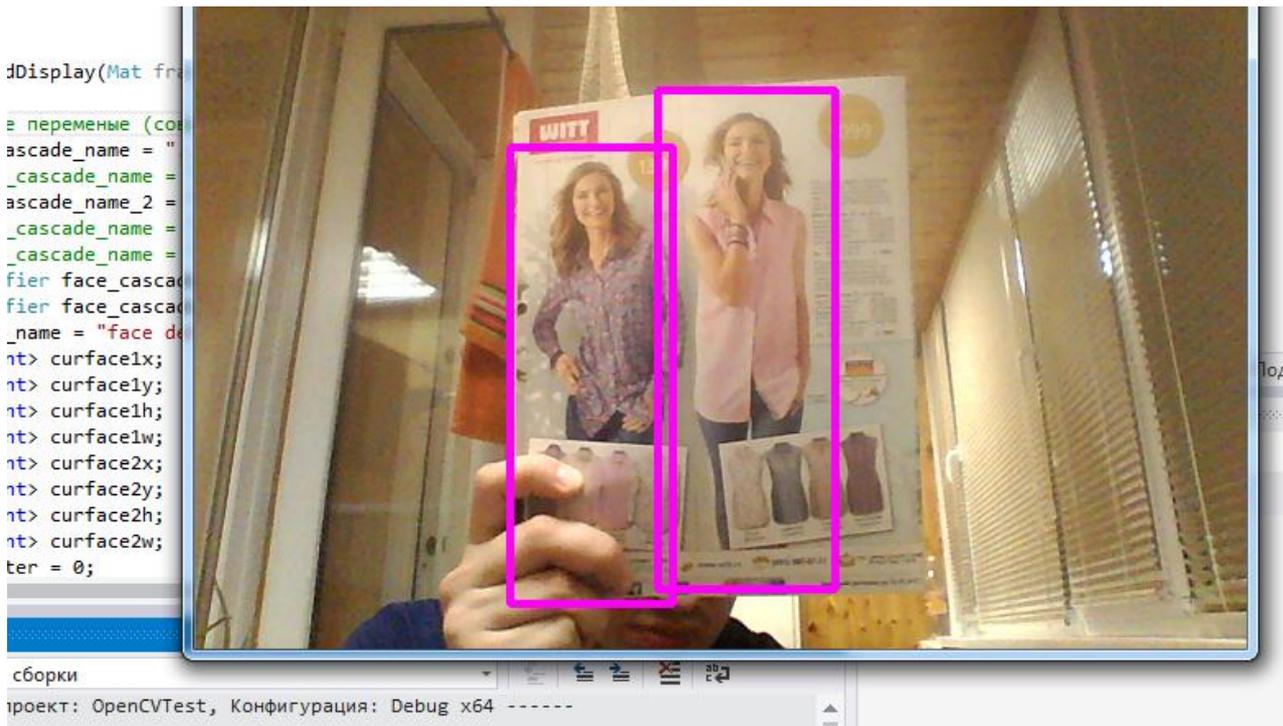


Рисунок 3.3 – Тестирование предложенного подхода

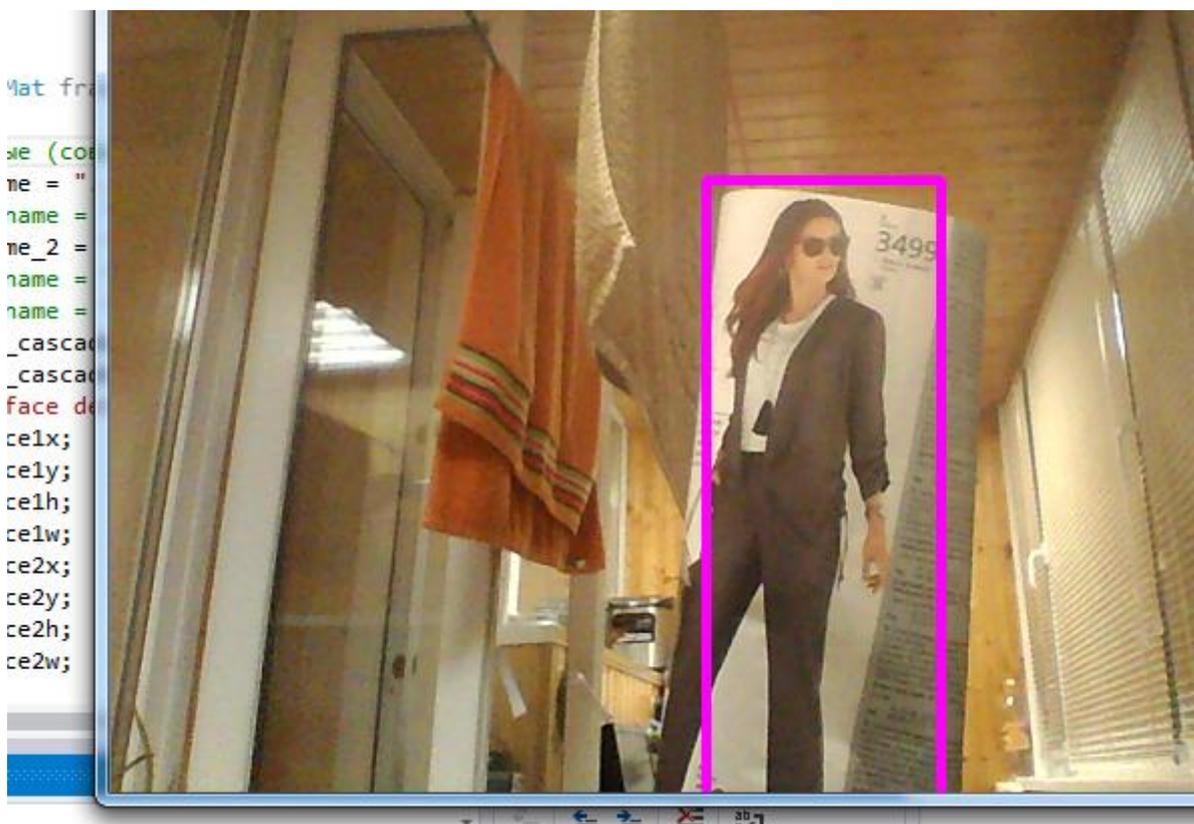


Рисунок 3.4 – Тестирование предложенного подхода



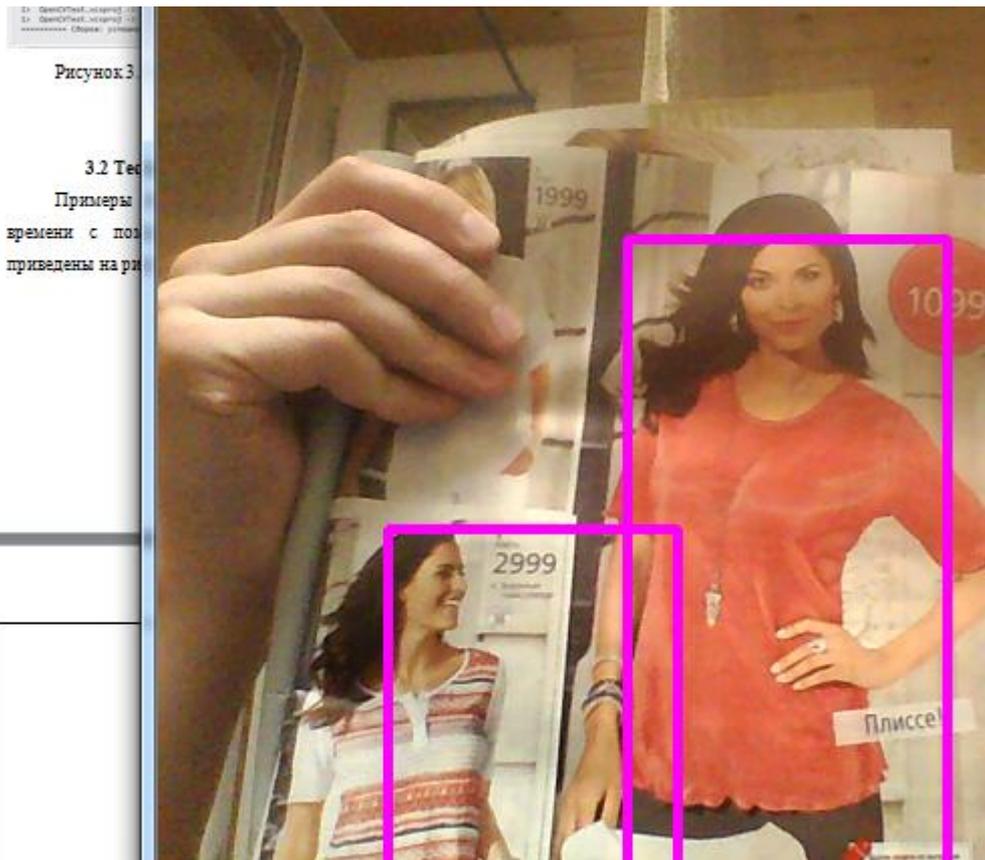


Рисунок 3.7 – Тестирование предложенного подхода

Результаты тестирования предложенного подхода (совместного использования двух каскадов) по сравнению с методами, описанными в первой главе, приведены в таблице 3.1.

Как видно из результатов тестирования использование одновременно двух каскадов, позволило значительно снизить ситуации, при которых возникали ошибки второго рода. Особенно это сказывается на изображениях, на которых человеческое тело частично скрыто препятствиями. Для этих изображений предложенный подход показывает даже лучшие результаты по сравнению с HOG SVM, GNG-FIS. Таким образом, можно считать, что поставленная цель работы достигнута.

Таблица 3.1 – Результаты апробации различных методов, на тестовой выборке, состоящей из 2100 изображений

Название метода	Количество не распознанных (ошибка 2-го рода)	Количество ложных распознаваний (ошибка 1-го рода)	Общая ошибка	Время обработки	Время обучения
HOG SVM	358	42	19,05%	40 мс	3-5 часов
Viola-Jones каскад fullbody	801	45	40,29%	10 мс	180-300 часов
GNG-FIS	76	112	8,95%	20 мс	5 часов
GNG-FIS с двухпроходной перемаркировкой	120	52	8,19%	30 мс	6 часов
Viola-Jones с каскадами fullbody + face	160	51	10,05%	20 мс	360-600 часов

При этом каскад, нацеленный на локализацию лица, человека внес незначительный вклад в количество ложных распознаваний людей (увеличение с 45 до 52 ошибок).

Слабой стороной предложенного подхода является неспособность локализации людей, когда наблюдение за человеком производится сверху-вниз.

Добавления дополнительного каскада замедляет работу системы локализации людей приблизительно в 2 раза. Однако конечное быстродействие предложенного подхода всё равно выше, чем у методов HOG SVM и GNG-FIS.

## ЗАКЛЮЧЕНИЕ

Основные выводы по проведенным в ВКР исследованиям:

1. На основе изучения литературных источников по теме исследования установлено, что совершенствование методов направленных на решение задач локализации людей на изображениях достигается как путем совершенствования алгоритмов компьютерного зрения, так путем их комбинирования.

2. Наиболее часто используемыми методами для решения задач локализации людей являются: HOG SVM, GNG-FIS, Viola-Jones. Произведено сравнения точности работы методов при решении задачи локализации человека. При этом метод Viola-Jones имеет самую низкую точность работы (общая ошибка 40%) с применением одного каскада классификаторов.

3. Показано, что задачу локализации человека на изображении можно свести к задаче бинарной классификации, если разбивать анализируемое изображение на прямоугольные области с помощью метода сканирующего окна.

4. Разработан подход одновременного использования нескольких каскадов классификаторов в методе Viola-Jones, что позволило снизить количество ошибок первого рода в задаче локализации людей с 40,29% до 10,05%.

5. Установлено, что добавления дополнительного каскада замедляет работу метода Viola-Jones по локализации людей приблизительно в 2 раза. Однако конечное быстроедействие предложенного подхода всё равно выше, чем у методов HOG SVM и GNG-FIS (детальное сравнение методов приведено в таблице в третьей главе).

6. На языке C++ (в среде Visual Studio Community) с использованием свободной библиотеки OpenCV разработана и протестирована программная реализация предложенного подхода. На практике доказана состоятельность предложенного подхода.

## СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Амосов, О.С. Алгоритм локализации номерных знаков транспортных средств / О.С. Амосов // Информатика и системы управления. – 2014. – Т. 39, № 1. – С. 127-140.
2. Мельников, И.И. Детектор движения на основе импульсных нейронных сетей / И.И. Мельников, К.А. Демиденков, И.А. Емельянов, И.А. Евсеенко // Информационные технологии. – 2013. – № 7. – С. 57-60.
3. Viola, P. Robust real-time face detection / P. Viola, M.J. Jones // International Journal of Computer Vision. – 2004. – Vol. 57(2). – P. 137-154.
4. Михаева, Е.И. Детектирование с использованием фрактального анализа в системах дорожной безопасности / Е.И. Михаева, А.В. Никанов // Компьютерная оптика. – 2012. – Т. 36, № 1. – С. 124-130.
5. Viola, P. Detecting Pedestrians Using Patterns of Motion and Appearance / P. Viola, J. Jones // International Journal of Computer Vision. – 2005. – Vol. 63(2). – P. 153-161.
6. Enzweiler, M. Monocular Pedestrian Detection: Survey and Experiments / M. Enzweiler, G.M. Dariu // IEE Transactions on Machine Intelligence. – 2010. – №12. – P. 2178-2196.
7. Dalal, N. Histograms of oriented gradients for human detection / N. Dalal, B. Triggs // IEE Conference on Computer Vision. – 2005. – P. 886-893.
8. Cristianini, N. An introduction to support Vector Machines / N. Cristianini, J. Shawe-Taylor. – Cambridge: Cambridge University Press, 2000. – P. 204.
9. Лекции по методу опорных векторов [Электронный ресурс] / К.В. Воронцов. – URL: <http://www.ccas.ru/voron/download/SVM.pdf> (дата обращения 11.05.2017).
10. Казаков, А. Быстрый алгоритм обнаружения пешеходов по видеоданным / А. Казаков, А. Бовырин // 22-nd International Conference on Computer Vision. – 2012. – С. 144-148.

11. Cho, H. Real-time pedestrian detection with deformable part models / H. Cho, P.E. Rybski, A. Bar-Hillel, W. Zhang // Intelligent Vehicles Symposium (IV), 2012 IEEE, Alcalá de Henares. – 2012. – P. 1035-1042. – DOI: 10.1109/IVS.2012.6232264.
12. Hua, Ch. Pedestrian detection by using spatio temporal histogram of oriented gradients / Ch. Hua, Y. Makihara, Y. Yagi // IEICE Transactions on Information and Systems. – 2013. – Vol. E96-D(6). – P. 1376-1386.
13. Vapnik, V.N. An overview of statistical learning theory / V.N. Vapnik // IEE Transaction on Neural Networks. – 1999. – Vol. 10(5). – P. 988-999.
14. Воронцов, К.В. Математические методы машинного обучения [Электронный ресурс] / К.В. Воронцов. – 2011. – URL: <http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf> (дата обращения 21.10.16).
15. Cireşan, D. Multi-column deep neural networks for image classification / D. Cireşan, U. Meier, J. Schmidhuber // IEE Conference on Computer Vision. – 2012. – P. 3642-3649. – DOI: 10.1109/CVPR.2012.6248110.
16. Cireşan, D. Flexible, high performance convolutional neural networks for image classification / D. Cireşan, U. Meier, J. Masci, L.M. Gambardella, J. Schmidhuber // Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. – 2011. – Vol. 2. – P. 1237-1242.
17. Karungaru, S.G. Face recognition in colour images using neural networks and genetic algorithms / S.G. Karungaru, M. Fukumi, N. Akamatsu // International Journal of Computational Intelligence and Applications. – 2005. – Vol. 5(1). – P. 55-67.
18. Солдатова, О.П. Применение свёрточной нейронной сети для распознавания рукописных цифр / О.П. Солдатова, А.А. Гаршин // Компьютерная оптика. – 2010. – Т. 34, № 2. – С. 252-259.
19. Verma, A. Pedestrian detection via mixture of CNN experts and thresholded aggregated channel features / A. Verma, R. Hebbalaguppe, L. Vig, S.

Kumar, E. Hassan // IEE International Conference of Computer Vision (ICCVW '15), Santiago. – 2015. – P. 555-563. – DOI: 10.1109/ICCVW.2015.78.

20. Ouyang, W. Joint Deep Learning for Pedestrian Detection / W. Ouyang, X. Wang // 2013 IEE Conference on Computer Vision. – 2013. – P. 2056-2063.

21. Хайкин, С. Нейронные сети: Полный курс: Пер. с англ. / С. Хайкин.; пер. с англ. – М.: Вильямс, 2008. – 1103 с.

22. Кохонен, Т. Самоорганизующиеся карты / Т. Кохонен; пер. с англ. – М.: БИНОМ. Лаборатория знаний, 2008. – 655 с.

23. Уоссермен, Ф. Нейрокомпьютерная техника: Теория и практика / Ф. Уоссермен. – М. : Мир. 1992. – 240 с.

24. Martinetz, T.M. Neural-gas network for vector quantization and its application to time series prediction / T.M. Martinetz, S.G. Berkovich, K.J. Schulten // IEE Transaction on NeuralNetworks. – 1993. – Vol. 4(4). – P. 557-568.

25. Амосов, О.С. локализация человека в кадре видеопотока с использованием алгоритма на основе растущего нейронного газа и нечёткого вывода / О.С. Амосов, Ю.С. Иванов, С.В. Жиганов // Компьютерная оптика. – 2017. – №1. – с.46 – 58.