

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Математико-механический факультет

Кафедра системного программирования

Батов Никита Станиславович

Обнаружение лиц  
на базе контроллера ТРИК

Курсовая работа

Научный руководитель:

ст. преп. Я. А. Кириленко

Санкт-Петербург

2015

## Оглавление

<b>Введение</b> .....	3
<b>Постановка задачи</b> .....	5
<b>Обзор предметной области</b> .....	6
Метод Виолы-Джонса.....	6
Алгоритмы обнаружения лица, основанные на выделении цвета кожи.....	8
Выбор цветового пространства .....	9
Вывод.....	10
<b>Реализация</b> .....	11
Архитектура.....	11
Работа с цветом .....	11
Алгоритм.....	14
Эксперименты .....	17
<b>Результаты</b> .....	19
<b>Список использованных материалов</b> .....	20

## Введение

По мнению аналитического агентства Gartner [1], к 2020 г. интернет вещей объединит 26 миллиардов устройств, а доход поставщиков продуктов и услуг рынка интернета вещей составит к этому времени 300 миллиардов долларов. Робототехника является неотъемлемой частью рынка интернета вещей.

По прогнозам Gartner в ближайшие годы эту отрасль ждет гигантский технологический скачок. Роботы станут незаменимыми помощниками людей и смогут взять на себя обеспечение большей части потребностей цивилизации. Робототехническим системам все чаще приходится взаимодействовать с человеком напрямую. А человеческое лицо достаточно яркий признак, по которому машина способна находить человека в пространстве. Поэтому обнаружения лица на кадре видеоизображения — актуальная и важная задача в робототехнике, которая позволит роботу реагировать на поведение человека.

Компания ООО «КиберТех» совместно с кафедрами системного программирования и кибернетики СПбГУ разрабатывают робототехнический контроллер ТРИК [2].

ТРИК — это миникомпьютер, содержащий все необходимое оборудование для создания на его базе автономных робототехнических систем. Контроллер может управлять двигателями прямого тока и сервоприводами, обрабатывать информацию как с цифровых датчиков, так и с аналоговых, работать с видео модулями и микрофонами, имеет интерфейсы Wi-Fi, Bluetooth 4.0 (включая LE), USB, Micro-SD и ANT.



Рисунок 1. «Робототехнический контроллер ТРИК»

ТРИК разработан на базе процессора OMAP-L138 C6-Integra™ DSP+ARM® SoC [3] производства Texas Instruments. Этот процессор имеет два вычислительных модуля:

- управляющее ядро ARM (ARM926EJ-S™ RISC MPU), обеспечивающее работу операционной системы Linux на контроллере;
- ядро DSP (C674x Fixed/Floating-Point VLIW DSP) специально разработанное для обработки большого объема данных, представляемых в виде векторов. Именно поэтому оно используется при обработке видео и аудио информации.

Управляющее ядро может передавать обработку векторных данных на специально предназначенный для этого вычислительный модуль.

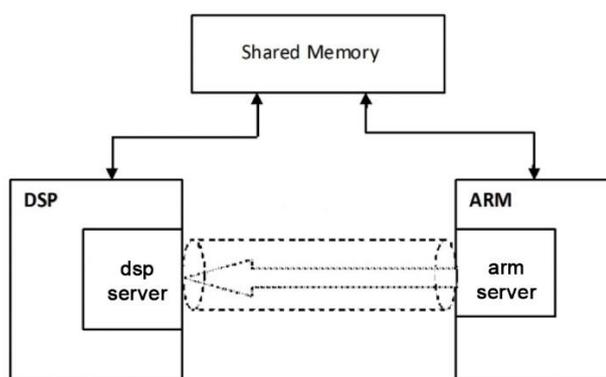


Рисунок 2. «Высокоуровневая архитектура процессора  
OMAP L138 (ARM+DSP)»

В работе так же был использован видео-модуль ov7670 [4], подключаемый к контроллеру по интерфейсу ВТ.656. Он используется в качестве камеры.



Рисунок 3. «Видео-модуль ov7670»

## Постановка задачи

Передо мной была поставлена задача реализации открытого алгоритма обнаружения человеческого лица на видеопотоке, выполняющегося полностью автономно на контроллере ТРИК.

Для выполнения поставленной задачи было необходимо выполнить следующие шаги.

- Провести исследование существующих методов и средств обнаружения, локализации и распознавания лиц на изображении.
- Познакомится с работами предшественников по аналогичной теме.
- Изучить устройство SoC-системы (англ. *System-on-a-Chip*) OMAP L-138 с разделенными процессорными ядрами и способы их взаимодействия.
- Выбрать и реализовать алгоритм обнаружения лиц, работающий на контроллере ТРИК. Важно было реализовать алгоритм, работающий автономно с приемлемой точностью.

## Обзор предметной области

В работе был очень важен выбор алгоритма, так как вычислительные мощности платформы накладывают определенные ограничения. Точные и мощные алгоритмы при работе на ГРИК будут перегружать процессор и в результате производительность будет недостаточным для адекватного восприятия видеоизображения.

Изученные мною подходы к распознаванию лиц можно разделить на следующие категории.

1. Методы, основанные на данных о человеческом лице.
2. Методы, основанные на поиске инвариантных черт лица на изображении, несмотря на угол и положение лица.
3. Шаблонный метод, сравнивающий изображение с базой шаблонов лиц или черт лица.
4. Обучаемые методы, это те же шаблонные методы, но набор базы шаблонов проводится с помощью машинного обучения.

Но каждый конкретный алгоритм не обязательно относится к одной из них, он может состоять из методов, принадлежавшим разным группам.

### Метод Виолы-Джонса

Общепризнанный и популярный алгоритм обнаружения и распознавания лиц — Метод Виолы-Джонса [5], описанный в 2001 году Полом Виолой и Майклом Джонсом. По мнению автора статьи [10], является одним из лучших по соотношению показателей эффективности распознавания/скорость работы.

Данный метод основан следующих принципах.

- Используются изображения в интегральном представлении, что позволяет вычислять быстро необходимые объекты.
- Используются признаки Хаара, с помощью которых происходит поиск нужного объекта (в данном контексте, лица и его черт).

- Используется бустинг (от англ. boost – улучшение, усиление) для выбора наиболее подходящих признаков для искомого объекта на данной части изображения.
- Все признаки поступают на вход классификатора, который даёт результат «истина» либо «ложь».
- Используются каскады признаков для быстрого отбрасывания окон, где не найдено лицо.

Данный алгоритм является стандартом де-факто в области обнаружения и распознавания человеческих лиц и в той или иной степени используется в миллионах устройств, связанных с распознаванием лиц [11]. Алгоритм является весьма точным благодаря системе машинного обучения каскадов, но одновременно с этим данный подход требователен к вычислительным ресурсам. Мой коллега, студент кафедры системного программирования СПбГУ Курбанов Р. Э., в своей курсовой работе [6] занимался запуском алгоритма Виолы-Джонса на контроллере ТРИК, но, к сожалению, полное портирование не было выполнено, по словам Рауфа Эльшадовича, из-за недостаточной вычислительной мощности процессора. Проанализировав работу своего коллеги, я принял решение выбрать алгоритм, который пусть и менее точен, но будет быстрее работать на контроллере ТРИК.

## Алгоритмы обнаружения лица, основанные на выделении цвета кожи

В работе «Hybrid Face Detection in Color Images» [9] автор разработал гибридный алгоритм, идея которого заключается в том, что обученные классификаторы исследуют не все изображение, а только выделенные по цвету фрагменты. Портитование данного алгоритма также вызовет сложности реализации на контроллере ТРИК, из-за недостаточной скорости работы классификаторов.

Алгоритм, описанный в статье «Face detection» [7], разделяется на несколько логических этапов:

Во-первых, алгоритм производит цветовую сегментацию цвета кожи в  $YCbCr$  цветовом пространстве, классифицируя все пиксели входного изображения как принадлежащие коже или нет. Следующим шагом является удаления шума, то есть замкнутых маленьких регионов, отсекаемых по определенном пороговому значению. Затем, чтобы отделить некоторые интегрированные регионы в отдельные лица, используется алгоритм обнаружения граней Роберта Кросса, вычисляющий плоское измерение градиента на изображении. Таким образом, подчеркиваются высокие градиенты, что часто соответствует краям. Наконец, предыдущие



*Рисунок 4. «Результат фильтрации»*

изображения интегрируются в один бинарный образ, при этом относительно маленькие черные и белые области удаляются. Разница между этим шагом и начальным фильтром в том, что ребра, соединенные с темными областями, остаются после этой фильтрации. Эти края очень важны, так как позволяют разделять лица между собой (Рисунок 4).

Последним шагом алгоритма идет сопоставление фрагментов изображения, с заранее подготовленными шаблонами человеческого лица.

В итоге, алгоритм с достаточной точностью обнаруживает и локализует человеческие лица на изображении. При этом стоит отметить, что у автора статьи данный алгоритм, обрабатывающий фотографию работает 90 секунд. Я принял решение видоизменить этот алгоритм для последующей реализации его на контроллере ТРИК.

### Выбор цветового пространства

В статье «A Hybrid Skin Color Model for Face Detection» [8], автор провел исследование по выявлению лучшего цветового пространства для построения в нем модели цвета кожи. В результате, лучшее цветовое пространство при комнатном освещении – HSV, поэтому мной было принято решение использовать именно его.

В той же статье описана модель цвета кожи человека:

$$\begin{aligned} 0 &\leq HUE \leq 50, \\ 0.1 &\leq SATURATION \leq 0.9 \end{aligned}$$

## Вывод

На основании исследования я принял решение разработать алгоритм, являющийся упрощенной версией алгоритма из статьи «Face detection» [7], заменив сопоставление шаблонов человеческого лица на валидацию, базирующуюся на пропорциях, и отказавшись от использования ресурсоёмкого оператора Кросса, сделав основной упор на обнаружение одного лица.

# Реализация

## Архитектура

Программа логически разделена на две части, размещенных на разных процессорных ядрах: ARM и DSP, управляющем и вычислительным соответственно.

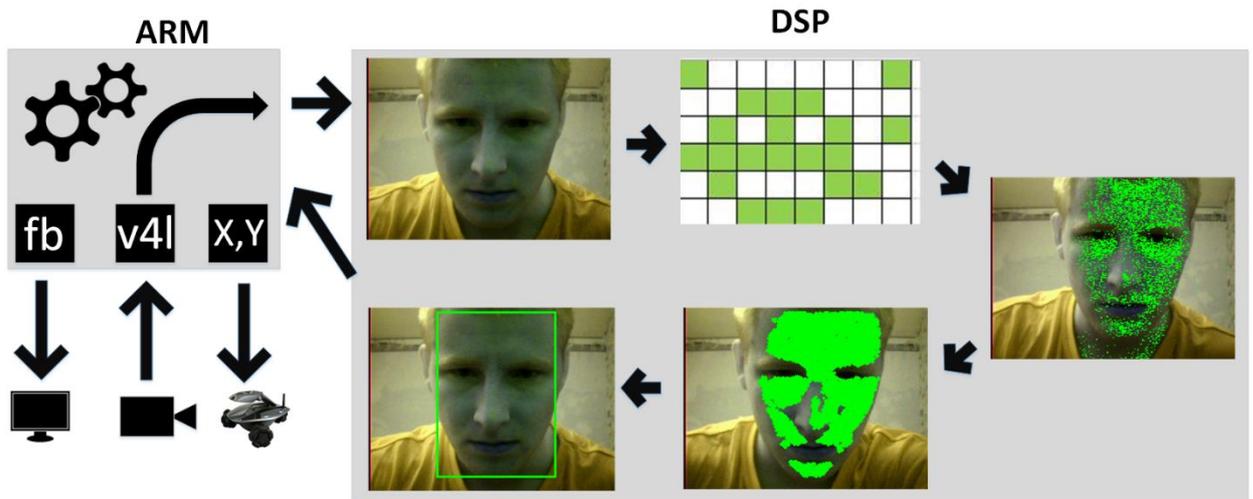
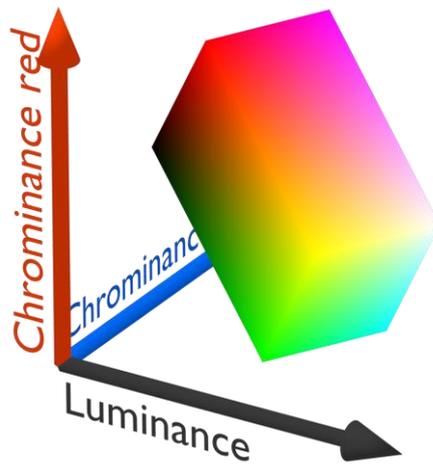


Рисунок 5. «Архитектура»

ARM получает изображение с камеры с помощью интерфейса видео захвата Video4Linux в формате YCbCr и отправляет полученное изображение к DSP-процессору. Тот в свою очередь производит обработку изображения и возвращает уже готовое изображение для вывода его на экран контроллера. А если удалось обнаружить лицо, то DSP-процессор отправляет координаты его центра, для последующего использования в прикладной управляющей программе робота.

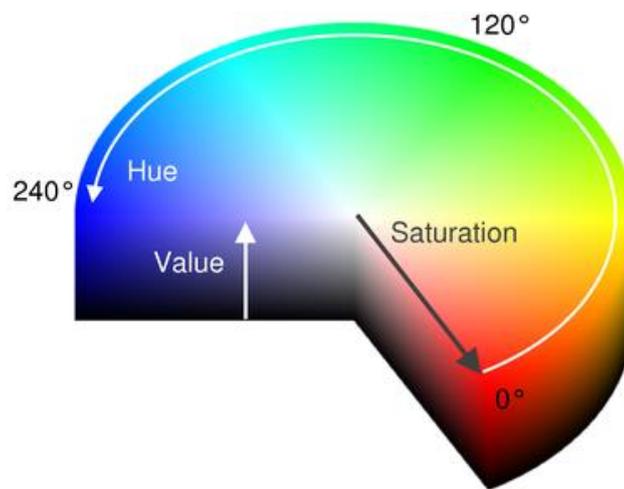
## Работа с цветом

DSP-программа получает изображение в YCbCr. YCbCr – это семейство цветовых пространств, используемых в том числе при передаче компонентных видеосигналов, когда компоненты цвета передаются по разным каналам.



*Рисунок 6. «YCbCr цветное пространство»*

YCbCr – аббревиатура описывающая компоненты, где Y – компонента яркости, Cb и Cr являются синей и красной цветоразностными компонентами соответственно.



*Рисунок 7. «HSV цветное пространство»*

Основная работа по нахождению лица ведется в HSV пространстве. HSV – аббревиатура описывающая цветовые компоненты, где Hue – цветовой тон, Saturation – насыщенность или чистота цвета, Value – яркость.

Цвета исходного изображения конвертируются из YCbCr сначала в RGB (цветное пространство, где компоненты R – красный, G – зеленый, B – синий),

а затем в целевой HSV. Такая цепочка необходима, так как при формировании результирующего изображения используется RGB цвет.

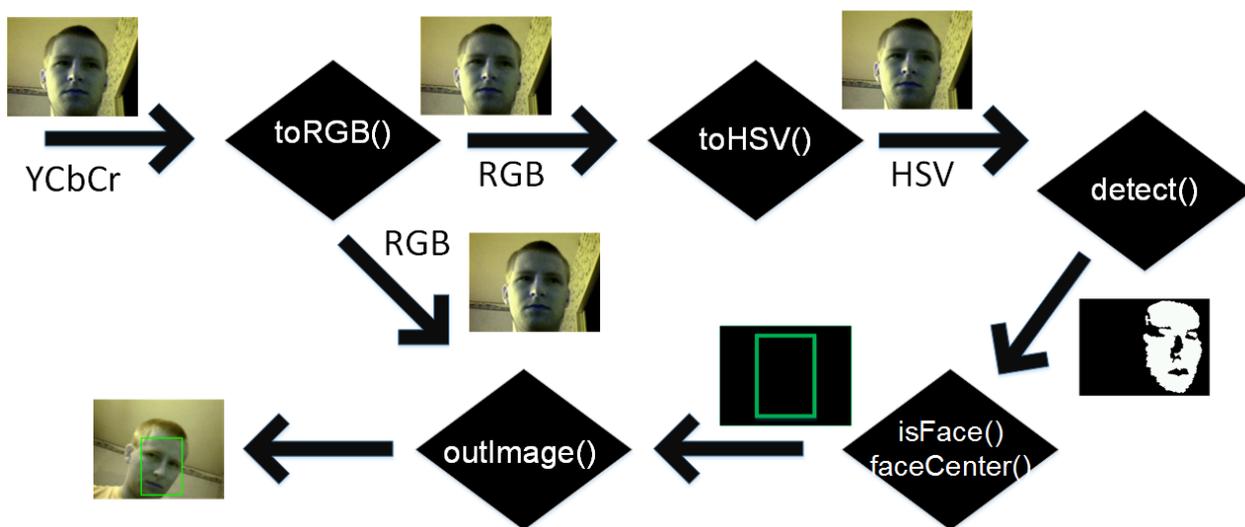


Рисунок 8. «Цепочка преобразования»

В результате мы получаем исходный кадр, готовый к работе алгоритма. Затем требуется определить диапазон значений в HSV-модели, соответствующий цвету человеческой кожи. В процессе исследования мной была обнаружена модель цвета кожи человека:

$$\begin{aligned} 0 &\leq \text{HUE} \leq 50, \\ 0.1 &\leq \text{SATURATION} \leq 0.9 \end{aligned}$$

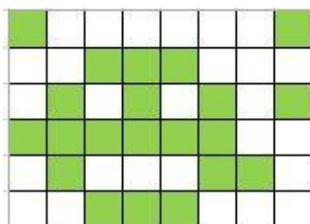
Где HUE определен от 0 до 359, а SATURATION от 0 до 1. Но данная модель при реализации на контроллере ТРИК работает некорректно и не выделяет соответствующие цвета. Причина такого поведения связана с тем, что цвет, соответствующий определенным HSV-значениям, зависит от конкретного устройства. Таким образом нахождения HSV-диапазона, удовлетворяющего цвету кожи необходимо проводить экспериментально.

В результате проведенных опытов HSV-диапазон, соответствующий цвету кожи человека для видео модуля ov7670, был определен следующим образом:

$$\begin{aligned} 120 &\leq HUE \leq 160, \\ 0.2 &\leq SATURATION \leq 0.9 \end{aligned}$$

### Алгоритм

Определим отображение  $f: M \rightarrow \{\text{ИСТИНА}, \text{ЛОЖЬ}\}$ , ставящее в соответствие к каждому пикселю кадра ИСТИНУ, если цвет пикселя попадает внутрь модели ограничивающей цвет кожи и ЛОЖЬ, если цвет пикселя другой. В результате применения отображения получим битовую матрицу размером с исходное изображение.



*Рисунок 9. «Битовая матрица»*

Если наложить эту матрицу на исходное изображение, то получим следующее изображение:



*Рисунок 10. «Изображение с шумом»*

К сожалению, в жизни достаточно предметов, подходящих под цветовой диапазон человеческого лица, например, древесина или песок. Так на Рисунке 20 показано изображение с экрана контроллера ТРИК, где зелеными точками отмечены пиксели, удовлетворяющие модели. Можно заметить, что на фоне проявились точки, явно не являющиеся фрагментами лица.

Для удаления этих точек используем следующую фильтрацию:

Введем отображение  $f: N_\delta \rightarrow \mathbb{R}$ , ставящие в соответствие некоторой окрестности пикселя число его «соседей», удовлетворяющих модели цвета кожи. Используя это отображение будем отсекаать шум по некоторому пороговому значению.

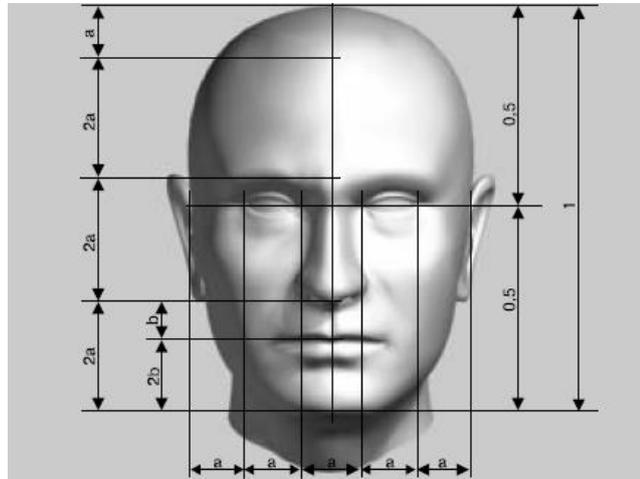


*Рисунок 11. «Результат фильтрации»*

На Рисунке 21 изображен пример, с наиболее подходящими коэффициентами фильтрации. Выделены только пиксели, находящиеся непосредственно на лице человека. Параллельно будем проводить наращивание «хороших» пикселей, по аналогии с морфологической операцией dilate [12].

В результате преобразования, получаем выделенным самый крупный предмет на кадре, удовлетворяющий модели цвета кожи человека.

Последним шагом алгоритма является валидация изображения именно как лица, а не другой части тела. Этот процесс производится, опираясь на известные пропорции лица, изображенные на Рисунке 12:



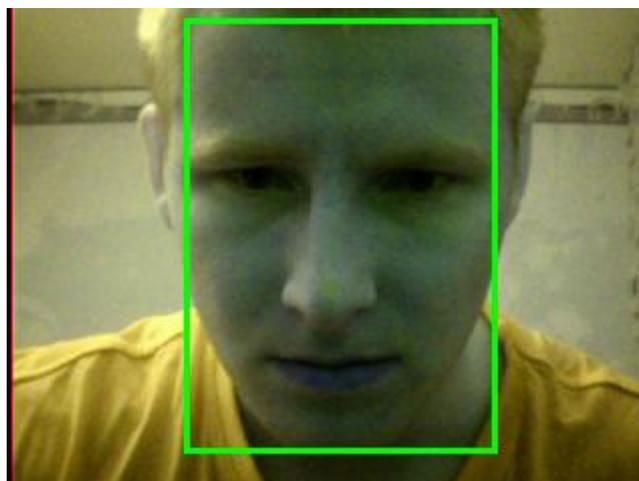
*Рисунок 12. «Пропорции человеческого лица»*

Базовой валидацией в алгоритме является проверка соотношений сторон прямоугольника, в который вписан овал человеческого лица,

Из Рисунка 12 можно заметить, что данное соотношение равно:

$$\frac{2a + 2a + 2a}{a + a + a + a + a} = \frac{6}{5} = 1,2$$

Соотношение сторон прямоугольника должно быть близко к этому значению. Соотношение сторон на Рисунке 13 примерно равно 1.33, поэтому найденный фрагмент является человеческим лицом.



*Рисунок 13. «Обнаружения лица»*

## Эксперименты

Для определения точности алгоритма были проведены эксперименты. Человеческое лицо попадало в поле видимости камеры, при этом в процессе эксперимента отслеживался контрольный кадр, по которому и происходил сбор статистики: если алгоритм находил лицо на этом кадре, то эксперимент признавался удачным. В случае с ложными обнаружениями, в кадр вносились различные предметы, предположительно, похожие на человеческое лицо. Если алгоритм находил лицо на кадре, то фиксировался факт ложного обнаружения.

Результаты экспериментов, для описанной в работе модели:

Условия эксперимента	Обнаружения	Ложные обнаружения
Искус. осв. Дальн. < 1м.	80 %	30 %
Естеств. осв. Дальн. < 1м.	20 %	20 %
Искус. осв. Дальн. > 1м.	0 %	0 %
Естеств. осв. Дальн. > 1м.	0 %	0 %

*Таблица 1. «Измерение точности»*

При проведении замеров видно, что при естественном освещении алгоритм работает недостаточно точно. Поэтому потребовалось провести калибровку. Для этого следует сменить модель кожи:

$$\begin{aligned} 10 &\leq HUE \leq 50, \\ 0.2 &\leq SATURATION \leq 0.9 \end{aligned}$$

И затем провести замеры повторно:

Условия эксперимента	Обнаружения	Ложные обнаружения
Искус. осв. Дальн. < 1м.	20 %	30 %
Естеств. осв. Дальн. < 1м.	70 %	20 %
Искус. осв. Дальн. > 1м.	0 %	0 %
Естеств. осв. Дальн. > 1м.	0 %	0 %

*Таблица 2. «Измерение точности»*

В результате проведения экспериментов мной сделан вывод о достаточной точности алгоритма при проведении предварительной настройки алгоритма для конкретного типа освещения.

## Результаты

В ходе работы полностью выполнена поставленная задача. Опираясь на знания, полученные в ходе исследования, был реализован алгоритм обнаружения и локализации человеческого лица на видео потоке, исполняющийся на контроллере ТРИК автономно без привлечения дополнительных вычислительных мощностей. Разработанный и реализованный алгоритм после настройки параметра тона определяет лицо с достаточно высокой точностью на расстоянии до 1 м от камеры.

## Список использованных материалов

[1] Gartner's Internet of Things Research.

URL: <http://www.gartner.com/technology/research/internet-of-things/>

(дата обращения: 14.05.2015).

[2] Блог компании ООО «КиберТех».

URL: <http://blog.trikset.com/> (дата обращения: 16.05.2015).

[3] Раздел сайта компании «Texas Instruments», посвященный процессору OMAP L138. URL: <http://www.ti.com/product/omap-1138> (дата обращения: 14.05.2015).

[4] Документация видео-модуля ov7670.

URL: <http://www.voti.nl/docs/OV7670.pdf> (дата обращения: 15.05.2015).

[5] *P. Viola, and M. J. Jones*, Robust Real-Time Face Detection, Int. Journal of Computer Vision, 2004.

[6] *Курбанов Р. Э.* Распознавание лиц на базе контроллера ТРИК. Курсовая работа СПбГУ. 2014.

[7] *Inseong Kim, Joon Hyung Shim, and Jinkyu Yang*, Face detection, Stanford University.

[8] *Vandana S. Bhat ,Dr. J.D. Pujari , Bhavana*, A Hybrid Skin Color Model for Face Detection, Int. Journal of Engineering Research and General Science, 2014.

[9] *M. Niazi, S. Jafari*, Hybrid Face Detection in Color Images, Int. Journal of Computer.

[10] Метод Виолы-Джонса как основа для распознавания лиц.  
URL: <http://habrahabr.ru/post/133826> (дата обращения: 25.05.2015)

[11] Новостной сайт, посвященный высоким технологиям.  
URL: <http://hi-news.ru/technology/razrabotan-novyj-bolee-sovershennyj-algorithm-raspoznavaniya-lic.html> (дата обращения: 28.05.2015)

[12] Википедия: математическая морфология.  
URL: [https://en.wikipedia.org/wiki/Mathematical\\_morphology](https://en.wikipedia.org/wiki/Mathematical_morphology) (дата обращения: 28.05.2015)