



Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет  
имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ \_\_\_\_\_

КАФЕДРА \_\_\_\_\_

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА  
К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ**

**НА ТЕМУ:**

**Распознавание лица и его элементов**

Студент РТ5-61Б  
(Группа)

\_\_\_\_\_  
(Подпись, дата) Сафиуллин А.П.  
(И.О.Фамилия)

Руководитель

\_\_\_\_\_  
(Подпись, дата) \_\_\_\_\_  
(И.О.Фамилия)

Консультант

\_\_\_\_\_  
(Подпись, дата) \_\_\_\_\_  
(И.О.Фамилия)

**Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)**

---

УТВЕРЖДАЮ

Заведующий кафедрой \_\_\_\_\_  
(Индекс)

\_\_\_\_\_  
(И.О.Фамилия)

« \_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ г.

**З А Д А Н И Е  
на выполнение научно-исследовательской работы**

по теме \_\_\_\_\_ Распознавание лица и его элементов \_\_\_\_\_

Студент группы \_\_\_\_\_ РТ5-61Б \_\_\_\_\_

\_\_\_\_\_ Сафиуллин Антон Павлович \_\_\_\_\_  
(Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.) \_\_\_\_\_

Источник тематики (кафедра, предприятие, НИР) \_\_\_\_\_

График выполнения НИР: 25% к \_\_\_\_ нед., 50% к \_\_\_\_ нед., 75% к \_\_\_\_ нед., 100% к \_\_\_\_ нед.

**Техническое задание** \_\_\_\_\_

**Оформление научно-исследовательской работы:**

Расчетно-пояснительная записка на \_\_\_\_ листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.) \_\_\_\_\_

Дата выдачи задания « \_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ г.

**Руководитель НИР**

**Студент**

\_\_\_\_\_  
(Подпись, дата)

\_\_\_\_\_  
(И.О.Фамилия)

\_\_\_\_\_  
(Подпись, дата)

**Сафиуллин А.П.** \_\_\_\_\_  
(И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

## Оглавление

Введение.....	4
Постановка задачи.....	5
Теоретическая часть.....	6
Практическая часть .....	12
Заключение .....	14
Список использованных источников информации.....	15

## Введение.

В данный момент для идентификации человека используются различные биометрические методы. Биометрией принято называть систему распознавания людей по одной или более физическим или поведенческим чертам (трёхмерная фотография лица и/или тела, образец голоса, отпечатки пальцев, рисунок вен руки, группа крови, специальное фото роговицы глаза и т.д.).[1]

В последние несколько лет наблюдается все больший интерес к распознаванию лица как одного из важнейших биометрических идентификаторов. Идентификацию по лицу всё чаще добавляют в банковские и иные критически важные для человека приложения. Задача распознавания лиц является одной из важнейших как в исследовательской среде, так и в системах безопасности. Использование электронно-вычислительных средств цифровой обработки изображений способствует расширению функциональных возможностей устройств различной направленности. Стабилизация, автоматическая настройка экспозиции, отслеживание нарушений, охрана правопорядка, оплата проезда, отслеживание лиц и многие другие подобные задачи теперь решаются различными устройствами в реальном времени.

Увеличение вычислительных мощностей процессоров позволяет разработчику реализовывать все более сложные алгоритмы и оснащать создаваемые устройства новыми функциями, практически граничащими с ожидаемыми возможностями искусственного интеллекта. Одной из фундаментальных основ современной интеллектуальной обработки изображений является захват движения некоторого объекта в видеопотоке и последующее его отслеживание. Среди наиболее успешных продуктов с открытым исходным кодом является библиотека OpenCV. Данная библиотека может свободно использоваться в различных целях на условиях лицензии BSD.

## Постановка задачи

Целью данной работы является создание программы для захвата, идентификации и отслеживания подвижных лиц в видеопотоке с видеокамеры в реальном масштабе времени при помощи фреймворка OpenCV. Для выполнения данной цели необходимо решить ряд задач.

- 1) Разобраться с методикой работы захвата движений и определения и отслеживания объектов в видеопотоке.
- 2) Изучить фреймворк OpenCV определить его место и назначение в программе.
- 3) Разработать программу, способную захватывать и отслеживать лица в видеопотоке в реальном времени.

Необходимо указать, что для разработки программы будет использоваться язык python 3, а также фреймворки OpenCV, numpy. Данный выбор обусловлен несколькими важными вещами. Python является наиболее распространенным языком для работы с библиотеками OpenCV, а также именно с применением данного языка зачастую проводятся различные вычисления в больших данных, что делает данный язык наиболее подходящим для нынешних целей и задач. Библиотека OpenCV выбрана т.к. в отличие от других библиотек, языков и ПО таких как PCL, ROS, MATLAB, SimpleCV, является наиболее распространенной, что позволяет опираться на различные примеры кода, а также разъяснения к ним, кроме того, OpenCV имеет удобное взаимодействие с библиотекой NumPy.

Актуальность данной задачи состоит в том, что подобные наработки можно будет использовать в других, более крупных проектах как их составную часть. В настоящий момент система распознавания лиц в реальном времени является популярной и многие крупные проекты используют данную систему.

## Теоретическая часть

Биометрические методы распознавания разнообразны, к ним относят дактилоскопию, аутентификацию по радужной оболочке и сетчатке глаза, распознавание по лицу и голосу, а также подерку (рукописному или клавиатурному), термографическое наблюдение (идентификация по лицевым артериям и венам), идентификация по венам рук и т.д.

Существует несколько принципов биометрических измерений[2]:

- 1) Универсальность – каждый человек должен обладать измеряемой характеристикой
- 2) Уникальность – насколько хорошо можно идентифицировать человека по определенному параметру
- 3) Постоянство – мера того, в какой степени могут изменяться биометрические данные человека в процессе жизни, например изменения лица в процессе старения
- 4) Взыскания – простота осуществления измерения
- 5) Производительность – точность, скорость, надежность идентификации при использовании технологии
- 6) Приемлемость – степень достоверности данных
- 7) Универсальность – простота использования замены

В распознавании лица можно выделить две основные задачи:

- 1) Верификация – сравнение лица с биометрическим шаблоном.
- 2) Идентификация – сравнение одного биометрического образца с многими. Т.е. нахождение в базе биометрических шаблонов необходимого для определенного процесса, в случаях если данный шаблон уже существует в базе.

Существует достаточно много методов распознавания человека по лицу. В основе данных методов лежит тот факт, что черты лица и форма черепа человека уникальны, а следовательно, могут являться идентификатором человека, да и с точки зрения человека это является наиболее понятным вариантом распознавания, ведь и сам человек использует данный тип распознавания.

Критериями оценки идентификации могут служить показатели FAR – ложного допуска, FMR - вероятность, что система неверно сравнивает входной

образец с несоответствующим шаблоном в базе данных, FRR - коэффициент ложного отказа, FNMR - вероятность того, что система ошибётся в определении совпадений между входным образцом и соответствующим шаблоном из базы данных, Коэффициент отказа в регистрации (FTE или FER) – коэффициент безуспешных попыток создать шаблон из входных данных (при низком качестве последних), Коэффициент ошибочного удержания (FTC) - вероятность того, что автоматизированная система не способна определить биометрические входные данные, когда они представлены корректно, Ёмкость шаблона - максимальное количество наборов данных, которые могут храниться в системе.[3]

Принято выделять два основных вида распознавания лица 2D и 3D. 2D распознавание появилось на заре биометрии и является малоэффективным методом распознавания, появился данный метод благодаря криминалистике, которая его развивала и способствовала существованию. В последствие данные методы были реализованы программно, метод стал более надёжным, однако, с каждым годом данные методы все больше уступают в безопасности и точности другим методам распознавания. В настоящее время данный метод из-за плохих статистических показателей применяется в основном там, где необходимо минимизировать количество вычислений, а высоких стандартов точности не требуется, например в социальных сетях (Для автоматической идентификации на фото и видео и т.д.), а также в случаях, когда осуществляется мультимодальная биометрия (т.е. комбинация различных методов биометрии).

Основными критериями оценки алгоритмов любого типа алгоритма биометрической идентификации являются FAR (False Acceptance Rate) – коэффициент ложного пропуска, т.е. ситуации, когда была осуществлена идентификация не пользователя и FRR (False Rejection Rate) – коэффициент ложного отказа, т.е. ситуации, когда не была осуществлена идентификация пользователя, имеющего биометрический шаблон в системе. Данные показатели имеют следующие формулы:  $FAR * N^2 \approx K \Rightarrow N \approx \sqrt{\frac{K}{FAR}}$ , где N – количество уникальных биометрических данных в выборке, а K – количество допускаемых ошибок, в лучшем случае должно равняться нулю, следовательно минимальный показатель ошибок примем за 1, следовательно, формула в данном случае будет

выглядеть так  $FAR * N^2 \approx 1 \Rightarrow N \approx \sqrt{\frac{1}{FAR}}$ . Так, например, для алгоритмов VeriLook на 0.10% FAR приходится 2,50% FRR, что для современных алгоритмов 2D метода распознавания лица являются стандартом, существует также ряд проектов, имеющих при аналогичном FAR FRR порядка 0,1%, однако зачастую данный показатель получается за счет специфичной выборки (отсутствие фона, одинаковые выражения лица, схожая причёска и освещение). Из ранее выведенной формулы получаем следующее, что численность, для которой система будет работать стабильно примерно 30 человек. За последнее время алгоритмы 2D распознавания практически не развиваются и не улучшаются, однако количество проектов, прибегающих к подобному подходу идентификации продолжает расти.

Преимуществом данного метода является тот факт, что 2D метод менее требователен к системе обработки, т.е. требует меньше финансовых вложений, а также возможность при соответствующем оборудовании распознавать лица на значительных расстояниях от камеры. Недостатками данной системы распознавания является его низкая по сравнению с другими методами достоверность, а также специфичная среда для идентификации, т.е. требования к освещению, фону, внешнему виду (наличие или отсутствие бороды, очков, причёсок и т.д.). Необходимость фронтального изображения лица с минимальными отклонениями. Многие из алгоритмов 2D распознавания не воспринимают изменение мимики, т.е. выражение лица должно быть нейтральным.

Распознавание лица 3D является более сложным методом. На данный момент существует достаточно много методов для распознавания 3D лица. Различные алгоритмы проблематично сравнить т.к. данные методы базируются на различных подходах, алгоритмах, имеют различное железо и базы.

Наиболее простым, фактически переходным методом является метод накопления информации. Данный метод является более эффективным чем 2D метод, хотя, как и он может применять лишь одну камеру. При таком методе субъект должен повернуть голову, а алгоритм соединяет группу 2D изображений, фактически создавая 3D шаблон. При распознавании



используется несколько кадров из видеопотока, данный метод является крайне специфическим и экспериментальным.

Наиболее распространенным вариантом реализации метода является метод проецирования шаблона. Данный метод заключается в том, что на лицо проецируется сетка, после чего камера делает снимки со скоростью в несколько десятков кадров в секунду после чего обрабатывает их. Луч падающий на искривлённую поверхность изгибается и чем больше кривизна поверхности тем сильнее изгиб луча. По полученным снимкам восстанавливается 3D модель лица, на который выделяются и удаляются ненужные помехи (Борода, усы, причёска и т.д.), после чего проводится анализ модели, где выделяются антропометрические особенности головы, после чего данные записываются в базу данных. Время захвата и обработки у оптимальных алгоритмов занимает порядка 3х секунд.

В последнее время получают распространение алгоритмы 3D сканеров, такие как сканеры Vocord. Данный метод даёт большую точность позиционирования, по заявлениям компании FAR 0,000001% на 10000 человек, а FRR 0,000001% на 10000.

Различные алгоритмы показывают различные FRR и FAR, а некоторые алгоритмы не проверялись на FRR и FAR. Но для моделей Bioscript (3D EnrolCam, 3D FastPass), работающих по методу проецирования шаблона при FAR = 0.0047% FRR составляет 0.103%.

Считается, что статистическая надёжность метода сравнима с надёжностью метода идентификации по отпечаткам пальцев.

Преимуществом данного метода является низкая зависимость от внешних факторов человека (борода, очки, причёска), так и его окружение (освещённость, поворот головы и т.д.) Высокий уровень надёжности, сравнимый со стандартами идентификации по отпечаткам пальцев. Недостатками являются дороговизна оборудования, изменения мимики лица ухудшают сатирическую надёжность метода.

Распознавание по геометрии лица причисляют к «трем большим биометрикам» вместе с распознаванием по отпечаткам пальцев и радужной оболочке. Надо сказать, что данный метод довольно распространен, и ему отдают пока предпочтение перед распознаванием по радужке глаза. Удельный

вес технологий распознавания по геометрии лица в общем объеме мирового биометрического рынка можно оценивать в пределах 13-18 процентов. В России к данной технологии также проявляется больший интерес, чем, например, к идентификации по радужной оболочке. Как уже упоминалось ранее, существует множество алгоритмов 3-D распознавания. В большинстве своем компании предпочитают развивать готовые системы, включающие сканеры, сервера и ПО. Однако есть и те, кто предлагает потребителю только SDK. На сегодняшний день можно отметить следующие компании, занимающиеся развитием данной технологии: Geometrix, Inc. (3D сканеры лица, ПО), Genex Technologies (3D сканеры лица, ПО) в США, Cognitec Systems GmbH (SDK, специальный вычислитель, 2D камеры) в Германии, Bioscrypt (3D сканеры лица, ПО) – дочернее предприятие американской компании L-1 Identity Solutions.

В России в данном направлении работают компании Artec Group (3D сканеры лица и ПО) – компания, головной офис которой находится в Калифорнии, а разработки и производство ведутся в Москве. Также несколько российских компаний владеют технологией 2D распознавания лица – Vocord, ITV и др.

В области распознавания 2D лица основным предметом разработки является программное обеспечение, т.к. обычные камеры отлично справляются с захвата изображения лица. Решение задачи распознавания по изображению лица в какой-то степени зашло в тупик – уже на протяжении нескольких лет практически не происходит улучшения статистических показателей алгоритмов. В этой области происходит планомерная «работа над ошибками».

3D распознавание лица сейчас является куда более привлекательной областью для разработчиков. В нём трудится множество коллективов и регулярно слышно о новых открытиях. Множество работ находятся в состоянии «вот-вот и выпустим». Но пока что на рынке лишь старые предложения, за последние годы выбор не изменился.

Цифровое изображение (кадр), получаемое видеокамерой в момент времени  $t$  и имеющее разрешение по вертикали  $h$  пикселей, а по горизонтали  $w$ , обозначим  $I_t(h, w)$ .

Видеопоток — это последовательность цифровых изображений (кадров)  $I_t(h, w), I_{t+1}(h, w), \dots, I_{t+k}(h, w)$ . Под объектом понимается прямоугольная область интересов  $Ob_t(X, Y)$ , представленная множеством пикселей цифрового изображения  $I_t(h, w)$ , очерчивающих искомый объект, содержащая  $X$  пикселей по вертикали и  $Y$  по горизонтали.

Захватом объекта называется выделение области интересов  $Ob_t(X, Y)$ , на цифровом изображении  $I_t(h, w)$  в момент времени  $t$ . Отслеживанием называется захват  $Ob_t(X, Y)$  на последовательности цифровых изображений. Под захватом объекта в реальном времени понимается обработка видеопотока с частотой не менее 10 кадров в секунду.

Обучением называется предварительная настройка и задание параметров алгоритма захвата и отслеживания для обнаружения интересующего объекта. Обучение может выполняться на основе обучающей выборки, состоящей из одного или нескольких обучающих изображений. Под обучающим цифровым изображением понимается изображение  $I_t(h, w)$  с дополнительной информацией о наличии или отсутствии на нем искомого объекта. Для решения задач захвата и отслеживания создано множество алгоритмов, позволяющих проектировать актуальное программное обеспечение для различных сфер применения.

Для анализа выбраны следующие наиболее часто используемые алгоритмы для захвата и отслеживания движущихся объектов  $Ob_t(X, Y)$ , исходный код которых свободно доступен на языке программирования python в сети Интернет для экспериментов и модификации:

- шаблонов движений (Motion Templates) — основан на поиске границ объектов в каждом кадре видеопотока  $I_t(h, w)$  [3]. Смещение границы на новом кадре  $I_{t+1}(h, w)$  относительно предыдущего задает вектор движения объекта. Данный алгоритм наиболее эффективен при движении крупных объектов и часто используется для распознавания динамических жестов в человеко-машинных интерфейсах;

- сдвига среднего (Mean-Shift) — основан на математической модели, которая заключается в том, что вычисляется локальный экстремум плотности распределения набора характерных точек, т. е. алгоритм отслеживает смещение центра масс точек, определяющих объект слежения, получая на выходе вектор

движения объекта [4]. Высокая эффективность достигается при ярко выраженном цветовом различии объекта и фона;

- непрерывно адаптирующегося сдвига (CamShift) — основан на алгоритме сдвига среднего, но отличается тем, что автоматически подстраивает границы и размер окна, в пределах которого расположены характерные точки [5]. Таким образом производится более точное отслеживание объекта, изменяющегося в размерах;

- Лукаса – Канаде (Lucas – Kanade) — основан на дифференциальном вычислении оптического потока с помощью анализа пикселей (предполагается, что оптический поток одинаков для пикселей, лежащих в окрестности центра окна слежения), при этом смещение пикселей между соседними кадрами  $I_t(h, w)$  должно быть невелико [6]. Данный алгоритм более двадцати лет активно используется в приложениях компьютерного зрения и уже доказал свою высокую эффективность для широкого круга задач;

- Виолы – Джонса (Viola – Jones) — основан на обнаружении в кадре  $I_t(h, w)$  наборов пикселей, совпадающих с заранее подобранными шаблонами, состоящими из белых и черных прямоугольников [7,8]. Для распознавания различных объектов требуется свой уникальный набор шаблонов, который создается путем обучения алгоритма на конкретном объекте. Правильно обученный алгоритм работает с высокой эффективностью, однако сам процесс обучения достаточно трудоемкий, требует от разработчика-исследователя специальных знаний о данном алгоритме и хорошо организованной обучающей выборки.

### Практическая часть

На данный момент практическая часть представлена набором программ, которые можно разделить на FaceDetection и FaceRecognition. Т.е. распознавание и идентификация. Программы FaceDetection используются для распознавания, они на базе набора шаблонов в каскадах в формате xml пытаются отследить в видеопотоке лицо, а также глаза, губы и т.д.

В свою очередь программы FaceRecognition решают задачу верификации лица, на базе шаблонов, созданных при помощи одной из программ face\_dataset и тренировки при помощи face\_training. Программа face\_recognition используется непосредственно для верификации. Face\_dataset формирует датасет из набора фотографий лиц и укладывает их в папку dataset в виде User{id}.{number} где id это идентификатор пользователя, number- номер фотографии. Обычно для верификации необходимо от 20 до 30 фото.

После программа тренируется при помощи рекогнайзера, LBPHFaceRecognizer из библиотеки OpenCV. Мы берем лица из ранее записанного датасета, после чего при помощи numpy раскладываем изображение в массив, после чего возвращаем идентификатор и массив.

После чего тренируем LBPHFaceRecognizer на лицах и массиве идентификаторов. После чего результат тренировки сохраняем в виде trainer.yml. Далее переходим к программе face\_recognition, данная программа захватывает изображение и ищет сходство лица перед собой с лицами в выборке, при этом чем больше лиц содержится в тренировочной выборке, тем большая точность существует. Программа выводит сходство лица, находящегося перед камерой с определенным шаблоном из выборки.

## Заключение

Рассмотренные в теоретической части алгоритмы отслеживания, а также их практическое подкрепление при помощи средств языка python и фреймворков OpenCV, pytorch позволили детальнее разобраться в теме Захвата, идентификации и отслеживании подвижных лиц в видеопотоке с видеокамеры в реальном масштабе времени. Данные знания, безусловно, будут полезны при дальнейшем их развитии. Т.к. захват, идентификация и отслеживание лиц в реальном времени является важной темой для развития. На базе данных технологий возможно обеспечение правопорядка, слежение и контроль за соблюдением ПДД, развитие систем умного дома. Поэтому, хотя и не получилось в полной мере реализовать поставленную цель, данная работа сыграла важную роль в развитии моих навыков и умений, а выполнение большей части задач позволяет и дальше развивать тему, заложенную в данной работе.

## Список использованных источников информации

- 1) <https://ru.wikipedia.org/wiki/%D0%91%D0%B8%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D1%8F>
- 2) Jain, A. K.; Ross, Arun & Prabhakar, Salil (January 2004), An introduction to biometric recognition, IEEE Transactions on Circuits and Systems for Video Technology T. 14th (1): 4—20
- 3) ГОСТ Р ИСО/МЭК 19795-1-2007 от 25.12.2007 с изм. и доп. От 25.11.2021. пункт 4.6 <https://docs.cntd.ru/document/1200067413>
- 4) Müller M., Röder T. Motion templates for automatic classification and retrieval of motion capture data. Symposium on Computer Animation — SCA. Vienna, Austria, 2006, pp. 137–146.
- 5) Comaniciu D., Meer P. Mean Shift: A Robust Approach Toward Feature Space Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence — PAMI, 2002, vol. 24, no. 5, pp. 603–619.
- 6) Wang Z., Yang X., Xu Y., Yu S. CamShift guided particle filter for visual tracking. Pattern Recognition Letters — PRL, 2009, vol. 30, no. 4, pp. 407-413.
- 7) Baker S., Matthews I. Lucas – Kanade 20 Years On: A Unifying Framework. International Journal of Computer Vision — IJCV, 2004, vol. 56, no. 3, pp. 221–255.
- 8) Kasinski A., Schmidt A. The architecture and performance of the face and eyes detection system based on the Haar cascade classifiers. Pattern Analysis and Applications — PAA, 2010, vol. 13, no. 2, pp. 197–211.