**Правительство Российской Федерации**

**Федеральное государственное автономное образовательное**

**учреждение высшего профессионального образования**

**"Национальный исследовательский университет**

**"Высшая школа экономики"**

***Отделение программной инженерии***

***Кафедра Управления разработкой программного обеспечения***

УТВЕРЖДАЮ

Зав. кафедрой УРПО

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ С.М. Авдошин

«\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2014 г.

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

**по направлению 231000.62 Программная инженерия**

**подготовки бакалавра**

На тему: Программа многофакторной биометрической идентификации человека

Студента группы № \_471 \_\_\_\_\_\_\_\_\_\_ \_\_\_Лапидуса\_Егора\_Владимировича

подпись (Ф.И.О.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Дата)

Научный руководитель

к.т.н., Доцент кафедры УРПО \_\_\_\_\_\_\_\_\_\_\_ \_\_Белова\_Наталья\_Сергеевна\_

(должность, звание) подпись ( Ф.И.О.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Дата)

Москва, 2014 г.

# Аннотация

Отчет 28 с., 5 глав, 3 рис., 2 табл., 10 источников, 4 прил.

Ключевые слова: идентификация, биометрические данные, eigenface, fisherface, mel-frequency cepstral coefficients, vector quantization.

Объектом работы является программа многофакторной биометрической идентификации человека по двум факторам: лицо и голос.

Основная цель работы – создать систему биометрической идентификации человека, которая использует изображение лица и аудиозапись голоса как факторы для идентификации, и разработать подход объединения результатов идентификации. Комбинация из двух факторов позволяет увеличить точность определения человека и расширить круг условий окружающей среды, в которых может производиться идентификация.

В ходе работы был проведен анализ существующих подходов к проблеме идентификации человека по лицу и голосу, выбраны алгоритмы для реализации, на основе выбранных алгоритмов была построена система идентификации человека, использующая веб-камеру и микрофон для сбора необходимых биометрических данных.

В результате было практическое увеличение точности идентификации человека на 7,5 % по сравнению с использованием идентификации только по изображению и на 2,5 % по сравнению с аудио-идентификацией.

Данная система идентификации может быть использована в различных сферах, начиная от естественных интерфейсов машина-человек, заканчивая быстрой идентификацией на предприятии.

Разработка имеет большой потенциал для расширения и переноса на дополнительные платформы. В первую очередь, предполагается возможность расширения системы до комплекса биометрической идентификации, который может включать дополнительные биометрические факторы, такие как отпечатки пальцев и изображение радужной оболочки глаза. С текущими темпами совершенствования мобильных устройств открывается и путь для адаптации системы для них.

Содержание

[Введение 4](#_Toc389138100)

[1. Обзор предметной области 6](#_Toc389138101)

[1.1. Направление исследования 6](#_Toc389138102)

[1.2. Существующие методы решения поставленной задачи 6](#_Toc389138103)

[1.2.1. Идентификация по изображению лица 6](#_Toc389138104)

[1.2.2. Идентификация по записи голоса 7](#_Toc389138105)

[1.3. Подходы к проблеме идентификации человека по изображению его лица 8](#_Toc389138106)

[1.3.1. Статья «Выделение лица на фотографии в реальном времени с помощью метода главных компонент» 8](#_Toc389138107)

[1.3.2. Статья «Eigenfaces for Recognition» 8](#_Toc389138108)

[1.3.3. Статья «Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection» 8](#_Toc389138109)

[1.4. Подход к идентификации человека по записи его голоса 9](#_Toc389138110)

[1.5. Аналоги 9](#_Toc389138111)

[1.6. Методика исследования 10](#_Toc389138112)

[2. Описание используемых алгоритмов 11](#_Toc389138113)

[2.1. Собственные вектора и числа матрицы 11](#_Toc389138114)

[2.2. Алгоритм Eigenface 11](#_Toc389138115)

[2.2.1. Этап обучения 11](#_Toc389138116)

[2.2.2. Этап распознавания 13](#_Toc389138117)

[2.3. Алгоритм Fisherface 14](#_Toc389138118)

[2.4. Методы аудио-идентификации 15](#_Toc389138119)

[2.4.1. Мел-частотные кепстральные коэффициенты 15](#_Toc389138120)

[2.4.2. Векторная квантизация 16](#_Toc389138121)

[2.4.3. Удаление участков тишины 17](#_Toc389138122)

[3. Методы реализации и практические результаты 19](#_Toc389138123)

[3.1. Инструменты реализации 19](#_Toc389138124)

[3.2. Структура приложения 19](#_Toc389138125)

[3.2.1. Модуль идентификации по изображению 20](#_Toc389138126)

[3.2.2. Модуль идентификации по голосу 21](#_Toc389138127)

[3.2.3. Модуль хранения данных и интерпретации результатов 22](#_Toc389138128)

[3.3. Практические результаты 23](#_Toc389138129)

[4. Результаты эксперимента 24](#_Toc389138130)

[4.1. Тестирование модуля идентификации по изображению лица 24](#_Toc389138131)

[4.2. Тестирование модуля идентификации по аудиозаписи голоса 24](#_Toc389138132)

[4.3. Тестирование объединения факторов 25](#_Toc389138133)

[4.4. Сводные результаты 26](#_Toc389138134)

[5. Заключение 27](#_Toc389138135)

[6. Список источников 28](#_Toc389138136)

# Введение

Текущие темпы развития технологии позволяют решать все более сложные задачи и строить более сложные вычислительные системы, которые могут лучше подходить под критерии удобства использования человеком. Естественные интерфейсы предполагают взаимодействие человека и машины способами привычными именно для человека – через голосовые команды, жесты и мимику. Однако стоит отметить, что значительного развития эта область достигла лишь в последние десятилетия, когда различные достижения компьютерных наук были применены в новой сфере. Тем не менее, многие задачи требуют дальнейших исследований и разработок.

Биометрическая идентификация, основанная на таких наиболее привычных человеку факторах как лицо и голос, применима в различных технологических сегментах, начиная с интерактивных видео-игр, и заканчивая идентификацией сотрудников на предприятии. Создание подобной системы является целью данного проекта.

В рамках теоритической части проекта ставится задача поиска и сравнения различных подходов к проблеме идентификации человека на основе изображения его лица. Помимо этого выбирается подход к решению задачи аудио-идентификации. Разрабатывается архитектура программы и методика объединения результатов идентификации.

В практической секции производится реализация выбранных алгоритмов, их компоновка в единую систему, реализация способа объединения результатов идентификации, тестирование полученной системы и разработка проектной документации.

# Обзор предметной области

# Направление исследования

В рамках данного проекта производится исследование методов и подходов к решению проблемы биометрической идентификации человека. В качестве биометрических факторов были выбраны изображение лица и запись голоса. Такое решение было принято в связи с тем, что идентификация по данным параметрам является наиболее естественной для человека и не требует сложного оборудования для получения образцов.

# Существующие методы решения поставленной задачи

# Идентификация по изображению лица

На сегодняшний день существует несколько методов идентификации человека по его лицу: Eigenface, Fisherface, нейронные сети. Данная работа фокусируется на первых двух ввиду их производительности и относительно высокой точности. [1]

В основе алгоритма Eigenface лежит метод главных компонент и вычисление собственных векторов и собственных значений ковариационной матрицы. Как и другие алгоритмы machine-learning метод Eigenface предполагает две основных процедуры: обучение и тестирование. На этапе обучения формируется матрица образцов изображений тех людей, которые в дальнейшем будут идентифицироваться алгоритмом. Каждое изображения представляются в виде одного вектора, получаемого с помощью последовательной конкатенации столбцов матрицы входного файла. По всей обучающей выборке рассчитывается средний вектор, который вычитается из каждого изображения. Полученные вектора хранят лишь уникальную информацию. Далее на основе новой матрицы изображений рассчитывается ковариационная матрица. На этом шаге также происходит уменьшение размерности данных, рассчитываются собственные вектора, значения и их веса. На этом процесс обучения заканчивается. И для тестирования передается новое изображение, которое аналогичным образом проецируется в новое пространство, производится кластеризация и определение ближайшего обучающего вектора к тестовому. [3,6]

Существует два варианта расчета расстояния между тестовым вектором и векторами обучающей выборки: евклидово расстояние и расстояние Махаланобиса. Проведение эксперимента показало, что квадрат Евклидова расстояния демонстрирует лучшее качество кластеризации.

Другим рассматриваемым подходом является алгоритм Fisherface, который имеет общие черты с Eigenface, однако использует линейный дискриминантный анализ Фишера для классификации изображений. Алгоритм более сложен в реализации, требует больших объемов памяти и большего объема вычислений, но по экспериментальным данным демонстрирует лучшие результаты при анализе изображений с различным освещением. [1]

# Идентификация по записи голоса

В общем, решение проблемы идентификации человека по голосу состоит из трех основных блоков: извлечение особенностей, квантизация и создание моделей.

Аудио-сигнал, получаемый с периферийного устройства, обычно представляется в виде амплитудного спектра, который в цифровом формате представлен большим набором точек. Такое представление избыточно в контексте рассматриваемой задачи, и для уменьшения размерности данных это амплитудное представление сигнала преобразуется в набор точек на Мел-шкале, которое имеет меньшую размерность, но, вместе с тем, хранит всю необходимую специфическую информацию о сигнале.

Векторная квантизация – один из подходов к проблеме идентификации диктора. Его суть заключается в кластеризации всего набора обучающих векторов - codebook. Каждый полученный класс характеризуется своим центральным вектором – centroid. На этапе распознавания рассчитываются расстояния между тестовым вектором и каждым классом. При условии получения расстояния меньшего порогового значения принимается решение об идентификации.

Моделирование диктора – другой подход к решению задачи тексто-независимой идентификации, при котором для каждого человека из обучающей выборки строится марковская или гауссовская модель. При распознавании происходит проверка тестового сигнала на предмет соответствия параметрам построенных моделей.

# Обзор литературы по проблеме идентификации человека по изображению его лица

# Статья «Выделение лица на фотографии в реальном времени с помощью метода главных компонент»

Данная статья авторов Е. В. Еремина, Ю. Х. Раджабовой и А. А. Тельных представляет краткий обзор подхода к проблеме поиска лица на фотографии и идентификации найденного человека с помощью метода главных компонент и алгоритма Eigenface. [3]

Статья описывает преимущества метода главных компонент в контексте поставленных задач, а именно: высокая скорость обработки и применимость подхода к идентификации лица при различных углах поворота и эмоционального выражения. Авторы статьи в доступной для понимания форме описывают этапы алгоритма и архитектуру своего решения. Важной особенностью статьи является математическое описание метода уменьшения размерности ковариационной матрицы выборки для нахождения её собственных векторов и собственных значений.

# Статья «Eigenfaces for Recognition»

Предметом работы исследователей M. Turk и A. Pentland Массачусетского Технологического Института также является применение метода главных компонент к проблемам идентификации человека по изображению его лица. Данная статья содержит более детальное и глубокое описание математического аппарата рассматриваемого подхода и краткое описание применения альтернативных методов, таких как нейронные сети. Помимо этого, авторы демонстрируют свои экспериментальные результаты в точности идентификации при различных параметрах. [6]

# Статья «Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection»

Отличительной особенностью данной работы является описание альтернативного алгоритма Fisherface для идентификации. Авторы Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman описывают подходы Eigenface, Linear Subspaces и Fisherface для решения задачи идентификации и выделяют основные особенности каждого из подходов. [1]

# Обзор литературы по проблеме идентификации человека по записи его голоса

Основным источником информации по проблеме идентификации человека по аудио-записи его голоса является книга Beigi Homayoon “Fundamentals of speaker recognition”. Прежде всего, она содержит детальный разбор подходов к решению поставленной задачи. Помимо этого автор затрагивает и другие близкие задачи, такие как верификация, сегментация, классификация и трекинг диктора. Книга содержит подробную теоритическую базу предлагаемых методологий и сравнительный анализ использования звука и других биометрический факторов в контексте задачи идентификации. Немаловажным аспектом является описание структуры речевой системы и системы восприятия звука человека. [4]

Другим источником, позволяющим получить базовое понимание принципа Мел-частотных кепстральных коэффициентов, является статья Vibha Tiwari “MFCC and its applications in speaker recognition”. Данная статья содержит обзор алгоритма получения кепстральных коэффициентов из аудио-сигнала. Будучи написанной доступным языком, статья становится подходящей отправной точкой исследования. [9]

# Аналоги

Аналогичные системы многофакторной идентификации на сегодняшний день чаще всего имеют распространение в сфере производства, входя в комплекс продуктов для аутентификации, например, в охранных системах. Такие системы обычно требуют дополнительного более точного оборудования для снятия образцов, примерами могут служить IFace202 и Multi-Bio 700, производимые ZKSoftware. Обе упомянутые системы представляют собой терминалы многофакторной идентификации, работают с фотографиями лиц, отпечатками пальцев и прочими атрибутами. Плюсом таких систем является высокая надежность, обусловленная специально разработанным оборудованием. К минусам можно отнести необходимость и габариты дополнительного устройства, ограниченность расширения функционала, стоимость.

Также существует набор программных продуктов, которые не требуют сложного оборудования, а используют стандартные веб-камеры и микрофоны. Однако обычно такие программы основываются на одном из факторов. Так, например, программа FastAccess от компании Logitech предлагает идентификацию и управление паролями на основе изображения лица. GritTec Laboratory предоставляет программный комплекс для идентификации по голосу.

Таблица . Сравнение аналогов

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название системы | Ценовая политика | Многофакторность | Необходимость дополнительного оборудования | Оценка простоты использования (0-5) |
| IFace202 | Коммерческая | Изображения, отпечатки пальцев, карта | Необходим дополнительный терминал | 4 |
| Multi-Bio 700 | Коммерческая | Изображение, отпечаток пальцев, карта, пароль | Необходим дополнительный терминал | 3 |
| FastAccess | Коммерческая | Изображение | Стандартная веб-камера | 5 |
| GraitTec lab. | Коммерческая | Звук | Стандартный микрофон | 2 |

# Методика исследования

Весь проект делится на три секции: теоритическая, аналитическая и практическая. Каждая из секций предполагает свои собственные цели, задачи и результаты, необходимые для последующих этапов. Целью первого этапа является детальное изучение алгоритмов идентификации человека по изображению лица и звуку голоса. На данном этапе формируется понимание подходов к решению поставленных задач и создается математическое описание, необходимое на следующем шаге. В рамках аналитической секции выбирается наиболее подходящие алгоритмы и подходы, формируется структура будущего программного средства. Необходимость такого анализа обуславливается тем, что как изображения, так и запись голоса имеют множество факторов влияющих на качество идентификации. Различное освещение, например, может сильно искажать результаты обработки изображения, в то время как посторонние шумы влияют на качество обработки звука. Таким образом, на аналитическом этапе выбираются подходы, которые минимизируют данные эффекты. Завершающим этапом является практическая секция, в рамках которой реализуются алгоритмы и разрабатывается пакет документации.

# Описание используемых алгоритмов

# Собственные вектора и числа матрицы

Вектор и число называются собственным вектором и числом квадратной матрицы *А*, если

(1)

# Алгоритм Eigenface

Данный алгоритм предполагает, что

1. Все изображения имеют одинаковые размерности по ширине и высоте;
2. Изображения представлены в оттенках серого;
3. Изображения должны быть центрированы, т.е. глаза, нос, губы должны находиться на приблизительно одном уровне.

# Этап обучения

На данном этапе создается обучающая выборка, строится ковариационная матрица, вычисляются собственные вектора и значения, определяются их веса. [3,6]

1. Создание матрицы обучающих векторов

Получаемое на вход изображение в обобщенном виде представляет собой матрицу . Это изображение преобразуется в один вектор путём конкатенации строк.

(1)

Где I – матрица исходного изображения,

- интенсивность оттенка серого в конкретном пикселе,

*r –* размер изображения в высоту в пикселях,

*с –* размер изображения в ширину в пикселях,

− результирующий вектор изображения.

Все изображения преобразовываются в вектора и образуют матрицу обучающей выборки, где *n –* количество изображений в обучающей выборке.

1. Нормирование обучающей матрицы

На данном шаге рассчитывается среднее изображение *M.*

(2)

После чего создается новый набор векторов, каждый из которых представляет собой один из первоначальных изображений за вычетом среднего.*.*

(3)

Этот шаг необходим для того, чтобы каждый вектор нес только уникальную информацию об изображении

1. Построение ковариационной матрицы и получение собственных векторов и значений.

Ковариационная матрица, получаемая по [формуле 4] имеет размерность , где , что предполагает в дальнейшем расчет собственных векторов и значений.

*,* (4)

где .

Для уменьшения количества обрабатываемых данных рассчитывается меньшая ковариационная матрица , размерность которой равна . Соотношение собственных векторов λ и значений *u* матрицы и собственных векторов и значений матрицы продемонстрировано на формуле:

(5)

1. Расчет весов собственных значений

На данном шаге рассчитывается матрица весов, с какими каждый собственный вектор влияет на матрицу.

, (6)

где - матрица весов, размерностью ,

- матрица собственных векторов ковариационной матрицы.

# Этап распознавания

На данном этапе используются матрицы весов и собственных векторов , полученные в ходе обучения, для проецирования нового изображения в новое пространство. [3,6]

1. Преобразование входного тестового изображения в новое пространство собственных векторов.

Тестовое изображение преобразуется в вектор аналогично обучающим по [формуле 1] и нормируется по среднему [формула 3]. После чего производится преобразование:

, (7)

где - вектор преобразованного тестового изображения,

*-* вектор входного изображения.

1. Кластеризация и поиск расстояний

После перехода в новое пространство необходимо определить, к какому из обучающих векторов ближе всего находится распознаваемое изображение. Для решения этой задачи могут использоваться две метрики: евклидово расстояние и расстояние Махаланобиса.

А) Евклидово расстояние

Расстояние *d* по Евклиду между двумя векторами *a* и *b* рассчитывается следующим образом:

(8)

Б) Расстояние по Махаланобису

Расстояние по Махаланобису помимо самих векторов учитывает и ковариационную матрицу *C.* Рассчитывается по формуле:

, (9)

где – инвертированная ковариационная матрица.

В контексте рассматриваемой задачи необходимо рассчитать расстояния между вектором *w* и каждым вектором матрицы весов :

(10)

где – расстояние по Евклиду между тестовым изображением и j-тым изображением обучающей выборки,

dist(a, b) – функция, рассчитывающая евклидово расстояние между векторами a и b.

- j-тый вектор матрицы весов .

Альтернативно рассчитывается расстояние по Махаланобису:

(11)

где – расстояние Махаланобиса между тестовым изображением и j-тым изображением обучающей выборки,

- функция расчета расстояния Махаланобиса между векторами a и b по ковариационной матрице *C.*

Таким образом формируется вектор расстояний .

1. Поиск вектора с наибольшим сходством и установка критерия отсечения изображений

В полученном векторе расстоянии *d* находится минимальное значение и его индекс:

(12)

где – вектор из двух элементов, хранящий минимально расстояние и его индекс в векторе расстояний,

min(a) – функция поиска минимально значения в векторе,

ind(x, a) – функция поиска индекса элемента x в векторе a.

В результате данных операций находится минимальное расстояние и индекс соответствующего изображения обучающей выборки. Однако он рассчитывается и для изображений, которые не принадлежат обучающей выборке, для отсечения таковых вводится эмпирический критерий *O,* расстояния большие которого идентифицируются как не принадлежащие выборке.

# Алгоритм Fisherface

Данный алгоритм является расширением алгоритма Eigenface. Отличием является кластеризация обучающих векторов, для чего рассчитываются внутри-классовые и между-классовые ковариационные матрицы. [1]

, (13)

где *с –* количество классов в выборке

-вектор, принадлежащий классу ,

-средний вектор i-го класса.

, (14)

где *с –* количество классов,

- количество элементов в i-том классе,

- средний вектор i-го класса,

- средний вектор всей выборки.

Следующим шагом рассчитывается матрица главных компонент:

, (15)

где *W –* вектор тестового изображения,

– ковариационная матрица.

Завершающим шагом является расчет итоговой матрицы:

. (16)

# Методы аудио-идентификации

# Мел-частотные кепстральные коэффициенты

Впервые мел-шкала была представлена в Ньюманом, Стивенсом и Волкманом в 1937, как новое логарифмическое представление звука, более точно отражающее его восприятие человеческим ухом. Значение в 1000 mel ставится в соответствие частоте 1 Khz. Перевод из частотного представления в мел производится по следующей формуле[5]:

(17)

где *f –* значение частоты звука.

Мел-частотные кепстральные коэффициенты – набор значений на мел-шкале, содержащий только уникальную и значимую информацию исходного сигнала.

Для получения мел-частотных кепстральных коэффициентов по сигналу необходимо произвести следующие действия[2,4,9]:

1. Сигнал разбивается на перекрывающиеся окна длиной 25 – 30 мс. Длительность окна выбирается таким образом, чтобы минимизировать колебания частоты внутри ее границ, но оставить возможность последующей обработки.
2. Производится преобразование Фурье для получения спектра сигнала.
3. Накладывается набор фильтров. На данном шаге спектр сигнала распределяется между треугольными мел-частотными окнами. Этот процесс симулирует невозможность человеческого уха различать близко расположенные частоты. Для получения такого распределения каждый вектор спектра умножается на оконную функцию. В результате получается набор значений энергии сигнала, соответствующих частотным окнам.
4. Производится логарифмирование каждого значения энергии.
5. Производится дискретное косинусное преобразование для получения искомых коэффициентов.

# Векторная квантизация

Метод векторной квантизации изначально использовался в задачах сжатия аудио и видео данных. В рамках задачи распознавания диктора эта техника распределяет k-мерные вектора особенностей векторного пространства по конечному набору областей Вороного. Распределение производится алгоритмом кластеризации, например, k-means. У каждого класса есть средний вектор – центроид или кодовое слово. Набор всех кодовых слов формирует кодовый справочник. Пример кластеризации приведен на Изображении 1. [4,8]

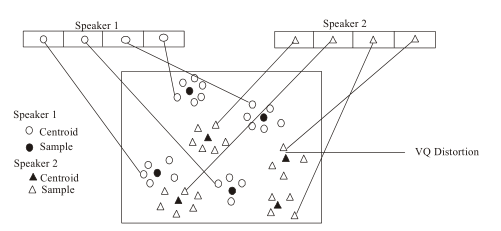


Рисунок . Пример кодового словаря [Srinivasan, 2012]

Кластеризация может производиться с помощью алгоритма k-means или алгоритмом Линде-Бузо-Грея [4]. После чего, на этапе идентификации, находится наиболее близкий кластер к тестовому сигналу.

# Удаление участков тишины

Для удаления участков записи, на которых нет значимых голосовых данных, используется алгоритм удаления тишины и определения конечных точек G. Saha, Sandipan Chakroborty, Suman Senapati [7]. В основе алгоритма лежит статистическая оценка параметров входного сигнала и разметка участков записи как содержащих запись голоса и не содержащих по определенным критериям. Данный алгоритм необходим для уменьшения влияния на качество идентификации участков тишины, когда диктор еще не начал говорить или уже закончил, и для устранения естественных участков тишины, производимых речевым аппаратом человека.

Алгоритм удаления участков тишины включает в себя следующие шаги:

1. Расчет среднего значения (форм. 18) и стандартного отклонения (форм. 19) для первых 1600 значений амплитуды входного сигнала.

, (18)

где – итоговое математическое ожидание,

*S(i)-* i-тое измерение амплитуды звукового сигнала.

, (19)

где – стандартное отклонение.

1. Для каждого значения амплитуды записи проверить, больше ли 3 значение функции расстояния Махаланобиса (формула 20).

(20)

где *R(i)*- флаг наличия голоса на i-том сэмпле,

*K-* количество измерений амплитуды в записи.

1. Разбить весь сигнал на неперекрывающиеся окна длиной 10 мс.
2. Для каждого окна рассчитать количество измерений отмеченных 1 и количество измерений отмеченных 0. Если единиц больше, чем нулей, инвертировать значения для данного окна. Данный шаг симулирует невозможность человеческого речевого аппарата изменить свои характеристики в течение такого короткого промежутка времени как 10 мс.
3. Результатом работы алгоритма являются те участки исходного сигнала, которые были помечены как вокализованные.

Блок-схема алгоритма представлена на рисунке 2.

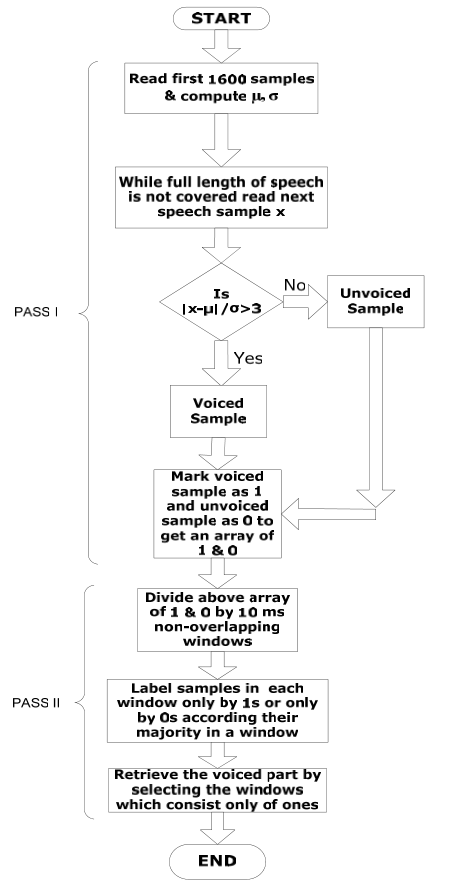


Рисунок . Алгоритм удаления тишины [7]

# Методы реализации и практические результаты

# Инструменты реализации

Данная программа многофакторной биометрической идентификации человека была разработана на языке Java, в среде разработки Intellij Idea. Для решения некоторых задач в процессе разработки были использованы следующие сторонние библиотеки:

* Sarxos Webcam – библиотека, обеспечивающая доступ к подключенным веб-камерам и захвата изображения с них;
* Slf4j, log4j – библиотека расширенных возможностей логирования, используется Sarxos webcam;
* Bridj – библиотека взаимодействия с драйверами веб-камеры, используется Sarxos webcam;
* Apache common math 3 – математическая библиотека;
* Алгоритм Fast Fourier Transform.[10]

# Структура приложения

Вся система идентификации разбита на три логических модуля: модуль идентификации по изображению, модуль идентификации по голосу, модуль хранения данных и интерпретации результатов. Взаимодействие пользователя с программой организовано через графический интерфейс. Структура приложения отражена на рис. 3.

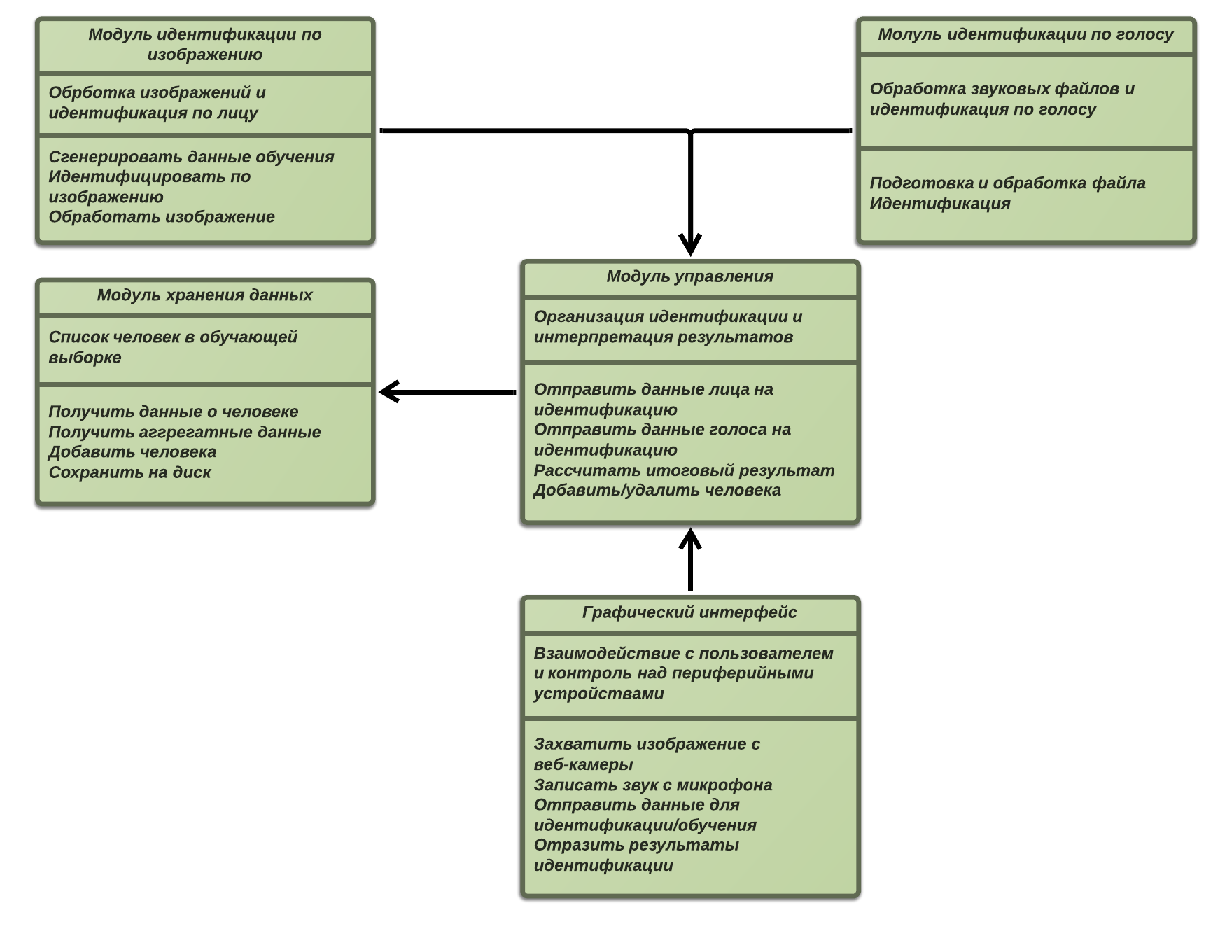


Рисунок . Схема архитектуры программы

# Модуль идентификации по изображению

Основой данного модуля является алгоритм Eigenface, который производит идентификацию на подготовленных данных в виде векторов изображений. Также данный модуль включает в себя алгоритмы предобработки растровых изображений в формате PGM и PNG.

1. Обучение

На вход алгоритма обучения поступает матрица, составленная из векторов всех изображений всех пользователей. Все векторы имеют одинаковую размерность, при работе с имеющейся выборкой изображений с размерами 92 x 112 точек, векторы имеют длину 10304 и хранят целочисленное значение интенсивности серого цвета от 0 до 255.

Обучение происходит каждый раз после добавления нового человека или изображения в базу. По завершении обучения матрица записывается на жесткий диск и выгружается из оперативной памяти.

1. Идентификация

На вход алгоритма идентификации подается целочисленный вектор тестового изображения, имеющий формат идентичный векторам обучающей выборки. Производится определение вектора, наиболее близко расположенного к данному тестовому вектору. При тестировании различных подходов к расчету расстояния были получены результаты, свидетельствующие о том, что квадрат Евклидова расстояния показывает лучшее качество выделения идентифицированного изображения среди всей выборки по сравнению с расстоянием Махаланобиса.

Эмпирически на основе 40 экспериментов было выбрано пороговое значение расстояния в 1\*10^6.

На выходе имеется индекс найденного изображения в обучающей матрице и расстояние до него от тестового вектора.

1. Предобработка изображений

Данный модуль позволяет подготавливать изображения для последующей их обработки. Так, имеются методы загрузки изображений в формате PNG и PGM. Модуль позволяет привести изображения размерностей больших, чем 92 x 112 к данным параметрам, с тем условием, что лица на изображении расположено точно по центру.

Имеется функция загрузки цветных PNG изображений с приведением их к формату оттенков серого.

# Модуль идентификации по голосу

В основе данного модуля лежит алгоритм вычисления мел-частотных кепстральных коэффициентов из аудиозаписи, кластеризации и вычисления расстояний до вектора тестового сигнала. Также данный модуль отвечает за загрузку аудиофайлов в формате WAV с частотой дискретизации 16 KHz и 16 бит на измерение.

1. Обучение

Обучение производится при добавлении нового человека в базу. В таком случае каждый выбранный аудиофайл разбивается на окна длиной 30 мс, для каждого окна рассчитываются мел-частотные кепстральные коэффициенты, формируется новый кластер, соответствующий этому человеку, для этого кластера вычисляется вектор-центроид, характеризующий речь данного человека.

1. Идентификация

На этапе идентификации также производится кадрирование сигнала, для каждого кадра рассчитываются коэффициенты, рассчитывается вектор-центроид тестового сигнала. После чего вычисляются квадраты Евклидовых расстояний от тестового вектора до каждого вектора обучающей выборки. Расстояние возводится в квадрат для получения более точных результатов идентификации – расстояния до тех кластеров, к которым не относится тестовый вектор, становится существенно большим по сравнению с расстоянием до кластера, к которому этот сигнал принадлежит.

Эмпирически на основе 40 экспериментов пороговое значение расстояния идентификации было установлено равным 5.

1. Определение тишины на записи

Каждый сигнал при подготовке к извлечению мел-частотных кепстральных коэффициентов проходит этап предобработки, на котором статистически вычисляются и удаляются участки тишины.

1. Загрузка аудиофайла

Модуль позволяет загружать аудиофайлы в формате WAV (16 KHz, 16 бит на сэмпл). Модуль проверяет соответствие файла данным параметрам, считывая информацию о кодировании из заголовка файла.

# Модуль хранения данных и интерпретации результатов

Данный модуль отвечает за хранение обучающих данных для каждого человека, организовывает взаимодействие между графическим интерфейсом и модулями идентификации, интерпретирует и объединяет результаты идентификации по обоим факторам.

1. Добавление нового человека в выборку

При добавлении нового человека в выборку производится получение входных данных от графического интерфейса: выбранные звуковые файлы и изображения или данные полученные с микрофона и веб-камеры, имя нового человека. Данные передаются в модули идентификации, где из них выделяются вектора особенностей, сохраняются в память. Для дальнейшей работы сами файлы, за исключением одного изображения, не нужны, в виду экономии системных ресурсов. Каждый человек представляется набором векторов изображений его лица, векторами мел-кепстральных частотных коэффициентов и одним изображением его лица для наглядного отражения результата идентификации.

1. Идентификация

Для идентификации производится аналогичная загрузка изображения и данных голоса.

Загружается матрица обучающей выборки изображений и вместе с тестовым вектором передается в модуль идентификации по изображению, где происходит идентификация.

Собираются вектора-центроиды голоса по всем кластерам, передаются вместе с тестовым вектором в модуль идентификации по голосу.

На выходе обоих модулей идентификации имеются данные об индексе и удалении наиболее близкого вектора изображения и наиболее близкого вектора звука. Данные результаты объединяются следующим образом:

А) Если результат идентификации по изображению выше порогового, а результат идентификации по голосу – ниже, то возвращается результат идентификации по голосу;

Б) Если результат идентификации по голосу выше порогового, а результат идентификации по изображению – ниже, то возвращается результат идентификации по изображению;

В) Если оба результата указывают на одного человека, то возвращается общий результат;

Г) Если оба результата ниже порогового, но указывают на разных людей, то возвращается результат отрицательной идентификации;

Д) Если оба результата выше порогового и указывают на разных людей, то возвращается результат отрицательной идентификации.

# Результаты эксперимента

# Тестирование модуля идентификации по изображению лица

Объект тестирования – модуль идентификации по изображению лица.

Входные данные – открытая выборка лиц Cambridge University Computer Laboratory, «**The Database of Faces**» (источн). Данная выборка содержит по 10 изображений для 50 человек в формате PGM (92 x 112 пикслей).

Обучающая выборка:

40 человек;

3 изображения на человека.

Тестовая выборка:

40 человек;

3 изображения на человека – 120 тестов.

Результаты идентификации:

93 успешных идентификаций;

12 неверно идентифицировано (ниже порогового значения)

15 неверно идентифицировано (выше порогового значения)

77,5 % точность идентификации.

# Тестирование модуля идентификации по аудиозаписи голоса

Объект тестирования – модуль идентификации по аудиозаписи голоса.

Входные данные выборка записей голосов людей CMU(). Также имеет свободный доступ для исследовательских целей, представляет собой наборы аудио коротких аудиозаписей (16 KHz, 16 бит) от различных людей.

Обучающая выборка:

40 человек

21 секунда записи голоса

Тестовая выборка:

40 человек

12 секунд записи голоса – 40 тестов.

Результаты идентификации:

33 верно идентифицировано

7 неверно идентифицировано (ниже порогового значения)

82,5 % верных идентификаций

# Тестирование объединения факторов

Обучающая выборка:

40 человек

3 изображения на человека

21 секунда аудиозаписи

Тестовая выборка:

40 человек

3 изображений на человека.

3 аудиозаписи по 5 секунд

120 тестов

Результаты идентификации:

102 верно идентифицировано

18 неверно идентифицировано (ниже порогового значения)

85 % верной идентификации.

Таблица . Сравнение результатов

|  |  |  |  |
| --- | --- | --- | --- |
| Идентификация | Количество тестов | Количество верных идентификаций | Процент ошибок |
| По изображению лица | 120 | 93 | 23.5 % |
| По аудиозаписи голоса | 40 | 33 | 17,5 % |
| Двухфакторная идентификация | 120 | 102 | 15 % |

# Заключение

Задачи идентификации человека по изображению лица и аудиозаписи голоса находятся среди наиболее сложных задач естественных интерфейсов машина-человек. Таким образом, разработка методик для решения этих задач имеет высокий научный интерес. Помимо этого, практическую важность биометрической идентификации сложно переоценить.

Тем не менее, не каждый биометрический фактор легко доступен для получения и обработки. Так, например, получение изображения радужной оболочки глаза требует камеру с высоким разрешением, получение отпечатков пальцев – дополнительное устройство для получения образцов. В свою очередь, фотографию лица и запись голоса можно получить при помощи стандартного оборудования, имеющегося практически у каждого пользователя ПК.

В процессе работы были проанализированы различные подходы к идентификации человека по лицу и голосу, принято решение о методах реализации, разработаны способы объединения результатов, реализована сама система, не уступающая по точности аналогичным разработкам других исследователей.

# Список источников

1. Belhumeur, P. & Hespanha, J. & Kriegman, D. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 711-721.*
2. Ejnarsson, M. & Nillson, M. (2002). *Speech Recognition using Hidden Markov Model*. Blekinge Institute of Technology.
3. Eremin, E. & Radzhabova, U. & Telnikh, A. (2003). Real-time face etection via principal component analysis. *Neuroinformatics-2003, 2*(5), 115-121.
4. Homayoon B. (2011). *Fundamentals of Speaker Recognition.* New York: Springer.
5. Newman, E. & Stanley, S. & Stevens, S. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America* *8*(3): 185–190.
6. Pentland, A. & Turk, M. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience* 3(1): 71-86.
7. Saha, G. & Sandipan, C. & Suman, S. (2004). *A New Silence Removal and Endpoint Detection Algorithm for Speech and Speaker Recognition Applications*. Indian Institute of Technology, Khragpur.
8. Srinivasan. A. (2012). Speaker Identification and Verification using Vector Quantization and Mel Frequency Cepstral Coefficients. *Research Journal of Applied Sciences, Engineering and Technology 4*(1), 33-40.
9. Tiwary, V. (2010). MFCC and its applications in speaker recognition. *International Journal on Emerging Technologies 1*(1), 19-22.
10. OrangeCow organization. OC Volume: [Электронный документ].- *(*[*http://ocvolume.sourceforge.net/ocvolume.php*](http://ocvolume.sourceforge.net/ocvolume.php)*).* Доступ 15.04.2014.