МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ «ГРОДНЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ ЯНКИ КУПАЛЫ»

На правах рукописи

УДК 004.93’1

ВЕРЕТИЛО

ЮРИЙ НИКОЛАЕВИЧ

СИСТЕМА ЗАЩИТЫ ПЕРСОНАЛЬНЫХ БИОМЕТРИЧЕСКИХ ДАННЫХ

Магистерская диссертация на соискание степени
магистра техники и технологий

по специальности
1-53 81 02 Методы анализа и управления в технических и экономических системах

Научный руководитель:
кандидат технических наук, доцент,
Ассанович Б.А.

Допущена к защите _____
(дата)

Кадан А.М. __ _____
(ФИО и подпись заведующего кафедрой)

Гродно, 2018

Посвящается моей сестре Анне

ОГЛАВЛЕНИЕ

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

БЧХ – Код Боуза – Чоудхури – Хоквингема.

СКУД – Система контроля и управления доступом.

CLNF – (англ. Constrained Local Neural Fields) – Ограниченное локальное нейронное поле.

CNF – (англ. Conditional Neural Fields) – Условное нейронное поле.

ERR – (англ. Equal Error Rate) – Коэффициент равного уровня ошибок.

FAR – (англ. False Acceptance Rate) – Коэффициент ложного пропуска.

FHD – (англ. Fractional Hamming Distance) – Дробное расстояние Хэмминга.

FRR – (англ. False Rejection Rate) – Коэффициент ложного отказа.

Fuzzy Commitment – «Нечёткое обязательство» – метод защиты биометрических шаблонов, представленных в виде двоичных строк фиксированной длины.

HD – (англ. Hamming Distance) – Расстояние Хэмминга.

HOG – (англ. Histogram of Oriented Gradients) – Гистограмма направленных градиентов.

Imprint – Цифровой отпечаток пользователя.

LNF – (англ. Local Neural Field) – Локальное нейронное поле.

PCA – (англ. Principal Component Analysis) – Метод главных компонент.

SVM – (англ. Support Vector Machine) – Метод опорных векторов.

ВВЕДЕНИЕ

В последнее десятилетие биометрия стала ценным средством автоматического распознавания человека, на основе его физиологических или поведенческих характеристик, из-за нескольких присущих им преимуществ, которые они предлагают по сравнению с обычными методами.

Системы распознавания на основе биометрических систем, основанные на личных чертах, как биологических, так и поведенческих, намного сложнее потерять, забыть, украсть, скопировать или подделать, чем традиционные идентификаторы. Недавние технологические разработки сделали возможным развертывание систем на основе биометрических систем, в которых используются биометрические данные, такие как лицо, радужная оболочка и отпечатки пальцев, в широком спектре применений: уголовные расследования, гражданская регистрация, пограничный контроль, проверка подлинности национальных удостоверений, коммерция, электронный банкинг, онлайн-платежи, физический и логический контроль доступа.

При разработке системы аутентификации на основе биометрических данных возникают различные проблемы, которые должны быть приняты во внимание. Как установлено в литературе, с идеальной точки зрения биометрия должна быть универсальной, уникальной, постоянной, собираемой и приемлемой. Кроме того, помимо выбора используемой биометрии, на этапе проектирования необходимо учитывать многие другие вопросы: точность системы, вычислительная скорость и стоимость являются важным параметром проектирования, особенно для систем, которые предназначены для больших групп населения.

Распознавание людей на основе биометрии ставит новые проблемы, связанные с защитой данных, не подкрепленной традиционными методами распознавания. Если биометрические данные похищаются злоумышленником, они могут быть воспроизведены и использованы неправильно. Биометрия пользователей не может быть изменена, если скомпрометирована, отличная от ПИН-кода или пароля, который может быть переиздан при необходимости. Более того, использование биометрии создает дополнительные проблемы конфиденциальности, поскольку биометрические данные могут выявлять конфиденциальную информацию о личности и здоровье человека, которые могут храниться, обрабатываться и распространяться без разрешения пользователей. Эта информация может использоваться для дискриминации людей, например, путем отказа в страховании людям со скрытыми проблемами со здоровьем. Более

того, уникальность биометрии среди отдельных лиц позволяет осуществлять перекрестное сопоставление с биометрическими базами данных, что позволяет осуществлять несанкционированное отслеживание деятельности субъектов. Кроме того, в сценариях, когда правительственные учреждения или частные компании могут собирать огромные базы данных о биометрии граждан, можно было бы предусмотреть некоторые риски для личной неприкосновенности и достоинства. Все это ведет к потере конфиденциальности пользователей.

Поэтому возникает необходимость защищать конфиденциальность и безопасность с процессуальной, юридической и технологической точек зрения.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Связь работы с крупными научными программами (проектами) и темами

Работа связана с европейскими проектами по защите персональных данных:

1. Проект COSTIC 1206 «Де-идентификация персональных данных в мультимедийном контенте».
2. COST проект CA16101 «Мультимодальная обработка изображений для криминалистики».

Цель и задачи

Цель: Разработка системы защиты персональных биометрических данных на основе обработки наборов изображений лиц человека.

Задачи:

1. Разработать алгоритм создания биометрического кода пользователя с использованием HOG-структур и маски, сокращающей длину представления кода.
2. Разработать процедуру аутентификации биометрического кода пользователя с применением корректирующих ошибки БЧХ-кодов.
3. Создать программное обеспечение, реализующее задачи создания биометрического кода и выполнения аутентификации, с использованием программных компонент OpenFace.
4. Провести апробацию предложенных и реализованных алгоритмов и процедур с использованием общедоступной базы изображений лиц человека Caltech Faces.

Положения, выносимые на защиту

1. Алгоритм создания биометрического кода пользователя из его цифровой записи (imprint) путем анализа HOG-структур.
2. Алгоритм аутентификации пользователя на основе применения помехоустойчивых кодов БЧХ с высокой избыточностью.
3. Программное обеспечение для считывания, обработки и анализа характеристик биометрических данных пользователя.

Личный вклад соискателя

1. Предложение структуры маски биометрического кода пользователя с использованием HOG-структур.
2. Предложение формата для представления данных кода БЧХ (511, 58, 91) и (511, 28, 111).
3. Реализация основных процедур считывания, анализа и обработки характеристик биометрических данных пользователя.

Апробация результатов диссертации

Результаты исследований, включенные в диссертацию, были доложены на конференциях:

1. «Физика конденсированного состояния» – XXV международная научно-практическая конференция аспирантов, магистрантов и студентов.
2. «Информационные технологии и системы 2017 (ИТС 2017) / Information Technologies and Systems 2017 (ITS 2017)» – международная научная конференция.

Опубликованность результатов диссертации

1. Создание биометрической базы данных лиц на основе HOG-структур / Ю. Н. Веретило // «Физика конденсированного состояния» – XXV международная научно-практическая конференция аспирантов, магистрантов и студентов.
2. Биометрическая база данных на основе HOG-структур и кодов БЧХ / Б.А.Ассанович / Ю.Н.Веретило // «Информационные технологии и системы 2017 (ИТС 2017) = Information Technologies and Systems 2017 (ITS 2017)» – международная научная конференция.

Структура и объем диссертации

Диссертация представлена в пяти главах, содержит 60 печатных страниц, 18 рисунков (7 страниц), 12 таблиц (3 страницы), 45 источников литературы и одно приложение.

ГЛАВА I
ОБЗОР ЛИТЕРАТУРЫ, МЕТОДОВ И ТЕХНОЛОГИЙ

1.1 Аналитический обзор литературы

Биометрические криптосистемы предоставляют средства для адаптации криптографических протоколов к биометрическим данным, которые являются по своей сути шумными данными. Они могут быть разделены на схемы генерации ключей, где двоичные ключи создаются непосредственно из приобретенной биометрии и схемы привязки ключей, которые хранят информацию, полученную путем объединения биометрических данных со случайно сгенерированными ключами.

Основная проблема, связанная с подходом к генерации ключей, связана с возможностью создания нескольких ключей из одной и той же биометрии без использования каких-либо внешних данных и стабильности полученного криптографического ключа. Более того, из-за трудностей в управлении внутриклассовой изменчивостью биометрических данных эффективность распознавания таких схем обычно значительно ниже, чем у их незащищенных аналогов [1].

Система привязки ключей может быть двоякой: ее можно использовать для защиты биометрического шаблона с помощью двоичного ключа; для обеспечения биометрической системы распознавания или для выпуска криптографического ключа только тогда, когда его владелец представляет конкретную биометрическую характеристику. В обоих случаях секретный ключ, независимо от рассматриваемой биометрии, объединяется во время регистрации с эталонным шаблоном для создания некоторых общедоступных данных, так называемых вспомогательных данных, из которых это должно быть невозможно или, по крайней мере, вычислительно трудно, извлекать информацию об исходном биометрическом признаке или ключе. Затем вспомогательные данные используются в сочетании с запросом биометрии во время распознавания для извлечения ключей. Как правило, эти подходы позволяют управлять внутри-пользовательскими вариациями биометрических данных, используя возможности кодов коррекции ошибок.

В сценарии генерации ключевых слов основная проблема проектирования связана с изменчивостью биометрических признаков. Поэтому многие усилия направлены на получение надежных ключей из шумных биометрических данных. В [2] и [3] криптографические ключи были созданы из голоса и лиц соответственно. Значительная активность была посвящена генерации ключей сигнатур. Как было предложено в [4] и более подробно изложено в [5], из каждой

динамической сигнатуры был выделен набор параметрических признаков, а для правильного распознавания использовался верхний и нижний допустимые пороговые значения. Аналогичный подход был предложен в [6]. Оба метода обеспечивают защиту шаблонов подписи. Тем не менее, изменчивость каждой функции должна быть явно доступна, и оба метода не обеспечивают возможности обновления шаблонов. В [7] сохранение и возобновление биометрической секретности было получено путем применения случайных токенов вместе с многобитовой дискретизацией и перестановкой функциональных функций, извлеченных из подписей. В [8] биометрические ключи были сгенерированы с использованием алгоритма генетического отбора и применены к динамической онлайн подписи.

В сценарии привязки ключей среди криптографических протоколов чаще всего упоминается нечеткое обязательство [9], в котором секретный ключ пользователя закодирован, а результат складывается с биометрическим шаблоном для обеспечения безопасности и конфиденциальности шаблона. Более подробно подход, предложенный в [9], описывается в [10], где исследуется роль кодов коррекции ошибок, используемых в рамках безопасного биометрического распознавания, и обеспечивает лучшую устойчивость к шумной биометрии. Чтобы справиться с набором неупорядоченных данных в [11] был введен протокол нечеткого хранилища. И нечеткое обязательство, и нечеткое хранилище широко использовались для биометрических систем, основанных на разных идентификаторах. Схема fuzzy commitment была применена к биометрике уха [12], отпечатку пальца [13, 14], 2D-лицам [15], 3D-лицам [16], радужной оболочке глаза [17, 18] и онлайн-подписи [19, 20].

В [21] введены два примитива - нечеткий экстрактор и безопасный эскиз. Первый извлекает равномерно случайную строку из ввода с допуском, устойчивым к ошибкам, то есть таким образом, что даже если фактический ввод отличается от исходного, который все еще остается закрытым, строка может быть точно восстановлена. Второй позволяет точно перестроить ввод, используя некоторую общедоступную информацию, извлеченную из него, а именно эскиз, который не отображает значительную информацию о самом входе и шумную реплику ввода, достаточно близкую к исходной. В [22] практические вопросы, связанные с разработкой безопасной системы эскиза, были проанализированы с конкретным применением для биометрии. В [23] нечеткие экстракторы использовались в настройках, где данные, полученные при регистрации и проверке, хранятся в разных представлениях. Доказательство концепции было дано с применением отпечатков пальцев. В [24] рассмотрены нечеткие экстракторы для непрерывных исходных данных, а в [25] предложены нечеткие экстракторы для непрерывной области с применением к лицам.

В последние несколько лет некоторые усилия были также посвящены разработке механизмов защиты шаблонов для мульти-биометрических систем. Хотя развитие темы все еще находится в зачаточном состоянии, некоторые интересные решения уже были предложены. В [26] шаблоны лица и отпечатков пальцев были объединены на уровне объектов и защищены с использованием схемы fuzzy commitment. В [27] предложена мульти-биометрическая система, основанная на слиянии на уровне признаков отпечатков пальцев и радужной оболочки глаза и защищены с использованием схемы нечеткого хранилища. В рамках конструкции нечеткого обязательства в [28] были исследованы различные формы слияния, в частности особенности оценки и слияние решений. В [29] предложена мульти-биометрическая система, объединяющая радужную оболочку глаза и лицо для получения длинного криптографического ключа с высокой энтропией. В [30] предложен подход слияния на уровне объектов для реализации мульти-биометрических криптосистем, основанный на использовании как нечеткого обязательства, так и нечеткого хранилища. В частности, используются одновременно отпечатки пальцев, радужная оболочка и лицо.

В последние годы алгоритмы обнаружения лиц были заметно улучшены. Однако они все еще борются с плохими условиями освещения и не четкими выражениями лица в кадре. В [31] представлена ограниченная локальная модель нейронного поля для обнаружения ключевых точек лица. Модель включает в себя две основные новинки. Во-первых, вводится вероятностный локальный детектор (детектор ключевых точек), который может обучаться на нелинейных и пространственных отношениях между входными пикселями и вероятностью выравнивания шаблона лицевых ключевых точек. Во-вторых, модель оптимизирована с использованием новой методики неравномерного регулируемого среднего сдвига, который учитывает надежность каждого локального детектора. Продемонстрирована польза подхода к ряду общедоступных наборов данных по сравнению с другими современными подходами при выполнении обнаружения лицевых точек в плохих условиях освещения и на фоне природы.

Разработка безопасных и конфиденциальных биометрических систем представляет собой сложную проблему, которая включает в себя несколько дисциплин, от законодательства и этики до обработки сигналов, распознавания образов, теории информации и криптографии. Поэтому на пути к вышеупомянутой цели существует несколько способов достижения ощутимых результатов.

Что касается безопасности, система обычно называется сильной системой, когда контроль атак больше, чем потенциальное преимущество

злоумышленника. Напротив, слабая система – это система, для которой контроль атак ниже соответствующего потенциального преимущества злоумышленника. До сих пор определение уровня безопасности в биометрических системах осуществлялось путем выявления возможных атак, уязвимостей, возможных контрмер и глобального анализа затрат. Нелегко определить безопасность, которая обеспечивается определенной системой и, в частности, биометрической системой в количественном, а не качественном виде. Поэтому необходимо предпринять значительные усилия для определения показателей, которые будут использоваться для оценки эффективности системы с точки зрения достигнутого уровня безопасности.

Со ссылкой на схемы защиты биометрических шаблонов к настоящему времени были предложены различные принципы и практики классификации и систематизации с риском потенциальной путаницы. Поэтому научное сообщество нуждается в гармонизации словаря. В настоящее время органах стандартизации проводятся некоторые мероприятия для достижения этой цели. Стоит отметить, что некоторые показатели, предназначенные для характеристики конкретных биометрических систем защиты шаблонов, уже были предложены. С другой стороны, когда трансформация, основанная на шаблоне различна, необходимо определить различные показатели оценки эффективности. Поэтому определение целостного подхода может стать значительным достижением

В недавнем прошлом мульти-биометрические системы вызвали интерес со стороны научного сообщества благодаря их внутренним возможностям, решению проблемы универсальности, лучше, чем унимодальным системам, и повышению уровня безопасности. Тем не менее, комплексный анализ возможных дополнительных угроз, атак, уязвимостей и контрмер, характерных для мульти-биометрических систем, по-прежнему необходимо систематически проводить. Кроме того, вопрос о разработке подходов к защите шаблонов для мульти-биометрических систем, все еще находящихся в зачаточном состоянии, является плодотворной областью исследований. Кроме того, оценка эффективности вышеупомянутых систем требует надлежащих процедур и показателей, которые еще предстоит разработать.

Стоит отметить, что в прошлом уделялось больше внимания обеспечению безопасности, а не разработке систем, отвечающих требованиям конфиденциальности. Только недавно конфиденциальность и безопасность были рассмотрены как два фактора, которые должны быть совместно оптимизированы, а не как два отдельных требования, мешающие друг другу. Это привело к необходимости включения требований конфиденциальности на раннем этапе разработки биометрической системы. Привлекательные темы

исследований включают анализ рисков конфиденциальности, определение необходимых требований для обеспечения конфиденциальности частной жизни, разработку надлежащих передовых методов, архитектур и систем с целью реализации необходимых ограничений конфиденциальности. Наконец, требуется этап тестирования для оценки того, были ли выполнены требования конфиденциальности. Моделирование и количественное определение свойств конфиденциальности, таких как анонимность, неприкосновенность и т.д., являются важными шагами в направлении глубокого понимания того, что предназначено для обеспечения конфиденциальности и определения показателей, необходимых для оценки уровня защиты конфиденциальности, предоставляемого различными биометрическими системами. Однако сохранение конфиденциальности - это многодисциплинарная область исследований, которая имеет соответствующие юридические, социальные, экономические, политические и культурные аспекты, которые необходимо понимать глубоко и применять для разработки эффективных подходов к защите частной жизни человека. Поэтому для успешного решения проблемы защиты частной жизни в биометрических системах необходим исследовательский опыт вне инженерных наук.

1.2 Обзор основных технологий и методов биометрической идентификации

Биометрическая идентификация – это предъявление пользователем своих уникальных биометрических параметров и процесс сравнения их со всей базой имеющихся данных. Для извлечения такого рода персональных данных используются биометрические считыватели.

Биометрические системы контроля доступа удобны для пользователей тем, что носители информации находятся всегда при них, не могут быть утеряны либо украдены. Биометрический контроль доступа считается более надежным, т.к. идентификаторы не могут быть переданы третьим лицам, скопированы.

1.2.1 Технологии биометрической идентификации

Методы биометрической идентификации:

1. Статические, основанные на физиологических признаках человека, присутствующих с ним на протяжении всей его жизни:

- Идентификация по отпечатку пальца;
- Идентификация по лицу;

- Идентификация по радужной оболочке глаза;
- Идентификация по геометрии руки;
- Идентификация по термограмме лица;
- Идентификация по ДНК;
- Идентификация на основе акустических характеристик уха;
- Идентификация по рисунку вен.

2. Динамические берут за основу поведенческие характеристики людей, а именно подсознательные движения в процессе повторения какого-либо обыденного действия:

- Идентификация по голосу;
- Идентификация по рукописному почерку;
- Идентификация по клавиатурному почерку;
- и другие.

Одним из приоритетных видов поведенческой биометрии – манера печатать на клавиатуре. При ее определении фиксируется скорость печати, давление на клавиши, длительность нажатия на клавишу, промежутки времени между нажатиями.

Отдельным биометрическим фактором может служить манера использования мыши. Помимо этого, поведенческая биометрия охватывает большое число факторов, не связанных с компьютером: походка, особенности того, как человек поднимается по лестнице.

Существуют также комбинированные системы идентификации, использующие несколько биометрических характеристик, что позволяет удовлетворить самые строгие требования к надежности и безопасности систем контроля доступа.

1.2.2 Критерии биометрической идентификации

Для определения эффективности СКУД на основе биометрической идентификации используют следующие показатели:

- FAR – коэффициент ложного пропуска;

- FMR – вероятность того, что система неверно сравнивает входной образец с несоответствующим шаблоном в базе данных;
- FRR – коэффициент ложного отказа;
- FNMR – вероятность того, что система ошибется в определении совпадений между входным образцом и соответствующим шаблоном из базы данных;
- График ROC – визуализация компромисса между характеристиками FAR и FRR;
- 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;
- Емкость шаблона – максимальное количество наборов данных, которые могут храниться в системе.

1.2.3 Сравнительный анализ основных методов биометрической идентификации

Сравнение методов биометрической аутентификации с использованием математической статистики (FAR и FRR). Главными для оценки любой биометрической системы являются два параметра:

- FAR - коэффициент ложного пропуска, т.е. процент возникновения ситуаций, когда система разрешает доступ пользователю, незарегистрированному в системе;
- FRR - коэффициент ложного отказа, т.е. отказ в доступе настоящему пользователю системы.

Обе характеристики получают расчетным путем на основе методов математической статистики. Чем ниже эти показатели, тем точнее распознавание объекта.

Для самых популярных на сегодняшний день методов биометрической идентификации значения FAR и FRR представлены в таблице 1.1.

Таблица 1.1 - Значения FAR и FRR популярных методов биометрической идентификации

Биометрическая система использует :	**FAR**	**FRR**
Отпечаток пальца	0,001%	0,6%
Распознавание лица 2D	0,1%	2,5%
Распознавание лица 3D	0,0005%	0,1%
Радужная оболочка глаза	0,00001%	0,016%
Сетчатка глаза	0,0001%	0,4%
Рисунок вен	0,0008%	0,01%

Для построения эффективной системы контроля доступа недостаточно отличных показателей FAR и FRR. Например, сложно представить СКУД на основе анализа ДНК, хотя при таком методе аутентификации указанные коэффициенты стремятся к нулю. Зато растет время идентификации, увеличивается влияние человеческого фактора, неоправданно возрастает стоимость системы.

Таким образом, для качественного анализа биометрической системы контроля доступа необходимо использовать и другие данные, получить которые, порой, возможно только опытным путем.

В первую очередь, к таким данным нужно отнести возможность подделки биометрических данных для идентификации в системе и способы повышения уровня безопасности. Во-вторых, стабильность биометрических факторов: их неизменность со временем и независимость от условий окружающей среды. Как логичное следствие, – скорость аутентификации, возможность быстрого бесконтактного снятия биометрических данных для идентификации. И, конечно, стоимость реализации биометрической СКУД на основе рассматриваемого метода аутентификации и доступность составляющих.

1.2.4 Сравнение биометрических методов по устойчивости к фальсификации данных и возможности строгой аутентификации

Фальсификация биометрических данных это в любом случае достаточно сложный процесс, зачастую требующий специальной подготовки и технического сопровождения. Но если подделать отпечаток пальца можно и в домашних

условиях, то об успешной фальсификации радужной оболочки - пока неизвестно. А для систем биометрической аутентификации по сетчатке глаза создать подделку попросту невозможно.

Повышение уровня безопасности биометрической системы контроля доступа, как правило, достигается программно-аппаратными методами. Например, технологии «живого пальца» для отпечатков, анализ непроизвольных подрагиваний – для глаз. Для увеличения уровня безопасности биометрический метод может являться одной из составляющих многофакторной системы аутентификации. Результаты сравнений приведены в таблице 1.2.

Таблица 1.2 - Возможности фальсификации и строгой аутентификации

Биометрическая система использует :	**Фальсификация**	**Строгая аутентификация (один фактор)**
Отпечаток пальца	Возможна	Возможна
Распознавание лица 2D	Возможна	Нет
Распознавание лица 3D	Проблематична	Нет
Радужная оболочка глаза	Безуспешна	Возможна
Сетчатка глаза	Невозможна	Возможна
Рисунок вен	Невозможна	Возможна

Включение в программно-аппаратный комплекс дополнительных средств защиты обычно довольно ощутимо увеличивает его стоимость. Однако, для некоторых методов возможна строгая аутентификация на основе стандартных составляющих: использование нескольких шаблонов для идентификации пользователя (например, отпечатки нескольких пальцев).

1.2.5 Сравнение методов аутентификации по неизменности биометрических характеристик и чувствительности к внешним факторам

Неизменность биометрической характеристики с течением времени понятие также условное: все биометрические параметры могут измениться вследствие медицинской операции или полученной травмы. Но если обычный бытовой порез, который может затруднить верификацию пользователя по отпечатку пальца – ситуация обычная, то операция, изменяющая рисунок радужной оболочки глаза – редкость.

Влияние параметров окружающей среды на эффективность работы СКУД зависит от алгоритмов и технологий работы, реализованных производителем

оборудования, и может значительно отличаться даже в рамках одного биометрического метода. Ярким примером подобных различий могут послужить считыватели отпечатков пальцев, которые в целом довольно чувствительны к влиянию внешних факторов.

Если сравнивать остальные методы биометрической идентификации – самым чувствительным окажется распознавание лиц 2D: здесь критичным может стать наличие очков, шляпы, новой прически или отпущенной бороды.

Системы, использующие метод аутентификации по сетчатке, требуют довольно жесткого положения глаза относительно сканера, неподвижности пользователя и фокусировки самого глаза.

Методы идентификации пользователя по рисунку вен и радужной оболочке глаза сравнительно стабильны в работе, если не пытаться использовать их в экстремальных условиях работы (например, бесконтактная аутентификация на большом расстоянии во время «грибного» дождя).

Наименее чувствительна к влиянию внешних факторов трехмерная идентификация по лицу. Единственным параметром, который может повлиять на работу подобной СКУД, является чрезмерная освещенность. Результаты сравнений приведены в таблице 1.3.

Таблица 1.3 - Неизменность характеристик и чувствительность к влиянию внешних факторов

Биометрическая система использует :	**Неизменность характеристики**	**Чувствительность к влиянию внешних факторов**
Отпечаток пальца	Низкая	Высокая
Распознавание лица 2D	Низкая	Высокая
Распознавание лица 3D	Высокая	Низкая
Радужная оболочка глаза	Высокая	Средняя
Сетчатка глаза	Средняя	Высокая
Рисунок вен	Средняя	Средняя

1.2.6 Сравнение по скорости аутентификации и возможности бесконтактной аутентификации

Скорость аутентификации зависит от времени захвата данных, размеров шаблона, объема ресурсов, отведенных на его обработку, и основных

программных алгоритмов, применяемых для реализации конкретного биометрического метода.

Бесконтактная аутентификация дает массу преимуществ использования биометрических методов в системах физической безопасности на объектах с высокими санитарно-гигиеническими требованиями (медицина, пищевая промышленность, научно-исследовательские институты и лаборатории). Кроме того, возможность идентификации удаленного объекта ускоряет процедуру проверки, что актуально для крупных СКУД с высокой поточностью. А также, бесконтактная идентификация может использоваться правоохранительными органами в служебных целях. Именно поэтому ученые стремятся разработать бесконтактные системы аутентификации по отпечатку пальца, но еще не достигли устойчивых результатов. Особенно эффективны методы, позволяющие захватывать биометрические характеристики объекта на большом расстоянии и во время движения. С распространением мегапиксельных камер видеонаблюдения реализация подобного принципа работы становится все более легкой. Результаты сравнений приведены в таблице 1.4.

Таблица 1.4 - Скорость аутентификации и возможность бесконтактной аутентификации

Биометрическая система использует :	**Скорость аутентификации**	**Бесконтактная аутентификация во время движения**
Отпечаток пальца	Высокая	Безуспешна
Распознавание лица 2D	Средняя	На большом расстоянии
Распознавание лица 3D	Низкая	На среднем расстоянии
Радужная оболочка глаза	Высокая	На большом расстоянии
Сетчатка глаза	Низкая	Невозможна
Рисунок вен	Высокая	На малом расстоянии

1.2.7 Сравнение биометрических методов по психологическому комфорту пользователя и стоимости реализации

Психологический комфорт пользователей – также достаточно актуальный показатель при выборе системы безопасности. Если в случае с двухмерным распознаванием лиц или радужной оболочкой – оно происходит незаметно, то сканирование сетчатки глаза – довольно неприятный процесс. А идентификация по отпечатку пальца, хоть и не приносит неприятных ощущений, может вызывать негативные ассоциации с методами криминалистической экспертизы.

Стоимость систем контроля и учета доступа в зависимости от используемых методов биометрической идентификации крайне различается между собой. Впрочем, разница может быть ощутимой и внутри одного метода, в зависимости от назначения системы (функциональности), технологий производства, способов, повышающих защиту от несанкционированного доступа и т.п. Результат сравнений показан в таблице 1.5.

Таблица 1.5 - Психологический комфорт пользователя и стоимость реализации

Биометрическая система использует :	**Комфорт пользователя**	**Стоимость**
Отпечаток пальца	Средний	Низкая
Распознавание лица 2D	Высокий	Средняя
Распознавание лица 3D	Средний	Высокая
Радужная оболочка глаза	Высокий	Высокая
Сетчатка глаза	Низкий	Высокая
Рисунок вен	Средний	Средняя

Безусловно, выбор метода биометрической аутентификации для системы контроля доступа в первую очередь зависит от предъявляемых к ней требований. Тем не менее, сравнение биометрических методов по совокупности факторов наглядно демонстрирует их преимущества в целом.

ГЛАВА II

ФОРМАЛИЗАЦИЯ МЕТОДОВ И СТРУКТУР ЗАЩИТЫ БИОМЕТРИЧЕСКИХ ДАННЫХ

2.1 Гистограммы направленных градиентов

Гистограмма направленных градиентов – метод информационной оценки особых точек изображения, основанный на подсчете количества направленных градиентов в пространстве этих точек. Используется в задачах распознавания изображения для выделения и получения признаков объекта интереса совместно с выбранным классификатором.

Впервые гистограмма направленных градиентов была представлена в работе Навнита Далала и Билла Триггса в июне 2005 г., в котором этот метод применялся для распознавания человека на статичных изображениях. В настоящее время этот метод широко используется не только для нахождения человека, но и для распознавания лиц, автомобилей и других объектов на видеопоследовательностях.

Основной идеей алгоритма является допущение, что внешний вид и форма объекта на участке изображения могут быть описаны распределением градиентов интенсивности или направлением краев [31]. Такое описание проводится путем разделения изображения на ячейки и построения гистограмм направленных градиентов пикселей ячеек (рисунок 2.1). Результатом работы алгоритма является дескриптор, включающий в себя комбинацию полученных гистограмм (рисунок 2.2).

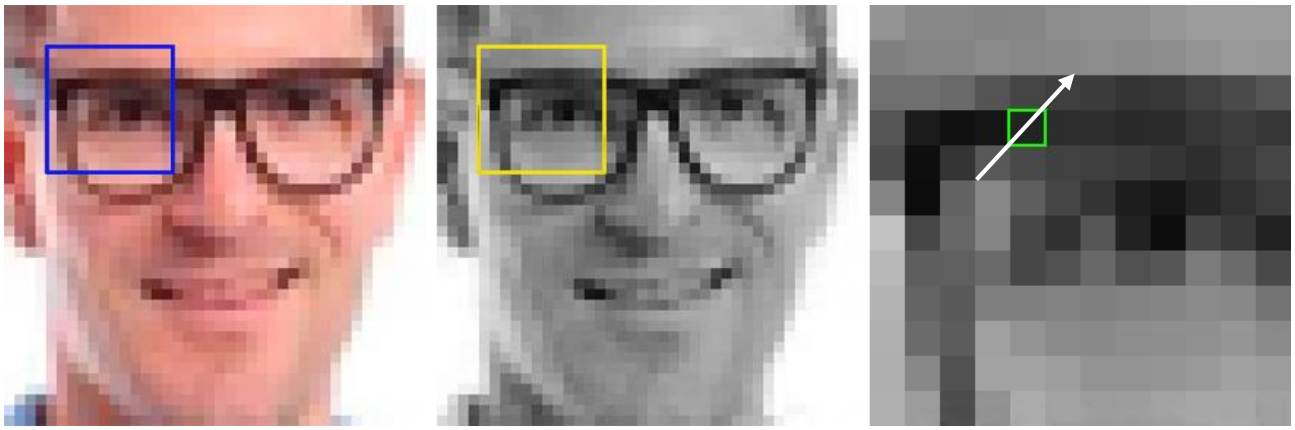


Рисунок 2.1 – Построение гистограмм направленных градиентов пикселей ячеек

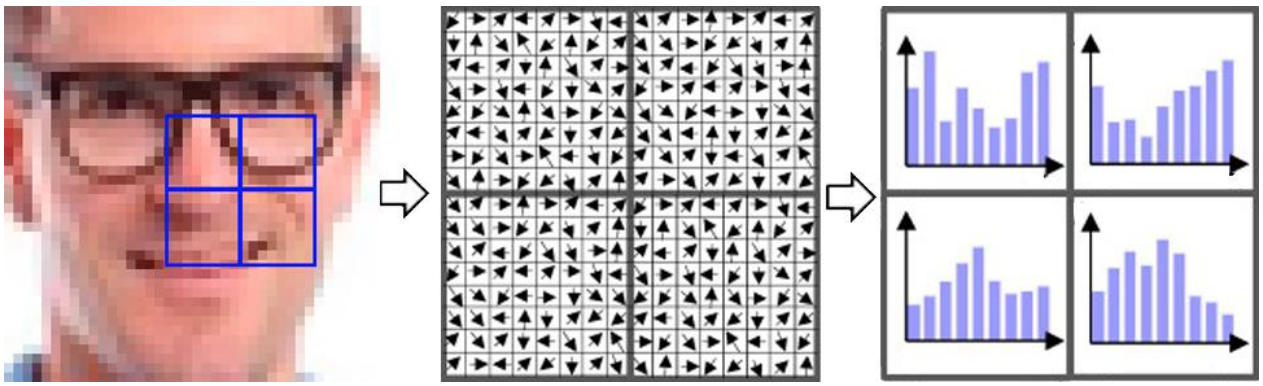


Рисунок 2.2 – Комбинации полученных гистограмм

Основные этапы работы алгоритма на примере алгоритма распознавания человека.

1.Вычисление градиента.

Вычисление градиента осуществляется для каждого пикселя изображения с помощью маски. Наиболее оптимальной маской для данного алгоритма является одномерная дифференцирующая маска.

2. Вычисление гистограмм ячеек изображения.

На этом шаге изображение делится на ячейки фиксированного размера (рисунок 2.3, а). В каждой ячейке производится расчет преобладающего направления градиента путем анализа значения градиента каждого пикселя ячейки. Каналы гистограммы равномерно распределяются от 0 до 180° или же от 0 до 360°, в зависимости от того, вычисляется «знаковый» или «беззнаковый градиент» [31].

3. Формирование и нормирование блоков дескрипторов.

Полученные ячейки группируются в более крупные связанные блоки для нормирования значения градиентов. Это позволяет учитывать яркость и контрастность каждой ячейки. Дескриптор HOG, таким образом, является вектором компонент нормированных гистограмм ячеек из всех областей блока (рисунок 2.3, б). Как правило, блоки перекрываются, то есть каждая ячейка входит более чем в один конечный дескриптор.

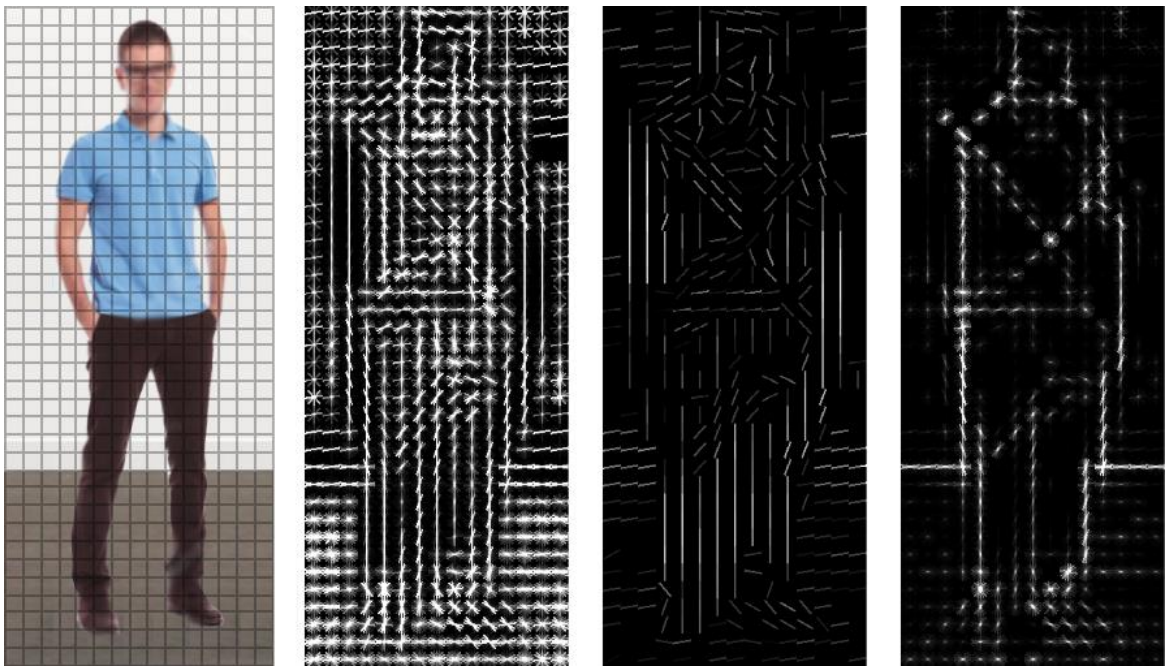
В основном используются следующие способы получения нормировочного множителя: L2-норма (2.1), L1-норма (2.2), корень из L1-нормы (2.3) [31].

$$f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \ . \tag{2.1}$$

$$f = \frac{v}{\sqrt{\|v\|_1 + e}} \ . \tag{2.2}$$

$$f = \sqrt{\frac{v}{\sqrt{\|v\|_1 + e}}} \ . \tag{2.3}$$

Конечным этапом является классификация HOG-дескрипторов с использованием системы обучения с учителем. В качестве классификатора в основном используется метод SVM, в основе которого лежит принцип изменения исходного векторного пространства в пространство с более высокой размерностью, а также поиск оптимальной гиперплоскости, разделяющей классифицируемые признаки [32, 33]. Результатом работы классификатора являются два образа объекта с положительными и отрицательными весами опорных векторов соответственно (рисунок 2.3, в, г). Положительные веса означают принадлежность признаков к человеку, отрицательные – к фону.



(а) (б) (в) (г)

Рисунок 2.3 - Пример распознавания с использованием HOG-дескрипторов

Дальнейшая работа по повышению точности алгоритма распознавания человека сводится к обработке большого количества схожих изображений для обучения классификатора различным вариациям как распознаваемых объектов, так и объектов фона. Для уменьшения размерности вектора гистограмм направленных градиентов возможно применение метода PCA.

Гистограммы направленных градиентов являются довольно эффективным методом распознавания объектов изображения. В общем случае, конечный результат распознавания будет зависеть от условий съемки (пересечение объекта другими объектами, ориентация объекта и др.), а также выбора и степени обученности классификатора.

2.2 БЧХ-коды

БЧХ-коды в теории кодирования — это широкий класс циклических кодов, применяемых для защиты информации от ошибок. Отличается возможностью построения кода с заранее определенными корректирующими свойствами, а именно, минимальным кодовым расстоянием.

БЧХ-коды составляют один из больших классов линейных кодов, исправляющих ошибки. Причем метод построения этих кодов задан явно.

Код БЧХ длины k, исправляющий $q_{\text{ис}}$-кратные ошибки, это циклический блочный код над полем $GF(p)$, корнями порождающего многочлена которого являются $\beta^v, \beta^{v+1}, \dots, \beta^{v+2q_{\text{ис}}-1}$, где β – элемент конечного поля $\mathrm{GF}(q^m)$; v – целое число.

В соответствии с этим определением порождающий многочлен кода БЧХ может быть представлен наименьшим общим кратным

$$\mathrm{g}(x) = HOK\left[M_v(x), M_{v+1}(x), \dots, M_{v+2q_{\text{ис}}-1}(x)\right],$$

где $M_j(x)$ – минимальные многочлены элементов β^j.

В [35, 36] доказано, что наличие $2q_{\text{ис}}$ корней полинома $\mathrm{g}(x)$, указанных в определении кода, гарантирует исправление всех ошибок кратности, меньшей или равной $q_{\text{ис}}$.

Коды БЧХ, имеющие длину $k = q^m - 1$ называются примитивными кодами БЧХ.

Для того, чтобы построить порождающий многочлен примитивного БЧХ-кода нужно:

1. Задать длину кода $k = q^m - 1$ и число t ошибок, которые необходимо исправлять.
2. Найти неприводимый многочлен степени m и построить поле $GF(q^m)$.
3. Найти примитивный элемент α в поле $GF(q^m)$.
4. Найти минимальные многочлены $f_i(x)$ для $\alpha^i, i = 1, \dots, 2t$ над $GF(q)$.
5. Взять в качестве $g(x) = HOK(f_1(x), f_2(x), \dots, f_{2t}(x))$.

2.2.1 Декодер Питерсона-Горенстейна-Цирлера

Рассматриваемый алгоритм был впервые предложен Питерсоном для двоичных кодов. Общий случай был разработан Горенстейном и Цирлером [37].

Описание алгоритма декодирования Питерсона-Горенстейна-Цирлера. Пусть α – элемент поля $GF(q^m)$, по которому строился БЧХ, а d – количество ошибок, исправляемых кодом.

1. На вход алгоритму поступает принятое слово $v(x)$.
2. Вычисляются компоненты синдрома $S_i = v(\alpha^i), i = 1, \dots, 2d$.
3. Полагают $v = d$.
4. Строится матрица

$$M = \begin{bmatrix} S_1 & S_2 & \dots & S_v \\ S_2 & S_3 & \dots & S_{v+1} \\ \vdots & \vdots & & \vdots \\ S_v & S_{v+1} & \dots & S_{2v-1} \end{bmatrix}.$$

5. Вычисляется определитель матрицы M. Если он равен нулю, v уменьшается на единицу и возвращаются к шагу 4.

6. Обращение матрицы M и вычисление коэффициентов многочлена $\Lambda(x)$:

$$\begin{bmatrix} \Lambda_v \\ \Lambda_{v-1} \\ \vdots \\ \Lambda_1 \end{bmatrix} = M^{-1} \begin{bmatrix} -S_{v+1} \\ -S_{v+2} \\ \vdots \\ -S_{2v} \end{bmatrix}.$$

7. Вычисление корней многочлена $\Lambda(x)$. Поскольку число элементов поля конечно, обычно корни ищут процедурой Ченя. Эта процедура заключается в последовательном вычислении $\Lambda(\alpha^i)$ для каждого i и проверки полученных значений на нуль.
8. Найдя корни, находят локаторы ошибок X_j (корни многочлена $\Lambda(x)$ являются обратными к локаторам ошибок).

9. Если код двоичный, то ошибки Y_j известны. В противном случае их вычисляют:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_v \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & \dots & X_v \\ X_1^2 & X_2^2 & \dots & X_v^2 \\ \vdots & \vdots & & \vdots \\ X_1^v & X_2^v & \dots & X_v^v \end{bmatrix}^{-1} \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_v \end{bmatrix}.$$

10. Исправление в полученном слове ошибки и получение на выходе алгоритма кодового слова.

К особенностям кодов БЧХ можно отнести тот факт, что с ростом длины k кода при фиксированном значении скорости кода отношение s/k стремится к нулю. В результате, несмотря на наличие у кодов БЧХ отмеченных положительных свойств, при больших длинах $(k > 1000)$ желательно отдавать предпочтение другим кодам [35, 37].

2.3 Модель CLNF

Модель CLNF использует новую неравномерную методику выравнивания средних значений, которая учитывает надежность детектора [38].

Эксперт по LNF, показанный на рисунке 2.4, вводит нелинейность в CNF, гибкость и непрерывный выход, характерный для непрерывных условных случайных полей.

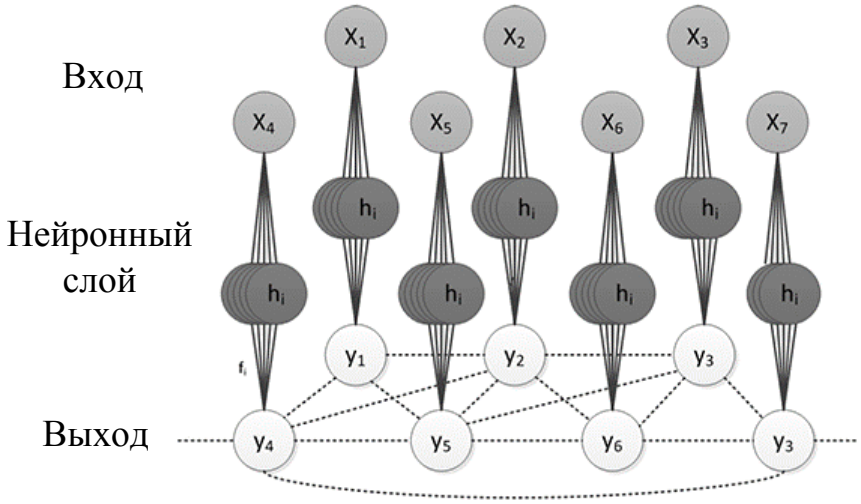


Рисунок 2.4 - Эксперт по локальному нейронному полю

Предлагаемый локальный детектор может описывать отношения между пикселями (соседними и расположенными на больших расстояниях), изучать схожести и ограничение разреженности удаленных элементов. LNF также включает слой нейронной сети, который может захватывать сложные нелинейные отношения между значениями пикселей и выходными откликами. Это модель с непрерывным, простым и эффективным логическим выводом, в которой определено два типа пространственных отношений, которые должен выдавать детектор. Во-первых, пространственное сходство, то есть пиксели поблизости должны иметь аналогичные вероятности выравнивания. Во-вторых, во всей оцениваемой области детектор должен выдавать только один пик.

Преимущества моделирования, изображенные на рисунке 2.5, показывают карты патч-откликов от SVR детекторов, LNF-детекторов без пространственных ограничений и полных LNF-детекторов с ограничениями сходства, и разреженности.

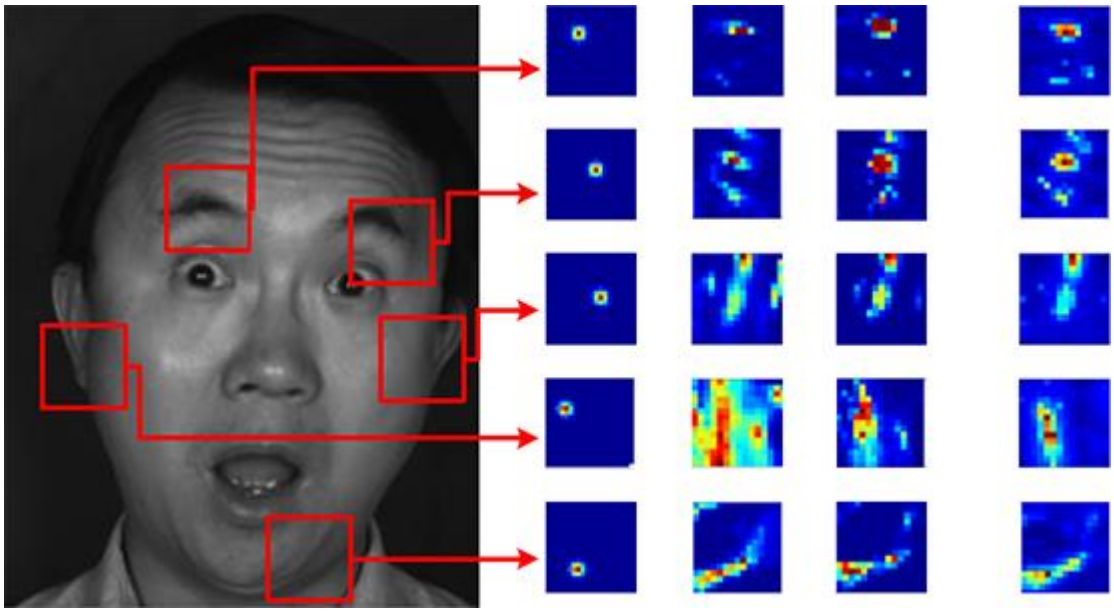


Рисунок 2.5 - Преимущества моделирования пространственных зависимостей и входные нелинейности

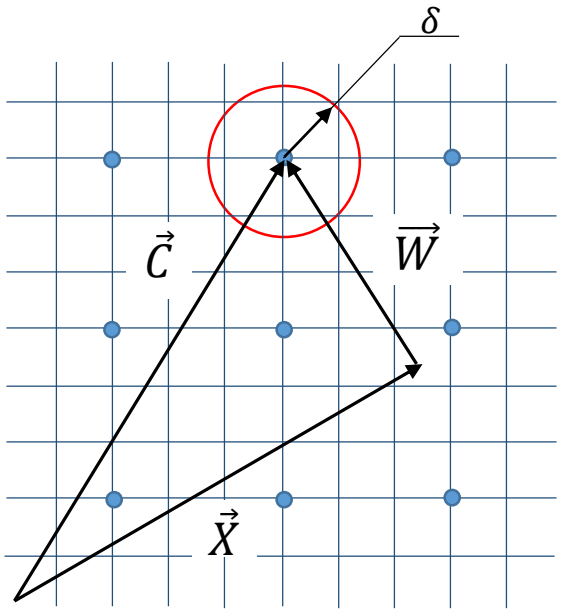
У представленного детектора меньше пиков на выходе и более плавный отклик, чем тот, у которого нет краевых признаков, и оба они более точны, чем у SVR-детектора. Эти пространственные ограничения предназначены для улучшения реакции патча, что приводит к более точной подгонке.

ГЛАВА III

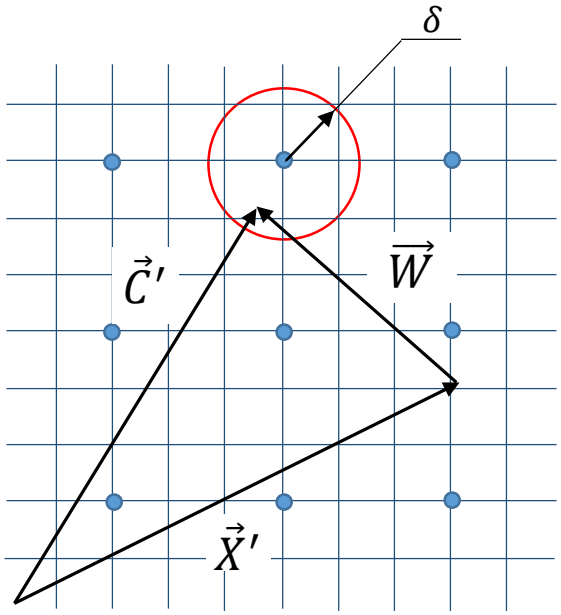
МОДЕЛЬ БИОМЕТРИЧЕСКОЙ СИСТЕМЫ

3.1 Основные принципы метода восстановления биометрических данных.

В алгоритме присутствуют две фазы: фаза регистрации (рисунок 3.1, а) и фаза проверки (рисунок 3.1, б) пользователей.



(а)



(б)

Рисунок 3.1 - Основные принципы алгоритма: а - фаза регистрации; б - фаза проверки

При регистрации пользователя выполняется одно или несколько биометрических измерений, что приводит к получению вектора признаков $\vec{X}$. Затем произвольно выбирается вектор кодового слова $\vec{C}$, который вычисляется при помощи кода с коррекцией ошибок. На рисунке 3.1 кодовые слова обозначаются как точки. Разница между вектором кодового слова и биометрическим измерением определяется как вспомогательный сигнал $\overrightarrow{W}$ (3.1).

$$\overrightarrow{W} = \vec{C} - \vec{X}. \tag{3.1}$$

Используя декодирование с исправлением ошибок из вектора кодового слова $\vec{C}$ получается случайный вектор S. Этот вектор или его защищенная производная используется для целей сопоставления.

Поскольку алгоритм работает с БЧХ-кодом, можно определить окружность с радиусом δ вокруг каждого кодового слова (рисунок 3.1) так, что функция декодирования отображает каждую точку в окружности на соответствующее кодовое слово. В целом это можно моделировать как (3.2).

$$\vec{X}' = \vec{X} + \vec{N}. \tag{3.2}$$

где N представляет собой шум измерений биометрии.

Если шум N достаточно мал (то есть $\left|\vec{N}\right| \leq \delta$), то X' находится внутри области δ вокруг $\vec{C}'$, так что его можно привести к $\vec{C}'$ в более точном порядке.

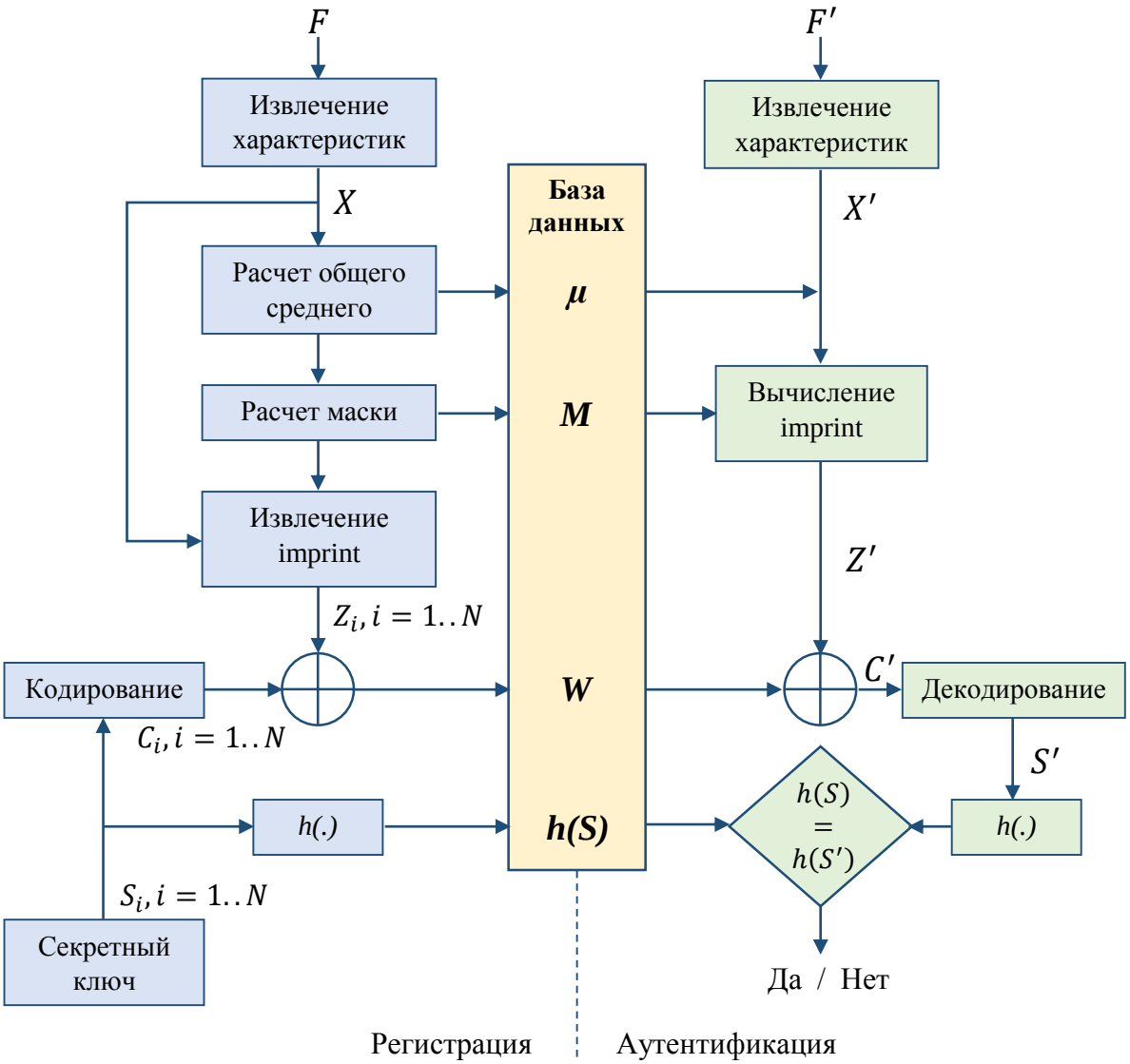
При переводе алгоритма в практическую биометрическую систему с защитой шаблонов можно выделить некоторые вопросы для рассмотрения:

• эффективность (с точки зрения FAR, FRR) должна быть сопоставимой с традиционными биометрическими системами;

• система должна иметь возможность работать с заданным кодом коррекции ошибок.

3.2 Система регистрации/аутентификации пользователей

Система регистрации пользователей (рисунок 3.2), основанная на предложенной в [39], требует перевода вещественных векторов признаков $\vec{X}$ в бинарные векторы (двоичные строки).



X и X' – входные данные для регистрации и аутентификации; μ – среднее значение межклассового распределения; M – маска (номера позиций наиболее «надежных» бит); W – кодовая запись пользователя; $h(S)$ – хешированное случайное число S.

Рисунок 3.2 - Схема системы регистрации/аутентификации пользователей с общими параметрами на основе HOG-структур и БЧХ-кодов

На этапе регистрации предполагается, что в системе будет зарегистрировано N пользователей с количеством G изображений для каждого. Тогда общее количество изображений датасета $N \cdot G$: $\{\vec{F}_{i,j}\}_{i=1..N, j=1..G}$.

Соответствующий набор признаков обозначается как $\{\vec{X}_{i,j}\}_{i=1..N,j=1..G}$, где $\vec{X}_{i,j} \in \mathbb{R}$ и содержит компоненты $(\vec{X}_{i,j})_t$, $t = 1..k$.

Для бинаризации векторов признаков применено квантование с использованием средних значений внутри класса $\vec{\mu}_i$ (3.3) и общего среднего по всему датасету $\vec{\mu}$ (3.4).

$$\vec{\mu}_i = \frac{1}{G}\sum_{j=1}^{G} \vec{X}_{i,j}. \tag{3.3}$$

$$\vec{\mu} = \frac{1}{N}\sum_{i=1}^{N} \vec{\mu}_i \ . \tag{3.4}$$

Для каждого i-го пользователя получен бинарный вектор Q_i с использованием (3.5).

$$(Q_i)_t = \begin{cases} 0 \ \Leftrightarrow \ (\vec{\mu}_i)_t \leq \ (\vec{\mu})_t \\ 1 \ \Leftrightarrow \ (\vec{\mu}_i)_t > \ (\vec{\mu})_t \end{cases}, \tag{3.5}$$

где $\Leftrightarrow$ – условие «если»,

t – номер элемента вектора.

3.2.1 Расчет надежности компонент

В случаях когда среднее значение внутри класса $\vec{\mu}_i$ близко к значению общего среднего $\vec{\mu}$ может возникать нестабильность элементов Q_i. В таких случаях небольшие вариации входных данных могут приводить к ошибкам в квантуемом векторе признаков Q_i. Кроме того, с точки зрения классификации эти компоненты менее дифференцируемы для класса i по сравнению с другими пользователями. Чтобы свести к минимуму эти эффекты, использован механизм извлечения наиболее «надежных» компонентов внутри одного вектора признаков. При расчете были использованы стандартные функции ошибок для оценки «надежности» $R_{i,t}$ (3.6) каждого бита t, каждого пользователя i:

$$R_{i,t} = \frac{1}{2}\left(1 + erf\left(\frac{(\vec{\mu}_i)_t - (\vec{\mu})_t}{\sqrt{2s_{i,t}^2}}\right)\right), \quad (3.6)$$

где erf – функция ошибки (3.7),

$s_{i,t}^2$ – дисперсия t-го элемента i-го пользователя (3.8).

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-p^2}dp\ . \quad (3.7)$$

$$s_{i,t}^2 = \frac{1}{G-1}\sum_{j=1}^{G}\left(\left(\vec{X}_{i,j}\right)_t - (\vec{\mu}_i)_t\right)^2\ . \quad (3.8)$$

3.2.2 Расчет маски и формирование цифровых отпечатков.

Маска представляет собой бинарный вектор M (3.9), в котором

$$M_t = \begin{cases} 1 \Leftrightarrow \left(R_{i,t} \geq \alpha\right) \wedge \left(Q_{i,t} \neq Q_{k=1..N(k\neq i),t}\right) \\ 0 \Leftrightarrow \left(R_{i,t} < \alpha\right) \vee \left(Q_{i,t} = Q_{k=1..N(k\neq i),t}\right) \end{cases}, \quad (3.9)$$

где $\Leftrightarrow$ – условие «если»,

$\wedge$ – логическое «и»,

$\vee$ – логическое «или»,

α – порог «надежности», подбираемый для всех элементов векторов Q_i под определенную длину k.

Маска M одна для всех пользователей базы данных, содержит номера позиций наиболее «надежных» бит векторов Q_i.

С помощью маски M формируются цифровые отпечатки пользователей «imprint» – бинарные векторы $Z_i \in \{0,1\}^k$, содержащие k наиболее «надежных» бит вектора Q_i.

3.2.3 Образование и защита кодовых записей пользователей.

Кодовую запись пользователя представляет собой бинарный вектор W_i – сумма по модулю два (3.10) бинарных векторов Z_i и C_i.

$$W_i = Z_i \oplus C_i \ . \tag{3.10}$$

Вектор C_i является кодовым словом и образуется путем кодирования секретного ключа S_i с помощью БЧХ-кода с параметрами (k, s, d), где k – длина кодового слова, s – количество информационных символов, d – количество исправляемы ошибок.

В свою очередь секретный ключ пользователя i – случайная бинарная последовательность $S_i \in \{0,1\}^s$ специфичная для пользователя, генерируется генератором случайных чисел, хэшируется с помощью хэш-функции $h()$.

Зная кодовую запись W_i и $h(S_i)$, невозможно извлечь цифровой отпечаток Z_i (так как $Z_i = C_i \oplus W_i$), потому что S_i и C_i нельзя извлечь из $h(S_i)$. Универсальность данного подхода в возможности выбора разных значений для S_i, что приводит к разной паре W_i и $h(S_i)$. Эта особенность позволяет получать из одного лицевого биометрического шаблона множество защищенных биометрических производных.

3.2.4 Аутентификация пользователей

Процедура аутентификации соответствует блокам схемы системы аутентификации на рисунке 3.2.

Пользователи предоставляют свои биометрические данные F', которые используются для получения соответствующих вещественных векторов признаков $\vec{X}_i'$.

С помощью (3.11) рассчитываются средние значения $\vec{\mu}_i'$

$$\vec{\mu}_i' = \frac{1}{G}\sum_{j=1}^{G} \vec{X}_{i,j}' \ . \tag{3.11}$$

С помощью вектора общего среднего $\vec{\mu}$, хранящегося в базе данных, производится бинаризация признаков, используя (3.5), в результате чего получается бинарный вектор Q_i'.

При наложении маски M , считываемой из базы данных, на бинарный вектор Q_i' формируются цифровые отпечатки Z_i' пользователей. В результате можно определить кодовое слово C_i' (3.12).

$$C_i' = Z_i' \oplus W_i \,, \tag{3.12}$$

где W_i – кодовые записи пользователей из базы данных.

Наконец S_i' и, следовательно, $h(S_i')$ можно получить путем декодирования C_i' при помощи БЧХ-кода с параметрами (k, s, d).

Аутентификация достигается путем сравнения обоих хешированных значений $h(S)$ и $h(S')$. Если оба значения идентичны, то человек аутентифицирован.

Успех аутентификации полностью зависит от HD между Z_i' и зарегистрированным в базе данных шаблоном Z_i. Учитывая параметры (k, s, d) БЧХ-кода, положительная аутентификация будет иметь место, если HD между Z_i' и Z_i меньше либо равно d.

Учитывая эти ограничения, можно сформировать некоторые требования для интеграции предлагаемой методики:

1. Максимизация размера количества информационных символов s БЧХ-кода. Поскольку в базе данных хранится $h(S)$, то практически невозможно получить S. Чтобы предотвратить исчерпывающий поиск по $h(S)$, размер S должен быть максимальным.
2. Возможность исправления большего количества ошибок d. К примеру код БЧХ с k = 511 и s = 58, имеет возможность гарантированной коррекции ошибок d = 91. Увеличение количества информационных символов, например, до s = 157 или s = 184 бит, снижает эту способность до d = 50 и d = 45 соответственно. Поэтому задача состоит в том, чтобы настроить конфигурацию, которая позволяет корректировать максимальное количество битов при достижении достаточно большого секретного ключа S.

ГЛАВА IV

РЕАЛИЗАЦИЯ СИСТЕМЫ ЗАЩИТЫ БИОМЕТРИЧЕСКИХ ДАННЫХ

4.1 Выбор инструментов и средств реализации

4.1.1 Конфигурация компьютера

Реализация алгоритмов выполнена на виртуальной машине с характеристиками, указанными в таблице 4.1.

Таблица 4.1 - Характеристики виртуальной машины

Характеристика	Значение
Тип машины	Виртуальная машина VMware
Операционная система	Windows 10 Pro x64 Insider Preview
Процессор	8×Intel Xeon E5606 2.13GHz
Оперативная память	16Gb
Накопитель	SSD 100Gb

4.1.2 Среда разработки

В качестве среды разработки выбран язык программирования Delphi из Embarcadero RAD Studio – среда разработки приложений фирмы Embarcadero Technologies.

Версия Embarcadero RAD Studio 10.2 Tokyo объединяет Delphi и C++ Builder в единую интегрированную среду разработки.

Основные новые возможности Delphi 10.2:

- Высокий уровень поддержки Windows 10. Поддерживаются компоненты Windows 10 и «родные» API и компоненты WinRT/UWP, элементы интерфейса Windows 10 VCL. Также обновлена поддержка Windows 10 FMX;
- Удвоенный размер проектов в IDE;
- Стабильность, качество и эффективная документация;
- поддержка параллельной компиляции C++;
- отладка iOS 64x приложений;
- поддерживается iOS 8.4;
- поддержка Android 6.0 (API Level 23);
- поддержка модульного тестирования DUnitX для Android и iOS;

- поддержка DirectX 12;
- поддержка вызова API WinRT;
- поддержка FireDAC для базы данных NoSQL MongoDB;
- новое поведение MultiView;
- новые компоненты VCL;
- новые компоненты для работы с Beacon;
- улучшен механизм стилей;
- улучшен диспетчер библиотек GetIt;
- улучшены возможности IDE;
- компилятор приложений под Linux (Ubuntu Server (LTS 16.04) and RedHat Enterprise (V7));
- включена поддержка СУБД MariaDB.

4.1.3 Извлечение биометрических характеристик

Для извлечения биометрических данных используется фрэймворк OpenFace [39].

OpenFace – инструмент, предназначенный для исследователей компьютерного зрения и машинного обучения, компьютерного сообщества и людей, заинтересованных в создании интерактивных приложений на основе анализа поведения лица.

OpenFace – инструментарий, способный распознавать ориентиры лица, оценку позиции головы, распознавание лица и оценку направления взгляда с доступным исходным кодом.

Алгоритмы компьютерного зрения, которые представляют собой ядро OpenFace, демонстрируют современные результаты во всех вышеупомянутых задачах. Кроме того, инструмент способен работать в режиме реального времени и может работать с простой веб-камерой без какого-либо специализированного оборудования. Наконец, OpenFace позволяет легко интегрироваться с другими приложениями и устройствами через легкую систему обмена сообщениями.

Код был написан главным образом Тадасом Балтрусайтисом в Университете Карнеги-Меллона и компьютерной лаборатории Кембриджского университета [40, 41].

OpenFace использует недавно предложенную CLNF [38] для обнаружения лица, его ориентации и трекинга. Модель CLNF выполняет обнаружение 68 лицевых точек, а также способна извлекать информацию о позиции головы и направлении взгляда.

4.1.4 Кодирование и декодирование

Для реализации кодирования/декодирования в FaceAnalyzer используется приложение MagicCoder [42].

MagicCoder – программная реализация алгоритмов кодирования и декодирования двоичных БЧХ-кодов с параметрами m и d, где параметр m задает длину кода $k = 2^m - 1$, а параметр d – количество исправляемых кодом ошибок. Параметры m и k задаются пользователем. Реализовано систематическое и несистематическое кодирование. Для декодирования используется декодер Питерсона-Горенстейна-Цирлера.

В целях автоматизации процессов кодирования/декодирования в исходный код MagicCoder были внесены некоторые дополнения: добавлена возможность запуска приложения с ключами (см.таблицу 4.2).

Таблица 4.2 - Ключи запуска приложения MagicCoder

Ключ	Параметры	Назначение
-bch=	m,d	Использовать БЧХ-код с параметрами m, d
-rm=	m,r	Использовать код Рида-Маллера с параметрами m,r
-encode	нет	Включить режим кодирования
-decode	нет	Включить режим декодирования
-input=	filename	Имя входного файла
-output=	filename	Имя выходного файла
-?	нет	Вызов справки

Пример использования ключей из командной строки:

```
MagicCoder.exe -bch=9,91 -encode -input=C:\original.txt -output=C:\encoded.enc,
```

команда запускает MagicCoder, который кодирует файл original.txt в файл encoded.enc при помощи БЧХ-кода с параметрами m=9 и d=91.

4.1.5 Датасет

Для проведения экспериментов и тестирования реализации алгоритма была использована база данных лиц Caltech Faces Калифорнийского технологического университета [43]. Информация о наборе данных представлена в Таблице 4.3.

Таблица 4.3 - Информация о датасете Caltech

Характеристика	Значение
Размер (пикселы)	896x592
Разрешение (точки на дюйм)	192
Представление цвета	sRGB
Глубина цвета	24
Количество человек	24
Количество кадров на человека	18,41 (от 5 до 25)

4.2 Прототип приложения, реализующего систему.

Для реализации предложенных алгоритмов разработано приложение FaceAnalyzer [1-A] (рисунок 4.1)

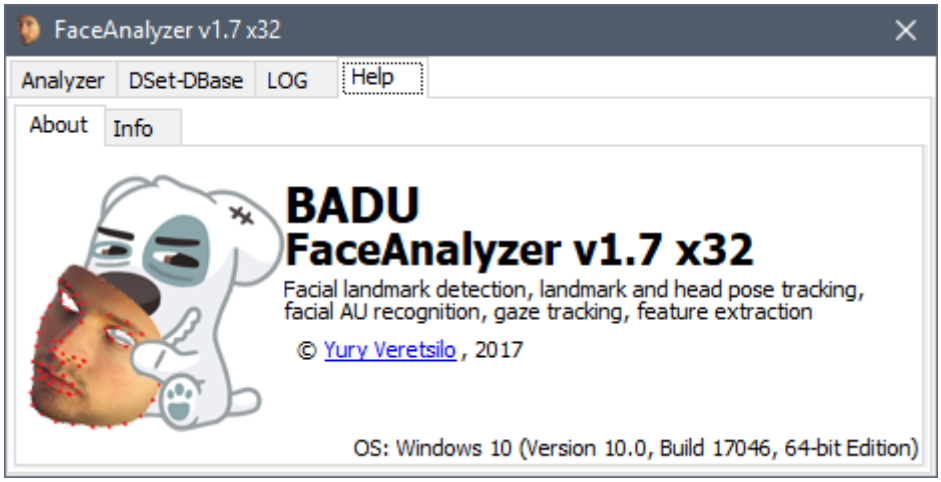


Рисунок 4.1 – Приложение FaceAnalyzer. Вкладка «Help» -> «About»

Схема взаимодействия приложения FaceAnalyzer и используемого программного обеспечения изображена на рисунке 4.2.

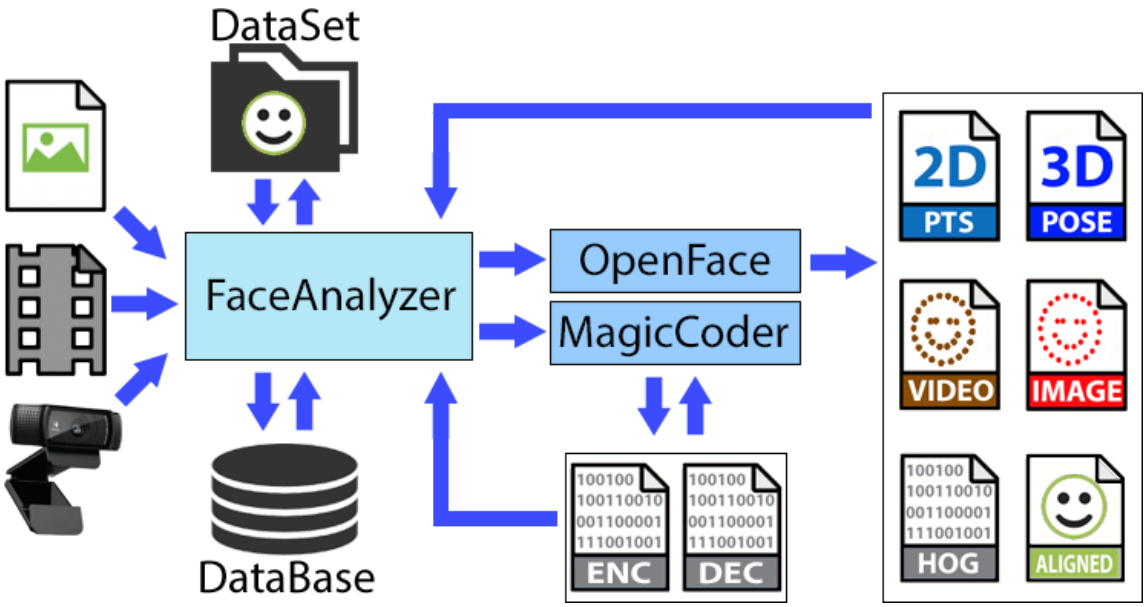


Рисунок 4.2 - Схема работы приложения FaceAnalyzer

Взаимодействие с OpenFace происходит путем передачи информации от источников данных (файл изображения, файл видео, IP/web-камера, датасет) и ключей обработки. В результате обработки изображения/пакета изображений/видео и в зависимости от выбранных параметров обработки формируются файлы с извлеченными характеристиками обнаруженных лиц:

- *.pts – файлы с 2D-координатами 68 лицевых точек и коэффициенты системы кодирования мимики;
- *.pose – файлы с 3D-координатами позиции головы и направление взгляда;
- *_det_0.bmp – изображения лица с нанесенными точками;
- _aligned_face\frame_det_XXXXX.bmp – изображение «вырезанного» лица, где XXXXX – номер кадра;
- *.hog – файл значений гистограмм направленных градиентов.

В процессе извлечения HOG фрэймворк OpenFace сохраняет результат в файл *.hog, имеющий следующую структуру: первые четыре блока по два байта целого типа содержат количество столбцов, строк матрицы HOG, количество элементов и индикатор присутствия лица в кадре соответственно; следующие 4464 блока по 2 байта вещественного типа содержат значения элементов вектора HOG-характеристик кадра.

Для извлечения данных из файлов *.hog реализован парсер (рисунок 4.3), работающий в ручном (для одного файла) и автоматическом (для пакета файлов) режимах.

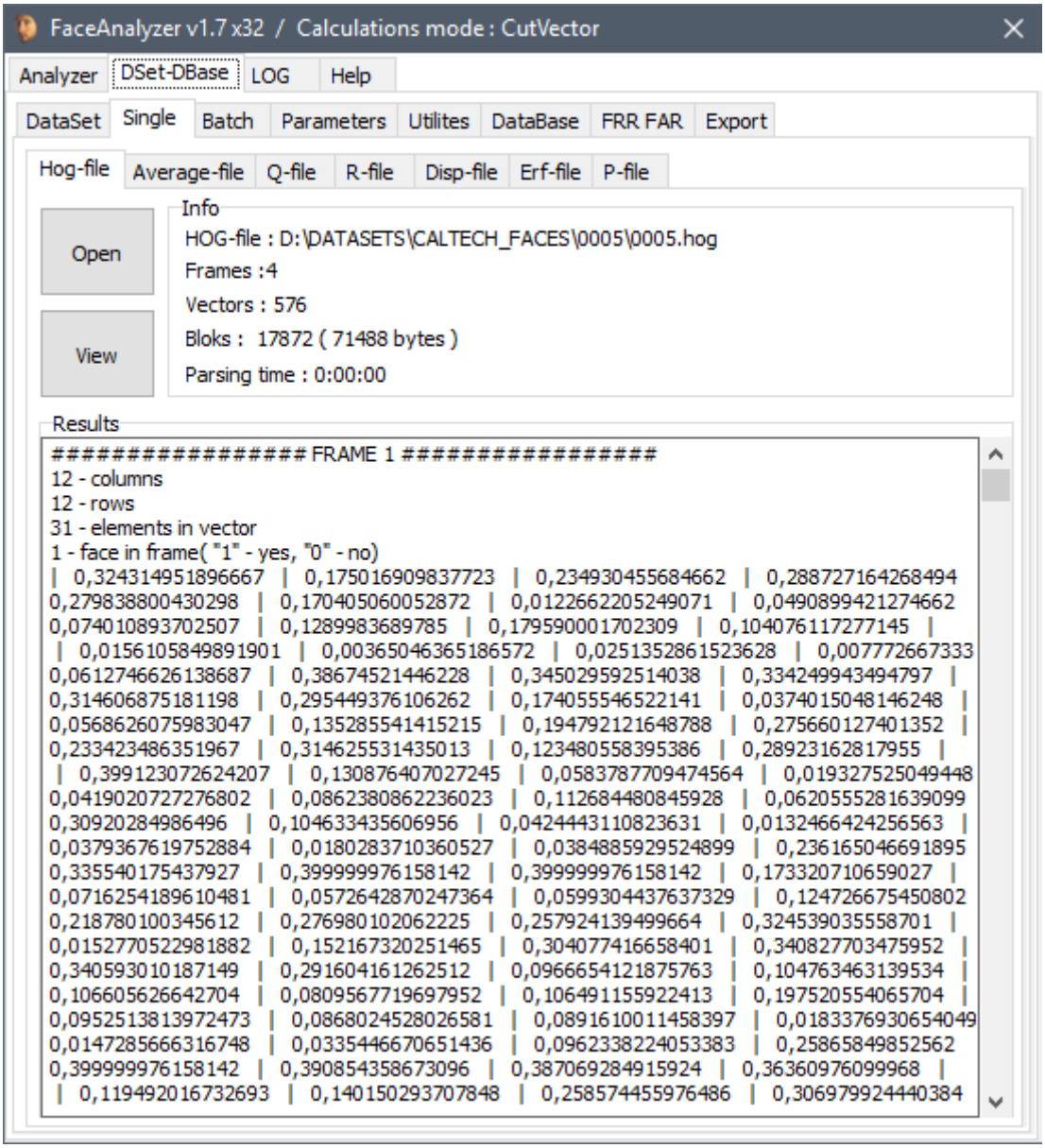


Рисунок 4.3 – Приложение FaceAnalyzer. Вкладка «Single» -> «Hog-file». Парсер HOG-фалов

С помощью парсера hog-файлов данные сгенерированных файлов *.hog подвергаются анализу и обработке для последующих расчетов средних значений выделенных характеристик, квантованных векторов, «надежности» элементов цифровых отпечатков и расчета маски.

Расчет маски производится путем подбора порога «надежности» R, такого, что можно выделить вектор наиболее «надежных» бит для каждого пользователя датасета, причем каждый вектор будет уникальным (отличается от остальных векторов хотя бы в одном бите). Процесс расчета маски показан на рисунке 4.4. Исходный код процедуры расчета маски представлен в приложении А.

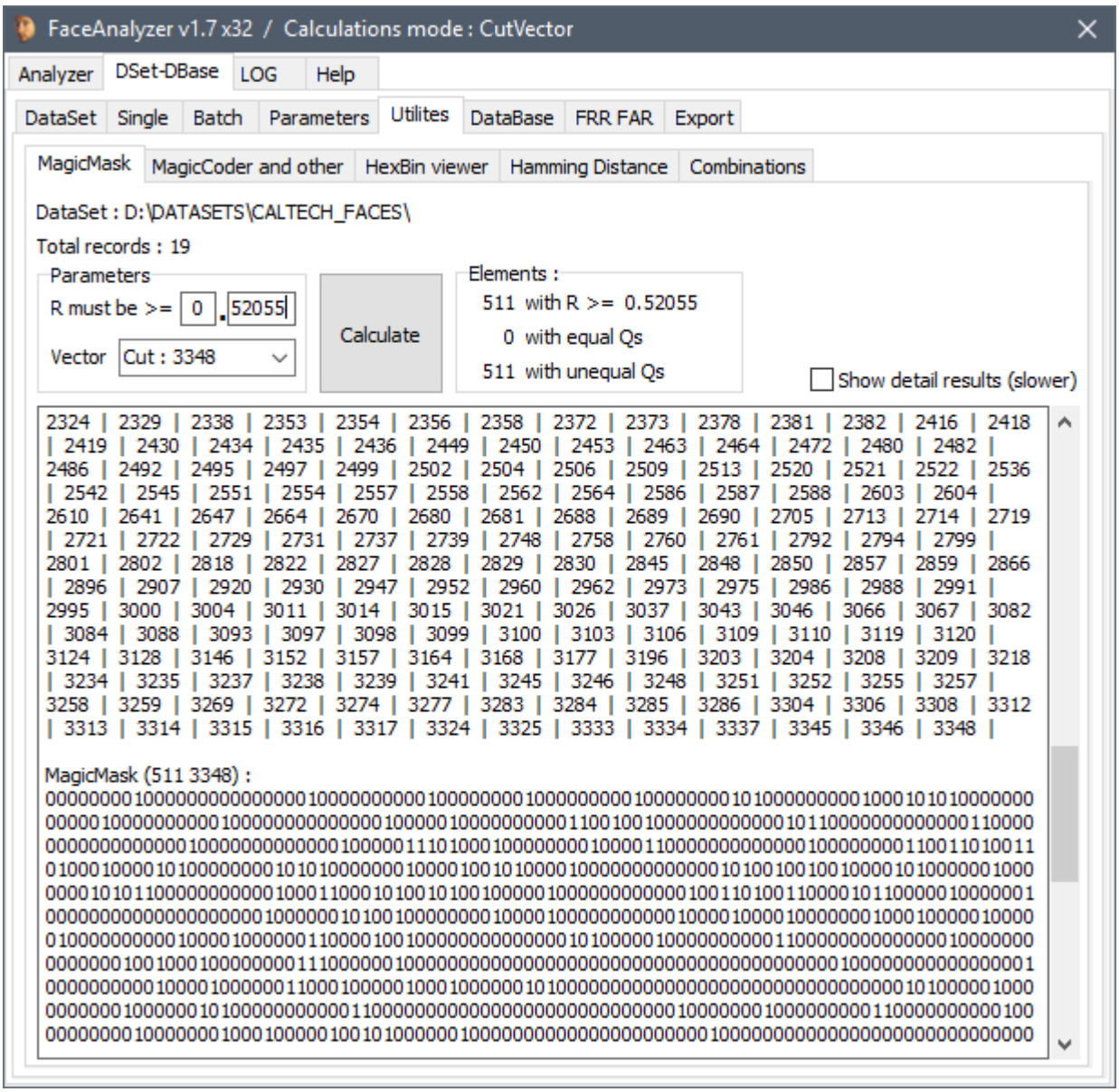


Рисунок 4.4 – Приложение FaceAnalyzer. Вкладки «Utilites» -> «MagicMask». Расчет маски

В процессе расчета датасета файлы со значениями векторов сохраняются в каталоге датасетов. Структура каталога представлена в таблице 4.4.

Таблица 4.4 - Структура каталога датасетов

<table>
<tr><td rowspan="2">DATASETS\</td><td rowspan="2">DATASET_NAME\</td><td>XXXX\</td><td>XXXXYY.jpg,
XXXX.hog,
XXXX.bfaa,
XXXX.bfaq,
XXXX.bfad,
XXXX.bfae,
XXXX.bfar,
XXXX.bfap</td></tr>
<tr><td colspan="2">General.bfaa,
MagicMask.bfam</td></tr>
</table>

Примечания:

1. DATASET – корневая директория датасетов;
2. DATASET_NAME – директория датасета;
3. XXXX – номер пользователя в датасете;
4. YY – номер фрэйма;
5. *.hog – файл значений HOG;
6. *.bfaa – файл средних значений;
7. *.bfaq – файл квантованных значений;
8. *.bfad – файл значений дисперсии;
9. *.bfae – файл значений функции ошибки;
10. *.bfar – файл значений репрезентативности;
11. *.bfap – файл номеров позиций наиболее репрезентативных элементов;
12. General.bfaa – файл общих средних значений датасета;
13. MagicMask.bfam – файл маски датасета.

В процессе создания базы данных (рисунок 4.5) для кодирования секретных ключей в кодовые слова пользователей используется приложение MagicCoder.

В результате для каждого пользователя создаются файлы: *.bfaw – файл со значением кодового слова пользователя; *.bfas – файл со значением секретного ключа. Структура каталога базы данных представлена в таблице 4.5.

Таблица 4.5 - Структура каталога баз данных

DBS\	DATABASE_NAME\	XXXX\	XXXX.bfaw, XXXX.bfas.
		General.bfaa, MagicMask.bfam.	

Примечания:

1. DBS – корневая директория баз данных;
2. DATABASE_NAME – директория базы данных;
3. XXXX – номер пользователя в датасете;
4. *.bfaw – файл значения кодового слова;
5. *.bfas – файл значения секретного ключа;
6. General.bfaa – файл общих средних значений датасета;
7. MagicMask.bfam – файл маски датасета.

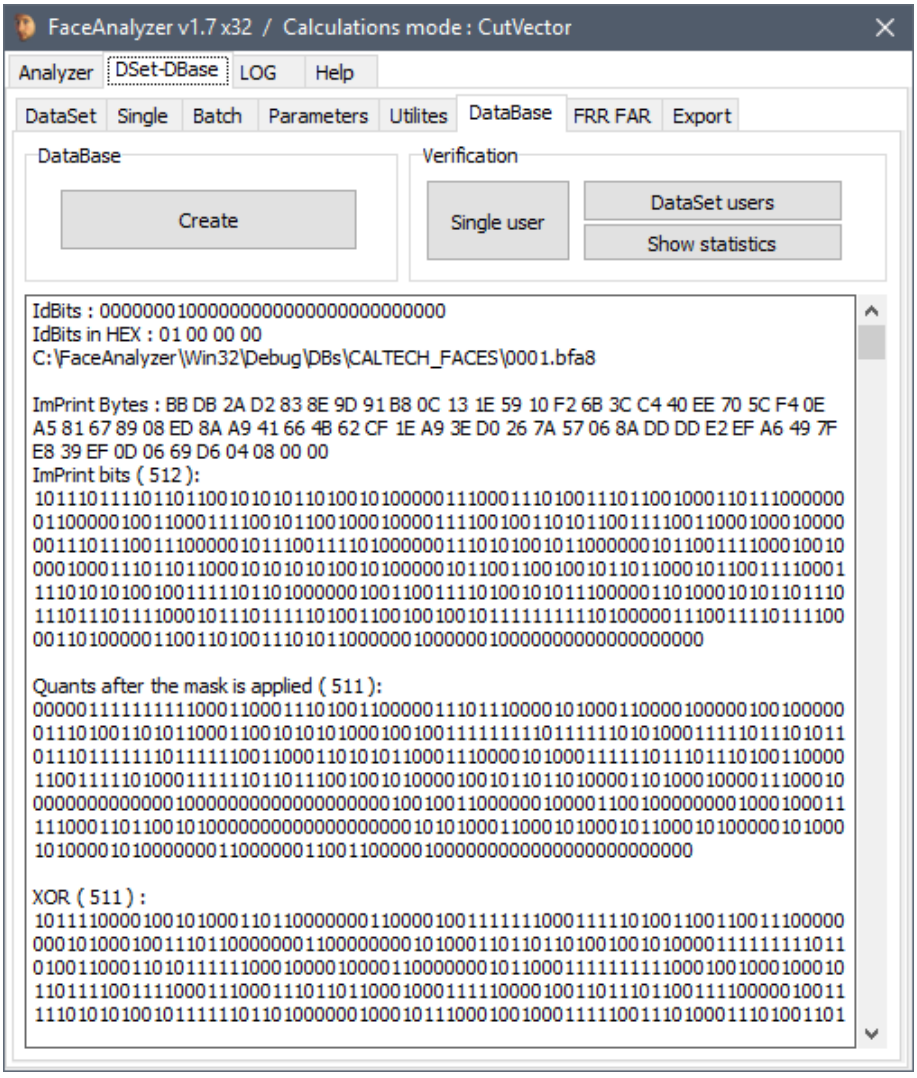


Рисунок 4.5 - Приложение FaceAnalyzer. Вкладка “DataBase”. Создание базы данных

ГЛАВА V
АПРОБАЦИЯ РЕАЛИЗОВАННОЙ СИСТЕМЫ

5.1 Подготовка датасета

Подготовка датасета заключается в сортировке и отборе позитивных кадров для каждого пользователя.

Подготовленный датасет должен быть помещен в директорию датасетов и строго соответствовать ее структуре (таблица 4.4).

В процессе подготовки датасета, из коллекции Caltech Faces были исключены пользователи с количеством кадров менее двенадцати, и низкокачественные (низкая освещенность) кадры (рисунок 5.1).

Рисунок 5.1 - Пример низкокачественных кадров

Для создания основного датасета случайным образом были отобраны по 4 кадра для каждого пользователя. Оставшиеся 8 кадров для каждого пользователя

сочетаниями по 4 кадра (1) были использованы при создании дополнительного датасета для проведения экспериментов сравнения.

$$C_n^k = \frac{n!}{(n-k)! * k!}. \quad (1)$$

$$C_8^4 = \frac{8!}{(8-4)! * 4!} = \frac{8!}{4! * 4!} = \frac{5 * 6 * 7 * 8}{1 * 2 * 3 * 4} = \frac{1680}{24} = 70$$

Таким образом из оригинального датасета Caltech Faces сформированы две коллекции:

1. Основной датасет – 19 пользователей, по 4 кадра у каждого;
2. Дополнительный датасет (для тестирования) – 19 пользователей, по 70 экземпляров (по 4 кадра) для каждого. Структура каталога датасета приведена в таблице 5.1.

Таблица 5.1 - Структура каталога датасета для тестирования

DATASETS\	TEST_DATASET_NAME\	XXXX\	ZZ\	XXXXYY.jpg, ZZ.hog

5.2 Извлечение биометрических характеристик

В качестве биометрических данных в данной работе используются гистограммы направленных градиентов (рисунок 5.2).

При выделении гистограмм направленных градиентов используются блоки 2 × 2 клеток, 8 × 8 пикселей, что приводит к 12 × 12 блоков 31 мерных гистограмм (4464-мерный вектор $\vec{X}$ действительных значений, описывающий лицо).

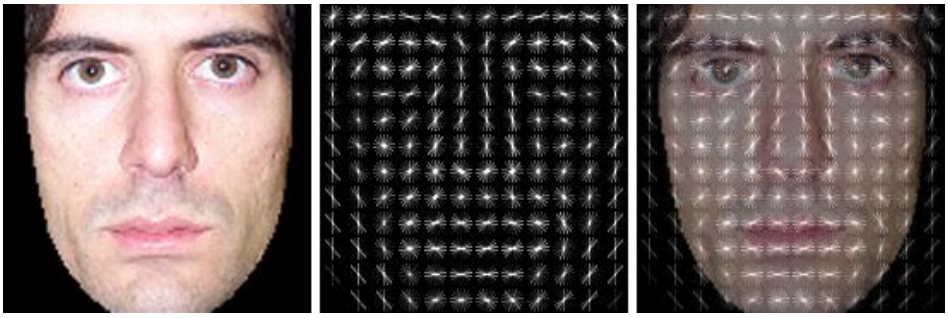


Рисунок 5.2 - Визуализация гистограмм направленных градиентов

Для увеличения точности расчетов и оптимизации отработки алгоритмов можно уменьшить размерность вектора $\vec{X}$ путем исключения из процедуры обработки блоков (рисунок 5.3) № 1, 12, 13, 24, 25, 36, 37, 48, 49, 60, 61, 72, 73, 84, 85, 96, 97, 108, 109, 110, 119, 120, 121, 122, 123, 130, 131, 132, 133, 134, 135, 136, 141, 142, 143, 144, что приводит к уменьшению размерности вектора входных данных с 4464 до 3348 элементов. При этом качество входных данных не уменьшается ввиду того, что исключенные из обработки блоки кадра содержат информацию приблизительно одинаковую для каждого пользователя и ≈ 0 – черный фон (рисунок 5.4).

1	2	3	4	5	6	7	8	9	10	11	12
13	14	15	16	17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72
73	74	75	76	77	78	79	80	81	82	83	84
85	86	87	88	89	90	91	92	93	94	95	96
97	98	99	100	101	102	103	104	105	106	107	108
109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132
133	134	135	136	137	138	139	140	141	142	143	144

Используемые блоки Исключенные блоки

Рисунок 5.3 - Исключение блоков из матрицы HOG

Следует отметить, что описанное сокращение размерности вектора биометрических характеристик без уменьшения качества входных данных применимо только для размера изображения «вырезанного» лица 112×112 пикселей с масштабом лица 70%.

После выделения векторов биометрических характеристик $\vec{X}$ в основном датасете для каждого пользователя i рассчитываются средние значения μ_i и общее среднее μ для всего датасета. Используя полученные μ_i и μ рассчитываются векторы квантованных значений Q_i и маска M для всех пользователей [2-А].

В дополнительном датасете для каждого пользователя $l_{c=1..70}$ выделяются векторы биометрических характеристик $\vec{X}$, рассчитываются средние значения $\mu_{l,c=1..70}$ и векторы квантованных значений $Q_{l,c=1..70}$.

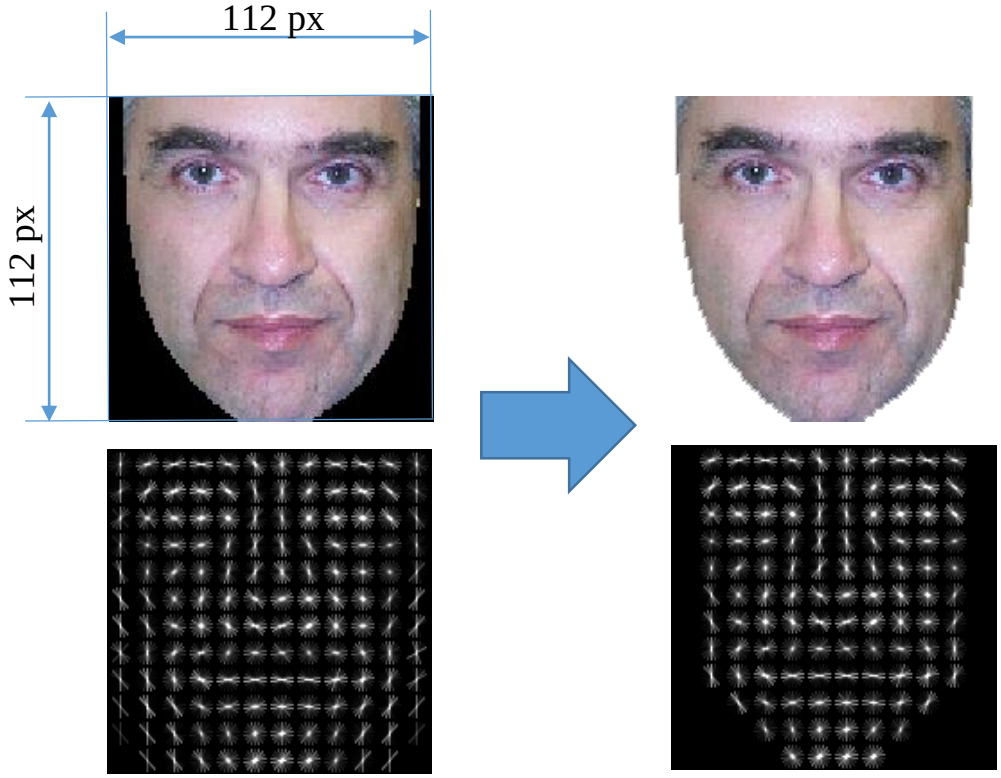


Рисунок 5.4 - Уменьшение размерности вектора биометрических характеристик с 4464 до 3348 элементов

5.3 Оценка распределения двоичных векторов признаков

Для внутри классовых сравнений определяется HD между векторами Q_i и $Q_{l,c=1..70}$, где $i = l$, основного и дополнительного датасетов после наложения маски M. Дополнительный датасет содержит по 70 экземпляров каждого i-го пользователя основного датасета, что приводит к 1330 сравнениям. Результат представлен на рисунке 5.5.

Для межклассовых сравнений HD определяется между векторами Q_i и $Q_{l,c=1..70}$, где $i <> l$, после наложения маски M, что приводит к 23940 сравнениям. Результат представлен на рисунке 5.5.

В вычислительных экспериментах использована длина кодового слова $k = 511$, что соответствует числу наиболее «надежных» бит, выделяемых маской M, поэтому результаты внутри- и межклассовых распределений двоичных векторов признаков представлены в нормированном виде FHD:

$$FHD = \frac{HD}{k} = \frac{HD}{511}$$

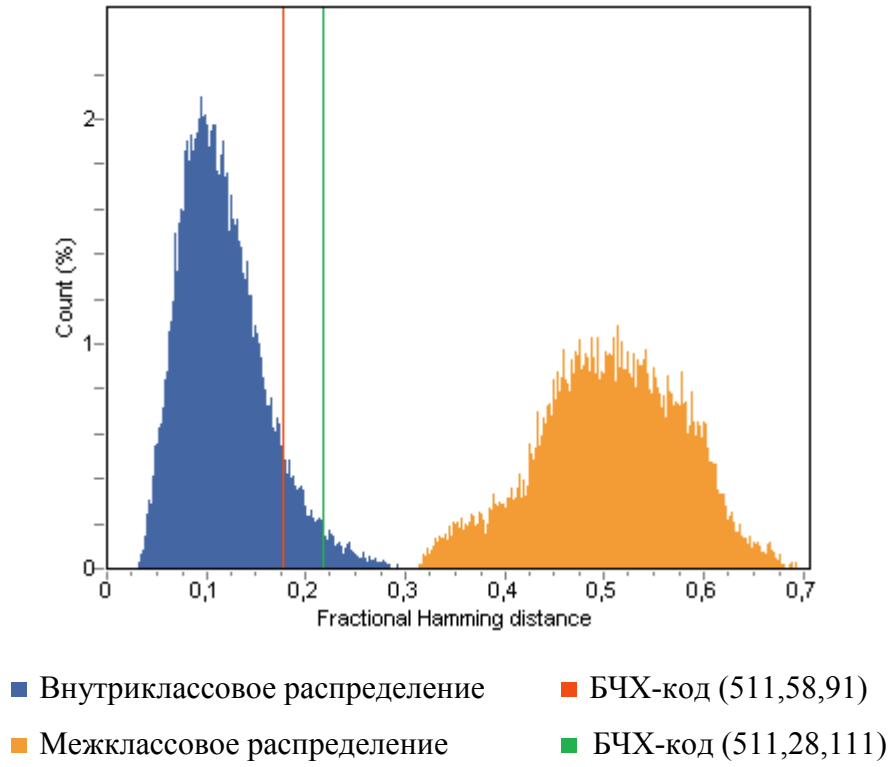


Рисунок 5.5 – Гистограммы внутри- и межклассовых распределений двоичных векторов биометрических характеристик

В процессе проведения тестовых верификаций проводились сравнения секретных ключей S_i, хранящихся в сгенерированной из основного датасета базе данных, с ключами $S'_{l,c=1..70}$, рассчитанными для дополнительного датасета с использованием БЧХ-кода с параметрами (511, 28, 111) по предлагаемому алгоритму.

Для тестирования внутри класса $(i = l)$ проведено 1330 операций сравнения, между классами $(i <> l)$ проведено 23940 операций сравнения. Результаты приведены в таблице 5.2.

Таблица 5.2 - Результаты внутри- и межклассовых верификаций

Пользователь	**Внутри класса**		**Меж-классово**	
	положительно	отрицательно	положительно	отрицательно
0001	68	2	13	1247
0002	70	0	0	1260
0004	70	0	0	1260
0005	70	0	0	1260
0006	64	6	0	1260
0007	70	0	18	1242
0009	70	0	13	1247
0013	57	13	0	1260
0014	70	0	3	1257
0015	70	0	5	1255
0016	70	0	2	1258
0018	70	0	0	1260
0019	58	12	0	1260
0020	54	16	0	1260
0022	70	0	0	1260
0023	70	0	16	1244
0024	70	0	0	1260
0026	70	0	15	1245
0027	70	0	0	1260
	Всего :	**48**	**85**	

$$FRR = \frac{48}{19 \cdot 70} 100\% = \frac{48}{1330} 100\% \approx 3{,}6\%$$

$$FAR = \frac{85}{19 \cdot 1260} 100\% = \frac{85}{23940} 100\% \approx 0$$

Как обсуждалось ранее (см. пункт 3.2), целью является то, что с учетом длины кодового слова k размер секретного ключа S должен быть максимальным, а также количество исправляемых ошибок d должно быть как можно больше. Так при выбранном БЧХ-коде (511, 28, 111) сформированная из датасета Caltech Faces база данных допускает приемлемую $FRR = 3{,}6\%$ и $FAR \approx 0$. Что

позволяет получить $ERR \approx 0{,}26\%$. Результаты классификации двоичных векторов биометрических характеристик показаны на рисунке 5.6.

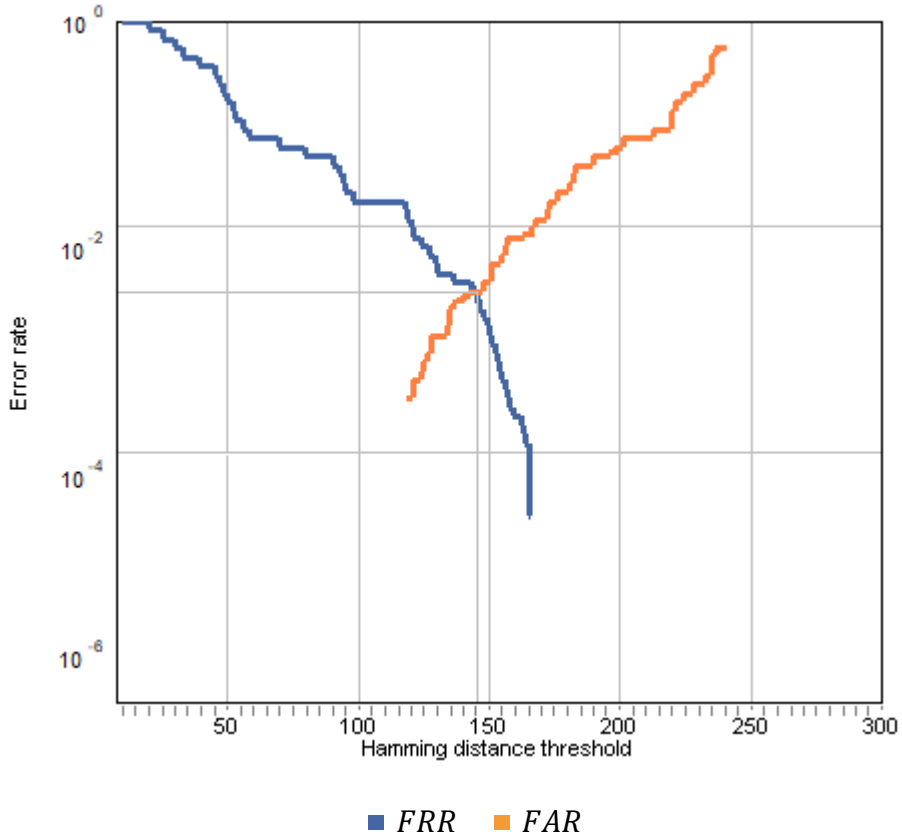


Рисунок 5.6 - Результаты классификации двоичных векторов биометрических характеристик

Тот факт, что ERR достигается на относительно большом $HD = 146$, в основном обусловлен большими вариациями внутри класса и ограниченным количеством биометрических измерений на пользователя. Это затрудняет выбор наиболее «надежных» компонент.

В предложенной системе корректирующая способность БЧХ-кода была увеличена до $d = 111$, что привело к уменьшению количества информационных бит $s = 28$ и уменьшению энтропии биометрического кода. Однако этот фактор можно улучшить путём дополнительного сложения по модулю два псевдослучайной последовательности с биометрическим кодом.

Дискриминативность системы обеспечивается путём применения предложенной маски, которая обеспечивает уникальность imprint каждого пользователя базы данных как минимум в одном бите.

ЗАКЛЮЧЕНИЕ

В ходе выполнения работы разработаны: алгоритм создания биометрического кода пользователя с использованием HOG-структур и маски, сокращающей длину его представления; процедура аутентификации пользователя по его биометрическому коду с применением БЧХ-кодов; приложение FaceAnalyzer, реализующее создание биометрического кода и выполнение аутентификации пользователей.

Апробация представленных алгоритмов и приложения была проведена на общедоступной базе данных лиц Caltech Faces и подтвердила заявленную в работе функциональность. Проведён ряд вычислительных экспериментов (50540 операций сравнения), в результате которых определены характеристики эффективности FRR=3.6%, FAR=0 и ERR=0.26% предложенной системы, что соответствует показателям известных на сегодняшний день методов биометрической идентификации, относящихся к классу «Fuzzy commitment». При этом предложенная система отличается применением маски выделения наиболее «надёжных» бит для расчёта биометрического кода и применением более помехоустойчивого корректирующего ошибки кода БЧХ (511, 28, 111).

Ожидается, что увеличение размерности квантования вещественных данных пользователя и большее количество биометрических измерений при регистрации (например, извлеченных из видеоряда), полученных в «контролируемых» (освещенность, углы наклона и поворота головы и т.п.) условиях, позволит обеспечить более надежный выбор компонент.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

Список использованных источников

1. Ballard L, Kamara S, ReiterM (2008) The practical subtleties of biometric key generation. In:17th Annual USENIX Security Symposium, San Jose, CA, USA, 28 July–1 August 2008.
2. Monrose F, Reiter M, Li Q, Wetzel S (2001) Cryptographic key generation from voice. In: IEEE Symp on Security and Privacy, Oakland, CA, USA, May 2001.
3. Goh A, Ngo D (2003) Computation of cryptographic keys from face biometrics. In: International Federation for Information Processing. Lecture Notes on Computer Science, vol 2828.
4. Vielhauer C, Steinmetz R, Mayerhoefer A (2002) Biometric hash based on statistical features of online signatures. In: 21st International Conference on Pattern Recognition, ICPR 2012, Tsukuba Science City, Japan, November 2012.
5. Vielhauer C, Steinmetz R (2004) Handwriting: feature correlation analysis for biometric hashes. EURASIP Journal on Applied Signal Processing 4:542–558. Special issue on biometric signal processing.
6. Feng H, Chan C (2002) Private key generation from on-line handwritten signatures. In: Information Management and Computer Security, pp 159–164.
7. Kuan Y, Goh A, Ngo D, Teoh A (2005) Cryptographic keys from dynamic hand-signatures with biometric secrecy preservation and replaceability. In: Proc Fourth IEEE Workshop on Automatic Identification Advanced Technologies, AUTO ID 2005, Buffalo, New York, USA, October 2005, pp 27–32.
8. Freire M, Fierrez J, Galbally J, Ortega-Garcia J (2007) Biometric hashing based on genetic selection and its application to on-line signatures. In: Lecture Notes on Computer Science, vol 4642, pp 1134–1143.
9. Juels A,WattenbergM(1999) A fuzzy commitment scheme. In: Proc ACM Conf on Computer and Communications Security, CCS99, Singapore, November 1999, pp 28–36.
10. Davida G, Frankel Y, Matt B, Peralta R (1999) On the relation of error correction and cryptography to an off line biometric based identification scheme. In: Proceedings of WCC99, Workshop on coding and cryptography, Paris, France, January 1999.
11. Juels A, Sudan M (2002) A fuzzy vault scheme. In: IEEE Intl Symp on Information Theory, ISIT 2002, Lausanne, Switzerland, 30 June–5 July 2002.
12. Tuyls P, Verbitsky E, Ignatenko T, Schobben D, Akkermans A (2004) Privacy protected biometric templates: acoustic ear identification. In: Proceedings SPIE, Biometric Technology for Human Identification, vol 5404, Orlando, FL, USA, April 2004, pp 176–182.

13. Tuyls P, Akkermans A, Kevenaar T, Schrijen G, Bazen A, Veldhuis R (2005) Practical biometric authentication with template protection. In: AVBPA, Rye Brook, NY, USA, pp 436–446.

14. Nandakumar K (2010) A fingerprint cryptosystem based on minutiae phase spectrum. In: IEEE International Workshop on Information Forensics and Security, WIFS10, Seattle, USA, December 2010.

15. Van der Veen M, Kevenaar T, Schrijen G-J, Akkermans T, Zuo F (2006) Face biometrics Brazil, with renewable templates. In: SPIE Proc on Security, Steganography, and Watermarking of Multimedia Contents, vol 6072, San Jose, CA, USA, January 2005.

16. Kelkboom E, Gökberk B, Kevenaar T, Akkermans AHM, Van der VeenM(2007) 3d face: biometrics template protection for 3d face recognition. In: Lecture Notes on Computer Science, vol 4642, pp 566–573.

17. Hao F, Anderson R, Daugman J (2006) Combining crypto with biometrics effectively. IEEE Transactions on Computers 55:1081–1088.

18. Rathgeb C, Uhl A (2009) Systematic construction of iris-based fuzzy commitment schemes. In: Proceedings of the Third International Conference on Advances in Biometrics, ICB'09, Alghero, Italy, June 2009.

19. Maiorana E, Campisi P, Neri A (2008) User adaptive fuzzy commitment for signature templates protection and renewability. SPIE Journal of Electronic Imaging 17(1), January–March. Special section on biometrics: advances in security, usability and interoperability.

20. Maiorana E, Campisi P (2010) Fuzzy commitment for function based signature template protection. IEEE Signal Processing Letters 17(3):249–252.

21. Dodis Y, Reyzin L, Smith A (2004) Fuzzy extractors: how to generate strong keys from biometrics and other noisy data. In: Eurocrypt. Lecture Notes on Computer Science, vol 3027. Springer, Berlin, pp 523–540.

22. Sutcu Y, Li Q, Memon N (2007) Protecting biometric templates with sketch: theory and practice. IEEE Transactions on Information Forensics and Security 2(3):503–512.

23. Li Q, Guo M, Chang E-C (2008) Fuzzy extractors for asymmetric biometric representations. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW'08, Anchorage, AK, USA, June 2008.

24. Buhan I, Doumen J, Hartel P, Veldhuis R (2007) Fuzzy extractors for continuous distributions. In: 2nd ACM Symposium on Information, Computer and Communications Security (ASIACCS), Singapore, March 2007, pp 353–355.

25. Sutcu Y, Li Q, Memon N (2009) Design and analysis of fuzzy extractors for faces. In: Proc SPIE Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI, vol 7306, Orlando, Florida, USA, April 2009.

26. Sutcu Y, Li Q,Memon N (2007) Secure biometric templates from fingerprint-face features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Workshop on biometrics, Minneapolis, MN, USA, June 2007.

27. Nandakumar K, Jain AK (2008) Multibiometric template security using fuzzy vault. In: 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, BTAS'08, Washington, DC, USA.

28. Kelkboom E, Zhou X, Breebaart J, Veldhuis R, Busch C (2009) Multi-algorithm fusion with template protection. In: 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems, BTAS'09, Washington, DC, USA.

29. Kanade S, Petrovska-Delacretaz D, Dorizzi B (2010) Obtaining cryptographic keys using feature level fusion of iris and face biometrics for secure user authentication. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Workshop on biometrics, San Francisco, USA, June 2010.

30. Nagar A, Nandakumar K, Jain AK (2012) Multibiometric cryptosystems based on feature level fusion. IEEE Transactions on Information Forensics and Security 7(1): 255–268.

31. Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency "OpenFace: an open source facial behavior analysis toolkit" in IEEE Winter Conference on Applications of Computer Vision, Lace Placid, NY, March 2016.

32. Гистограмма направленных градиентов // Академик. Словари и энциклопедии [Электронный ресурс]. – URL: https://dic.academic.ru/dic.nsf/ruwiki/1700368 (дата обращения: 23.03.2017).

33. Метод опорных векторов // MachineLearning.ru - Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. [Электронный ресурс]. URL: http://www.machinelearning.ru/wiki/index.php?title=SVM (Дата обращения: 21.03.2017).

34. Dala N., Triggs B. Histograms of oriented gradients for human etection. [Электронный ресурс]. – URL: http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf (Дата обращения: 22.03.2017).

35. Мак-Вильямс Ф.Дж., Слоэн Н.Дж.А. Теория кодов, исправляющих ошибки. – М.: Связь, 1979. – 744с.

36. Питерсон У., Уэлдон Э. Коды, исправляющие ошибки / Пер. с англ. под ред. РЛ. Добрушина и С.И. Самойленко. – М.: Мир, 1976. – 596с.

37. Кларк Дж. мл., Кейн Дж. Кодирование с исправлением ошибок в системах цифровой связи / Пер. с англ. под ред. Б.С. Цыбакова. – М.: Радио и связь, 1987. – 392с.

38. Constrained local neural fields for robust facial landmark detection in the wild / Tadas Baltrusaitis, Peter Robinson / Louis-Philippe Morency // University of

Cambridge Computer Laboratory, 15 JJ Thomson Avenue, USC Institute for Creative Technologies, 12015 Waterfront Drive.

39. Face recognition with renewable and privacy preserving binary templates / T. A. M. Kevenaar, G. J. Schrijen, M. van der Veen, A. H. M. Akkerman // Philips Research Prof. Holstlaan 4, 5656 AA, Eindhoven, the Netherlands.

40. University of Cambridge, Computer Laboratory [Электронный ресурс]: OpenFace: an open source facial behavior analysis toolkit. – Режим доступа: http://www.cl.cam.ac.uk/research/rainbow/projects/openface/. – Дата доступа: 20.02.2017.

41. The world’s leading software development platform [Электронный ресурс]: Tadas Baltrusaitis \ OpenFace. – Режим доступа: https://github.com/TadasBaltrusaitis/OpenFace. – Дата доступа: 16.02.2017.

42. Сайт Славы Антонова [Электронный ресурс]: Программная реализация БЧХ-кодов переменной длины. – Режим доступа: http://deadbeef.narod.ru/work/ecc2/index.htm. – Дата доступа: 21.05.2017.

43. M. Weber, Frontal face dataset 1999 [Electronic resource]: California Institute of Technology. - Mode of access: http://www.vision.caltech.edu/html-files/archive. - Date of access 21.04.2017.

Список публикаций соискателя

1-А. Веретило, Ю. Н. Создание биометрической базы данных лиц на основе HOG-структур / Ю. Н. Веретило // Физика конденсированного состояния: материалы XXV междунар. науч.-практ. конф. аспир., магистр. и студ., Гродно, 20 апр. 2017 г. ГрГУ им. Я. Купалы: тез. докл. /физ.-техн. фак.; редкол.: В.Г.Барсуков [и др.] - Гродно, 2017. - С. 127-129.

2-А. Ассанович, Б.А., Веретило, Ю.Н. Биометрическая база данных на основе HOG-структур и кодов БЧХ / Б.А.Ассанович / Ю.Н.Веретило // Информационные технологии и системы 2017 (ИТС 2017) = Information Technologies and Systems 2017 (ITS 2017) : материалы междунар. науч. конф. Минск, 25 окт.2017 г., тез. докл. / редкол. : Л.Ю.Шилин [и др.] – Минск, 2017. – С. 286-287.

ПРИЛОЖЕНИЕ А

Процедура расчета маски

```
procedure TMainForm.MaskCalcClick(Sender: TObject);
var i,
    j,
    k,
    Equal,
    Unequal,
    CountM          : integer;

    QFileName,
    RFileName,
    PFileName,
    Str,
    MagicFileName : string;

    QArray,
    PArray,
    MagicArray      : BoolArray;

    StatArray       : array of Statistic;

    IndexPArray,
    IndexQArray     : array of word;

    Flag            : boolean;

    MagicFile       : file;

    Block           : byte;
begin
//длина массива маски = длине биометрического вектора
SetLength(MagicArray,BioVectorLength);
```

```
//заполнение массива маски нулями (false)
for i := 0 to Length(MagicArray) - 1 do
    MagicArray[i] := false;
if DataSetsBox.text <> ''
then begin
        LogMemo.Lines.Add(TimeTostr(now) + ' >> Start to calculation of statistics');
        //расчитываю файлы позиций, где R>=порога
        for i := 0 to UsersBox.Items.Count - 1 do
        begin
        RFileName := dataset_path + UsersBox.Items[i] + '\' + UsersBox.Items[i]
        + RFileType;
        PFileName := dataset_path + UsersBox.Items[i] + '\' + UsersBox.Items[i]
        + PFileType;
        CreatePFile(RFileName, StrToFloat(Edit1.Text + ',' + Edit2.Text),
        PFileName);
        end;
        //сбрасываю массив статистики
        StatArray := nil;
        SetLength(StatArray, 0);
        //заполняю массив статистики векторами позиций и квантований
        for i := 0 to UsersBox.Items.Count - 1 do
        begin
        FileName := dataset_path + UsersBox.Items[i] + '\' + UsersBox.Items[i]
        + QFileType;
        PFileName := dataset_path + UsersBox.Items[i] + '\' + UsersBox.Items[i]
        + PFiletype;
        SetLength(StatArray,Length(StatArray) + 1);
        StatArray[i].Positions := GetBoolArray(PFileName);
        StatArray[i].Quantizations := GetBoolArray(QFileName);
        end;
        //длина массива индексов позиций R>=порога  = 0
        SetLength(IndexPArray, 0);
        k := -1;
        for i := 1 to BioVectorLength do
          begin
          Flag := true;
```

```
        for j := 0 to Length(StatArray) - 1 do
          if not StatArray[j].Positions[i - 1]
          then begin
                  Flag := false;
                  Break;
                  end;
        if Flag
        then begin
                SetLength(IndexPArray,Length(IndexPArray) + 1);
                inc(k);
                IndexPArray[k] := i;
                end;
        end;
SetLength(IndexQArray, 0);
k := -1;
for i := 0 to Length(IndexPArray) - 1 do
   begin
   Flag := true;
   for j := 0 to Length(StatArray) - 1 do
      if not StatArray[j].Quantizations[IndexPArray[i] - 1]
      then begin
              Flag := false;
              Break;
              end;
   if Flag
   then inc(equal)
   else begin
          inc(unequal);
          SetLength(IndexQArray,Length(IndexQArray) + 1);
          inc(k);
          IndexQArray[k] := IndexParray[i];
          Flag := true;
          end;
   end;

for k := 0 to Length(IndexQArray) - 1 do
```

```
    MagicArray[IndexQArray[k] - 1] := true;
CountM := 0;
Str := '';
for i := 0 to Length(MagicArray) - 1 do
   if MagicArray[i]
   then begin
            Str := Str + '1';
            inc(CountM);
            end
   else Str := Str + '0';
MagicFileName := dataset_path + 'MagicMask' + MagicFileType;
AssignFile(MagicFile, MagicFileName);
ReWrite(MagicFile,1);
Str := '';
if BioVectorLength = FullVector
then for i := 0 to FullVector - 1 do
          begin
          if MagicArray[i]
          then Str := Str + '1'
          else Str := Str + '0';
          if i mod 8 = 0
          then begin
                  block := BinToHex(Str);
                  BlockWrite(MagicFile, block, SizeOf(block));
                  Str := '';
                  end;
          end;
if BioVectorLength = CutVector
then  begin
        for i := 1 to CutVector - 4 do
          begin
          if MagicArray[i - 1]
          then Str := Str + '1'
          else Str := Str + '0';
          if i mod 8 =0
          then begin
```

```
                    block := BinToHex(Str);
                    BlockWrite(MagicFile, block, SizeOf(block));
                    Str := '';
                    end;
                  end;
                Str := '';
                for i := 3345 to CutVector do
                if MagicArray[I - 1]
                then Str := Str + '1'
                else Str := Str + '0';
                Str := Str + '0000';
                Block := BinToHex(Str);
                BlockWrite(MagicFile, block, SizeOf(block));
                end;
    CloseFile(MagicFile);
    LogMemo.Lines.Add(TimeTostr(now) + ' >> End of statistics calculation');
    Application.MessageBox(PChar('Calculation of statistics is complete' ),
                           PChar(APP_NAME + ' :) / Complete ! '),
                           mb_OK+MB_ICONINFORMATION);
    MagicArray := nil;
    QArray := nil;
    PArray := nil;
    IndexPArray := nil;
    IndexQArray := nil;
    StatArray := nil;
    end
else
Application.MessageBox(PChar('DataSet not selected. There is nothing to calculate
:(' ), PChar(APP_NAME + ' :( / No DataSet '), mb_OK+MB_ICONWARNING);
end;
```