Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.93'12

Григорьев
Андрей Владимирович

Средство обеспечения безопасности для системы электронного голосования

АВТОРЕФЕРАТ
на соискание степени магистра информатики и вычислительной техники
по специальности 1-40 81 01 «Информатика и технологии разработки программного обеспечения»

_____

Научный руководитель
Теслюк Владимир Николаевич
доцент, кандидат физико-математических наук

_____

Минск 2018

КРАТКОЕ ВВЕДЕНИЕ

С развитием информационных технологий и проникновением сети Интернет во все сферы общественной жизни возможность выбора и голосования стали неотъемлемой частью повседневной рутины.

Всеобщая информатизация и компьютеризация привели к тому, что к проблеме тайного голосования обратилась криптография. Первый протокол тайного голосования был предложен Д. Чаумом в 1981 году. Его протокол был защищён от внешнего вмешательства, однако строился на полном доверии сторон. Существенным недостатком протокола была невозможность проверки участниками голосования, как были учтены их голоса и даже получил ли организатор их бюллетень. Для решения этих проблем в 1991 году Нурми, Салома и Сантин предложили разделить единого организатора голосования на два агентства: собственно, организатора голосования и регистратора.

В 1992 году была представлена схема Фудзиоко-Окамото-Охта. Она также основывается на протоколе двух агентств, но предлагает использовать маскирующее (ослепляющее) шифрование для проведения т.н. "подписи вслепую" со стороны регистратора. В 1998 году Ци Хэ и Джунминь Су представили улучшенную версию схемы Фудзиоко-Окамото-Охта: они предложили регистратору подписывать лишь ключ участника голосования, что позволило участникам изменять свой выбор в любой момент до окончания голосования.

В 2016 году в рамках работы над дипломным проектом была разработана система электронного голосования PKN Voting, основанная на протоколе Хэ-Су. Несмотря на то, что сам протокол тайного голосования был в ней успешно реализован и протестирован, в системе не были реализованы меры, позволяющие проводить безопасный обмен сообщениями по сети Интернет, а также не была реализована безопасная система идентификации пользователей.

Наибольшей проблемой в безопасности системы была идентификация пользователей. Этот вопрос был оставлен за рамками производимой тогда работы, и в итоге пользователь для идентификации должен был вводить свой ключ (например, номер паспорта). Однако, номер паспорта не является криптографически скрытой информацией, и введение паролей вряд ли смогло бы полностью решить проблему, т.к. база паролей могла бы быть похищена, либо же пароль мог быть утерян самим пользователем.

Диссертационная работа посвящена разработке средства обеспечения безопасности для системы электронного голосования, необходимого для надёжной идентификации пользователей.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью данного исследования является разработанное средство обеспечения безопасности для системы электронного голосования, с помощью которого будет осуществляться надёжная и безопасная аутентификация пользователей.

Для достижения поставленной цели необходимо решить следующие задачи:

1 Произвести обзор предметной области биометрических способов аутентификации.

2 Произвести исследование выбранного способа биометрической аутентификации пользователей.

3 Спроектировать модуль биометрической аутентификации.

4 Разработать и интегрировать модуль биометрической аутентификации в систему электронного голосования.

Объектом исследования является система электронного голосования. Предметом исследования является аутентификация пользователей системы электронного голосования.

Гипотеза исследования состоит в положении о возможности реализации модуля биометрической аутентификации и интеграции его в существующую систему электронного голосования.

Связь работы с приоритетными направлениями научных исследований

Надёжность аутентификации пользователей в различных системах является ключевым направлением исследований в пользовательских системах. Одним и простых и эффективных способов является двухфакторная аутентификация, которая, впрочем, лишь увеличивает сложность взлома системы. Биометрическая аутентификация обладает несравнимо большим удобством для пользователей, а также вносит принципиальной иной характер аутентификации, что существенно усложняет взлом. Как показывает рынок мобильных устройств, именно биометрическая аутентификация является наиболее перспективным видом аутентификации, в особенности аутентификация по отпечатку пальца и с недавнего времени аутентификация по геометрии лица. С развитием технологий необходимое для биометрической аутентификации становится всё более мобильным, что расширяет

применимость данных методов. Кроме того, рост вычислительных мощностей позволяет применять всё более и более сложные алгоритмы.

Личный вклад соискателя

Результаты, приведённые в диссертации, получены соискателем лично. Вклад научного руководителя, Теслюка В.И., заключается в формулировке целей и задач исследования.

Структура и объём диссертации

Диссертация состоит из оглавления, общей характеристики работы, введения, основной части, состоящей из 4 глав, заключения и списка использованной литературы.

Первая глава содержит обзор и анализ предметной области, описание различных методов безопасной аутентификации пользователей и выбор наиболее перспективного по таким показателям, как надёжность, простота реализации и эффективность. Также в главе содержится постановка задач на дальнейшие исследования исходя из выбранного пути решения.

Вторая глава содержит теоретическое описание различных методов биометрической аутентификации с указанием преимуществ и недостатков, анализ алгоритмов аутентификации по геометрии лица: метод Виолы-Джонса для обнаружения лица и алгоритм распознавания лиц Eigenfaces. Также в этом разделе даётся краткое описание библиотеки OpenCV.

В третьей главе описывается процесс проектирования модуля биометрической аутентификации пользователей для системы электронного голосования.

В четвёртой главе описывается процесс реализации модуля, его основных составляющих и общих схем работы при распознавании лиц. Также даётся описание результатов работы модуля.

В заключении подводятся итоги и делаются выводы по работе, а также описывается план дальнейшего развития проекта.

Общий объём работы составляет 57 с., 16 рис., 0 табл., 18 источников.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследований, показана актуальность темы диссертационной работы, дана

краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В **первой главе** проведено исследование способов аутентификации пользователей в электронных системах. Приведён анализ трёх типов аутентификации. На основании проведённого анализа выполнена постановка задачи.

Даны определения понятиям идентификация, аутентификация и авторизация.

Идентификация – это заявление о том, кем является пользователь. К примеру, логин, электронная почта, номер паспорта и т.д.

Аутентификация – процесс доказательства того, что пользователь является тем, кем идентифицируется (от англ. «authentic» – подлинный).

Авторизация – процесс проверки, что у пользователя есть права на доступ к определённому ресурсу, запрашиваемому пользователем.

По базису системы аутентификации выделено три типа систем:

1 Основанные на чём-то, что известно пользователю (к примеру, пароль).

2 Основанные на чём-то, чем владеет пользователь. Это должен быть предмет с уникальными характеристиками или содержимым, например, USB-токен или смарт-карта.

3 Основанные на чём-то, чем является сам пользователь, или что является неотъемлемой частью пользователя. К этому типу относятся биометрические методы аутентификации.

Наиболее широкое распространение получила форма аутентификации, основанная на вводе логина и пароля. Она проста и понятна для пользователей. В простейшем виде она представляет из себя реестр пар логин-пароль. По введённому пользователем логину находится пара в реестре, далее введённым пароль сравнивается с найденным паролем. Если пароли совпадают, пользователь аутентифицирован, иначе запрос на аутентификацию отклоняется.

Существенным недостатком таких систем является тот факт, что за создание и хранение паролей ответственны сами пользователи. Р. Моррис и К. Томсон из Bell Labs в 1979 году провели показательный опыт. Они составили список из вероятных паролей пользователей: фамилий, имён, улиц, городов, слов из словарей и слов в обратном написании, регистрационных номеров и коротких строк из случайных символов. После этого они сравнили этот список с системный файлом паролей в системе UNIX. Более 86% всех паролей оказались в списке.

Таким образом, было принято решение не использовать простую систему логин-пароль в качестве средства обеспечения безопасности.

Следующий метод аутентификации заключается в проверке наличия у пользователей каких-либо материальных объектов вместо проверки знания какого-либо пароля. В наши дни в качестве физического объекта часто используются смарт-карты и USB-токены. Смарт-карты могут вставляться в специальное считывающее устройство, либо быть считаны бесконтактно. USB-токены необходимо вставлять USB-порт компьютера.

Аутентификация со смарт-картами часто проводится по простой схеме «оклик-отзыв». Например, сервер посылает 512-разрядное случайное число на смарт-карту, которая добавляет к нему 512-разрядный пароль. Затем сумма возводится в квадрат, и серверу отсылаются обратно средние 512 бит результата. Т.к. сервер знает пароль пользователя, то может произвести те же операции и проверить подлинность результата. На рисунке 1 приведена схема такой системы.

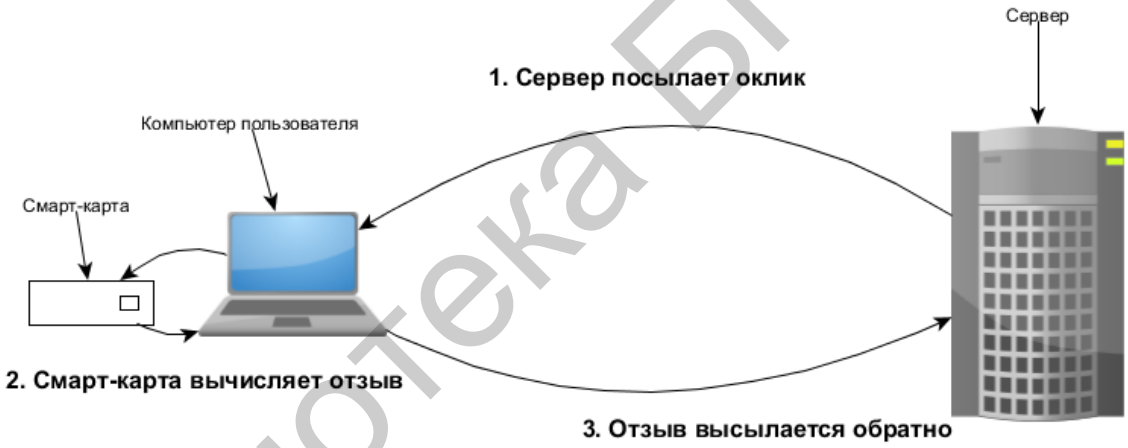

Рисунок 1 – Схема «оклик-отзыв»

Вследствие необходимости наличия специального оборудования у пользователей для использования систем со смарт-картами, от них было решено также отказаться.

Последний метод – биометрический. Этот метод основан на измерении физических параметров пользователя, которые тяжело или невозможно подделать. Они называются биометрическими параметрами. Например, для аутентификации пользователей может использоваться устройство для считывания отпечатков пальцев или тембра голоса.

После измерения какой-то характеристики из неё вычисляются характерные признаки, которые затем сохраняются в базе данных. Кроме того, биометрические характеристики могут храниться на смарт-карте, которая находится у пользователя. Для идентификации пользователя необходимо его

регистрационное имя и новое измерение тех же биометрических признаков. Регистрационное имя необходимо потому, что измерения биометрических характеристик не имеют точного результата, отчего их довольно сложно индексировать для проведения поиска.

Таким образом, проанализировав основные типа аутентификации пользователей, был сделан выбор в пользу биометрических методов. Это было сделано по нескольким причинам:

- во-первых, широта выбора возможных вариантов реализации позволяет выбрать тот способ, который окажется приемлемым как для разработчика, так и для пользователей;
- во-вторых, биометрические методы сочетают в себе высокую надёжность и относительную простоту реализации;
- в-третьих, биометрические методы очевидно дешевле во внедрении, чем схожие по надёжности методы, основанные на физических носителях информации;
- в-четвёртых, биометрические методы являются наиболее перспективными в данный момент, т.к. всё больше и больше компаний внедряют их в свои продукты.

Вторая глава посвящена исследованию теоретической базы систем биометрической аутентификации, выбору подходящего метода и описанию алгоритмов работы.

Биометрические методы аутентификации поделены на две группы: *физиологические*, или статические, и *психологические*, или динамические.

К физиологическим относятся методы, основанные на физиологических характеристиках, т.е. таких, которые даны ему от рождения, и являются неотъемлемыми и уникальными.

К психологическим относят методы, основанные на поведенческой (динамической) характеристике человека. Эти методы используют особенности, которые характерны для подсознательных движений в процессе выполнения каких-либо действий.

Все системы биометрической аутентификации имеют общую схему работы, представленную на рисунке 2.

Приводится описание наиболее популярных методов биометрической аутентификации: по отпечатку пальца, по сетчатке глаза, по геометрии лица, по голосу и по рукописной подписи.

Для разработки модуля биометрической аутентификации пользователей в системе электронного голосования изначально был выбран метод дактилоскопической аутентификации. Однако, в ходе исследований выяснилось, что приобретение дактилоскопического датчика не является

тривиальным. Было принято решение отказаться от данного способа и обратиться к аутентификации по геометрии лица, для чего использовались технологии компьютерного зрения.

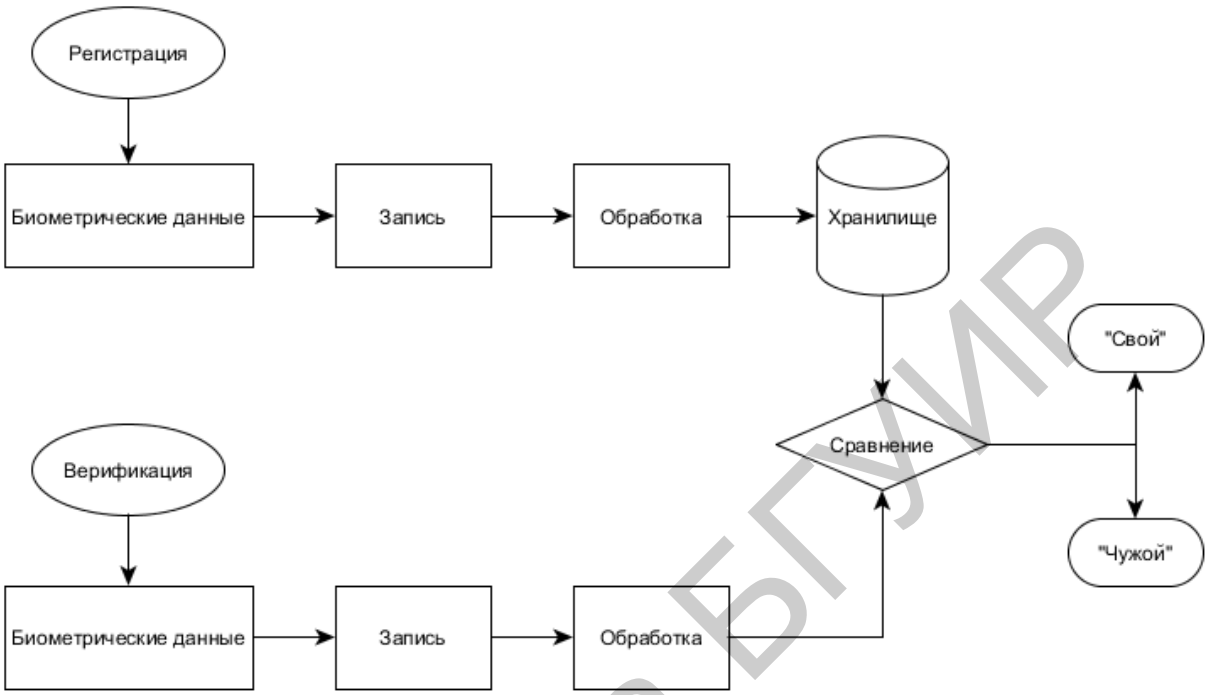

Рисунок 2 – Схема системы биометрической аутентификации

Компьютерное зрение – теория и технология создания машин для обнаружения, отслеживания и классификации объектов на изображениях. Для эмуляции зрения в реальном времени используются видео данные, которые могут быть представлены в виде последовательности изображений с различных камер или трёхмерных датчиков, например, медицинского сканера.

Аутентификация пользователей по геометрии лица будет состоять из нескольких этапов: *определение* лица на изображении и *распознавание* пользователя, которому принадлежит лицо. Основными алгоритмами, выбранными для каждого из этапов, стали:

- метод Виолы-Джонса для определения лиц на изображении;
- алгоритм Eigenfaces для распознавания принадлежности полученного на предыдущем шаге лица.

Первый метод был разработан и представлен в 2001 году Полом Виолой и Майклом Джонсом. На сегодняшний день он является основным для поиска объектов на изображении в реальном времени. Основные принципы метода таковы:

- используется принцип сканирующего окна;

- использование интегрального представления изображений для быстрого вычисления необходимых объектов;
- для поиска нужного объекта используются признаки Хаара;
- для выбора наиболее весомых признаков для искомого объекта используется бустинг (от англ. boost – улучшение, усиление);
- использование классификаторов для обработки признаков;
- использование каскадов признаков для быстрого отсеивания окон, где не найден объект.

Принцип сканирующего окна состоит в сканировании изображения при помощи окна поиска и применения классификатора к каждому полученному участку изображения.

Итак, интегральное представление – это матрица такого же размера, что и исходное изображение. Каждый элемент этой матрицы – сумма интенсивности пикселей левее и выше данного элемента. Элементы матрицы рассчитываются по формуле (1):

$$L(x, y) = \sum_{i=0, j=0}^{i \le x, j \le y} I(i, j), \quad (1)$$

Признаки Хаара представляют собой наборы прямоугольных областей, разделённых на 2 цвета: чёрный и белый. Значение признака представляет собой разность между суммами интенсивностей пикселей в белых и чёрных областях. Для вычисления сумм интенсивностей как раз используется интегральное представление изображений, описанное выше.

Для построения сильных комплексных классификаторов из множества слабых была предложена идея цепочки классификаторов (называемой каскадом), каждый из которых обучается на ошибках предыдущего. Эта идея и называется бустингом.

В настоящее время бустинг, т.е. усиление слабых классификаторов является наиболее эффективным методом классификации ввиду сочетания скорости, эффективности и простоты реализации.

Для распознавания необходимо вычислять особенные свойства, присущие конкретному человеку. Именно подобная идея и легла в основу алгоритма Eigenfaces. Его название можно перевести как «собственные лица» или «характеристические лица», по аналогии с англ. eigenvector, eigenvalue – собственный вектор, собственное значение.

В 1987 году Л. Сирович и М. Кирби показали, что для того, чтобы сформировать набор базисных характеристик лиц можно использовать метод главных компонент над коллекцией изображений лиц. Эти базисные

изображения можно линейно скомбинировать, чтобы восстановить изображения из исходной выборки.

Задача алгоритма при обучении – представить каждое изображение как сумму базисных изображений:

$$I_i = \sum_{j=1}^{K} w_j * u_j + \mu \quad (2)$$

Таким образом, мы получим представление изображения в виде суммы «среднего» лица и некоторых долей «стандартных особенностей» лиц.

Для распознавания, когда на вход подаётся произвольное изображение, производятся похожие преобразования. Сначала изображение преобразуется в вектор. После этого по формуле вычисляется вектор, отражающий переведённое в новый базис и усреднённое изображение. Затем нужно определить, к какому из существующих изображений новое расположено ближе всего. Для этого используется расстояние Махалонобиса.

Эти и многие другие методы реализованы в библиотеке OpenCV. OpenCV – библиотека компьютерного зрения с открытым исходным кодом. Библиотека написана на C и C++ и работает на компьютерах под управлением Linux, Windows, Mac OS X. Так же активно развиваются интерфейсы библиотеки для Python, Ruby, Matlab и других языков программирования.

Библиотека OpenCV была разработана с целью повышения вычислительной эффективности и с уклоном на приложения реального времени. OpenCV написана с использованием оптимизированного C и может использовать многоядерные процессоры.

Третья глава посвящена проектированию модуля биометрической аутентификации для системы электронного голосования.

Первая задача – разработать архитектурное решение для интеграции модуля. Результат обновления архитектуры системы электронного голосования представлен на рисунке 3.

Подобная архитектура позволит при внесении минимальных изменений в уже существующую систему добавить новую функциональность. При этом архитектура системы останется достаточно слабосвязанной, что серьёзно упрощает внесение изменений в будущем.

Для поддержки модуля биометрической аутентификации необходимо внести изменения в базу данных валидатора, т.к. существующая модель не предполагает пользователя как единую сущность.

Между сущностями Voter и VoterSet теперь установлено отношение многие-ко-многим. Это сделано с помощью промежуточной таблицы VoterSetVoter. В этой таблице помимо ссылок на таблицы Voter и VoterSet

также хранятся поля TagId и HasVoted. TagId – ссылка на таблицу Tag, где хранятся сгенерированные при создании набора голосующих специальные тэги.

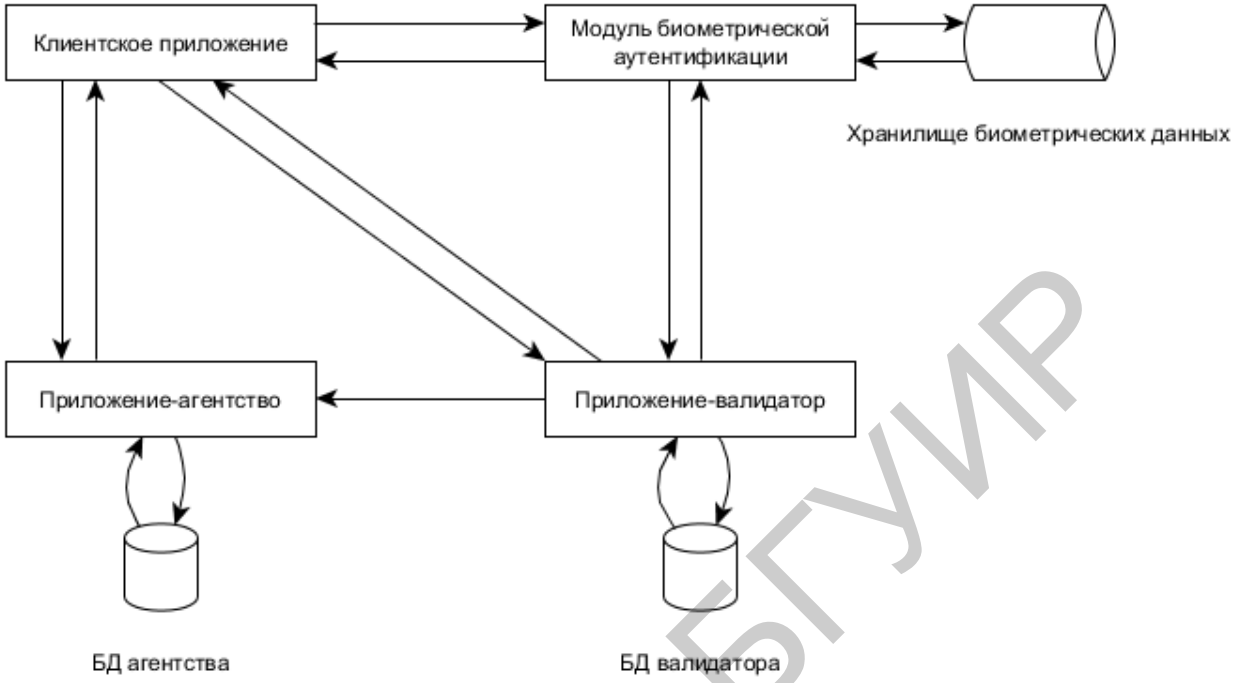

Рисунок 3 – Обновлённая архитектурная схема системы электронного голосования

Для более удобного использования и следования принципам разделения ответственностей логика работы с хранилищем биометрической аутентификации и логика проведения обнаружения и распознавания лиц отделены от непосредственно веб-сервиса, предоставляющего интерфейс для использования другими приложениями системы. Таким образом, архитектура модуля будет соответствовать принципам трёхслойной архитектуры, общей для всей системы. Наличие единой архитектуры позволяет приобщить новую кодовую базу к уже существующей, что упрощает последующее внесение изменений.

Чтобы не нарушать общность кодовой базы системы, необходимо использовать те же подходы к построению новых страниц в клиентской части, что были использованы при разработки самой системы электронного голосования.

Четвёртая глава описывает реализацию модуля биометрической аутентификации.

Как было отмечено выше, вообще говоря, для распознавания лиц нужно сделать два шага: определить лицо на изображении и затем распознать лицо на ней. Из этого следует, что главными в модуле являются два класса: FaceDetector

и FaceRecognizer. Эти два класса составили основу модуля биометрической аутентификации.

Класс FaceDetector отвечает за обнаружение лиц на изображениях, поступающий на вход. Для обнаружения используется метод Виолы-Джонса, описанный в главе 2. Готовая тренированная модель каскада признаков Хаара хранится в приложении и используется при создании объекта типа CascadeClassifier.

При создании матрицы из байтов, содержащих изображения, они вычитываются в режиме gray scale. Это необходимо для того, чтобы все элементы матрицы укладывались в промежуток от 0 до 255, т.к. в этом режиме изображения переводятся в чёрно-белый режим, в котором каждый пиксель занимает 1 байт и представляет градацию серого, где 0 – чёрный цвет, а 255 – белый. Перед тем, как вернуть вырезанные изображения с лицами, над ними проводится операция ResizeImage. Это необходимо для корректной работы метода Eigenfaces.

Класс FaceRecognizer отвечает за распознавание лиц, а также за тренировку и сохранения модели. Для распознавания лиц используется метод Eigenfaces, также описанный в главе 2. Для того, чтобы сохраняться состояние модели (т.е. вычисленные собственные векторы и коэффициенты), используется встроенное сохранение в файл, путь к которому передаётся в конструктор класса. Далее проверяется наличие этого файла, и, если он есть, модель вычитывается оттуда.

Предполагается, что метод распознавания будет принимать сразу приведённое к стандарту изображение, на котором лицо было обнаружено при помощи класса FaceDetector. Для тренировки используются изображения, полученные при помощи класса FaceDetector и сохранённые в базе данных. Результатом распознавания является объект класса RecognitionResult, содержащий поля Id и Confidence.

После разработки модуля распознавания лиц необходимо произвести интеграцию данного модуля в систему с целью создания возможности регистрации пользователями изображений со своим лицом, а также входа в систему при помощи полученного в режиме реального времени изображения их лица с веб-камеры.

При внесении изменений в базу данных, для того, чтобы избежать этого, был совершён переход на миграции Code-First, которые позволяют применять (в автоматическом или ручном режиме) изменения к текущей базе данных, не теряя данных пользователей. Автоматическая миграция базы данных в таком случае должна быть отключена, так как это предупредит нежелательное применение изменений в автоматическом режиме, даже без создания явных

миграций. Код миграции применяет лишь необходимые изменения, не затрагивая остальную базу. Кроме того, встроенные механизмы проверок следят за сохранением целостности данных.

Во всех приложениях системы уровень представления реализован с использованием технологий ASP.NET MVC 5 и библиотеки knockout.js. Приводится более подробное рассмотрение данных технологий.

Ключевым элементом инфраструктуры сервера на ASP.NET MVC 5 является контроллер. Контроллер объединяет набор методов-действий, каждый из которых возвращает либо ViewResult, либо JsonResult (в случае методов для обслуживания AJAX-запросов). Отличительной особенностью использования MVC в данной системе стало повсеместное использование асинхронных действий.

Клиентская часть уровня представления реализована с использованием библиотеки knockout.js. Данная библиотека подразумевает использование двунаправленных привязок данных к разметке для обеспечения их автоматического отображения и обновления. Для создания таких привязок используются т.н. «наблюдаемые» переменные (observable).

Далее описывается метод захвата изображений с веб-камеры пользователя для последующей обработки на сервере.

Для доступа к видео с веб-камеры пользователя используется механизм Media Devices. Эта технология является достаточно новой и пришла на замену устаревшему методу браузера getUserMedia().

В новом стандарте объявлено пространство mediaDevices, содержащее новый метод getUserMedia, а также методы enumerateDevices, getSupportedConstraints и событие ondevicechange. Они используются при работе с медиа-устройствами пользователя, в частности, микрофоном и веб-камерой.

Наиболее надёжный поход для захвата изображения с камеры состоит в следующем:

1 Создать в документе элемент canvas (холст), однако не вставлять его в DOM, чтобы не отображать его.

2 Установить созданному элементу canvas ширину и высоту, соответствующую ширине и высоте видео.

3 Получить контекст элемента canvas с помощью метода getContext(‘2d’).

4 Воспользоваться методом drawImage у контекста, передав ему ссылку на элемент video и ширину и высоту, а также позицию начала отрисовки изображения.

5 Вызвать у созданного элемента canvas метод toDataURL с параметром 'image/jpeg', обозначающим MIME-тип изображения в формате JPEG.

6 Полученную таким образом ссылку на изображение необходимо поместить в атрибут src элемента img, расположенного на странице.

ЗАКЛЮЧЕНИЕ

В рамках данной работы было разработано средство обеспечения безопасности для системы электронного голосования, решающее проблему аутентификации пользователей при помощи методов биометрической аутентификации по геометрии лица.

В работе были рассмотрены и проанализированы основные методы биометрической аутентификации, приведён обзор алгоритма обнаружения лиц Виолы-Джонса и алгоритма распознавания лиц Eigenfaces.

Был спроектирован модуль биометрической аутентификации, а также необходимые изменения в архитектуру системы электронного голосования, которые позволили провести интеграцию нового модуля наиболее оптимальным образом. Был реализован и внедрён модуль биометрической аутентификации для системы электронного голосования, а также реализован пользовательский интерфейс в двух приложениях системы, позволяющий регистрировать биометрические данные пользователей в приложении-валидаторе и производить аутентификацию по биометрическим данным в приложении-клиенте.

В результате разработки было внедрено слабосвязанное средство обеспечения безопасности, которое может быть легко заменено модулем, реализующим тот же интерфейс. Модуль обеспечивает надёжную биометрическую аутентификацию, удобную для пользователей и предотвращающей проникновение злоумышленников в систему.

В рамках дальнейшей работы следует исследовать более сложные алгоритмы распознавания лиц, позволяющие учитывать повороты головы, а также предусматривающие анализ трёхмерной модели, снятой при помощи датчиков.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

[1-А.] Григорьев А.В. Обзор биометрических способов аутентификации / А.В. Григорьев // Технические науки: проблемы и решения: сб. ст. по материалам V Международной научно-практической конференции «Технические науки: проблемы и решения». – № 5(4). – М., Изд. «Интернаука», 2017.

[2-А.] Григорьев А.В. Алгоритм распознавания лиц Eigenfaces / А.В. Григорьев // Инновационные подходы в современной науке: сб. ст. по материалам XIII Международной научно-практической конференции

«Инновационные подходы в современной науке». – № 1(13). – М., Изд. «Интернаука», 2018.