МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ

Кафедра дискретной математики и алгоритмики

МЕТОДЫ ПРОТИВОДЕЙСТВИЯ СПУФИНГУ В БИОМЕТРИЧЕСКИХ СИСТЕМАХ ИДЕНТИФИКАЦИИ, ОСНОВАННЫХ НА РАСПОЗНАВАНИИ ЛИЦ

Курсовой проект

Ематинова Кирилла Александровича
студента 3 курса, 3 группы
специальность "Информатика"

Научный руководитель:
Шур Наталья Александровна
ассистент кафедры БМИ

Минск, 2018

ОГЛАВЛЕНИЕ

Введение **3**

1 КЛАССИФИКАЦИЯ АТАК СИСТЕМ БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ, ИСПОЛЬЗУЮЩИХ РАСПОЗНАВАНИЕ ЛИЦ **5**

1.1 Атаки, использующие фотографии 5

1.2 Атаки, использующие видеозаписи 6

1.3 Атаки, использующие 3d-маски 7

2 АНАЛИЗ СУЩЕСТВУЮЩИХ МЕТОДОВ ПРОТИВОДЕЙСТВИЯ СПУФИНГУ **9**

2.1 Методы, основанные на текстурном анализе 10

2.1.1 Local Binary Patterns 11

2.1.2 Haralick Textural Features 14

2.2 Методы, использующие исскуственные нейронные сети 16

2.2.1 Краткое описание структуры искусственной нейронной сети 16

2.3 Методы, основанные на обнаружении биофизических признаков 19

3 ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ **21**

3.1 Оценка алгоритма 21

3.2 Подготовка данных 22

3.3 Разработка и тестирование алгоритмов 23

3.3.1 Стартовая модель: LBP, Haralick, SVM 23

3.3.2 Правило конъюнкции 23

3.3.3 Стэкинг 24

3.4 Выводы 24

4 ЗАКЛЮЧЕНИЕ **26**

Список использованной литературы **27**

ВВЕДЕНИЕ

Методы идентификации личности, основанные на различных биометрических данных, в настоящее время получили очень широкое распространение. Существуют популярные подходы, использующие отпечатки пальцев, радужную оболочку или сетчатку глаз, лицо человека или его голос. Однако с развитием биометрических систем идентификации связано и развитие методов их обмана, или, по-другому, биометрического спуфинга (spoofing) [9]. Спуфинг — это обход системы идентификации путём предоставления биометрическому сенсору сфабрикованной версии оригинальных биометрических данных. Поддельную версию биометрических данных также называют артефактом (artefact). Примерами артефактов являются поддельные фотографии, видеозаписи, различные муляжи, заранее записанные звуки речи и т. д. Целью спуфинга является предоставление незарегистрированного в системе пользователя как зарегистрированного. Противодействие атакам спуфинга является трудоёмкой задачей, так как злоумышленник имеет непосредственный контакт с сенсором, а в связи с этим невозможно использовать методы шифрования и цифрового кодирования.

Системы, основанные на распознавании лиц, в силу некоторых преимуществ, считаются одними из наиболее перспективных систем биометрической идентификации в данный момент. Они не требуют прямого контакта, являются достаточно быстрыми и простыми, а для реализации многих из них не используется специализированная техника. Однако, они так же подвержены атакам с ипользованием биометрического спуфинга (спуфинг-атакам). В качестве артефактов в таких атаках могут применятся распечатанные фотографии лиц объектов, цифровые видеозаписи и фотографии, транслируемые с различных устройств, или 3d-маски.

Целью данной работы является изучение существующих методов противодействия спуфингу (antispoofing) в биометрических системах, основанных на распознавании лиц, а так же определение направлений дальнейших исследований по повышению эффективности методов антиспуфинга.

Задачи исследования:

- Изучение различных типов атак в биометрических системах контроля доступа, основанных на распознавании лиц;
- Изучение и анализ существующих методов противодействия спуфингу

в биометрических системах контроля доступа, использующих распознавание лиц;

- Определение направлений исследований по повышению эффективности методов антиспуфинга.

ГЛАВА 1
КЛАССИФИКАЦИЯ АТАК СИСТЕМ БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ, ИСПОЛЬЗУЮЩИХ РАСПОЗНАВАНИЕ ЛИЦ

Как было сказано выше, методы биометрической идентификации уязвимы для атак, использующих спуфинг. Для биометрических систем, использующих распознавание лиц, условно, спуфинг-атаки можно разделить на три типа (по способу создания артефакта лица пользователя):

- Атаки, использующие фотографии (фото-атаки).
- Атаки, использующие видеозаписи (видео-атаки).
- Атаки, использующие 3d-маски.

1.1 Атаки, использующие фотографии

Атаки такого типа используют фотографии зарегистрированного в системе пользователя, для получения несанкционированного доступа к ней. Фотографии лица достаточно высокого качества могут быть получены злоумышленниками разными способами. Например, они могут быть опубликованы сами пользователем, в интернете (в частности, в популярных социальных сетях) [11], получены с помощью несанкционированного доступа к устройствам пользователя, имеющим фронтальную камеру, или же вручную сделаны злоумышленником с помощью цировых камер. Далее фотография может быть распечатана на бумаге и представлена системе (биометрическому сенсору) в качестве артефакта (Рисунок 1.1). Иногда такие атаки рассматриваются как отдельный тип, в силу того, что именно они были наиболее часто применящимися до распространения мобильных электронных устройств с достаточно крупными экранами высокого разрешения [25]. Также фотография может быть представлена системе с экрана некоторого электронного устройства, например, мобильного телефона, ноутбука или электронного планшета[9]. Более сложным способом фото-атаки является использование «фото-маски». Так называют фотографию, распечатанную на бумаге в высоком разрешении, с вырезанными глазами и ртом. Во время атаки биометрической системы

идентификации злоумышленник надевает маску. Это позволяет сымитировать часть движений лица (например, моргание глаз), на которых основаны некоторые методы противодействия спуфинг-атакам [25].

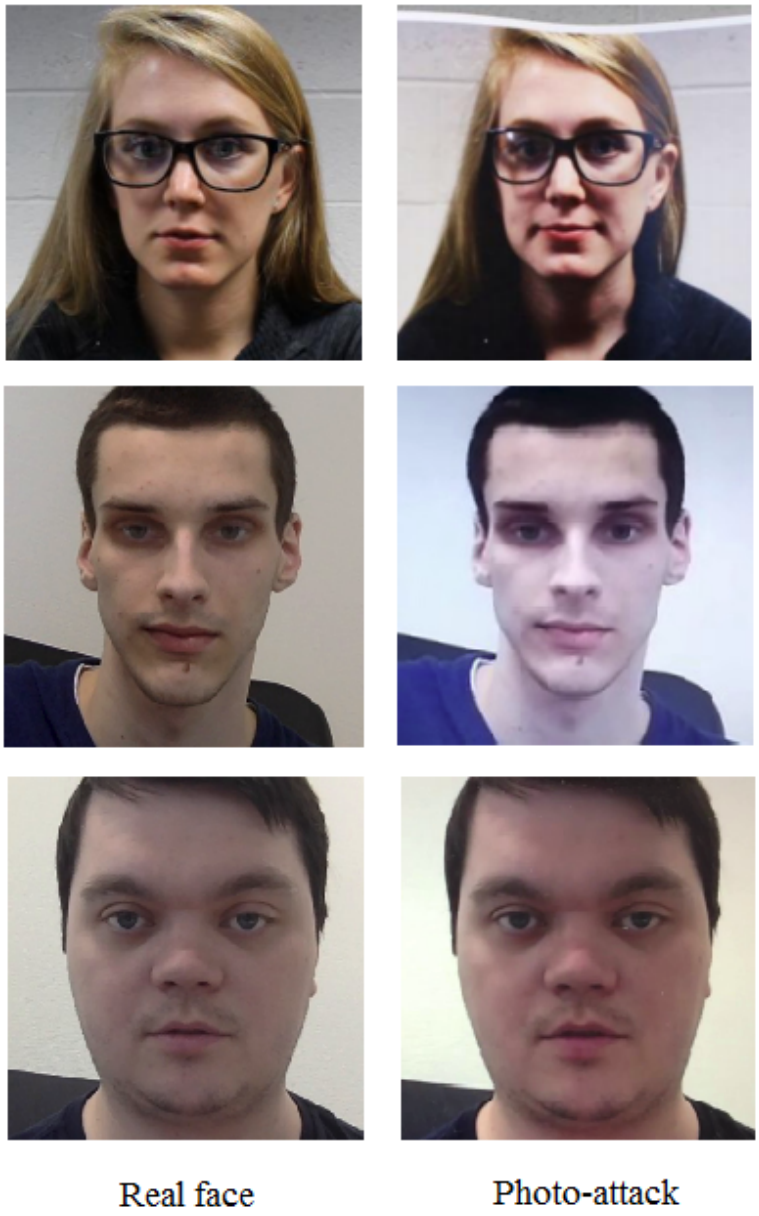

Рисунок 1.1: Пример фото-атак [22]. Слева представлены оригинальные лица пользователей, а справа - артефакты.

1.2 Атаки, использующие видеозаписи

Несколько более продвинутым типом атак являются видео-атаки. При данных атаках в качестве биометрического артефакта используется видеозапись лица пользователя. Пути получения артефакта для спуфинг-атак, использующих видеозаписи, в целом, совпадают с таковыми для фото-атак.

Полученная видеозапись представляется биометрическому сенсору с экрана некоторого электронного устройства. В отличие от статического изображения, используемого при фото-атаке, которое имитирует только 2d-структуру реального лица, при видео-атаке также имитируется значительная часть движений лица пользователя [32].

Рисунок 1.2: Пример кадров видео-атак, проведённых с помощью экранов различных устройств, на которых отображались фотографии пользователя [22].

1.3 Атаки, использующие 3d-маски

В данном случае в качестве артефакта для спуфинг-атаки используется 3d-маска, копирующая реальное лицо пользователя. Как и в случае с «фото-масками» во время атаки биометрической системы идентификации злоумышленник надевает маску. Однако, в данном случае имитируется значительно большая часть движений лица пользователя, а также его полная трёхмерная структура, что значительно усложняет обнаружение спуфинга такого типа. В частности, методы, основанные на использовании датчиков глубины, хорошо справляющиеся с обнаружением фото- и видео-атак, теряют свою эффективность в случае спуфинг-атак с использованием 3d-масок.

Идеи о проведении атак подобного типа существовали достаточно давно [16], однако лишь в последнее время, в связи с развитием технологий 3d-печати, появилась относительно доступная возможность их практической реализации (в частности, относительно доступная возможность создания достаточно качественного артефакта) [19]. Именно по этой причине серьёзное изучение такого вида атак, создание специализированных баз данных и разработка особых способов противодействия начались совсем недавно [10].

Однако, даже при наличии практической возможности создания качественной 3d-маски, данный метод всё ещё является труднореализуемым. Связано это в первую очередь со сложностью получения биометрических данных

пользователя, нужных для создания достаточно качественной трёхмерной модели лица. Зачастую получить такие биометрические данные без согласия пользователя практически невозможно.

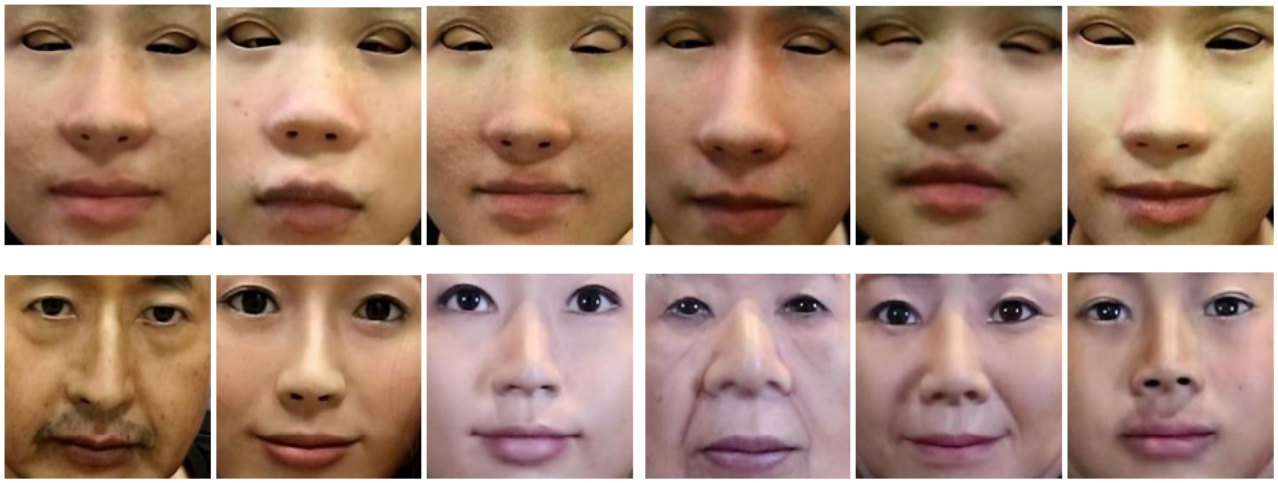

Рисунок 1.3: Примеры различных 3d-масок [21].

На основании изученных данных можно сделать вывод о том, что в настоящее время наиболее вероятным вариантом атаки биометрических систем контроля доступа является применение методов, основанных на фото- и видеозаписях лица пользователя. В связи с этим было принято решение сконцетрироваться именно на этих типах атак.

ГЛАВА 2
АНАЛИЗ СУЩЕСТВУЮЩИХ МЕТОДОВ ПРОТИВОДЕЙСТВИЯ СПУФИНГУ

В общем случае, методы противодействия спуфингу в системах, использующих распознавание лиц можно разделить на два типа (Рисунок 2.4):

1. Методы, требующие дополнительной модификации биометрического сенсора (камеры), или прямого взаимодействия пользователя с ним.

2. Методы, основанные на изучении некоторых характеристик имеющихся биометрических данных (в частности фото- и видеозаписей лица пользователя).

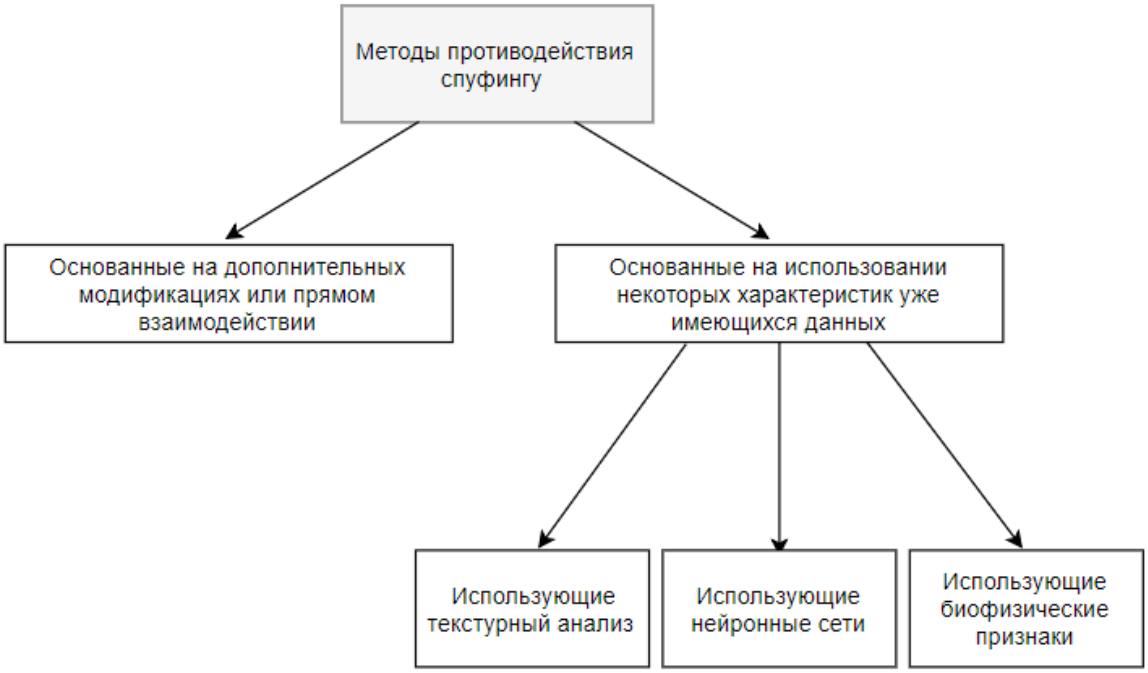

Рисунок 2.4: Классификация основных методов, применяющихся для противодействия спуфинг-атакам.

Способы, относящиеся к первому типу, чаще всего основаны на некоторых признаках, анализируемых с помощью дополнительных устройств.

В работе [17] для подтверждения подлинности распознаваемого объекта используются изображения зрачков пользователя, подвергаемых воздействию света. В частности, защита основана на свойстве человеческого зрачка — сужаться при повышении интенсивности света. Чтобы узнать, получено

ли изображение пользователя с глаза живого человека, варьируют интенсивность освещения. Система защиты от спуфинга отслеживает реакцию зрачка на модуляцию освещения, поэтому время регистрации несколько увеличивается.

В качестве примеров методов, для которых требуется непосредственное взаимодействие с пользователем, можно привести: [5]. Для подтверждения подлинности распознаваемого объекта используются изображения, полученные в обычном состоянии, и в состоянии, когда пользователю требуется выполнить предложенное системой действие (применяется командное воздействие). Набор действий не повторяется, действия могут быть следующими: открыть или закрыть глаза, закрыть один глаз, закрыть или открыть рот и т. д.

Подобные методы имеют ряд существенных минусов, вследствие которых в настоящее время теряют свою актуальность:

- использование дополнительных специализированных устройств (что может существенно повысить стоимость реализации системы).
- физическое или командное воздействие на пользователя.

В силу этих причин, было принято решение не рассматривать эти методы в дальнейшем (в контексте данной работы).

Этих недостатков лишены методы второго типа: признаки, требуемые для проверки, извлекаются напрямую из полученных фото- или видеозаписей лица пользователя. Такие методы, в свою очередь, так же можно разделить на несколько основных групп, каждая из которых будет рассмотрена подробнее ниже.

2.1 Методы, основанные на текстурном анализе

Данные методы основываются на выделении некоторых информативных характеристик, описывающих текстуру заданного изображения. Затем полученные характеристики используются в качестве признаков, по которым некоторая предварительно обученная модель способна определить класс изображения (реальное изображение или спуфинг-изображение). Чаще всего, в качестве классификатора используется, в силу относительной простоты, метод опорных векторов (**support vector machines, SVMs**), однако приме-

няются и более сложные методы классификации [6]. Ниже представлены основные способы, используемые для выделения текстурных признаков.

2.1.1 Local Binary Patterns

Локальные бинарные шаблоны (**Local Binary Patterns, LBP**) [24] — оператор, применяемый для классификации текстур в компьютерном зрении. Каждому пикселю изображения оригинальный оператор LBP ставит в соответствие описание его локальной окрестности в виде двоичного вектора (local binary pattern) следующим образом:

Рассматриваемый пиксель далее будем называть *центральным*. Значение каждого пикселя, смежного с центральным, сравнивается со значением центрального пикселя. Пиксели, которые имеют значения больше, чем центральный (или равное ему), обозначаются «1». Пиксели, которые имеют значения меньше центрального, обозначаются «0». Далее, путём конкатенации полученных значений (по часовой стрелке, начиная с верхнего левого пикселя) создаётся восьмиразрядный двоичный вектор, который описывает окрестность центрального пикселя.

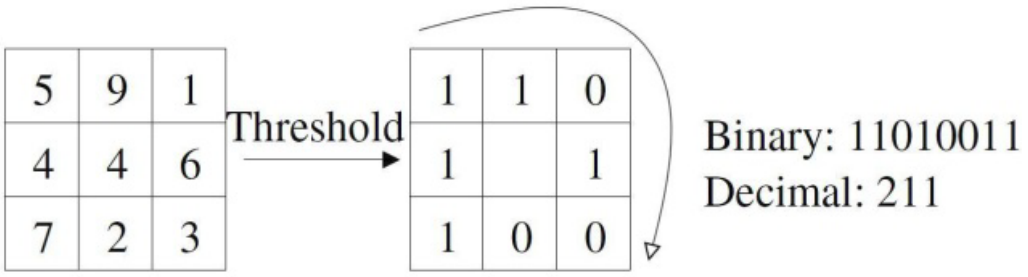

Рисунок 2.5: Пример оригинального оператора LBP [13].

Одним из ограничений оригинального оператора LBP является использование достаточно малых 3x3 окрестностей, которые не позволяют получить информацию о крупномасштабных структурах, присутствующих на изображении. Для устранения данного ограничения оператор LBP был обобщён для работы с окрестностями разных размеров [23]. Локальная окрестность пикселя определяется как набор из P точек, равномерно распределённых на окружности радиуса R с центром в рассматриваемом пикселе. При этом точки, которые не попадают внутрь пикселей, интерполируют с использованием билинейной интерполяции.

Вектор, полученный с помощью оператора LBP (LBP-код), может рассматриваться в качестве двоичного числа. Так же может использоваться его

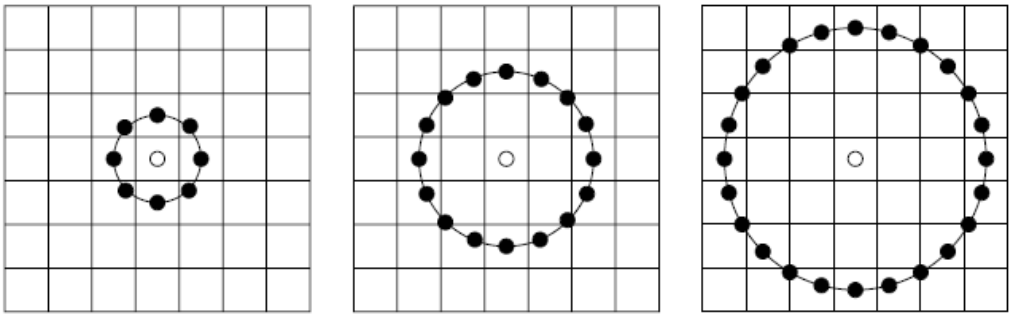

Рисунок 2.6: Пример обобщённого оператора LBP [13] с окрестностями: (8, 1), (16, 2), (24, 3).

представление в десятичной системе исчисления.

Формально, для пикселя (x_c, y_c) результат применения оператора LBP в десятичной системе исчисления может быть получен с использование следующей формулы:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_c - i_p)2^p \tag{1}$$

где i_c and i_p значения центрального пикселя и p-го пикселя из рассматриваемой окрестности соответственно, P - количество пикселей в рассматриваемой окрестности, R – радиус рассматриваемой окрестности, а $s(x)$ функция, определённая как:

$$s(x) = \begin{cases} 1, & \text{if } x >= 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{2}$$

Из формулы (1) следует, что оператор LBP является инвариантным относительно преобразований, сохраняющих порядок. Так же нетрудно видеть, что оператор LBP может принимать 2^P различных значений, соответствующих 2^P различным двоичным векторам, сформированным пикселями рассматриваемой окрестности. Полученные значения используются для построения гистограммы, которая в дальнейшем используется в качестве дескриптора текстуры изображения [13].

Было показано, что некоторые LBP-коды содержат больше информации, чем другие [23]. Таким образом, для получения достаточного количества информации о текстуре изображения достаточно использовать некоторое подмножество всего 2^p-элементного множества LBP-кодов. LBP-коды, входящие

в это подмножество, назвали «униформными» (uniform patterns) или значимыми. LBP-код называют значимым, если он содержит не более двух переходов от 0 к 1 или от 1 к 0. Например, 00000000 (0 переходов) и 01110000 (2 перехода) являются значимыми, а 11001001 (4 перехода) и 01010011 (6 переходов) таковыми не являются. Авторами работы был проведён эксперимент, который показал, что на исследуемых изображениях в среднем 90% полученных LBP-кодов являлись значимыми при использовании окрестности $(8, 1)$ и около 70% при использовании окрестности размера $(16, 2)$. Для построения гистограмм предлагается все LBP-коды, не являющиеся значимыми в один блок, что существенно уменьшает количество блоков гистограммы (признаков).

Первоначально оператор LBP использовался только для изображений, представленных в оттенках серого. Однако впоследствии были предложены различные вариации оператора LBP, эффективно применимые для цветных изображений [31].

В [3] был предложен эффективный метод противодействия спуфингу, использующий оператор LBP, а также цветовые модели HSV и $YCbCr$ (Рисунок 2.7).

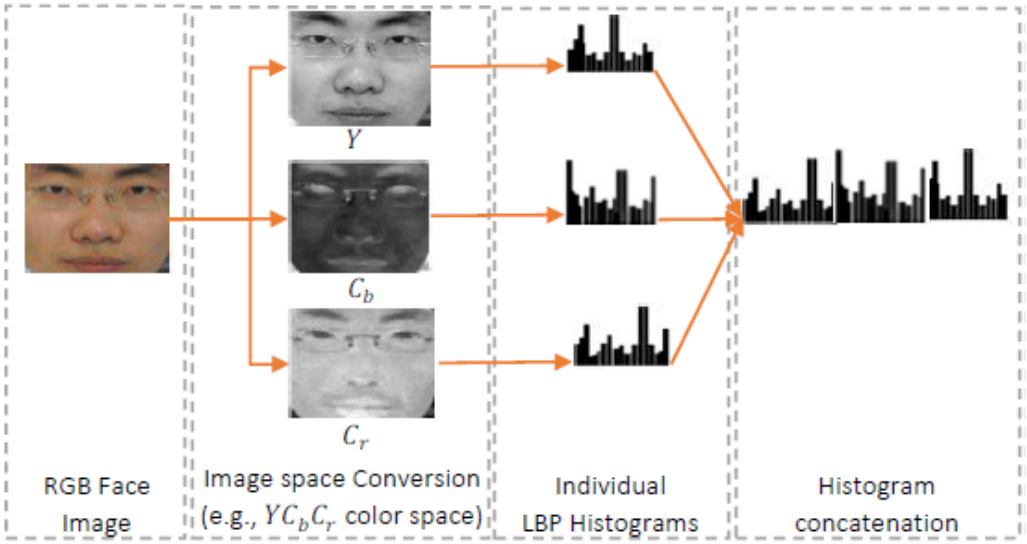

Рисунок 2.7: Архитектура предложенного в [3] метода.

Обе цветовые модели основаны на разделении цветовых и яркостных составляющих. В HSV компоненты H (Hue) и S (Saturation) определяют цветовую составляющую изображения, а компонента V (Value) — яркостную. В $YCbCr$ Y — компонента яркости, Cb и Cr являются синей и красной цветоразностными компонентами. Две копии сходного RGB изображения независимо

преобразуются в HSV и в $YCbCr$ соответственно. LBP оператор применяется к каждой компоненте обоих цветовых моделей, формируются отдельные гистограммы. Полученные гистограммы соединяются, и формируется конечный вектор признаков изображения. Таким образом признаки выделяются для некоторой выборки, состоящей из фотографий реальных лиц и спуфинг-изображений. Затем полученную матрицу объектов–признаков используют для обучения выбранного классификатор. В [3] используется метод опорных векторов.

2.1.2 Haralick Textural Features

Признаки Харалика (**Haralick Textural Features**) — набор признаков, описывающих текстуру, часто применяемый для классификации изображений. Предложен Robert M. Haralick в 1973 [12].

Описанный метод выделения признаков выглядит следующим образом: Для исходного изображения, представленного в оттенках серого, строится четыре различных матрицы совместного появления (**Gray Level Co-occurrence matrix, GLCM**). Матрица совместного появления (Рисунок 2.8) является квадратной и имеет размеры N_g, где N_g — это число возможных оттенков серого в представлении изображения (чаще всего $N_g = 2^8 = 256$). Элемент матрицы $P(i, j)$ есть колличество раз, когда в оригинальной матрице изображения пиксель со значением i смежен с пикселем со значением j. Далее матрицы принято нормализовать. Таким образом, каждое значение $p(i, j)$ в полученной нормализованной матрице является вероятностью того, что пиксель со значением i будет смежным с пикселем, имеющим значение j.

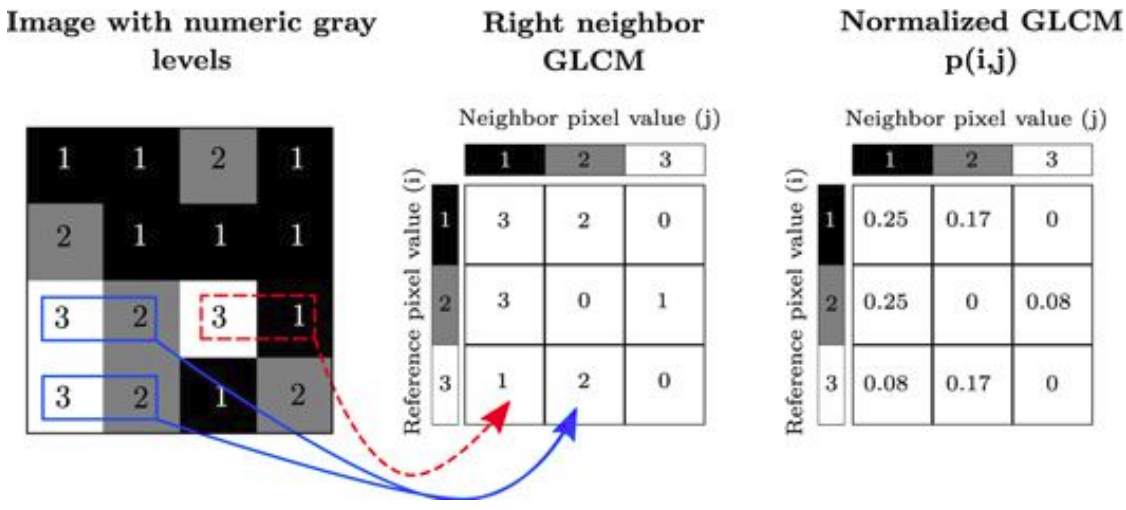

Рисунок 2.8: Пример построения нормализованной GLCM матрицы [4].

При этом существует несколько разлиных типов смежности, для каждого из которых строится отдельная нормализованная GLCM матрица:

- смежность с правым элементом
- смежность с нижним элементом
- смежность с правым нижним элементом
- смежность с левым нижним элементом

Следующим шагом является вычисление тринадцати различных статистик (Таблица 2.1) для каждой из полученных матриц. Полученные значения и называют признаками Харалика. Также принято усреднять каждый из полученных признаков по всем четырём GLCM матрицам с разными типами смежности.

f	Название	Формула
1	Angular Second Moment	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(\frac{P(i,j)}{R}\right)^2 = \sum_i\sum_j p(i,j)^2$
2	Contrast	$\sum_{k=0}^{N_g-1} k^2\left\{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\delta_{\lvert i-j\rvert,k}p(i,j)\right\} = \sum_{k=0}^{N_g-1} k^2 p_{x-y}(k)$
3	Correlation	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(ij)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$
4	Sum of Squares: Variance	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu)^2 p(i,j)$
5	Inverse Difference Moment	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{1}{1+(i-j)^2}p(i,j)$
6	Sum Average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$
7	Sum Variance	$\sum_{i=2}^{2N_g}(i-f_8)^2 p_{x+y}(i)$
8	Sum Entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i)\log(p_{x+y}(i))$
9	Entropy	$-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log(p(i,j))$
10	Difference Variance	variance of p_{x-y}
11	Difference Entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i)\log(p_{x-y}(i))$
12	Measure of Correlation	$\frac{f_9-HXY1}{\max(HX,HY)}$
13	Measure of Correlation	$[1-\exp(-2(HXY2-f_9))]^{1/2}$

Таблица 2.1: Признаки Харалика.

Пример использования признаков Харалика в контексте противодействия спуфингу предложен в [1]. Описываемый подход предлагается использовать в системах, обеспечивающих непрерывный видеопоток. Полученные видеозаписи лица разбиваются на кадры, затем для каждого кадра подсчитываются признаки Харалика. Далее полученные для каждого кадра признаки объединяются в один вектор признаков, описывающий исходное видео. В силу

большого количества используемых кадров, размерность вектора также получается достаточно большой. Для уменьшения размерности полученного вектора в работе применяется метод главных компонент (**principal component analysis, PCA**) [14]. Кроме того, авторами было показано, что вычисление признаков Харалика для каждого канала RGB изображения значительно увеличивает эффективность метода. Способ, описанный выше, применяется для выделения признаков на некоторой выборке, состоящей из видеозаписей реальных лиц и спуфинг-видеозаписей . Затем полученную матрицу объектов–признаков используют для обучения выбранного классификатор. В рассматриваемой работе так же применяется метод опорных векторов.

2.2 Методы, использующие исскуственные нейронные сети

В настоящее время широкое распространение получили методы, использующие искусственные нейронные сети (далее *нейронные сети*). Они так же как и методы, основанные на текстурном анализе, используются для выделения некоторых информативных признаков изображения. Однако, зачастую, при использовании нейронных сетей появляется возможность выделять более сложные зависимости в данных. Существуют подходы, при которых нейронная сеть выступает в качестве автокодировщика, применяемого, например, к картам LBP [29]. Однако куда более распространено применение *свёрточных нейронных сетей* напрямую к изображению.

2.2.1 Краткое описание структуры искусственной нейронной сети

Основу нейронной сети составляют минимальные вычислительные единицы, которые называют нейронами. В общем случае нейронная сеть состоит из слоя входных нейронов, нескольких слоёв скрытых нейронов и слоя выходных нейронов. «Входными» и «выходными» называют нейроны, отвечающие за получение входного сигнала и выдачу выходного сигнала (результатов обработки соответственно). «Скрытыми» называют нейроны, принимающие информацию от других нейронов в сети, и, преобразуя её некоторым образом, передают следующим нейронам. Связи между нейронами называют синапсами.

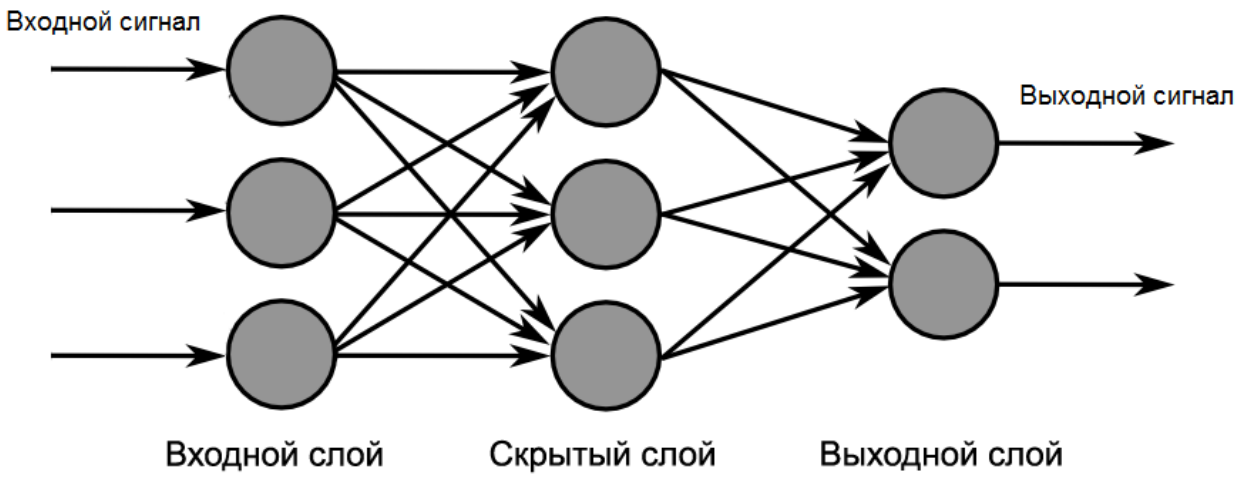

Рисунок 2.9: Пример архитектуры нейронной сети с одним скрытым слоем.

Выделяют понятие полносвязных нейронных сетей прямого распространения (**fully connected feed-forward neural networks**). В такой реализации каждый нейрон некоторого слоя имеет связь с каждым нейроном предыдущего слоя. Такие сети имеют несколько существенных недостатков:

- Большое количество обучаемых параметров.
- Линейность преобразований.

Первую проблему частично решают свёрточные нейронные сети (**convolutional neural network, CNN**), в частности слои, которые применяют к входным данным операцию «свёртки» (свёрточные слои), а так же слои, которые реализуют операцию подвыборки (**subsampling** or **pooling**). С помощью метода скользящего окна (**sliding window**) операция свёртки последовательно применяется к областям входных данных некоторых размеров (Рисунок 2.10). При каждой отдельной свёртке происходит перемножения входных данных со значениями некоторой матрицы, общей для текущего слоя, называемой ядро или фильтром. Результат свёртки часто называют картой признаков.

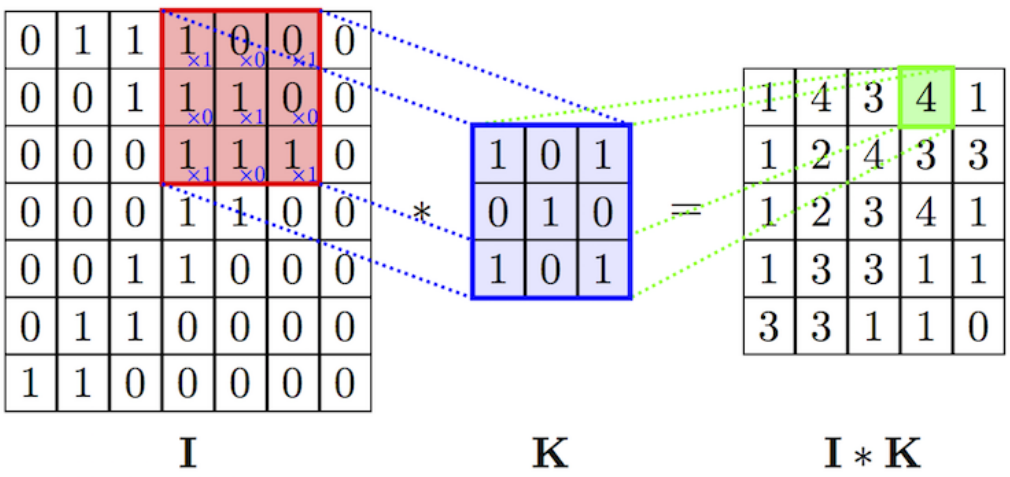

Рисунок 2.10: Иллюстрация операции свёртки. I — входные данные (выходная карта свойств предыдущего слоя), K — ядро свёртки, $I * K$ — результат операции свёртки (выходная карта свойств).

Слой подвыборки так же применяется с помощью метода скользящего к областям входных данных (карты признаков), заданным образом уменьшая их размер, тем самым уменьшая размер исходной карты и при этом сохраняя информацию в некоторой (достаточной) степени. Для этого используются различные (Рисунок 2.11) методы (выбор максимума из области, подсчёт среднего).

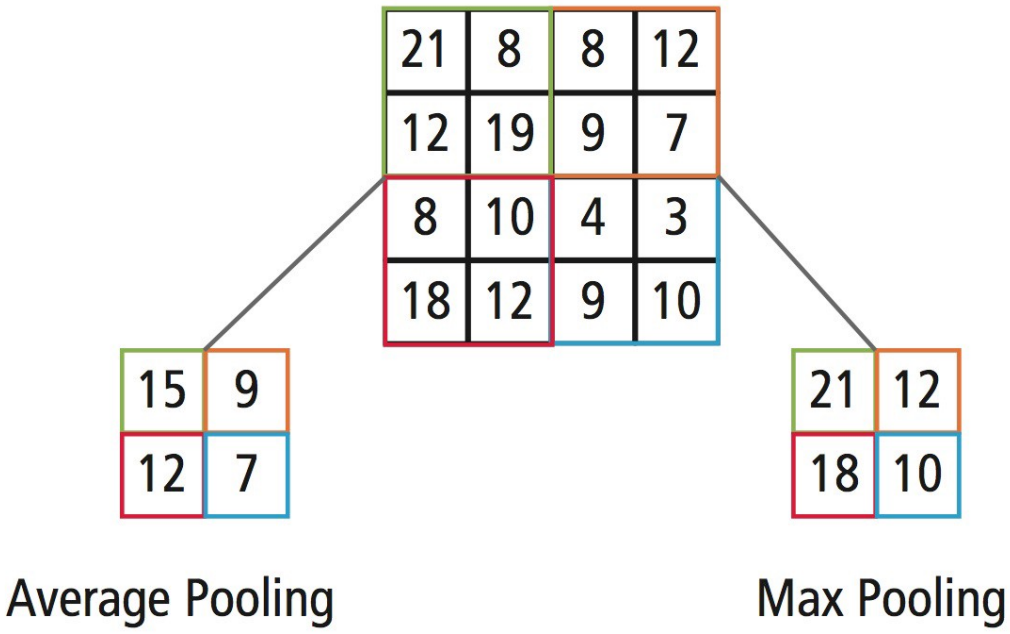

Рисунок 2.11: Популярные способы реализации операции «подвыборки»: подсчёт среднего значения и выбор максимума.

Для получения нелинейных преобразований данных в сети используют

различные слои активации. Такие слои содержат в себе некоторые нелинейные функции, которые применяются к выходным сигналам предыдущего слоя. Наиболее часто применяемой в настоящее время является функция ReLU.

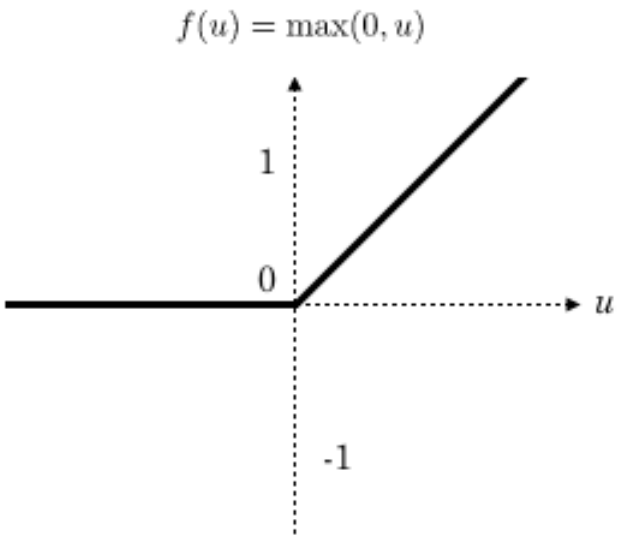

Рисунок 2.12: Функция активации ReLU: $f(u) = max(0, u)$.

Авторы [20] используют в своей работе ансамбль CNN, каждая из которых обучается распознавать спуфинг различных типов. В частности ансамбль двух свёрточных сетей, одна из которых классифицирует атаки с использованием распечатанных на бумаге фото, а вторая — атаки, с использованием видеозаписи.

Достаточно интересный подход предлагается в [22]. Нейронные сети используются для воссоздания карты глубины изображения лица, а затем предсказания сердечного ритма пользователя по 2d-видео файлу. На основании этих признаков делаются выводы о том, является ли видео спуфинг-атакой.

2.3 Методы, основанные на обнаружении биофизических признаков

К биофизическим признакам относят: движение глаз, моргание, мимику, движение головы относительно фона, кровеносные сосуды, термограмму. Большинство методов, основанных на обнаружение определённых биофизических признаков объекта, может применяться только в биометрических системах, обеспечивающих непрерывный видеопоток. Следует уточнить, что

методы, описанные в данном разделе, используют только биофизические признаки, для обнаружения которых не требуются дополнительные сенсоры. В частности, используются только те признаки, которые можно выделить по видеозаписи, содержащей лицо пользователя.

Ранее был исследован ряд способов, основанных на скорости моргания [25], анализе движения глаз [18], рта или головы относительно фона [8]. Однако в [9] была продемонстрирована возможность успешного взлома систем, использующих способы противодействия спуфингу такого типа. В частности, была обнаружена неустойчивость подобных методов к атакам, основанным на демонстрации фото и видео фрагментов с экранов различных устройств.

В [2] была представлена группа методов, основанных на анализе термограммы или рисунка поверхностных кровеносных сосудов распознаваемого объекта. Преимущество данных способов заключается в их относительной нечувствительны к изменению освещения. Подтверждение подлинности основано на том, что реальные объекты, в отличие от искусственных, излучают различный диапазон инфракрасного излучения или имеют поверхностные кровеносные сосуды.

ГЛАВА 3
ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

На основании данных, представленных во второй главе, были выбраны наиболее перспективные для дальнейшего исследования методы: методы, основанные на текстурном анализе и методы, использующие нейронные сети. Далее была проведена серия практических экспериментов, с использованием некоторых методов, описанных в главе 2.

3.1 Оценка алгоритма

Для оценки алгоритмов в большинстве работ на данную тематику используется метрика **Equal Error Rate (ERR)**. EER называют точку пересечения **False Acceptance Rate (FAR)** и **False Rejection Rate (FRR)** прямых. Оценка многих алгоритмов, представленных в главе 2, была проведена с использованием данной метрики.

Approach	Replay-attack-dev(EER)%	Replay-attack-test(HTER)%	CASIA-FASD-test-EER (%)
IQA based (Galbally and Marcel, 2014)	---	---	32.40
Motion (Pereira et al., 2013)	11.60	11.70	26.60
LBP + SVM (Yang et al., 2013)	8.55	11.75	18.50
LBP-TOP + SVM (Pereira et al., 2012)	7.88	7.60	10.00
SBIQF+NN (Feng et al., 2016)	3.83	6.13	15.5
YCbCr + HSV + LBP (Boulkenafet et al., 2015)	0.4	2.9	6.2
CNN+LSTM (Xu et al., 2016)	---	---	5.17

Рисунок 3.13: Оценка некоторых методов противодействия спуфингу [20]. Использованные базы данных: REPLAY-ATTACK [6], CASIA-FASD [32].

3.2 Подготовка данных

Для оценки собственных реализаций алгоритмов было решено использовать датасет REPLAY-ATTACK [6], который содержит более 1000 коротких видеозаписей 50-ти различных людей. Атаки в датасете разделены на три типа: фото-атаки с использованием распечатанной фотографии, видео и фото-атаки с использованием экрана электронного устройства. Видео сняты в различных условиях освещения, с использованием камер различного качества. REPLAY-ATTACK используется во множестве работ, посвящённых изучению методов противодействия спуфинга. В силу этого, он очень удобен для сравнения различных алгоритмов. Однако, датасет был создан в 2012 году, поэтому многие видеозаписи являются недостаточно качественными.

Рисунок 3.14: Пример исходного кадра видеозаписи из REPLAY-ATTACK [6].

Данные были предобработаны следующим образом: из каждой видеозаписи были извлечены кадры (2 кадра на каждую секунду видеозаписи). На каждом полученном кадре с помощью метода Виолы-Джонса [26] была произведена детекция лица пользователя. Далее лица пользователей на полученных кадрах были выровнены [27]. Всего было получено порядка 6000 и 5000 кадров для тренировочного и тестового датасетов соответственно.

Рисунок 3.15: Пример полученного после предобработки кадра видеозаписи из REPLAY-ATTACK [6].

3.3 Разработка и тестирование алгоритмов

В качестве экспериментального метода для извлечения признаков из изображения было решено опробовать комбинацию LBP и Haralick texture features. Каждое изображение тренировочной и тестовой выборки было преобразовано из RGB в HSV и YCbCr. Для каждого канала полученных изображений были извлечены LBP-гистограммы и признаки Харалика, из которых затем был сформирован вектор признаков. Таким образом, для каждого изображения был сформирован вектор из 139 различных признаков.

3.3.1 Стартовая модель: LBP, Haralick, SVM

Далее, как и в «Face anti-spoofing based on color texture analysis» [3], полученные данные были использованы для обучения классификатора на основе метода опорных векторов. Для нахождения оптимальных гиперпараметров была проведена кросс-валидация. **ERR** полученной модели составила 0.47% процента, тогда как в оригинальной работе — 0.40. Небольшая потеря точности может быть связана с предобработкой данных, которая в [3] проводилась другим способом, отличным от применённого в данной работе.

3.3.2 Правило конъюнкции

Было решено испробовать упрощённую эвристику объединения двух моделей. Способом, описанным выше, на различных данных были обучены две одинаковые модели: первая модель обучалась на датасете, состоящем из ре-

альных лиц и фото-атак, а вторая на датасете, состоящем из реальных лиц и видео-атак. Далее каждой из моделей делалось предсказание для всего тестового датасета. Если хотя бы одна модель относила изображение к классу «спуфинг», то изображение получало класс «спуфинг». Суть идеи заключалась в снижении внутриклассовых различий для спуфинг-изображений, в соотвествии с этим были выделены два подкласса, имеющие достаточно большие различия. Однако, результат **ERR** полученной модели ухудшился до 1.46%. Можно предположить, что это связано с сильным повышением вероятности ложного отказа в доступе (False Rejection Rate).

3.3.3 Стэкинг

Стэкинг [28] является несколько более продвинутым методом объединения нескольких различных моделей. Была применена эвристика разделения фото- и видео-атак, описанная выше. Далее каждый из тренировочных датасетов, содержащий реальные лица и артефакты одно из типов, был поделен на выборку для обучения моделей нулевого уровня и выборку для обучения модели первого уровня в соотношении 3 к 1. Моделями нулевого уровня являлись классификаторы на основе метода опорных векторов. Каждый классификатор обучался на сообственной выборке для моделей нулевого уровня, предсказывал вероятность принадлежности каждого объекта тестовой выборки к классу «спуфинг-атаки», а затем также предсказывал вероятность принадлежности каждого объекта из обоих выборок для обучения модели первого уровня к классу «спуфинг-атаки». Таким образом, для объектов выборки, сформированной для обучения модели первого уровня, и объектов тестовой выборки были созданы два мета-признака (вероятности, предсказаные классификаторами нулевого уровня). На выборке, сформированной для обучения модели первого уровня, был обучен стандартный линейный классификатор. Который затем был использован для классификации объектов тестовой выборки по их мета-признакам. **ERR** полученной двухуровневой модели составила 0.3%, что стало некоторым улучшение результатов, полученных выше, а также улучшение результата работы [3].

3.4 Выводы

Использованная комбинация классических методов текстурного анализа (LBP + Haralick features) показывает достаточно хорошие результаты, но не

даёт существенного улучшения качества классификации даже в сравнении с каждым из методов в отдельности.

Однако, эвристика минимизации внутриклассовых различий, путем выделения «разных» типов спуфинг-атак в отдельные классы, показывает достаточно перспективные результаты сама по себе и требует дополнительных исследований.

Реализацию описанных выше методов можно найти в GitHub репозитории [15] (`https://github.com/Ematinovkey/course_work`).

ГЛАВА 4
ЗАКЛЮЧЕНИЕ

В процессе выполнения данного курсового проекта были получены следующие результаты:

- Проанализированы основные принципы работы биометрических систем контроля доступа в целом и, в частности, систем идентификации, основанных на распознавании лиц.
- Изучены различные типы спуфинг-атак, а также методы противодействия, применяемые к ним.
- На основе исследованных методов было реализовано несколько простых и в то же время достаточно точных алгоритмов.

Полученные теоретические знания и практические навыки позволили определить направления дальнейших исследований по повышению эффективности методов антиспуфинга.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

[1] Akshay Agarwal, Richa Singh, and Mayank Vatsa. Face anti-spoofing using haralick features. pages 1–6, 09 2016.

[2] Faris Baker. Thermal face recognition using moments invariants. *Journal of Signal Processing Systems*, 3:94–99., 12 2015.

[3] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face anti-spoofing based on color texture analysis. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 2636–2640, Sept 2015.

[4] Patrik Brynolfsson, David Nilsson, Turid Torheim, Thomas Asklund, Camilla Thellenberg-Karlsson, Johan Trygg, Tufve Nyholm, and Anders Garpebring. Haralick texture features from apparent diffusion coefficient (adc) mri images depend on imaging and pre-processing parameters. *Scientific Reports*, 7:4041, 06 2017.

[5] Shaxun Chen, Amit Pande, and Prasant Mohapatra. Sensor-assisted facial recognition: An enhanced biometric authentication system for smartphones. *MobiSys 2014 - Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, 06 2014.

[6] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. September 2012.

[7] Corinna Cortes and VN Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 01 1995.

[8] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel. Can face anti-spoofing countermeasures work in a real world scenario? In *2013 International Conference on Biometrics (ICB)*, pages 1–8, June 2013.

[9] N.M. Duc and B.Q. Minh. Your face is not your password. *Black Hat Conference*, pages 1–16, 01 2009.

[10] Nesli Erdogmus and Sébastien Marcel. Spoofing face recognition with 3d masks. *Information Forensics and Security, IEEE Transactions on*, 9:1084–1097, 07 2014.

[11] J. Galbally, S. Marcel, and J. Fierrez. Biometric antispoofing methods: A survey in face recognition. *IEEE Access*, 2:1530–1552, 2014.

[12] Robert Haralick, K Shanmugam, and Ih Dinstein. Textural features for image classification. *IEEE Trans Syst Man Cybern*, SMC-3:610–621, 01 1973.

[13] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen. Local binary patterns and its application to facial image analysis: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(6):765–781, Nov 2011.

[14] Ian Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374:20150202, 04 2016.

[15] Ematinov K.A. Facial antispoofing course work repository.

[16] Youngshin Kim, Jaekeun Na, Seongbeak Yoon, and Juneho Yi. Masked fake face detection using radiance measurements. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 26 4:760–6, 2009.

[17] Mehmet Kıllıoğlu, Murat Taskiran, and Nihan Kahraman. Anti-spoofing in face recognition with liveness detection using pupil tracking. 01 2017.

[18] Klaus Kollreider, Hartwig Fronthaler, Maycel Isaac Faraj, and Josef Bigün. Real-time face detection and motion analysis with application in "liveness" assessment. *IEEE Transactions on Information Forensics and Security*, 2:548–558, 2007.

[19] Neslihan Kose and Jean-Luc Dugelay. On the vulnerability of face recognition systems to spoofing mask attacks. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2357–2361, 2013.

[20] H. Li, P. He, S. Wang, A. Rocha, X. Jiang, and A. C. Kot. Learning generalized deep feature representation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 13(10), Oct 2018.

[21] S. Liu, B. Yang, P. C. Yuen, and G. Zhao. A 3d mask face anti-spoofing database with real world variations. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1551–1557, June 2016.

[22] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *CVPR*, 2018.

[23] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, July 2002.

[24] T. Ojala, Matti Pietikäinen, and D. Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. volume 1, pages 582 – 585 vol.1, 11 1994.

[25] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, Oct 2007.

[26] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I, Dec 2001.

[27] Yichen Wei. Face alignment by explicit shape regression. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '12, pages 2887–2894, Washington, DC, USA, 2012. IEEE Computer Society.

[28] David Wolpert. Stacked generalization. *Neural Networks*, 5:241–259, 12 1992.

[29] Z. Xu, S. Li, and W. Deng. Learning temporal features using lstm-cnn architecture for face anti-spoofing. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 141–145, Nov 2015.

[30] Dong Yi, Zhen Lei, Zhiwei Zhang, and Stan Li. *Face Anti-spoofing: Multi-spectral Approach*, pages 83–102. 07 2014.

[31] Jae Young Choi, Konstantinos Plataniotis, and Yong Man Ro. Using colour local binary pattern features for face recognition. pages 4541–4544, 09 2010.

[32] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. A face antispoofing database with diverse attacks. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 26–31, March 2012.