Обнаружение атак на биометрическое предъявление на системы аутентификации лиц при помощи специальных устройств съёмки

А.Ю. Денисова [1,2], В.А. Федосеев [1,2]
[1] Самарский национальный исследовательский университет имени академика С.П. Королёва, 443086, Россия, г. Самара, Московское шоссе, д. 34;
[2] ИСОИ РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, 443001, Россия, г. Самара, ул. Молодогвардейская, д. 151

Аннотация

В статье предлагается система признаков, предназначенная для обнаружения атак на биометрическое предъявление на системы аутентификации, использующие лицевую биометрию. При таком типе атаки злоумышленник маскируется под авторизованного пользователя, используя его изображение. Предложенная система признаков предполагает возможность использования одного или нескольких изображающих сенсоров в дополнение к базовой RGB-камере (тепловизоры, дальномеры, инфракрасные камеры). Использование предложенной системы признаков в сочетании с одной из классических моделей бинарной классификации составляет предлагаемый в работе метод обнаружения атак на биометрическое предъявление. Данный метод продемонстрировал низкий уровень ошибок на наборе данных WMCA, при этом эксперименты показали его способность оставаться эффективным в условиях нехватки обучающих данных. Проведённые сравнительные эксперименты показали, что предложенный метод превзошёл алгоритм RDWT-Haralick-SVM и приблизился к результатам алгоритма MC-CNN, основанного на глубоком обучении и требующего значительно больший объём обучающих данных.

Ключевые слова: атака на биометрическое предъявление, спуфинг, аутентификация, распознавание лиц, биометрия, тепловизионные данные, данные глубины.

Цитирование: **Денисова, А.Ю.** Обнаружение атак на биометрическое предъявление на системы аутентификации лиц при помощи специальных устройств съёмки / А.Ю. Денисова, В.А. Федосеев // Компьютерная оптика. – 2022. – Т. 46, № 4. – С. 612-620. – DOI: 10.18287/2412-6179-CO-1054.

Citation: Denisova AY, Fedoseev VA. Detection of presentation attacks on facial authentication systems using special devices. Computer Optics 2022; 46(4): 612-620. DOI: 10.18287/2412-6179-CO-1054.

1. Введение

1.1. Атаки на биометрическое предъявление и методы их обнаружения

С точки зрения процедуры сбора данных технология биометрии лица является самой простой среди других биометрических технологий [1, 2]. По этой причине она широко используется во многих приложениях обеспечения безопасного доступа к данным, а рынок систем биометрии лица постоянно растет [3]. Однако недавние исследования показали, что биометрия лица уязвима для так называемых атак на биометрическое предъявление (presentation attacks), ранее также называемыми атаками спуфинга в биометрии. Данный тип атаки заключается в подделке входных данных с целью их восприятия системой как биометрических данных легальных пользователей. В простейшем случае злоумышленник демонстрирует фотографию легального пользователя при доступе к системе [4]. Очевидно, что в отличие от таких биометрических данных, как отпечатки пальцев и узоры вен, качественные снимки лиц конкретных людей зачастую нетрудно найти благодаря социальным сетям и интернету.

За последние двадцать лет разработано множество методов защиты от атак на биометрическое предъявление. Они используют различные подходы, среди которых можно выделить главные [5 – 7]:

– использование текстурных признаков изображения;
– взаимодействие человека с машиной;
– использование биологических характеристик живых людей;
– использование параметров качества изображения;
– глубокое обучение;
– использование информации с дополнительных устройств.

Подход с использованием текстурных признаков изображения предполагает, что микротекстурные свойства поддельного изображения отличаются от реального изображения лица. Это означает, что злоумышленник демонстрирует изображение клиента, снятое в других условиях или воспроизведенное с более низким качеством, чем реальное изображение лица. Данный подход основан на обработке RGB-изображений, снятых в оптическом диапазоне, и достаточно распространён при защите от атак на биометрическое предъявление. Однако успех конкретного метода сильно зависит от качества устройства, при

помощи которого воспроизводится поддельное изображение. При высоком качестве подделки анализа текстурных свойств недостаточно.

Методы человеко-машинного взаимодействия регистрируют обратную связь от клиента или другие поведенческие особенности, соответствующие живому человеку, такие как моргание глаз [2, 8]. Например, система может регистрировать речь и анализировать соответствие ожидаемого движения губ зарегистрированному видео [8]. Главный недостаток таких систем – увеличивающееся время регистрации изображений и более сложная логика аутентификации.

Подход, основанный на использовании биологических характеристик живых людей, использует такие характеристики, как сердцебиение, кровоток и микродвижение лицевых мышц, чтобы различать настоящие и поддельные изображения лица. Эти методы включают в себя контактные методы с использованием дополнительных устройств для измерения биологических характеристик в реальном времени и бесконтактные методы, в которых используются камеры высокого разрешения и соответствующие условия освещения для получения эталонных оценок биологических характеристик по изображениям [9,10]. Главный недостаток таких методов – жесткие требования к процессу регистрации изображений.

В методах, основанных на измерении качества изображений, подделка рассматривается как метод искажения исходного подлинного изображения. Эффективность таких методов зависит от способа подделки, поскольку в этих методах учитывается различие свойств отражения различных материалов для подделки, таких как бумага, видеоэкран, латексные маски и т.д. Исследователями предложено несколько методов данного класса, включающих анализ искажений на изображениях [11] и анализ качества оборудования для съёмки [12]. Однако данные методы по-прежнему подвержены атаке с использованием высококачественных устройств воспроизведения изображений.

Подход на основе глубокого обучения является наиболее актуальным и активно развивающимся в последние годы. Он основан на использовании глубоких свёрточных нейронных сетей для извлечения признаков и обнаружения факта атаки. Основным и едва ли не единственным недостатком данного подхода является огромный объём размеченных данных, необходимых для обучения модели, используемой для принятия решений. Для каждого конкретного устройства и условий съемки изображения требуются тысячи изображений.

Наконец, заключительный из выделенного выше списка подход основан на использовании видеоданных с дополнительных устройств съёмки, тогда как ранее рассмотренные подходы в основном рассчитаны на использование только одной RGB-камеры, осуществляющей съёмку в оптическом диапазоне. Исключение составляет контактная информация, которая измеряется в реальном времени некоторыми медицинскими устройствами для измерения пульса или сердцебиения, но эти показатели не являются видеоданными. Наиболее часто в качестве дополнительных устройств съёмки используются инфракрасные, тепловизионные камеры и дальномеры. Изображения в ближнем и коротковолновом инфракрасном диапазонах позволяют отличать кожу от материалов, используемых в масках, применяемых для подделки изображения лица, из-за различных отражательных свойств материалов [13]. Тепловизионные изображения позволяют выявить возможное несоответствие температуры лица естественным показателям [14]. Устройства определения глубины сцены характеризуют трёхмерные свойства сцены и, следовательно, определяют трехмерные свойства лица [15]. Существующие методы обнаружения атак на биометрическое представление с использованием дополнительных устройств обычно либо оценивают корреляционные свойства между RGB-изображением и изображением, полученным с помощью дополнительного сенсора [16, 17], либо используют для совокупности изображений один из подходов, перечисленных выше в контексте анализа одного источника данных (например, использование текстурных признаков [18] или моделей глубокого обучения [19, 20]). Среди всех этих методов есть лишь единичные примеры учёта природы дополнительного видеоизображения при разработке метода обнаружения атаки [21, 22].

1.2. Постановка задачи

В настоящей статье рассматривается задача обнаружения атак на биометрическое предъявление в условиях следующих ограничений:

1) мы можем использовать дополнительные съёмки для повышения качества решения задачи;
2) разрешение дополнительного датчика может быть низким относительно RGB-камеры;
3) в наличии имеется лишь ограниченное количество обучающих образов (до нескольких сотен).

Эти ограничения воспроизводят практическую ситуацию, в которой для недопущения презентационных атак на некоторую систему делается попытка внедрить в подсистему аутентификации дополнительную стадию проверки, использующую недорогие датчики. В этой ситуации обучающие данные необходимо формировать самому ответственному за внедрение подобного решения. Очевидно, что у него, скорее всего, нет времени и возможностей для подготовки десятков тысяч обучающих образцов.

Для решения задачи, поставленной таким образом, мы предлагаем набор простых признаков, основанных на характере сигнала и содержании сцены. Предложенная система признаков в совокупности с одним из классификаторов (линейный SVM, Random Forest или SVM-RBF) составляют предлагаемый в статье

метод обнаружения атак на биометрическое предъявление. В данной статье предполагается, что дополнительные устройства съёмки могут включать сенсоры, регистрирующие сцену в ближнем инфракрасном диапазоне или тепловом диапазоне, а также дальномеры. Однако если некоторые из этих источников недоступны, это не должно влиять на работоспособность метода, а лишь должно сокращать используемый набор признаков, что может, в свою очередь, приводить к снижению качества обнаружения атаки.

Также в завершение вводной части отметим, что использование сенсоров трёх выбранных выше типов для решения задачи обнаружения атаки на биометрическое предъявление рассматривалось в недавней работе [23]. В этой статье был предложен отобранный авторами набор обучающих данных, а также рассмотрены два решения: MC-CNN на основе глубокого обучения и RDWT-Haralick-SVM на основе текстурных признаков Харалика. В ходе наших исследований мы использовали тот же набор данных, а также тот же порядок проведения основных экспериментов, чтобы сравнить результаты нашего решения с результатами методов из работы [23].

2. Описание предлагаемого метода

2.1. Предлагаемая система признаков

Обозначим за $X, Y_1, Y_2, ..., Y_N$ исходные изображения области лица, где X – изображение яркости в оттенках серого, полученное по изображению, зарегистрированному RGB-камерой, а $Y_1, Y_2, ..., Y_N$ – одноканальные изображения, полученные с помощью дополнительных сенсоров, $n = 1, ..., N$ (далее для краткости будем называть их *сенсорными изображениями*). Будем предполагать, что изображения были сняты одновременно и для них была произведена геометрическая калибровка таким образом, что все изображения $X, Y_1, Y_2, ..., Y_N$ имеют одинаковый размер $I \times J$ пикселей и одинаковое расположение лица в кадре. Также будем предполагать, что все изображения принимают значения яркости от 0 до 255.

Извлечение признаков производится для каждой пары изображений X, Y_n, в результате чего формируется вектор признаков $f_n \in R^M$. Результирующий вектор признаков составляется как конкатенация извлечённых векторов признаков для всех сенсорных изображений

$$\left(f_1^T, f_2^T, ..., f_N^T\right)^T \in R^{NM}.$$

В настоящем исследовании используются два основных предположения для формирования признаков:

– изображение с RGB-камеры и изображение дополнительного сенсора должны согласоваться по контурам лица, если атаки не было;
– статистические и текстурные характеристики яркости для дополнительного изображения существенно отличаются для реального лица и поддельных изображений.

На рис. 1 показаны примеры тепловизионной съёмки, обосновывающие эти закономерности: рис. 1*а* содержит изображение реального человека, а рис. 1*б-в* – два примера атак. Изображение человека характеризуется неравномерным распределением температуры в области лица. Напротив, распределение температуры поддельных изображений почти постоянно, при этом на них отсутствуют контуры лица.

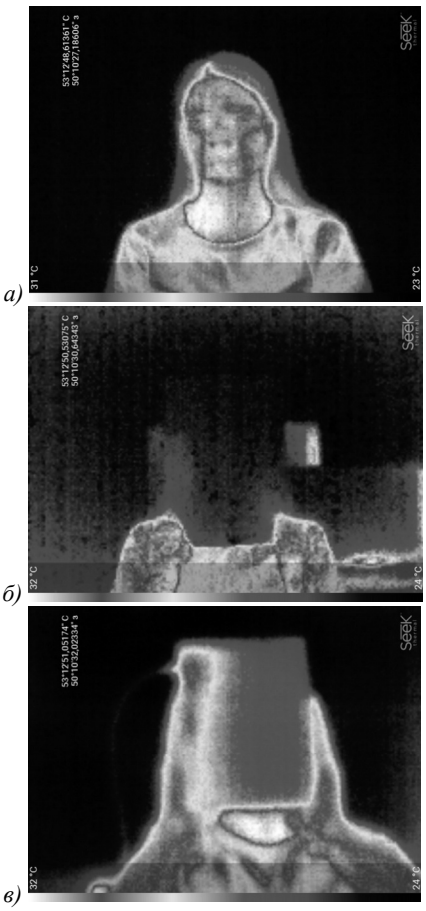

Рис. 1. Примеры тепловых изображений: а) подлинное изображение, б) атака при помощи распечатанного фото, в) атака при помощи фото, отображаемого на экране планшета

Предлагаемая система признаков включает в себя семь групп признаков, представленных в табл. 1. Группы признаков 1 – 6 были предложены и исследованы нами ранее [24] и показали высокую эффективность. Дополнительно к ним в настоящей статье добавлена седьмая группа признаков Харалика.

Рассмотрим далее процедуру извлечения вектора признаков $f_n = (f_{n1}, \ldots, f_{nM})^T$ для пары изображений X, Y_n.

Первая группа признаков (Var) – это собственно один признак – дисперсия сенсорного изображения. Для расчёта второй группы признаков (AvgArr) сенсорное изображение Y_n разбивается на $K \times K$ непересекающихся областей (в данной работе использовалось $K = 3$). Далее для каждой из областей вычисляется среднее значение яркости. Из этих значений формируется результирующий вектор данной группы признаков.

Табл. 1. Структура предлагаемой системы признаков

Группа признаков	Обозначение	Изображения для расчёта	Число признаков в группе
Дисперсия изображения	Var	Y_n	1
Средние значения по областям	AvgArr	Y_n	$K^2=9$
Гистограмма градиента	GradYH	Y_n	$Q=20$
Коэффициенты корреляции	CorrY	Y_n	4
Гистограмма градиента по чужим контурам	GradXH	X,Y_n	$Q=20$
Взаимная корреляция двух градиентов	CorrXY	X,Y_n	1
Признаки Харалика	H13	Y_n	13

Третья группа признаков (GradYH) – это гистограмма $H_{Yn}(q), q=1,\ldots,Q$ градиента сенсорного изображения G_{Yn}. В настоящей статье гистограмма строится только для диапазона значений градиента от 0 до 30 с количеством столбцов Q, равным 20. Такой выбор обусловлен малым динамическим диапазоном сенсорных изображений и, как следствие, малым диапазоном значений градиента. Данный признак позволяет эффективно различать случаи демонстрации фотоизображений или экрана от подлинных биометрических предъявлений, так как на изображениях первых практически отсутствуют малые значения градиента, соответствующие контурам в области лица.

Четвертая группа признаков (CorrY) – это коэффициенты корреляции значений сенсорного изображения в четырёх направлениях. Пусть $B_n(p_1,p_2)$ – автокорреляционная функция сенсорного изображения Y_n [25]. Тогда в данную группу признаков включены значения $B_n(-1,0)$, $B_n(1,0)$, $B_n(0,1)$ и $B_n(0,-1)$.

Данный признак позволяет различать высоко коррелированные 2D-изображения подделки и достоверные изображения. Первые дают более высокие значения корреляции.

Пятая группа признаков (GradXH) – гистограмма $H_{X(n)}(q), q=1,\ldots,Q$ значений градиента G_X изображения лица в оттенках серого X, вычисленная в контурных точках изображения Y_n. Пусть $i,j\in\Omega_{Y(n)}$ – координаты пикселей, принадлежащих множеству пар координат контурных точек $\Omega_{Y(n)}$ на сенсорном изображении Y_n. Под контурными точками будем понимать точки, определенные как контурные с помощью детектора контуров Canny [26]. Тогда $H_{X(n)}(q)$ – это столбцы гистограммы для значений $G_X(i,j),(i,j)\in\Omega_{Y(n)}$, причём гистограмма строится только для диапазона яркости от 0 до 128. Пиксели $G_X(i,j),(i,j)\in\Omega_{Y(n)}$ со значениями выше чем 128 соответствуют последнему столбцу гистограммы $H_{X(n)}(Q)$. Q также было принято равным 20. Данная группа признаков позволяет определить тот факт, что контуры лица расположены на RGB-изображении и сенсорном изображении в одних и тех же местах.

Шестая группа признаков (CorrXY), как и первая, состоящая из одного значения, – это взаимная корреляция изображений градиента G_X и G_{Yn}.

Седьмая группа признаков (H13) формируется конкатенацией векторов признаков из 13 текстурных признаков Харалика [27], вычисляемых независимо по $W\times W$ участкам сенсорного изображения лица Y_n (в работе использовалось значение $W=4$). Список используемых признаков Харалика приведён в табл. 2 (названия соответствуют статье [27]).

Табл. 2. Список используемых признаков Харалика

Название признака	Название признака
Angular Second Moment	Contrast
Correlation	Sum of Squares: Variance
Inverse Difference Moment	Sum Average
Sum Variance	Sum Entropy
Entropy	Difference Variance
Difference Entropy	Info. Measure of Correlation 1
Info. Measure of Correlation 2	

Признаки Харалика показали свою эффективность для решения рассматриваемой задачи в методе RDWT-Haralick-SVM, предложенном в работе [28] для случая RGB-изображений и повторно реализованном в работе [23] для случая использования изображения в оттенках серого и изображений глубины, а также инфракрасного и теплового изображений лица. В настоящей статье в отличие от реализаций данных признаков, использованных в статьях [28] и [23], вычисление признаков Харалика производится только по участкам сенсорного изображения Y_n, тогда как в оригинальном методе к этим признакам добавляются признаки Харалика, рассчитанные по результатам RDWT вейвлет-преобразования соответствующего участка лица. Кроме того, в настоящей статье данные признаки не вычисляются для изображения X, полученного в видимом диапазоне спектра.

В совокупности все предлагаемые признаки просты в вычислении и могут быть рассчитаны даже для изображений, полученных датчиками невысокого разрешения.

2.2. Классификация по полученным признакам

Полученные признаки использовались далее для бинарной классификации набора изображений (класс 1 – реальные лица, класс 0 – подделки) с использованием нескольких алгоритмов: метод опорных векторов (Support Vector Machine – SVM) с линейной функцией ядра [29], алгоритм Random Forest (RF) [30] и метод опорных векторов с радиальными базисными функциями (SVM-RBF). Причины выбора данных алгоритмов – высокая дискриминирующая способность

и относительно низкие требования к размеру обучающей выборки.

Параметры классификаторов подбирались методом решетчатого поиска путем оптимизации качества классификации на паре случайным образом сформированных обучающих (training) и валидационных (validation или development) наборов данных. Для измерения качества классификации использовались три показателя, рекомендуемых стандартом ISO/IEC 30107-3 [31] для детектирования атак на биометрическое предъявление:

– вероятность ошибки классификации предъявления при атаке (attack presentation classifcation error rate – APCER, соответствует значению False Positive Rate);
– вероятность ошибки классификации подлинных биометрических предъявлений (bonafide presentation classification error rate – BPCER, соответствует значению False Negative Rate);
– средняя частота ошибок классификации (average classification error rate – ACER), рассчитываемая как среднее арифметическое APCER и BPCER.

Для линейного классификатора SVM оптимизировались значение штрафного коэффициента C за неправильную классификацию в диапазоне [0,1] с шагом 0,0001 и матрица стоимости классов A. Матрица стоимости класса выбиралась из следующих трех вариантов:

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, A_3 = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix},$$

где стоимость класса изображений реальных лиц составляет 1, 2 и 4 соответственно.

Для алгоритма Random Forest оптимизировались количество деревьев классификации N в диапазоне от 10 до 200 с шагом 10, а также матрица стоимости классов A.

Для алгоритма SVM-RBF оптимизировались те же параметры, что и для линейного SVM, а также параметр масштаба функции ядра K в диапазоне от 1 до 10 с шагом 1.

Диапазоны изменения параметров классификаторов и матрицы A были определены в результате предварительных экспериментов с обучающими и валидационными наборами данных.

3. Экспериментальные исследования

3.1. Исходные данные для экспериментов

Для экспериментального исследования была использована база данных Wide Multi-Channel Presentation Attack Database (WMCA) [23]. Она содержит 1679 10-секундных видео, из которых 347 содержат реальных людей, а 1332 являются примерами атак на биометрическое предъявление следующих видов:

- атака «фальшивые очки», когда на человеке надеты бумажные очки или очки с фальшивым изображением глаз;
- атака «фальшивая голова», когда на камеру демонстрируется голова манекена,
- атака «печатное фото» соответствует случаю, когда на камеру демонстрируется бумажная фотография человека,
- атака «экран», когда изображение человека отображается на экране iPad,
- атака «жесткая маска», когда на человека надета напечатанная на 3D-принтере маска с лицом другого человека,
- атака «гибкая маска», когда на человека надета мягкая латексная маска с лицом другого человека,
- атака «3D бумажная маска», когда на человеке надета маска ручной работы, сделанная по фотографии другого человека.

В статье [23] авторы использовали в среднем 50 кадров для каждого видео, что в сумме дало 83950 изображений для формирования обучающей, валидационной и тестовой выборок. Подобная аугментация им потребовалась для обучения свёрточной нейронной сети MC-CNN. В настоящей статье мы использовали только один кадр из каждого видеофайла, поскольку цель предлагаемого метода – качественная работа в условиях ограниченного объёма обучающих данных.

Каждый кадр в используемом наборе данных был доступен в четырёх каналах: яркость в оптическом диапазоне (C), глубина (D), инфракрасный (I) и тепловой (T). Каналы C, D, I были получены с помощью камеры Intel RealSense SR300, канал T – тепловизором низкого разрешения Seek Thermal Compact PRO. Изображения во всех каналах были геометрически выровнены по области лица и приведены к размеру 128×128 пикселей. Пример изображений для атаки на биометрическое предъявление типа «жёсткая маска» в каждом из каналов показан на рис. 2.

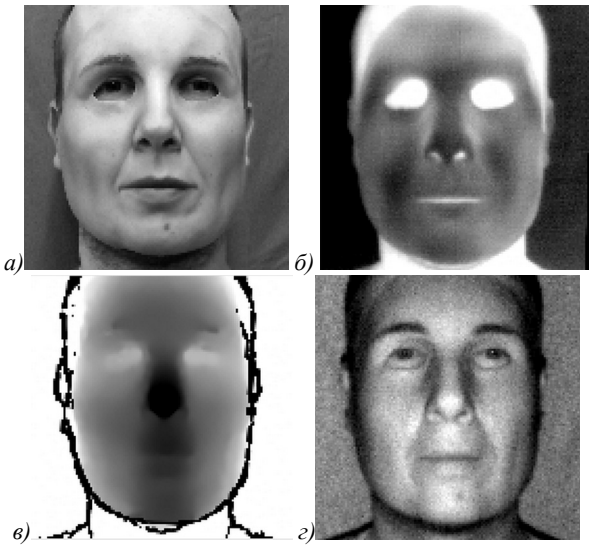

Рис. 2. Пример данных для атаки на биометрическое предъявление типа «жесткая маска» в различных каналах: а) цвет, б) тепловой, в) глубина, г) инфракрасный [23]

3.2. Методы, с которыми осуществлялось сравнение

Для сравнения с предложенным алгоритмом были выбраны методы RDWT-Haralick-SVM [28] и MC-CNN [23], показавшие наилучшие результаты обнаружения атак при тестировании на базе данных WMCA на 83950 изображениях.

Для реализации более близких условий сопоставления алгоритмов мы выполнили собственную реализацию алгоритма RDWT-Haralick-SVM в следующем виде:

1. Для каждого из каналов входного изображения вычислялся вектор признаков по следующей схеме:
   – изображение разделяется на 4×4 участка размерами 32×32 пикселя;
   – для каждого участка изображения рассчитываются 13 признаков Харалика, перечисленных в табл. 2, и RDWT-преобразование, как первый уровень двумерного вейвлет-преобразования Хаара. В результате вычисления RDWT-преобразования формируются четыре изображения участка, соответствующие четырём компонентам R_a, R_v, R_h, R_d: аппроксимирующей, горизонтальной, вертикальной и диагональной соответственно;
   – для каждого участка и каждой из четырёх компонент R_a, R_v, R_h, R_d также вычисляются 13 признаков Харалика,
   – в результате все полученные признаки Харалика по всем участкам и всем их компонентам RDWT-преобразования объединяются в единый вектор признаков размерности 1040.

2. Вектора признаков для каждого канала изображения объединяются в единый вектор признаков размерности 4160.

3. Полученные вектора признаков классифицируются с помощью линейного классификатора SVM.

Алгоритм MC-CNN представляет собой свёрточную нейронную сеть, основанную на использовании предварительно обученной на большом числе изображений для распознавания лиц нейронной сети LightCNN [32]. Тем не менее, MC-CNN требует дополнительного обучения некоторых частей сети и содержит дополнительные слои, необходимые для конкатенации векторов признаков, полученных для каждого из каналов изображения, и несколько полносвязных слоев для вычисления результата бинарной классификации. Ввиду отсутствия реализации MC-CNN провести эксперимент в тех же условиях, что и для предложенного алгоритма, не удалось. Поэтому для MC-CNN используются значения ошибок, полученные на наборе из 83950 изображений [23].

3.3. Протокол проведения эксперимента

В настоящей статье в экспериментах мы следуем протоколу grandtest, описанному в [23], чтобы получить более сопоставимые результаты с MC-CNN. Весь набор из 1679 изображений был разделен на три подмножества: обучающий набор, валидационный набор и тестовый набор в соотношении 3:3:4. Разделение на наборы данных в точности соответствовало разделению, используемому в работе [23]. Наборы для обучения и валидации использовались для оптимизации параметров классификатора как для реализации предложенного метода, так и для реализации RDWT-Haralick-SVM.

Целями экспериментов являлись:
– определение наиболее информативных каналов для классификации атак на биометрическое предъявление с использованием предложенных признаков;
– сравнение результатов работы с использованием различных классификаторов и предложенных признаков с базовыми алгоритмами;
– оценка влияния размера обучающей выборки на качество классификации для предложенных признаков;
– оптимизация состава предлагаемых признаков в соответствии с процедурой последовательного отбора признаков.

3.4. Результаты экспериментов

В табл. 3 приведены результаты определения наиболее информативных каналов для обнаружения атак с использованием предложенных признаков и линейного SVM-классификатора. Результаты классификации для алгоритмов RF и SVM-RBF продемонстрировали аналогичную зависимость от состава каналов. Табл. 3 показывает, что наилучшая ошибка соответствует использованию всех каналов. Средняя ошибка в этом случае равна 2,91 %, что значительно меньше, чем при использовании любой другой комбинации каналов.

В табл. 4 представлены результаты работы трех классификаторов с использованием предложенных признаков, а также метода RDWT-Haralick-SVM. Из табл. 4 видно, что наилучший результат достигнут при помощи двух классификаторов (SVM и SVM-RBF) на предложенном наборе признаков. В то же время важно отметить, что достигнутый результат (ACER = 2,91 %) всё же значительно уступает результату MC-CNN, полученному в [23] (ACER = 0,30 %, см. табл. 5). Однако этот результат достигается нейронной сетью при использовании для обучения в 50 раз больше данных. Таким образом, мы можем заключить, что предложенный метод предпочтителен для ситуаций, когда получение больших наборов данных для обучения затруднительно.

Для сопоставления предложенного метода с MC-CNN в идентичных условиях проведён отдельный эксперимент на наборе из 83950 изображений, результаты которого представлены в табл. 5.

Табл. 3. Классификация линейным SVM при различных наборах каналов для предложенных признаков

Каналы	Оптимальные гиперпараметры		Валидационная выборка		Тестовая выборка		
	C	*A*	APCER, %	BPCER, %	APCER, %	BPCER, %	ACER, %
CDIT	**0,0055**	A_3	**3,34**	**1,85**	**4,07**	**1,74**	**2,91**
CDT	0,0119	A_3	4,90	0	7,01	6,96	6,99
CTI	0,0111	A_3	4,45	0,93	5,20	1,74	3,47
CDI	0,0015	A_3	11,58	7,41	9,28	0	4,64
CT	0,0439	A_3	7,13	0	10,86	1,74	6,30
CD	0,0319	A_3	10,02	13,89	12,90	6,09	9,49
CI	0,0856	A_3	6,46	3,70	6,11	8,70	7,40

Табл. 4. Ошибки классификации CDIT данных на тестовой выборке для различных методов классификации (объём выборки – 1679 изображений)

Метод	Оптимальные гиперпараметры	APCER, %	BPCER, %	ACER, %
Предложенные признаки + SVM	$\boldsymbol{C = 0{,}0055, A_3}$	**4,07**	**1,74**	**2,91**
Предложенные признаки + RF	$N = 80, A_2$	5,88	7,83	6,85
Предложенные признаки + SVM-RBF	$\boldsymbol{K = 9, C = 0{,}3382, A_3}$	**4,07**	**1,74**	**2,91**
RDWT-Haralick+SVM	$C = 0{,}0002, A_3$	8,37	0	4,18

Табл. 5. Ошибки классификации CDIT данных на тестовой выборке для различных методов классификации (объём выборки – 83950 изображений)

Метод	Оптимальные гиперпараметры	APCER, %	BPCER, %	ACER, %
Предложенные признаки + SVM	$C = 1{,}1327, A_3$	0,11	2,24	1,18
Предложенные признаки + RF	$N = 80, A_2$	3,23	6,50	4,87
Предложенные признаки + SVM-RBF	$K = 1, C = 1{,}23, A_3$	0,11	2,24	1,18
RDWT-Haralick+SVM (результат из [23])	-	6,39	0,49	3,44
MC-CNN (результат из [23])	**-**	**0,60**	**0,0**	**0,30**

Табл. 6. Ошибки классификации предложенных признаков при различном размере обучающей выборки

Классификатор	Соотношение данных в обучающей, валидационной и тестовой выборках	Оптимальные гиперпараметры	APCER, %	BPCER, %	ACER, %
SVM	5:5:90	$C = 0{,}0014, A_3$	12,79	0,95	6,87
SVM	10:10:80	$C = 0{,}0041, A_3$	6,02	4,30	5,16
SVM	20:20:60	$C = 0{,}0051, A_3$	4,26	1,91	3,09
RF	5:5:90	$N = 200, A_2$	6,77	8,57	7,67
RF	10:10:80	$N = 60, A_2$	5,83	7,89	6,86
RF	20:20:60	$N = 40, A_2$	8,02	8,13	8,08
SVM-RBF	**5:5:90**	$\mathbf{K = 8, C = 0{,}5300, A_3}$	**7,36**	**5,08**	**6,22**
SVM-RBF	**10:10:80**	$\mathbf{K = 10, C = 0{,}4400, A_3}$	**6,02**	**4,30**	**5,16**
SVM-RBF	**20:20:60**	$\mathbf{K = 10, C = 0{,}3562, A_3}$	**4,89**	**0,96**	**2,92**

Из табл. 5 видно, что наилучшее значение ACER для предложенного набора признаков упало в 2,5 раза – до 1,18 %, значительно приблизившись к MC-CNN, а также увеличив отрыв от RDWT-Haralick+SVM. Несмотря на этот результат, следует повторно подчеркнуть, что использование столь большого набора данных не является рабочим сценарием использования разработанного метода.

Для тестирования предложенного метода в условиях малых обучающих выборок был проведен эксперимент, в котором для обучения использовалось только 5, 10 и 20 % от общего числа (1679) изображений в наборе. Табл. 6, иллюстрирующая результаты данного эксперимента, демонстрирует, что наилучшие результаты соответствуют классификатору SVM-RBF.

Поскольку предложенный набор признаков был сформирован из эвристических соображений, вполне возможна его избыточность. Поэтому был проведён эксперимент с отбором групп признаков в рамках семи предложенных групп с использованием линейного SVM для классификации. Результаты оптимизации состава признаков методом последовательного отбора признаков приведены в табл. 7. При этом группы признаков 1 – 7 вычислялись для всех используемых каналов CDIT. Результаты в табл. 7 означают, что при использовании для классификации только одного признака из всех предложенных наивысшую точность классификации алгоритм демонстрирует для группы признаков Харалика. Для всех наборов из двух групп признаков, содержащих группу H13, наилучшие ре-

зультаты показал набор признаков Харалика, дополненный гистограммой градиента изображения в оттенках серого, рассчитанной в контурных точках сенсорных изображений. Наконец, наилучший набор признаков, дальнейшее расширение которого не приводит к уменьшению средней ошибки классификации, включает дополнительно группы признаков AvgArr, GradYH, Var. В результате отбора признаков удалось уменьшить ошибку BPCER до 0,47 %, а среднюю ошибку классификации ACER – до 2,47.

Табл. 7. Результаты предложенного алгоритма для различных наборов признаков

Признаки	ACER %	APCER %	BPCER %
H13	5,18	3,39	6,96
H13, GradXH	4,10	3,85	4,35
H13, GradXH, AvgArr	3,11	3,62	2,61
H13, GradXH, AvgArr, GradYH	2,91	4,07	1,74
H13, GradXH, AvgArr, GradYH, Var	2,47	4,07	0,87

5. Заключение

В статье рассматривается использование данных от различных изображающих сенсоров для обнаружения атак на биометрическое предъявление при использовании лицевой биометрической аутентификации. Разработана универсальная по отношению к используемым сенсорам система признаков, позволяющая с высокой точностью обнаруживать презентационные атаки и не требующая большого объёма данных для обучения. В статье исследование осуществлялось на наборе данных WMCA, содержащем изображения лица, полученные датчиком глубины, инфракрасной камерой и тепловой камерой в дополнение к изображению, полученному с помощью RGB-камеры.

Предложенная система признаков в сочетании с любой классической моделью бинарной классификации (в работе рассмотрены линейный SVM, Random Forest и SVM-RBF), обучаемой при помощи этих признаков, составляют предлагаемый в статье метод обнаружения атак на биометрическое предъявление. В результате сравнения метода на основе предложенной системы признаков с алгоритмами RDWT-Haralick-SVM и MC-CNN, показавшими в работах других исследователей наилучшие результаты классификации на наборе WMCA, было установлено, что предложенный метод позволяет уменьшить среднюю ошибку обнаружения атак на биометрическое предъявление по сравнению с алгоритмом RDWT-Haralick-SVM и приблизиться к результату на основе свёрточной нейронной сети MC-CNN, обученной на в 50 раз большем наборе данных.

Таким образом, можно заключить, что предложенный метод улучшает точность классификации по сравнению с методами, также реализующими традиционный подход к решению задач классификации, состоящий из выбора признаков и обучения одной из классических моделей бинарной классификации. При этом метод уступает MC-CNN, реализующему сценарий глубокого обучения, однако в отличие от методов этого класса может быть использован в ситуации, когда получение больших наборов обучающих данных затруднительно или невозможно.

Благодарности

Работа выполнена при поддержке РФФИ (проект 19-29-09045 в части параграфа 2.1, 3, проект 19-07-00357 в части параграфа 2.2) и Министерства науки и высшего образования РФ в рамках выполнения работ по Государственному заданию ФНИЦ «Кристаллография и фотоника» РАН (соглашение 007-ГЗ/Ч3363/26) в части параграфа 1.

References

[1] Mahmood Z, Muhammad N, Bibi N, Ali T. A review on state-of-the-art face recognition approaches. Fractals 2017; 25(2): 1750025.

[2] Kalinovskiy IA, Lavrentyeva GM. Face anti-spoofing for biometric systems [In Russian]. 28th Int Conf on Computer Graphics and Vision (GraphiCon) 2018: 204-207.

[3] Facial recognition market to grow at 12 percent CAGR to 2024, technavio forecasts. Biometric update. Source: ⟨https://www.biometricupdate.com/202011/facial-recognition-market-to-grow-at-12-percent-cagr-to-2024-technavio-forecasts⟩.

[4] Bhattacharjee S, Mohammadi A, Anjos A, Marcel S. Recent advances in face presentation attack detection. In Book: Marcel S, Nixon MS, Fierrez J, Evans N, eds. Handbook of biometric anti-spoofing. Cham: Springer; 2019: 207-228.

[5] Zhang M, Zeng K, Wang J. A survey on face anti-spoofing algorithms. Journal of Information Hiding and Privacy Protection 2020; 2(1): 21-34.

[6] Nikitin MYu, Konushin VS, Konushin AS. Face anti-spoofing with joint spoofing medium detection and eye blinking analysis. Computer Optics 2019; 43(4): 618-626. DOI: 10.18287/2412-6179-2019-43-4-618-626.

[7] Gorbatsevich VS, Moiseenko AS, Vizilter YV. FaceDetectNet: Face detection via fully-convolutional network. Computer Optics 2019; 43(1): 63-71. DOI: 10.18287/2412-6179-2019-43-1-63-71.

[8] Wang T, Yang J, Lei Z, Liao S, Li SZ. Face liveness detection using 3D structure recovered from a single camera. Int Conf on Biometrics (ICB) 2013: 1-6.

[9] Li X, Komulainen J, Zhao G. Generalized face anti-spoofing by detecting pulse from face videos. Proc IEEE 23rd Int Conf on Pattern Recognition 2016; 4239-4244.

[10] Bao W, Li H, Li N, Jiang W. A liveness detection method for face recognition based on optical flow field. Proc Int Conf on Image Analysis and Signal Processing 2009: 233-236.

[11] Li HL, Wang SQ, Kot AC. Face spoofing detection with image quality regression. Proc 6th Int Conf on Image Processing Theory Tools and Applications 2016; 1-6.

[12] Yi D, Lei Z, Zhang ZW, Li SZ. Face anti-spoofing: Multi-spectral approach. In Book: Marcel S, Nixon MS, Li SZ, eds. Handbook of biometric anti-spoofing. London: Springer; 2014: 83-102.
[13] Mohamed S, Ghoneim A, Youssif A. Visible/infrared face spoofing detection using texture descriptors. MATEC Web of Conferences 2019; 292: 04006.
[14] Sun L, Huang WB, Wu MH. TIR/VIS correlation for liveness detection in face recognition. Int Conf on Computer Analysis of Images and Patterns 2011: 114-121.
[15] Erdogmus N, Marcel S. Spoofing 2D face recognition systems with 3D masks and antispoofing with kinect. Int Conf of the BIOSIG Special Interest Group (BIOSIG) 2013: 1-8.
[16] Sun X, Huang L, Liu C. Multispectral face spoofing detection using VIS–NIR imaging correlation. Int J Wavelets Multiresolut Inf Process 2018; 16: 1840003.
[17] Sun X, Huang L, Liu C. Multimodal face spoofing detection via RGB-D images. 2018 24th Int Conf on Pattern Recognition (ICPR) 2018: 2221-2226.
[18] Wang Y, Nian F, Li T, Meng Z, Wang K. Robust face anti-spoofing with depth information. J Vis Commun Image Represent 2017; 49: 332-337.
[19] Kowalski M. A study on presentation attack detection in thermal infrared. Sensors 2020; 20: 3988.
[20] Ewald KE, Zeng L, Yao Z, Mawuli CB, Abubakar HS, Victor A. Applying CNN with extracted facial patches using 3 modalities to detect 3D face spoof. 2020 17th Int Computer Conf on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) 2020: 216-221.
[21] Singh M, Arora AS. Computer aided face liveness detection with facial thermography. Wireless Pers Commun 2020; 111: 2465-2476.
[22] Tang Y, Chen L. 3D Facial geometric attributes based anti-spoofing approach against mask attacks. 2017 12th IEEE Int Conf on Automatic Face & Gesture Recognition (FG 2017) 2017: 589-595.
[23] George A, Mostaani Z, Geissenbuhler D, Nikisins O, Anjos A, Marcel S. Biometric face presentation attack detection with multi-channel convolutional neural network. IEEE Trans Inf Forensics Secur 2019; 15: 42-55.
[24] Denisova A, Fedoseev V. Presentation attack detection in facial authentication using small training data obtained by multiple devices. 2021 Int Conf on Information Technology and Nanotechnology (ITNT) 2021: 1-5. DOI: 10.1109/ITNT52450.2021.9649390.
[25] Soifer VA, ed. Computer image processing. VDM Verlag Dr Müller; 2007. ISBN: 978-3-639-16837-2.
[26] Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 1986; PAMI-8(6): 679-698.
[27] Haralick R, Shanmugam K, Dinstein I. Textural features for image classification. IEEE TSMC 1973; 3(6): 610-621.
[28] Agarwal A, Singh R, Vatsa M. Face anti-spoofing using Haralick features. 8th Int Conf on Biometrics Theory, Applications and Systems (BTAS) 2016: 1-6.
[29] Christianini N, Shawe-Taylor J. An introduction to support vector machines and other Kernel-Based learning methods. Cambridge, UK: Cambridge University Press; 2000.
[30] Kulkarni VY, Sinha PK. Pruning of random forest classifiers: A survey and future directions. Int Conf on Data Science & Engineering (ICDSE) 2012: 64-68.
[31] ISO/IEC 30107-3. Information technology – Biometric presentation attack detection – Part 3: Testing and reporting. Source: ⟨https://www.iso.org/obp/ui/#iso:std:iso-iec:30107:-3:ed-1:v1:en⟩.
[32] Wu X, He R, Sun Z, Tan T. A light cnn for deep face representation with noisy labels. IEEE Trans Inf Forensics Secur 2018; 13(11): 2884-2896.

Сведения об авторах

Денисова Анна Юрьевна, 1988 года рождения, в 2011 году окончила Самарский государственный аэрокосмический университет имени академика С.П. Королёва (ныне Самарский университет) по специальности «Прикладная математика и информатика». В 2014 году защитила диссертацию на соискание степени кандидата технических наук. Работает доцентом кафедры геоинформатики и информационной безопасности, Самарский университет. Область научных интересов: обработка изображений, геоинформационные системы.
E-mail: *denisova_ay@geosamara.ru* .

Федосеев Виктор Андреевич, 1986 года рождения, в 2009 году окончил Самарский государственный аэрокосмический университет имени академика С.П. Королёва (ныне – Самарский университет) по специальности «Прикладная математика и информатика», кандидат физико-математических наук (2012). В настоящее время работает доцентом кафедры геоинформатики и информационной безопасности Самарского университета и научным сотрудником в ИСОИ РАН – филиале ФНИЦ «Кристаллография и фотоника» РАН. Области научных интересов: анализ изображений, цифровые водяные знаки, стеганография. E-mail: *vicanfed@gmail.com* .

ГРНТИ: 28.23.15
Поступила в редакцию 24 сентября 2021 г. Окончательный вариант – 21 февраля 2022 г.

Detection of presentation attacks on facial authentication systems using special devices

A.Y. Denisova [1], V.V. Fedoseev [1,2]
[1] Samara National Research University, 443086, Samara, Russia, Moskovskoye Shosse 34,
[2] IPSI RAS – Branch of the FSRC "Crystallography and Photonics" RAS,
443001, Samara, Russia, Molodogvardeyskaya 151

Abstract

The article proposes a feature system designed to detect presentation attacks on facial authentication systems. In this type of attack, an attacker disguises as an authorized user using his image. The feature system assumes the possibility of using one or more special imaging sensors in addition to the basic RGB camera (thermal cameras, depth cameras, infrared cameras). The method has demonstrated a low error rate on the WMCA dataset, while experiments have shown its ability to remain effective in the case of the lack of training data. The comparative experiments carried out showed that the proposed method surpassed the RDWT-Haralick-SVM algorithm, and also approached the results of the MC-CNN algorithm, based on deep learning, which requires a significantly larger amount of training data.

Keywords: presentation attack, authentication, face recognition, thermal data, depth data.

Citation: Denisova AY, Fedoseev VA. Detection of presentation attacks on facial authentication systems using special devices. Computer Optics 2022; 46(4): 612-620. DOI: 10.18287/2412-6179-CO-1054.

Acknowledgements This work was supported by the Russian Foundation for Basic Research under projects Nos. 19-29-09045, 19-07-00357 and state contract 007-GZ/Ch3363/26.

Authors' information

Anna Yurievna Denisova (b. 1988). Graduated from Samara National Research University (Samara University) in 2011 as Master of Mathematics and Computer Science. She received a degree of Candidate in Technical Sciences in 2014. Now she works at Samara University. The area of research interest includes image processing and geoinformational systems. E-mail: *denisova_ay@geosamara.ru* .

Victor Andreevich Fedoseev (b. 1986) graduated (2009) from Samara State Aerospace University (presently, Samara National Research University), majoring in Applied Mathematics and Computer Science. Candidate degree in Computer Science (2012). Currently he is an associate professor at the Geoinformatics and Information Security department at Samara University and a research scientist at the Image Processing Systems Institute of RAS – Branch of the FSRC "Crystallography and Photonics" RAS. His scientific interests include image processing and analysis, digital watermarking and steganalysis. E-mail: *vicanfed@gmail.com* .

Received September 24, 2021. The final version – February 21, 2022.