

УДК 004.932.72'1

DOI: [10.26102/2310-6018/2023.40.1.019](https://doi.org/10.26102/2310-6018/2023.40.1.019)

Реконструкция лица как метод повышения точности идентификации человека в видеокадре

И.О. Причко✉, А.Д. Афанасьев

*Иркутский национальный исследовательский технический университет,
Иркутск, Российская Федерация
i@prichko.ru*

Резюме. В статье рассмотрен подход, направленный на повышение точности идентификации человека по изображению его лица в системах видеонаблюдения, основанный на методе реконструкции на базе генеративно-сопоставительных сетей. В ходе расследования правонарушений часто встречаются видеозаписи представляющих интерес для следствия людей с низким разрешением или содержащие визуальные помехи различного происхождения, которые ограничивают применения методов идентификации человека при помощи нейронных сетей глубокого обучения. Это создает проблему как обнаружения по лицу определенного человека в видеоданных, так и поиск выбранного в кадре человека по базе данных. Известно, что реконструкция лица на основе генеративно-сопоставительных сетей позволяет существенно улучшать изображения лица низкого качества, однако, этот метод требователен к содержимому исходного изображения, т. к. любые окклюзии и помехи многократно усиливаются. В статье приводится подход, заключающийся в предобработке изображений на основе известного свойства видеозаписей – наличия версионности изображений объектов. Предложенный алгоритм позволяет устранить большинство визуальных помех и в дальнейшем восстановить изображение лица с высоким качеством. В ходе проведения экспериментов нами также найден метод реставрации элементов лица, который позволяет повысить узнаваемость неизвестного лица человеком, что может иметь важное значение в ходе опознания свидетелями.

Ключевые слова: идентификация человека, реконструкция лица, видеоаналитика, суперразрешение изображения, генеративно-сопоставительные сети, компьютерное зрение.

Для цитирования: Причко И.О., Афанасьев А.Д., Реконструкция лица как метод повышения точности идентификации человека в видеокадре. *Моделирование, оптимизация и информационные технологии.* 2023;11(1). URL: <https://moitvivr.ru/ru/journal/pdf?id=1303> DOI: 10.26102/2310-6018/2023.40.1.019

Face reconstruction as a technique for enhancing accuracy of human recognition in a video frame

I.O. Prichko✉, A.D. Afanasyev

*Irkutsk National Research Technical University,
Irkutsk, Russian Federation
i@prichko.ru*

Abstract. The article considers an approach aimed at enhancing the accuracy of human identification by facial image in video surveillance systems, using the reconstruction method based on generative-adversarial networks. During the investigation of offenses, one often encounters video recordings of people of interest for the investigation with a low resolution or containing visual disturbances of different genesis, which limits the implementation of techniques for identifying the person by means of deep learning neural networks. This causes two problems: one pertaining to face detection of a certain person in the video data and another regarding the search for a selected person in the frame contained in the database. The reconstruction of a face using generative adversarial networks is known to

significantly improve low-quality face images, but this method is demanding of the content of the original image as any occlusions and disturbances are multiply amplified. The paper presents an approach composed of image preprocessing on the basis of the known property of video recordings – the presence of object image versioning. The proposed algorithm helps to correct much of the visual noise and subsequently reconstruct the face image with high quality. During the experiments, we have also found a method of facial elements restoration which enables the increase in the recognizability of an unknown face by a person, which can be important during the identification by witnesses.

Keywords: face recognition, face restoration, video analytics, superresolution image, generative and adversarial networks, computer vision.

For citation: Prichko I.O., Afanasyev A.D., Face reconstruction as a technique for enhancing accuracy of human recognition in a video frame. *Modeling, Optimization and Information Technology*. 2023;11(1). URL: <https://moitvvt.ru/ru/journal/pdf?id=1303> DOI: 10.26102/2310-6018/2023.40.1.019 (In Russ.).

Введение

Видеоаналитика как технология автоматизированного анализа информации из систем видеонаблюдения является одним из наиболее активных направлений исследований в разделе компьютерного зрения и находит применение в широком спектре практических внедрений в различные сферы жизни человека. Одним из примеров социально значимых применений видеоаналитики является обеспечение безопасности граждан на основе систем видеонаблюдения, в том числе расследование уже свершившихся преступлений. Рост интереса правоохранительных органов к видеоаналитике закономерно формируется из эффективности такого инструмента в задаче обнаружения значимой информации, но также он обусловлен распространением стационарных систем видеофиксации. Это обязывает сотрудников органов правопорядка исследовать большие объемы данных, что без применения современных методов анализа вызывает очевидные трудности. В современном мире данные из стационарных систем видеонаблюдения имеют особую ценность в ходе предотвращения правонарушений и изучения обстоятельств их совершения, а обнаруженная криминалистически значимая информация во многих случаях является неопровержимым доказательством вины или невиновности человека.

Возникающие при расследовании преступлений следственные ситуации обобщенно можно разделить на две обособленные категории по критерию наличия информации о фигурантах преступления [1] (подозреваемого или иных участников события). В случае совершения правонарушения в условиях неочевидности применяются особые методы их обнаружения, которые не рассматриваются в данной работе, тогда как имея сведения ориентирующего значения об участниках преступления и владея видеоданными из систем наблюдения, открывается возможность использовать технологию видеоаналитики в части извлечения признаков и последующего поиска по ним. Благодаря постоянному развитию методов глубокого обучения, накоплению наборов данных для тренировки нейронных сетей и увеличению вычислительных мощностей технология идентификации человека по лицу достигла значимых результатов, которые широко используются правоохранительными органами во всем мире. Так V. L. Raposo в своей работе [2], помимо проблемы правового обеспечения современных технологий в правоохранительной деятельности и принципа «осторожного подхода», подчеркивает неоспоримую значимость идентификации людей в видеопотоках с применением искусственного интеллекта.

В качестве основной цели видеоаналитики декларируется обнаружение и определение временных и пространственных событий, а также их анализ. Еще в

2002 году коллектив авторов под редакцией N. Dimitrova [3] описали четыре основных этапа в построении прикладного решения для анализа видеоконтента: извлечение признаков, структурный анализ видеоряда, абстрагирование и индексирование. Таким образом, применение рассматриваемой технологии сводится к выполнению четырех взаимосвязанных процессов для каждого объекта: обнаружение (детекция), отслеживание (трекинг), выделение признаков (идентификация) и прогнозирование (анализ ситуаций), которые выполняются многократно, что обеспечивает непрерывное уточнение гипотез о количестве, расположении, типах объектов и связывающих их событиях. Каждый обозначенный выше процесс создает множество сложных исследовательских проблем, но в данной работе анализу подвергается подзадача выделения признаков объекта – идентификация человека по изображению лица в кадре.

Описанный коллективом авторов [4] типовой подход в проектировании end-to-end решения по обнаружению определенного человека в видеоданных по изображению его лица строится на трех ключевых процессах: обнаружение в кадре, выравнивание и преобразование лица. В рамках первого этапа производится определение расположений лиц в кадровом пространстве выбранным методом, опираясь на ограничения и допущения предметной области. В дальнейшем производится процедура выравнивания лиц с целью их калибровки до общепринятого представления «анфас», в ходе которой производится кадрирование с нормализацией пиксельных размеров и аффинного преобразования. Затем из приведенных к нормали лиц извлекаются дискриминантные признаки (дескриптор) для дальнейшего этапа сопоставления.

Автор работы разделяет позицию исследователей [5-8] в части добавления к обозначенным выше этапам важного дополнительного – предобработка изображения лица. Вне сомнений лабораторные исследования отличаются от натуральных испытаний наличием множества случайных помех, возникающих ввиду воздействия факторов окружающей обстановки, и которые не могут быть в полной мере симитированы и спрогнозированы. В качестве распространенных источников шумов, с которыми автор работы часто сталкивается в ходе трудовой деятельности, являются: загрязнение линзы камеры (искажение лица); электромагнитные помехи (артефакты в виде полос, пикселизации, потеря фокуса); быстрые движения объектов (эффект «желе»); нахождение объектов вне фокусного расстояния (размытие лица); низкая освещенность (неравномерное цветовое распределение лица) и прочие. Привнесенные в видеоданные искажения во множестве случаев нивелируют преимущества автоматического анализа с использованием компьютерной техники, тогда как в тех же самых случаях человеческий мозг воспринимает лицо в видеозаписи в динамике, что позволяет выявлять значимые признаки. Таким образом, с целью повышения точности идентификации человека в видеокadre исследователями предлагается добавить предобработку, включающую в себя устранение последствий привнесенных шумов различными методами.

Тогда как автор предлагает углубиться в процесс подготовки изображения лица основываясь на гипотезе, что время перекрытия «полезных» пикселей лица человека пикселями, относящимися к шумам, в большинстве случаев занимает меньше половины времени выборки. В этом случае, сформировав из видеозаписи набор изображений лиц одного человека (на основании отслеживания объекта в кадре) и применив к нему процедуру расчета медианы для каждого пикселя, можно получить модель лица, содержащую наименьшее количество пикселей шумов. Более того, «усреднение» таким образом изображений лица открывает возможность применения современных методов реконструкции элементов лица на основании генеративно-состязательных сетей (GAN). Полученная в результате обозначенных действий модель лица имеет важную для расчета дескриптора особенность – восстановление частей лица с сохранением расположения опорных точек [4].

Таким образом, предложенный коллективом авторов [4] типовой подход в разработке прикладных решений с реализацией функции идентификации человека в кадре предлагается дополнить не менее значимым процессом – «предобработкой изображения лица», тогда этапы построения приложения строятся в следующем порядке: обнаружение в кадре, выравнивание, предобработка и преобразование лица. Безусловно каждая добавленная функция увеличивает вычислительную стоимость проекта, и для каждой прикладной задачи формируются допущения и ограничения. В рамках текущей работы исследуются пути повышения точности идентификации человека по лицу на примере ретроспективного анализа видеоданных с целью выявления криминалистически значимой информации, в связи с чем в качестве допущения принимается превалирование точности результата над скоростью и стоимостью исполнения, но описанный подход может также быть применен в иных сферах. Целью работы является создание эффективного метода повышения точности идентификации человека в видеокadre, основанного на алгоритме усреднения изображения лица и последующем процессе его реконструкции с применением генеративно-состязательных сетей.

Материалы и методы

В качестве одного из решений задачи повышения точности идентификации человека по изображению его лица в данной статье предлагается добавление этапа предобработки, состоящего из трех элементов: усреднение изображений лица, реконструкция лица (построение модели, наиболее близкой к усредненному лицу), реставрация частей лица. Таким образом, реализация такого подхода подразумевает эффективное выполнение следующей последовательности этапов:

1. Формирование набора изображений лиц по определенному человеку.
2. Построение усредненного лица.
3. Реконструкция лица методом построения наиболее близкой модели.
4. Восстановление частей лица методом слепой реставрации.
5. Извлечение дескриптора лица.

1. Формирование набора изображений лиц

Как было изложено ранее, в ходе работы с видеоданными специалист сталкивается с шумами различного происхождения, но в большинстве своем это временные искажения (артефакты). Большинство таких искажений перекрывают лицо или его части на время, меньшее чем такие части видны. Таким образом, пока пиксель лица виден в кадре больше 50 % времени, медиана такого пикселя вернет значение самого лица без учета пикселя помех. Для проведения этой процедуры необходима выборка кадров L , содержащая k изображений. Для каждого пикселя мы имеем k оценок состояния и, соответственно, по получившемуся дискретному ряду можем рассчитать значение медианы (квантиль распределения уровня 0,5). Математическое представление этого механизма по одному пикселю (x, y) отражено в (1), где $f(L(x, y))$ – вариационный ряд $L(x, y)$.

$$F(x, y) = \begin{cases} f(L(x, y))_{\frac{k+1}{2}}, & k - \text{нечетное} \\ \frac{f(L(x, y))_k + f(L(x, y))_{\frac{k+1}{2}+1}}{2}, & k - \text{четное} \end{cases} \quad (1)$$

Для реализации обозначенной процедуры необходимо построить набор изображений лица, например, взяв за основу последовательно идущие кадры. Методы и

подходы детекции лиц на изображении постоянно совершенствуются, при этом в ходе создания реальных проектов разработчики не отказываются и от классических методов. В частности, исследователи [9] указывают как на использование классических численных методов обнаружения лиц в кадре, так и с применением машинного обучения. Различие в подходах использования того или иного детектора основано на учете множества факторов, препятствующих или снижающих обнаружение: окклюзии (частичное закрытие лица), неравномерность освещения, положение головы человека, выражение эмоций, аксессуаров и макияж, а также масштаб лица. В связи с тем, что в случае реализации end-to-end решения анализу подвергается не весь кадр, а лишь обнаруженный детектором [10] объект, требуется применение современных моделей с высокой вычислительной стоимостью. Одним из прогрессивных и эффективных методов обнаружения лиц автор выделяет модель «Sample and Computation Redistribution for Efficient Face Detection» (SCRFD) [11]. В случае приоритетности скорости над точностью автором предлагается использовать модель «SCRFD-2.5GF», в ином случае – «SCRFD-10G». Пример набора обнаруженных лиц приведен на Рисунке 1.



Рисунок 1 – Набор изображений лица одного человека (последнее – медиана)
Figure 1 – Set of images of one person's face (the last is the median)

Исходя из изображений на Рисунке 1, становится очевидным, что процедура усреднения кадров в выборке принесет желаемый результат только при соотношении каждого пикселя к соответствующему пикселю единого кадрового пространства (центрирование).

2. Построение усредненного лица

Для эффективного решения задачи усреднения изображений динамического объекта требуется определить опорные точки на нем для дальнейшего сведения набора изображений к единому макету. В случае с лицом человека общеприняты 7 основных видов наборов опорных точек [4] (4, 5, 29, 68, 98, 106, 194 точки), каждый из которых отличается как количеством и точностью определения точек, так вычислительной стоимостью их обнаружения на изображениях. В настоящее время наиболее эффективными для этих целей являются обученные нейронные сети. В связи с тем, что в задаче усреднения необходима коррекция всех лиц к единому положению, в качестве опорных точек может быть взят набор из 5 типовых (две точки, расположенные на глазах, две – на носу и две – на губах), так как они чаще всего занимают центральную часть изображения. Рассмотрим пример на Рисунке 1: обратите внимание на наклон линии между глазами (желтым цветом) и горизонтом, а также на отрицательный результат построения усреднения без предобработки (см. последнее фото на Рисунке 1). Можно указать на то, что в выборке имеются изображения лица в достаточном качестве и в данном случае не требуется усреднение, но пример выбран специально без окклюзий с целью показать влияние фактора положения головы на процессы усреднения.

Наиболее простое решение задачи уравнивания лиц в выборке – наклон изображения и сдвиг, для чего достаточно рассчитать угол между линией глаз и

горизонтом через арктангенс высоты и ширины получившегося треугольника, и в дальнейшем произвести наклон на полученный угол, а также сместить по вертикали и горизонтали каждое изображение до соотношения с единой координатной системой (например, принятой точки для левого глаза).

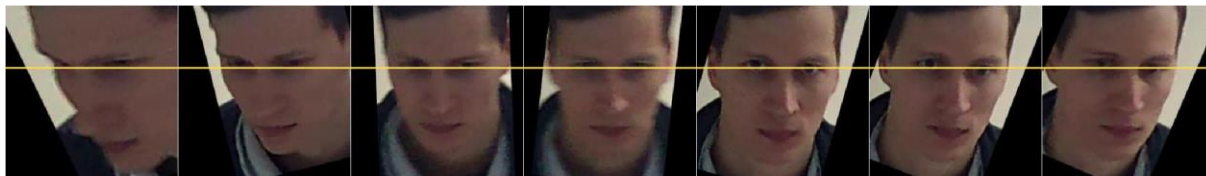


Рисунок 2 – Выравнивание набора лиц по оси глаз
Figure 2 – Face set alignment along the eye axis

В рассматриваемом на Рисунке 2 примере используется сдвиг оси глаз на треть размера изображения (поворот, сдвиг вверх-вниз и влево-вправо). Немного усложненный описанному механизму – аффинное преобразование (например, «`warpAffine()`» библиотеки «`OpenCV`»), который в дальнейшем и будет использован в примерах одномерной трансформации. Описанный механизм приведения лиц к единому положению в плоскости частично решает проблему, но при этом на Рисунке 2 целенаправленно показана одна из проблем идентификации человека – различие положения головы в пространстве. В случае применения указанного метода, в усредненном изображении будут искажения по контуру лица (см. второе фото на Рисунке 4).

Распознавание лиц основано на построении дескриптора (уникальный вектор, описывающий взаимную удаленность опорных точек лица), при этом разными авторами не так давно была замечена зависимость алгоритмов идентификации от точности определения расположения глаз (точность определения расположения глаз на изображении имела большое значение в эффективности процесса идентификации, чем корректность определения расположения иных точек), тогда как сейчас модели эволюционировали до полноценного учета 68 и более точек. То есть, в случае приведения положения лица до пяти точек в итоговом усредненном изображении именно эти элементы будут наиболее детализированы, тогда как современные модели построения дескриптора ориентированы в том числе на контур лица. С целью снижения такого рода искажений необходима корректировка в двумерном пространстве (поворот, сдвиг, наклон, ширина, высота), для чего по каждому изображению лица производится поиск 2D координат 106 опорных точек (см. левую часть Рисунка 3). Далее на основании полученных групп точек формируется усредненное расположение опорных точек (см. среднюю часть Рисунка 3), имеющее свойство равного удаления максимального и минимального значения каждой соответствующей опорной точки в выборке.

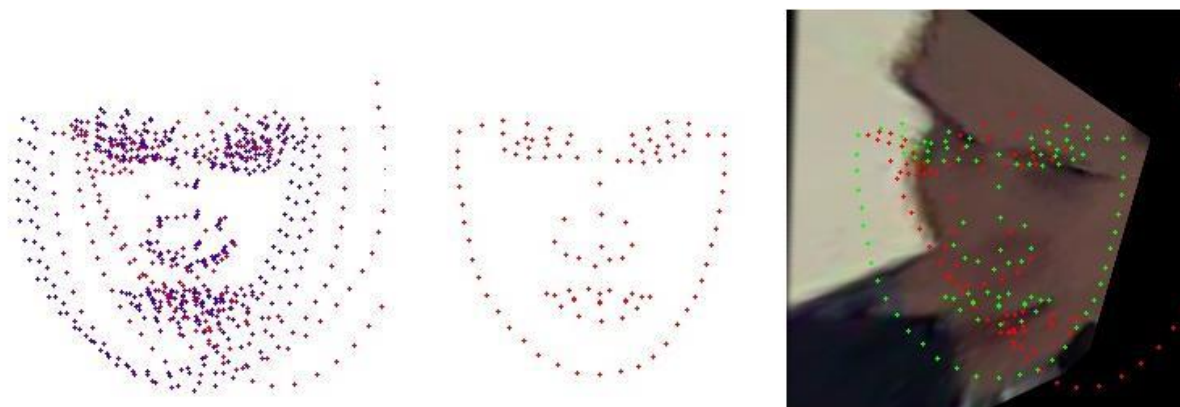


Рисунок 3 – Приведение лица с большим отклонением (справа) к эталонной маске (в центре), а также сводное изображение масок опорных точек в выборке (слева)
Figure 3 – Facial conversion of the face with a large deviation (right) to the reference mask (center), and the summary image of the reference point masks in the set (left)

На левой части Рисунка 3 заметна одна группа точек (в дальнейшем группа опорных точек лица будет именоваться «маской»), которая нетипична по сравнению с эталонной группой (см. центральная часть Рисунка 3). В математической статистике такие нехарактерные наблюдения к общей выборке называются «выбросами» [12], которые необходимо исключать из наблюдения. В качестве набора для поиска отклонений от нормы используется стандартное отклонение каждой маски в выборке от эталонной маски (к которой в дальнейшем будет произведено приведение изображений). Обнаружение выбросов в этом случае возможно, в том числе, посредством расчета межквартильного диапазона (IQR) по формуле (2), где в результате будут получены границы типичных значений (Q_1 и Q_3 можно получить функцией percentile из библиотеки numpy).

$$[Q_1 - 1.5 \times \text{IRQ}; Q_3 + 1.5 \times \text{IRQ}], \text{ где } \text{IRQ} = Q_3 - Q_1 \quad (2)$$

Метод поиска выбросов посредством расчета межквартильного диапазона по набору значений стандартных отклонений (от каждой маски-выборки к эталонной маске) выбран из-за того, что на него не влияют экстремальные выбросы (которые в предметной области возникают чаще равномерных выбросов). Таким образом на небольших выборках снижается риск утраты важных кадров в связи с ошибочным их исключением из-за перекоса диапазона. На правой части Рисунка 3 показан пример изображения-выброса (первое на Рисунке 1), маска которого достаточно сильно отклонена от эталонной группы (зеленым показана эталонная маска, красным – группа опорных точек). Приведение такого изображения к эталонному вносит неконтролируемые помехи (как в нашем случае с нехарактерным носом и контуром).

После процедуры исключения нежелательных изображений лиц, если такие были выявлены, производится процедура приведения всех частей изображения на основе их удаленности от эталонной группы точек. В исследовании применялись методы «findHomography()» и «warpPerspective()» библиотеки «OpenCV».

По результатам создания трансформированных изображений возможна генерация усредненного лица человека обозначенным ранее методом медианы каждого пикселя (Рисунок 4).

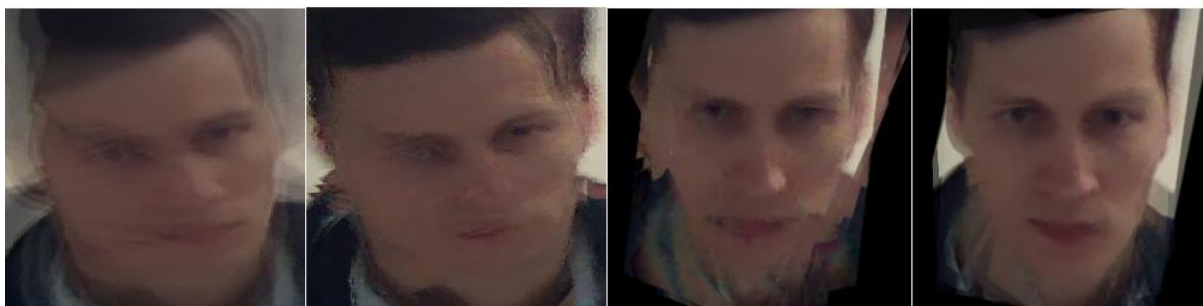


Рисунок 4 – Виды усредненного лица, слева направо: среднее, медиана, медиана после одномерных корректировок, медиана после двумерных корректировок
Figure 4 – Views of the average face, from left to right: average, median, median after 1-dimensional corrections, median after 2-dimensional corrections

Рассмотрим работу алгоритма в более сложном примере, содержащем как существенные окклюзии и низкое разрешение, так и наклон / поворот головы (Рисунок 5).



Рисунок 5 – Апробация усреднения лиц при окклюзиях, построчно сверху вниз: исходные изображения (~32x44); выравненные в двумерном пространстве (6 изображение – выброс); среднее, медиана, медиана при 1D трансформации, медиана при 2D трансформации
Figure 5 – Practice of facial approximation with occlusions, top-down row: original images (~32x44); adjusted in 2-dimensional (6 image – outlier); average, median, median at 1-dimensional corrections, median at 2-dimensional corrections

Вернемся к Рисунку 3, где в центральной части изображена классическая форма из 106 опорных точек лица, в том числе применяемая в подавляющем большинстве алгоритмов идентификации лиц, и точность определения расположения каждой из указанных точек на интересующем изображении напрямую влияет на конечный результат идентификации. На Рисунках 4 и 5 заметны различия одномерной и двумерной коррекции, а именно обратите внимание на контуры лица, губ и бровей: максимально точное их восстановление позволяет добиться большей точности идентификации. Кроме

методов усреднения в настоящее время разработаны механизмы реконструкции лица, т. е. создание несуществующего лица на основании максимального подобия представленному на вход изображению лица.

3. Создание модели лица методом реконструкции

Методы реконструкции лица появились не так давно и основаны на генеративно-сопоставительных сетях (GAN-сети). Задачей таких сетей является построение изображения лиц людей, которые визуально не отличимы от настоящих, при этом не принадлежат таковым. В рассматриваемой предметной области интерес представляет результат применения таких сетей, при котором создается изображение вымышленного лица и уменьшение его до разрешения поданного на вход изображения (подвергаемого улучшению). В дальнейшем созданное изображение сопоставляется с «входным» и, в случае наличия различий между ними, итерации повторяются.

Опуская этапы развития методов реконструкции (хорошо описаны и сравнены создателями сети «GFP-GAN» в статье [13]), в ходе испытаний на реальных следственных ситуациях наиболее подходящим для целей идентификации выбрана «Generative Facial Prior Generative Adversarial Network» (GFP-GAN). В отличие от общепринятого подхода, метод «GFP-GAN» строится на слепом восстановлении с предварительной оценкой геометрии лица, а также ориентирован на восстановление его деталей и элементов [13]. Таким образом, чем точнее у изображения на входе восстановлены контуры элементов, тем выше точность реконструкции их деталей, что безусловно повышает общую точность идентификации человека на изображении. Модель «GFP-GAN» в настоящее время имеет в общем доступе три версии: «v2», «v3» и «v4». При этом в ходе реальных испытаний нельзя однозначно указать на необходимость выбора лишь последней версии, хотя она и является приоритетной, с точки зрения автора (см. раздел «Результаты»).



Рисунок 6 – Применение трех версий «GFP-GAN» при реконструкции к усредненному лицу, слева направо: усредненное, GFP-GANv2, GFP-GANv3, GFP-GANv4

Figure 6 – Applying three versions of "GFP-GAN" in reconstruction to the average face, from left to right: median, GFP-GANv2, GFP-GANv3, GFP-GANv4

На Рисунке 6 показано применение трех версий рассматриваемой сети к предобработанному изображению (медиана по 2D корректированным изображениям). На указанном рисунке заметно, что в ходе работы метода «GFP-GAN» сеть стремится к максимальному сохранению положения элементов лица и предсказания их деталей. Это именно та особенность, которая необходима для повышения точности идентификации, так как в зависимости от того, насколько точно распознающая сеть (для нахождения опорных точек лица) определит положение точек на изображении, будет зависеть итоговое значение вектора-дескриптора. Фактически в данном случае сгенерированное лицо является математической моделью лица существующего человека. Насколько

точно полученная модель предсказывает положение элементов и деталей (и не обязательно его внешние черты, такие как цвет, контрастность атрибуты лица) настоящего лица, настолько ближе будут полученные в ходе идентификации результаты. Таким образом, перед алгоритмом идентификации человека нам необходимо предварительно предсказать наиболее вероятные расположения деталей основных элементов лица и их расположение, что в том числе возможно с применением генеративно-состязательных сетей, так как строение лица человека закономерно и имеет незримые зависимости.

4. Восстановление частей лица методом реставрации

В качестве дополнительного метода предобработки изображения перед процедурой расчета дескриптора и последующего процесса идентификации предлагается рассмотреть недавно опубликованный метод реконструкции лица посредством его реставрации [14]. Метод «RestoreFormer» не является продолжением разработки «GFP-GAN», он ориентирован на применение высококачественного словаря для восстановления контекстной информации при пространственном внимании, а не на использовании локальных операторов. В упрощенном виде метод подразумевает повышение детализации изображений и их разрешения на основе базы высокого разрешения. При этом на практике повышение точности происходит не во всех случаях и только после применения сетей реконструкции. На Рисунке 7 показана работа метода «RestoreFormer» (второе) к усредненному лицу (первое), а также применение реконструкции «GFP-GANv4» (третье) с последующей реставрацией (четвертое). Подход в использовании реконструкции «GFP-GANv4» и последующей реставрации «RestoreFormer» показывает бóльшую визуальную узнаваемость человека, тогда как методы программной идентификации по лицу только в ряде случаев показывают повышение точности.



Рисунок 7 – Применение реставрации лица, слева направо: исходное (усредненное), реставрация («RestoreFormer»), реконструкция («GFP-GANv4»), реконструкция + реставрация («GFP-GANv4» + «RestoreFormer»)

Figure 7 – Face restoration techniques, left to right: baseline (median), restoration ("RestoreFormer"), reconstruction ("GFP-GANv4"), reconstruction + restoration ("GFP-GANv4" + "RestoreFormer")

Результаты

В целях апробации изложенных в работе решений рассмотрено пять выборок кадров (показаны на Рисунке 9) из видеозаписей различных стационарных систем наблюдения (со средним разрешением лица в выборке 31x42, 30x40, 141x210, 36x49 и 45x63 соответственно), на которых содержится изображение лица автора в различных условиях съемки. На Рисунке 9 в каждой строке последовательно показаны результаты работы различных методов улучшения: линейная интерполяция; медианное усреднение после 2D трансформации (см. Раздел 2); сеть «GFP-GAN» вер. 2 по усредненному лицу; сеть

«GFP-GAN» вер. 3 по усредненному лицу; сеть «GFP-GAN» вер. 4 по усредненному лицу; сеть «RestoreFormer» после сети «GFP-GAN» вер. 4. В качестве эталонного изображения лица взята фотография высокого качества (Рисунок 8).



Рисунок 8 – Эталонное изображение лица для процесса идентификации
Figure 8 – Reference face image for the identification process

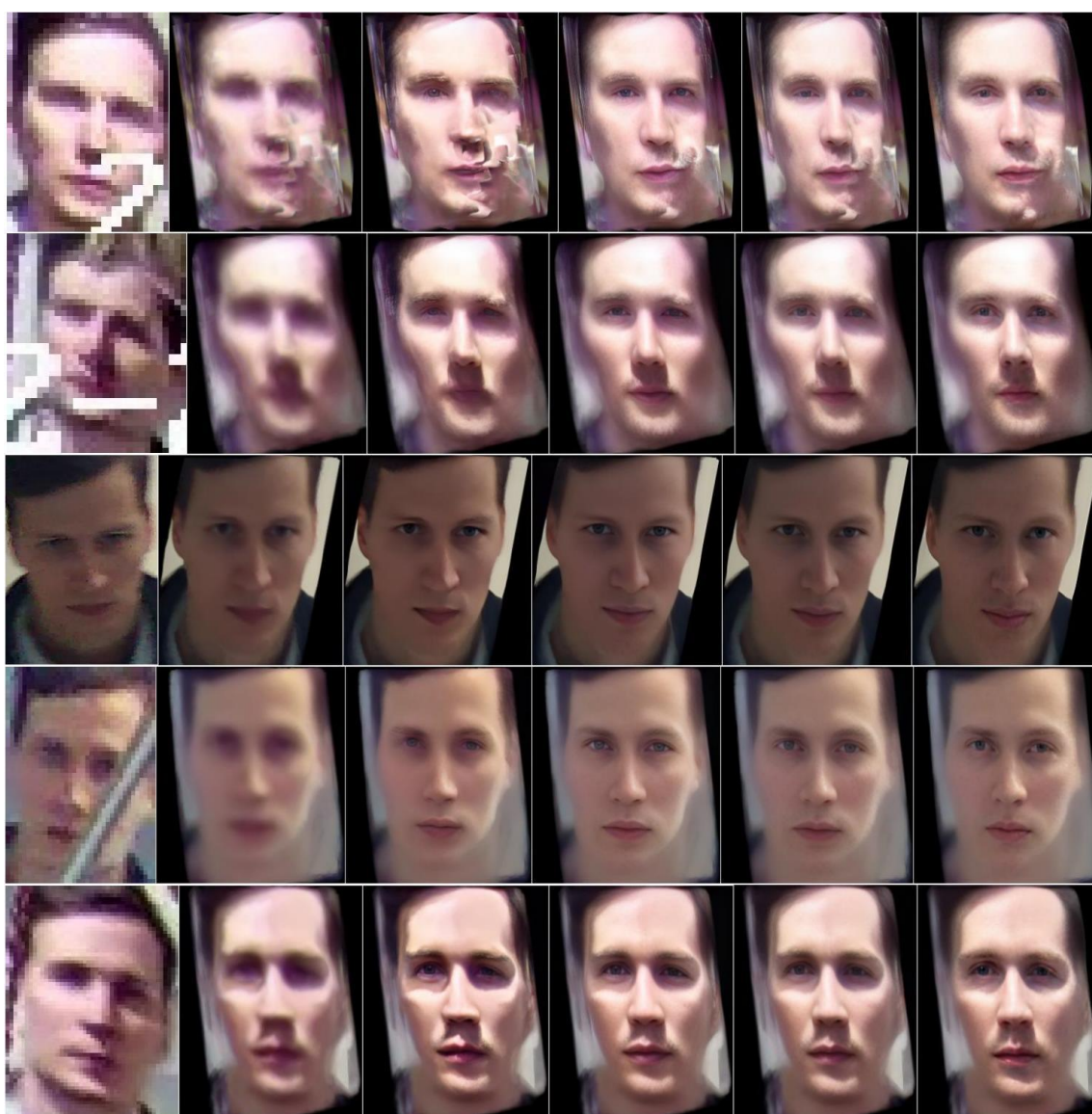


Рисунок 9 – Пять выборок изображений лиц (построчно), слева направо: условные номера изображений 2, 6, 7, 8, 9, 10 (согласно Таблице 1)
Figure 9 – Five sets of face images (line by line), from left to right: image numbers 2, 6, 7, 8, 9, 10 (according to Table 1)

Для определения удаленности дескрипторов (степень идентичности лиц) использовалась библиотека DeepFace с набором предобученных моделей, при этом в качестве наиболее приемлемых и современных для построения дескрипторов (face recognition models) были выбраны «VGG_FACE» и «ARCFACE». Также в качестве детектора лиц наиболее стабильные результаты отмечены в «OpenCV» (ResNet-10), а для расчета расстояния между векторами (дескрипторами) принята Евклидова метрика с нормализацией L2-вектора (Евклидова норма). В ходе испытаний Евклидова метрика (L2) показала более стабильные оценки сходимости векторов-дескрипторов.

В Таблице 1 для удобства восприятия используется градиент от зеленого (успешная идентификация), до красного (отрицательная идентификация), жирным шрифтом выделены наилучшие значения в наборе.

Таблица 1 – Результаты идентификации лиц в выборках (Евклидова метрика с L2-нормализацией)

Table 1 – Results of identifying individuals in the sets (Euclidean metric with L2-normalization)

Усл. номер	НАБОР					
	1	2	3	4	5	
VGG-FACE (порог 0,86)						
Изображение	1	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>
	2	0,893	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>
	3	0,921	1,097	0,985	1,077	0,883
	4	0,879	0,989	0,941	1,050	0,964
	5	1,005	1,036	0,685	0,847	0,723
	6	0,967	1,046	0,677	0,884	0,721
	7	0,849	0,938	0,694	0,781	0,723
	8	0,809	0,860	0,703	0,782	0,649
	9	0,820	0,915	0,642	0,756	0,687
	10	0,759	0,840	0,675	0,773	0,681
	11	1,237	0,834	0,751	0,717	0,706
ARCFACE (порог 1,13)						
Изображение	1	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>
	2	1,375	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>	<i>Не обнаруж.</i>
	3	1,363	1,388	1,352	1,364	1,301
	4	1,357	1,384	1,316	1,358	1,308
	5	1,351	1,365	0,985	1,277	1,083
	6	1,297	1,361	1,032	1,280	1,063
	7	1,179	1,157	0,997	1,196	1,031
	8	1,108	1,043	1,047	1,029	1,015
	9	1,150	1,069	0,972	1,065	0,979
	10	1,016	1,113	1,033	1,165	1,029
	11	1,403	1,108	1,142	1,136	1,029

В Таблице 1 сведена информация о проведенных испытаниях с применением двух моделей сетей для идентификации с расчетом метрики (чем меньше значение, тем больше вероятность принадлежности двух лиц одному человеку, при этом в случае перехода граничного значения 0,86 для VGG-FACE и 1,13 для ARC-FACE лица считаются идентифицированными с вероятностью не менее 94 %), при этом под условными номерами изображений (столбец 1, 2) понимаются следующие:

- 1 – исходное изображение в первоначальном разрешении;
- 2 – исходное (усл. номер 1) с увеличением разрешения (линейная интерполяция);
- 3 – усредненное лицо (методом арифметического среднего);
- 4 – усредненное лицо (методом медианы);
- 5 – усредненное лицо (методом медианы) после алгоритма 1D трансформации;
- 6 – усредненное лицо (методом медианы) после алгоритма 2D трансформации;
- 7 – GFP-GAN версии 2 по усредненному лицу (усл. номер 6);
- 8 – GFP-GAN версии 3 по усредненному лицу (усл. номер 6);
- 9 – GFP-GAN версии 4 по усредненному лицу (усл. номер 6);
- 10 – RestoreFormer после GFP-GAN версии 4 (усл. номер 9);
- 11 – RestoreFormer (со «scale=2») после GFP-GAN версии 4 (усл. номер 9).

Обсуждение

В рассматриваемой Таблице 1 результаты испытаний намеренно не переводились в процентное значение в связи с тем, что в настоящее время нет доказанных сведений о функции зависимости метрики от достоверности идентификации в процентах (зависимость вероятности от расстояния), тогда как принято считать, что при достижении порога (в рассматриваемом случае и для Евклидовой метрики L2: 0,86 для «VGG-FACE» и 1,13 для «ARCFACE») точность идентификации равна точности сети на испытаниях при обучении (т. е. не менее 94 %) [15]. В большинстве случаев в рассматриваемой группе выборок процесс идентификации не выдавал дескриптор для всех первоначальных изображений лиц в выборках (см. строки с изображениями 1 и 2 в Таблице 1), это связано с низким качеством изображений и наличием различных помех (выборки целенаправленно были собраны в реальных условиях и с помехами различного происхождения, при этом изображения взяты без разрыва последовательности кадров).

В части влияния усреднения на точность идентификации (см. строки с изображениями 3-6 в Таблице 1) можно сделать вывод, как минимум, о появлении возможности построения дескриптора, при этом со значениями, достаточно близкими к граничным, а в ряде случаев, позволяющим успешно идентифицировать человека. Таким образом, опираясь на данные испытаний, можно сделать вывод о достаточности выравнивания в одномерном пространстве для устранения окклюзий, двумерных трансформаций – для устранения искажений положения головы, при этом возникает закономерный вопрос об эффективности выравнивания в трехмерном пространстве, при котором производится построение 3D модели лица на основе усреднения по 68 опорным точкам (3D координаты), что может быть объектом дальнейших исследований. Также отдельной темой исследования может быть выявление частей лица в выборке кадров, которые наиболее качественно зафиксированы камерой (имеют лучшую детализацию или контрастность) и будут использоваться в усредненном лице, т. е. метод составления усредненного лица по частям (Триангуляция Делоне и диаграмма Вороного [16]).

Применение методов реконструкции лица на основе GAN-сетей (см. строки с изображениями 7-9 в Таблице 1) во всех случаях повысило точность, при этом однозначно выбрать единую версию «GFP-GAN» нельзя. Исходя из особенностей выборки, можно сделать вывод о том, что для более качественных изображений подходит «GFP-GAN» версии 4, тогда как для лиц низкого качества – «GFP-GAN» версии 3 и 4.

Алгоритмы реставрации в некоторых случаях показали как самый лучший результат (см. строки с изображениями 10 и 11 в Таблице 1) повышения точности, так и полностью исказили лицо (с точки зрения метрики идентификации, а не видимой узнаваемости). Отдельно отмечается испытание на выборке № 1 особо низкого

разрешения, при котором реставрация лица без модуля суперразрешения (параметр «scale=1») показала наивысший результат среди испытаний в выборке, тогда как суперразрешение в два раза (параметр «scale=2») полностью исказило лицо. Важно отметить, что метод «RestoreFormer» не во всех случаях повышает точность «машинной» идентификации, тогда как для человека, в небольшом отличии от результатов «GFP-GAN», результирующее изображение после обработки «RestoreFormer» более узнаваемо. Эта особенность может иметь важное значение с точки зрения восстановления лица и последующего опознания свидетелями преступления. Исходя из испытаний и понимания этапов работы описанных методов, можно сделать вывод о том, что для повышения точности необходимо восстановление пространственного расположения элементов лица, тогда как низкая узнаваемость человеком не является помехой для методов идентификации, основанных на относительности опорных точек лица.

Заключение

Данная работа направлена на исследование методов повышения точности идентификации реального человека на изображениях лица в кадрах систем видеонаблюдения. В ходе исследования были рассмотрены как методы усреднения изображений лица в целях устранения помех, так и современные методы реконструкции на основе генеративно-состязательных сетей. Множество проведенных испытаний с выборками различного характера (различного разрешения исходных изображений, наличия разнообразных видов помех, пространственного положения лица и т. д.) показал эффективность применения отраженного в статье подхода.

СПИСОК ИСТОЧНИКОВ

1. Себякин А.Г. Анализ информации о соединениях между абонентами, использование его результатов в раскрытии и расследовании преступлений. *Полицейская и следственная деятельность*. 2018;4:29–38. Доступно по: https://nbpublish.com/library_read_article.php?id=27992. DOI: 10.25136/2409-7810.2018.4.27992 (дата обращения: 10.02.2023).
2. Raposo V.L. The Use of Facial Recognition Technology by Law Enforcement in Europe: a Non-Orwellian Draft Proposal. *European Journal on Criminal Policy and Research*. 2022;1–19. Доступно по: <https://link.springer.com/content/pdf/10.1007/s10610-022-09512-y.pdf>. DOI: 10.1007/s10610-022-09512.
3. Dimitrova N., Zhang H.J., Shahraray B., Sezan I., Huang T., Zakhor A. Applications of video-content analysis and retrieval. *IEEE multimedia*. 2002;9(3):42–55. Доступно по: http://www-video.eecs.berkeley.edu/papers/Dimitrova/IEEE_MM2002.pdf. DOI: 10.1109/MMUL.2002.1022858.
4. Du H., Shi H., Zeng D., Zhang X.P., Mei T. The elements of end-to-end deep face recognition: A survey of recent advances. *ACM Computing Surveys (CSUR)*. 2022;54(10s):1–42. Доступно по: <https://arxiv.org/pdf/2009.13290v3.pdf>. DOI: 10.1145/3507902.
5. Насонов А.В., Крылов А.С., Ушмаев О.С. Применение метода суперразрешения для биометрических задач распознавания лиц в видеопотоке. *Системы высокой доступности*. 2009;1:26–34. Доступно по: https://imaging.cs.msu.ru/pub/2009.SVD.Nasonov_Krylov.SRVidRec.ru.pdf (дата обращения: 10.02.2023).
6. Пахирка А.И. Применение метода улучшения изображений для систем распознавания лиц. *Сибирский аэрокосмический журнал*. 2010;3:25–29. Доступно по: https://www.elibrary.ru/download/elibrary_15204085_84528176.pdf (дата обращения: 10.02.2023).

7. Морнева А.Е., Масленников А.А. Методы повышения точности работы алгоритмов распознавания лиц на изображении. *Вестник молодёжной науки России*. 2020;1:14–18. Доступно по: https://www.elibrary.ru/download/elibrary_44487166_63470132.pdf (дата обращения: 10.02.2023).
8. Бордюжа В., Умняшкин С.В. Предобработка изображений для повышения качества распознавания лиц на цифровых фото. *Академическая наука как фактор и ресурс инновационного развития*. 2022;26–38. Доступно по: <https://sciencen.org/assets/Kontent/Konferencii/Arhiv-konferencij/KOF-578.pdf#page=26> (дата обращения: 10.02.2023).
9. Minaee S., Luo P., Lin Z., Bowyer K. Going deeper into face detection: A survey. *arXiv preprint*. 2021;2103(14983). Доступно по: <https://arxiv.org/pdf/2103.14983.pdf>. DOI: 10.48550/arXiv.2103.14983.
10. Афанасьев А.Д., Причко И.О. Детектор объектов непостоянного движения в задаче обнаружения криминалистически значимой информации. *Моделирование, оптимизация и информационные технологии*. 2021;9(2):23–24. Доступно по: <https://moitvivr.ru/ru/journal/pdf?id=928>. DOI: 10.26102/2310-6018/2021.33.2.007. (дата обращения: 10.02.2023).
11. Guo J., Deng J., Lattas A., Zafeiriou S. Sample and computation redistribution for efficient face detection. *arXiv preprint*. 2021;2105(04714). Доступно по: <https://arxiv.org/pdf/2105.04714.pdf>. DOI: 10.48550/arXiv.2105.04714.
12. Положинцев Б.И. *Теория вероятностей и математическая статистика. Введение в математическую статистику: Учебное пособие*. СПб.: ЮНИТИ; 2016. 95 с.
13. Wang X., Li Y., Zhang H., Shan Y. Towards real-world blind face restoration with generative facial prior. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021;9168–9178. Доступно по: <https://arxiv.org/pdf/2101.04061.pdf>. DOI: 10.48550/arXiv.2101.04061.
14. Wang Z., Zhang J., Chen R., Wang W., Luo P. RestoreFormer: High-Quality Blind Face Restoration From Undegraded Key-Value Pairs. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022;17512–17521. Доступно по: <https://arxiv.org/pdf/2201.06374.pdf>. DOI: 10.48550/arXiv.2201.06374.
15. DeepFace – The Most Popular Open Source Facial Recognition Library. Доступно по: <https://viso.ai/computer-vision/deepface/> (accessed on 10.02.2023).
16. Adeshara K., Elangovan V. Face recognition using PCA integrated with Delaunay triangulation. *arXiv preprint*. 2020;2011(12786). Доступно по: <https://arxiv.org/ftp/arxiv/papers/2011/2011.12786.pdf>. DOI: 10.48550/arXiv.2011.12786.

REFERENCES

1. Sebyakin A.G. Analiz informatsii o soedineniyakh mezhdu abonentami, ispol'zovanie ego rezul'tatov v raskrytii i rassledovanii prestuplenii. *Politseiskaya i sledstvennaya deyatel'nost'*. 2018;4:29–38. URL: https://nbpublish.com/library_read_article.php?id=27992 DOI: 10.25136/2409-7810.2018.4.27992 (accessed on 10.02.2023). (In Russ.).
2. Raposo V.L. The Use of Facial Recognition Technology by Law Enforcement in Europe: a Non-Orwellian Draft Proposal. *European Journal on Criminal Policy and Research*. 2022;1–19. Available from: <https://link.springer.com/content/pdf/10.1007/s10610-022-09512-y.pdf>. DOI: 10.1007/s10610-022-09512.
3. Dimitrova N., Zhang H.J., Shahraray B., Sezan I., Huang T., Zakhor A. Applications of video-content analysis and retrieval. *IEEE multimedia*. 2002;9(3):42–55. Available from: http://www-video.eecs.berkeley.edu/papers/Dimitrove/IEEE_MM2002.pdf. DOI: 10.1109/MMUL.2002.1022858.

4. Du H., Shi H., Zeng D., Zhang X.P., Mei T. The elements of end-to-end deep face recognition: A survey of recent advances. *ACM Computing Surveys (CSUR)*. 2022;54(10s):1–42. Available from: <https://arxiv.org/pdf/2009.13290v3.pdf>. DOI: 10.1145/3507902.
5. Nasonov A.V., Krylov A.S., Ushmaev O.S. Primenenie metoda superrazresheniya dlya biometricheskikh zadach raspoznavaniya lits v videopotoke. *Sistemy vysokoi dostupnosti*. 2009;1:26–34. Available from: https://imaging.cs.msu.ru/pub/2009.SVD.Nasonov_Krylov_SRVIDRec.ru.pdf (accessed on 10.02.2023). (In Russ.).
6. Pakhirka A.I. Primenenie metoda uluchsheniya izobrazhenii dlya sistem raspoznavaniya lits. *Sibirskii aerokosmicheskii zhurnal = The Siberian Aerospace Journal*. 2010;3:25–29. Available from: https://www.elibrary.ru/download/elibrary_15204085_84528176.pdf (accessed on 10.02.2023). (In Russ.).
7. Morneva A.E., Maslennikov A.A. Metody povysheniya tochnosti raboty algoritmov raspoznavaniya lits na izobrazhenii. *Vestnik molodezhnoi nauki Rossii*. 2020;1:14–18. Available from: https://www.elibrary.ru/download/elibrary_44487166_63470132.pdf (accessed on 10.02.2023). (In Russ.).
8. Bordyuzha V., Umnyashkin S.V. Predobrabotka izobrazhenii dlya povysheniya kachestva raspoznavaniya lits na tsifrovyykh foto. *Akademicheskaya nauka kak faktor i resurs innovatsionnogo razvitiya*. 2022;26–38. Available from: <https://sciencen.org/assets/Kontent/Konferencii/Arhiv-konferencij/KOF-578.pdf#page=26> (accessed on 10.02.2023). (In Russ.).
9. Minaee S., Luo P., Lin Z., Bowyer K. Going deeper into face detection: A survey. *arXiv preprint*. 2021;2103(14983). Available from: <https://arxiv.org/pdf/2103.14983.pdf>. DOI: 10.48550/arXiv.2103.14983.
10. Afanas'ev A.D., Prichko I.O. Detektor ob"ektov nepostoyannogo dvizheniya v zadache obnaruzheniya kriminalisticheskoi znachimoi informatsii. *Modelirovanie, optimizatsiya i informatsionnye tekhnologii = Modeling, optimization and information technology*. 2021;9(2):23–24. Available from: <https://moitvvt.ru/ru/journal/pdf?id=928>. DOI: 10.26102/2310-6018/2021.33.2.007 (accessed on 10.02.2023). (In Russ.).
11. Guo J., Deng J., Lattas A., Zafeiriou S. Sample and computation redistribution for efficient face detection. *arXiv preprint*. 2021;2105(04714). Available from: <https://arxiv.org/pdf/2105.04714.pdf>. DOI: 10.48550/arXiv.2105.04714.
12. Polozhintsev B.I. *Teoriya veroyatnostei i matematicheskaya statistika. Vvedenie v matematicheskuyu statistiku: Uchebnoe posobie*. SPb.: YuNITI; 2016. 95 s. (In Russ.).
13. Wang X., Li Y., Zhang H., Shan Y. Towards real-world blind face restoration with generative facial prior. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021;9168–9178. Available from: <https://arxiv.org/pdf/2101.04061.pdf>. DOI: 10.48550/arXiv.2101.04061.
14. Wang Z., Zhang J., Chen R., Wang W., Luo P. RestoreFormer: High-Quality Blind Face Restoration From Undegraded Key-Value Pairs. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022;17512–17521. Available from: <https://arxiv.org/pdf/2201.06374.pdf>. DOI: 10.48550/arXiv.2201.06374.
15. DeepFace – The Most Popular Open Source Facial Recognition Library. Available from: <https://viso.ai/computer-vision/deepface/> (accessed on 10.02.2023).
16. Adeshara K., Elangovan V. Face recognition using PCA integrated with Delaunay triangulation. *arXiv preprint*. 2020;2011(12786). Available from: <https://arxiv.org/ftp/arxiv/papers/2011/2011.12786.pdf>. DOI: 10.48550/arXiv.2011.12786.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Причко Илья Олегович, эксперт Судебно-экспертного центра Следственного комитета Российской Федерации, аспирант, институт Информационных технологий и анализа данных, Иркутский национальный исследовательский технический университет, Иркутск, Российская Федерация.

e-mail: i@prichko.ru

ORCID: [0000-0001-6261-4102](https://orcid.org/0000-0001-6261-4102)

Афанасьев Александр Диомидович, доктор физико-математических наук, профессор института Информационных технологий и анализа данных, Иркутский национальный исследовательский технический университет, Иркутск, Российская Федерация.

e-mail: aad@istu.edu

ORCID: [0000-0001-8745-1725](https://orcid.org/0000-0001-8745-1725)

Pricha Olegovich Prichko, Expert of the Forensic Expert Center of the Investigative Committee of the Russian Federation, Postgraduate Student, The Institute of Information Technologies and Data Analysis, Irkutsk National Research Technical University, Irkutsk, Russian Federation.

Aleksandr Diomidovich Afanasyev, Doctor of Physical and Mathematical Sciences, Professor at The Institute of Information Technologies and Data Analysis, Irkutsk National Research Technical University, Irkutsk, Russian Federation.

Статья поступила в редакцию 03.01.2023; одобрена после рецензирования 15.02.2023; принята к публикации 03.03.2023.

The article was submitted 03.01.2023; approved after reviewing 15.02.2023; accepted for publication 03.03.2023.