

ПРИНЦИП РАБОТЫ СИСТЕМ РАСПОЗНАВАНИЯ ЛИЦ И ОСОБЕННОСТИ ИХ ПРИМЕНЕНИЯ ПРИ ОБРАБОТКЕ ВИДЕОПОТОКА В РЕАЛЬНОМ ВРЕМЕНИ

OPERATING PRINCIPLE OF FACE RECOGNITION SYSTEMS AND PECULIARITIES OF THEIR APPLICATION IN REAL TIME VIDEO STREAM PROCESSING

**O. Lukyanchikov
M. Trushin
A. Rusakov
A. Lukyanchikov
V. Filatov
S. Dolzhenkov**

Summary: This article discusses the use of human face recognition (identification) technology in real-time frame stream processing systems. In particular, the principles of functioning of face recognition systems are discussed, as well as various features that arise as a result of real-time image processing, and the problems arising from them. In addition, this paper presents several methods for increasing the speed of the face identification algorithm using the example of a recognition system used during video conferencing.

Keywords: face recognition, face identification, real time face recognition, computer vision.

Введение

Компьютерное зрение — стремительно развивающаяся технология, при помощи которой можно создавать системы, автоматизирующие процессы обнаружения, отслеживания и классификации объектов, что делает их незаменимыми в различных областях, от безопасности до медицины. Одна из ведущих групп в области IT-аналитики, компания Tractica, в 2018 году отчиталась о развитии нескольких новых областей применения этих технологий (рис. 1) [1].

Компьютерное зрение тесно пересекается с предметной областью машинного обучения. Использование обученных моделей для решения задач обработки и анализа изображений позволило значительно улучшить результаты по сравнению с уходящими в прошлое

Лукьянчиков Олег Игоревич

Руководитель группы разработки CommuniGate Systems;
к.т.н., доцент, ПТУ МИРЭА
lukoilo@communigate.ru

Трушин Максим Михайлович

ПТУ МИРЭА
maxmtrushin@gmail.com

Русаков Алексей Михайлович

старший преподаватель, ПТУ МИРЭА
rusal@bk.ru

Лукьянчиков Андрей Игоревич

Ведущий инженер-программист АО «НПО Ангстрем»;
Соискатель ПТУ МИРЭА
dronluk@yandex.ru

Филатов Вячеслав Валерьевич

Доцент, ПТУ МИРЭА
filv@mail.ru

Долженков Сергей Сергеевич

Ассистент, ПТУ МИРЭА
dolzhenkov@mirea.ru

Аннотация: В данной статье рассматривается использование технологии распознавания (идентификации) человеческих лиц в системах обработки потока кадров, работающих в режиме реального времени. В частности, обсуждаются принципы функционирования систем распознавания лиц, а также различные особенности, возникающие вследствие обработки изображений в реальном времени, и вытекающие из них проблемы. Кроме того, в данной работе приведены несколько методов повышения скорости работы алгоритма идентификации лиц на примере системы распознавания, использующейся во время видеоконференцсвязи.

Ключевые слова: распознавание лиц, идентификация лиц, распознавание в реальном времени, компьютерное зрение.

статистическими методами и привело к возникновению новых возможностей в области машинного зрения [2–3].

Одной из наиболее востребованных задач компьютерного зрения является распознавание человеческих лиц. Чаще всего данная задача применяется при авторизации пользователей, например, всем известная технология Face ID [4], используемая для разблокировки телефонов.

Благодаря большим наработкам в этой области [5–7] современные алгоритмы могут довольно точно идентифицировать личности по изображению лица, однако существует множество внешних факторов, возникновение которых значительно снижает точность при распознавании. В то же время учет этих особенностей заметно снижает производительность системы, что играет

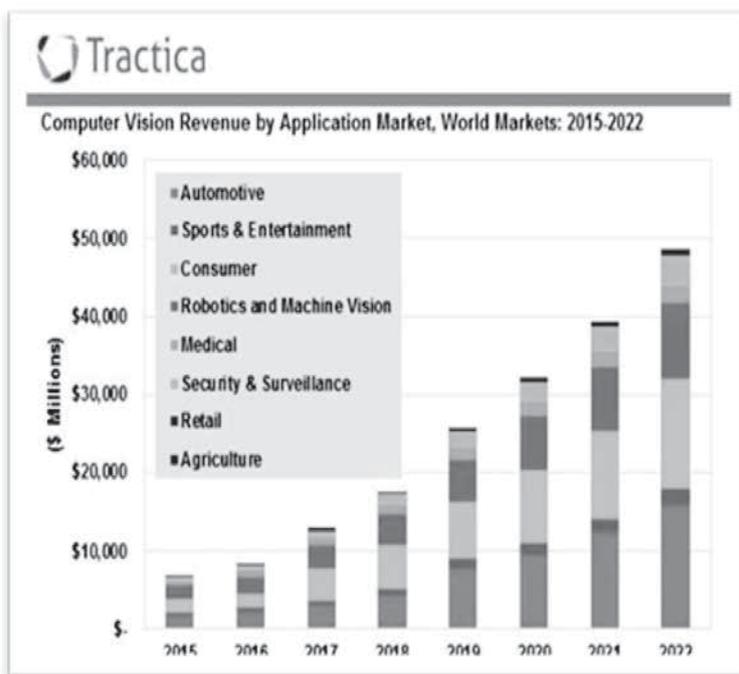


Рис. 1. Области применения технологий компьютерного зрения согласно отчету компании Tractica



Рис. 2. Общая структура алгоритма распознавания лиц

существенную роль при распознавании лиц в системах обработки видеопотока в реальном времени, из-за чего необходимо находить баланс между точностью распознавания и скоростью обработки изображений.

Алгоритм распознавания лиц

В общем виде процесс распознавания представляет из себя последовательность шагов, в ходе которых по представленному изображению или видеофрагменту, содержащему лицо, идентифицируется личность человека (рис.2).

В первую очередь на полученном из видеопотока кадре необходимо определить местоположение лиц. Для решения этой задачи применяются различные алгоритмы детектирования, результатом работы которых является множество координат прямоугольников, обозначающих границы лиц.

На следующем этапе лица вырезаются из общего изображения по полученным координатам, после чего осуществляется их нормализация с целью повышения точности при распознавании. Процесс нормализации

включает в себя различные этапы, необходимость которых определяется степенью устойчивости алгоритма распознавания к воздействию внешних факторов. К основным способам предобработки изображений лиц можно отнести следующие:

1. Уменьшение шума: для уменьшения шума (дефектов изображения в виде хаотично разбросанных пикселей различных цветов) применяются различные фильтры (фильтр Гаусса, медианный фильтр [8] и пр.);
2. Повышение контрастности: повышения контрастности изображений можно добиться при помощи алгоритмов эквализации гистограмм [9], а также адаптивных алгоритмов локальной контрастности.
3. Выравнивание: для выравнивания лица осуществляется поиск ключевых точек (к примеру, углы глаз, губ, кончик носа), после чего при помощи геометрических преобразований выполняется наклон изображения лица в необходимую сторону, в результате которого лицо располагается в анфас.

Этап выделения признаков включает в себя последовательность шагов, в ходе которых для каждого преобразованного изображения лица определяется набор репрезентативных характеристик, при помощи которых можно идентифицировать личность человека. К наиболее распространенным характеристикам относятся:

1. Контуры лица (форма и размеры);
2. Расстояние между ключевыми точками;
3. Относительное положение элементов лица;
4. Уникальные текстуры (родинки, шрамы, рубцы, морщины и пр.);
5. Термограммы (инфракрасные изображения лица).
6. Динамические особенности (движение глаз, изменение формы рта и пр.)

После выделения признаков лица необходимо выявить наличие или отсутствие совпадений на имеющемся множестве признаков. В основе любой системы распознавания присутствует алгоритм, определяющий соответствия между считываемыми характеристиками распознаваемого объекта и заранее заданными шабло-

нами, хранящимися в базе данных. На сегодняшний день существует большое множество алгоритмов, решающих задачу кластеризации, при помощи которых определяется «схожесть» объектов (к-средних, выделение связанных компонент и т.д.). Выбор той или иной меры расстояния, лежащей в основе, а также самого алгоритма во многом зависит от требований к системе (производительность, точность, аккуратность и пр.).

В любой системе идентификации личности по изображению лица вне зависимости от лежащих в ее основе алгоритмов детектирования лиц и выделения их признаков точность распознавания сильно зависит от качества лица на исходном изображении. Под качеством в данном случае подразумевается степень воздействия различных внешних факторов, непосредственно влияющих на результат работы системы распознавания. К таким факторам относятся:

Освещенность лица на изображении [10]. В зависимости от расположения источника света относительно лица, на изображении могут возникать яркие пятна и тени, из-за которых ухудшается точность распознавания (рис. 3). Это обусловлено тем, что при их возникновении могут быть ошибочно определены отдельные признаки, а в худшем случае — лицо вовсе не будет обнаружено.

Искажение изображения. Данная проблема возникает в тех случаях, когда лицо на изображении оказывается смазанным, либо меняется угол обзора, из-за чего может измениться форма лица и его пропорции. В таких случаях система может ошибочно идентифицировать личность человека.

Положение головы относительно камеры [11]. Наклоны и повороты головы на изображении (рис. 4) очень сильно влияют на результат работы системы. Наибольшее влияние на точность распознавания этот фактор оказывает в тех случаях, когда имеющаяся база данных эталонных изображений не является репрезентативной (к примеру, присутствуют только изображения, на которых лицо расположено в анфас).



Рис. 3. Влияние положения источника света на изображение лица



Рис. 4. Вариативность положения головы в пространстве на изображениях при распознавании

Низкое качество изображения. В отдельных случаях на вход системе распознавания могут поступать изображения в низком качестве, возникающем вследствие шума, размытия или низкого разрешения.

Присутствие посторонних объектов на изображении. Снимки людей могут выполняться в различных условиях, к примеру, может изменяться фон изображения, также возможно появление в кадре различных предметов, частично заслоняющих лицо.

Изменение внешнего вида. Возникновение этой проблемы может быть обусловлено старением, изменением причёски, ношением очков, появлением растительности на лице, родинок, шрамов и т.д. Кроме того, к проблеме изменения внешности можно также отнести мимику лица, поскольку при различном эмоциональном состоянии и, как следствие, изменении мимики, пропорции лица могут также поменяться.

Распознавание лиц в реальном времени

Все описанные факторы проявляются в зависимости от условий, в которых осуществляется распознавание. К примеру, если предполагается, что система будет осуществлять идентификацию в идеальных условиях, разумеется, ни одна из перечисленных проблем не возникнет, однако при использовании системы распознавания для обработки потока кадров в реальном времени важно удостовериться, что используемые алгоритмы достаточно устойчивы к внешним факторам.

На сегодняшний день наиболее успешно с задачей распознавания лиц в реальном времени справляются системы, в которых используются методы машинного обучения. Одним из наиболее распространенных методов является использование нейронных сетей, которые обучаются на большом количестве данных о лицах. При обучении используются наборы данных, содержащие изображения лиц людей из разных возрастных, этнических групп и рас, благодаря чему нейронные сети способны достигать высокой точности даже в условиях изменения

образа человека или ухудшении качества изображения, обеспечивая в то же время сравнительно высокую производительность. Эти нейронные сети обычно имеют несколько слоев, которые могут обрабатывать различные аспекты изображения лица. Например, первый слой может распознавать форму и размер лица, второй слой — расположение глаз, носа и рта, а третий слой — цвет кожи. В результате получается точное представление лица, которое затем используется для идентификации человека.

Одним из примеров нейронных сетей, решающих задачу распознавания человеческих лиц, является MobileFaceNet. Она была разработана в 2018 году и представлена в статье [12] авторами из компании Tencent. Данная сеть была спроектирована специально для создания быстрых и точных систем распознавания, способных работать на устройствах с ограниченными вычислительными ресурсами. Архитектура данной нейронной сети основана на сверточных слоях, сравнительно небольшой размер модели и высокая производительность достигается благодаря использованию нескольких техник оптимизации:

1. Квантизация весов (weights quantization). Веса модели квантизированы до 8-битных чисел. Квантизация — особая техника сжатия, благодаря которой достигается уменьшение количества бит, затрачиваемых для представления чисел.
2. Обрезка (pruning). Из модели удалены сверточные фильтры, незначительно влияющие на итоговую точность распознавания.
3. Кластеризация (clustering). Веса модели кластеризованы. Вместо того, чтобы использовать уникальные значения для каждого веса, модель использует среднее значение весов, которые были объединены в кластеры.
4. Сжатие весов (weights compression). Веса модели сжаты с использованием метода декомпозиции, который разбивает веса на несколько небольших матриц. В результате во время работы модели осуществляется меньшее количество операций матричного умножения.

MobileFaceNet была протестирована на нескольких популярных наборах данных, таких как LFW [13] и AgeDB [14], и показала очень впечатляющие результаты (99.28 % и 93.05 % соответственно), сравнимые с результатами нейронных сетей, обладающих более сложной архитектурой, но в то же время работающих более медленно.

В 2022 году авторами из института распознавания образов, Китайской академии наук была опубликована работа [15], в которой они представили функцию потерь SFace для обучения нейронных сетей, решающих задачу распознавания лиц. Основная идея данной функции потерь заключается в минимизации расстояния внутри одного класса признаков и в то же время максимизации расстояния между различными классами. Одновременно с этим SFace учитывает возможность присутствия в обучающем наборе низкокачественных изображений, влияющих на дальнейшую точность при распознавании, благодаря чему достигается баланс между уменьшением расстояния внутри класса при корректных изображениях и предотвращением переобучения при возникновении различного рода шумов.

Множественные сравнения моделей, обученных с использованием различных функций потерь, показали превосходство SFace на всевозможных наборах данных. К примеру, модель ResNet50 [16], обученная с использованием SFace, продемонстрировала результат в 99.57 % процентов точности на наборе данных LFW и 95.30 % на наборе AgeDB, что является лучшим результатом среди всех функций, участвовавших в эксперименте.

SFace состоит из двух компонентов (рис. 5):

Гиперсферный компонент, который описывает геометрическую структуру пространства признаков. Цель этого компонента заключается в том, чтобы убедиться, что вектора признаков находятся на гиперсфере фиксированного радиуса.

Сигмоидный компонент, который осуществляет управление масштабированием признаков. Данный

компонент контролирует размер шкалы между классами и регулирует расстояние в пространстве признаков.

Произведем оценку скорости работы системы распознавания лиц путем замера времени, затрачиваемого системой для выделения признаков каждого лица и поиска соответствий в базе данных эталонов. Для оценки производительности будет использовано приложение, реализованное средствами библиотеки OpenCV [17]. Задачу выделения признаков будет решать модель MobileFaceNet, обученная с использованием функции потерь SFace. В ходе каждого теста системе на вход будет циклично подаваться одно и то же изображение разрешением 640x480 пикселей, содержащие одно лицо, пять и десять лиц соответственно. Для каждого из изображений будет осуществлено 3 эксперимента, в которых поиск совпадений будет осуществляться в базе данных из 100, 1000 и 10000 изображений, при этом среди этих изображений будут равномерно расположены эталонные для распознаваемых лиц. Все вычисления выполняются на процессоре Ryzen 5 3600x на частоте 4.0 ГГц (табл. 1).

Таблица 1.

Результаты тестирования производительности системы распознавания лиц

Количество распознаваемых лиц	Количество записей в БД	Время на первом кадре, ms	Время на последующих кадрах, ms
1	100	26.2406	26.6412
	1000	30.4283	30.6011
	10000	60.6948	60.8152
5	100	130.8841	130.6395
	1000	141.6522	140.9604
	10000	233.4025	233.0442
10	100	260.1924	260.2259
	1000	275.5398	277.8126
	10000	446.4057	446.6831

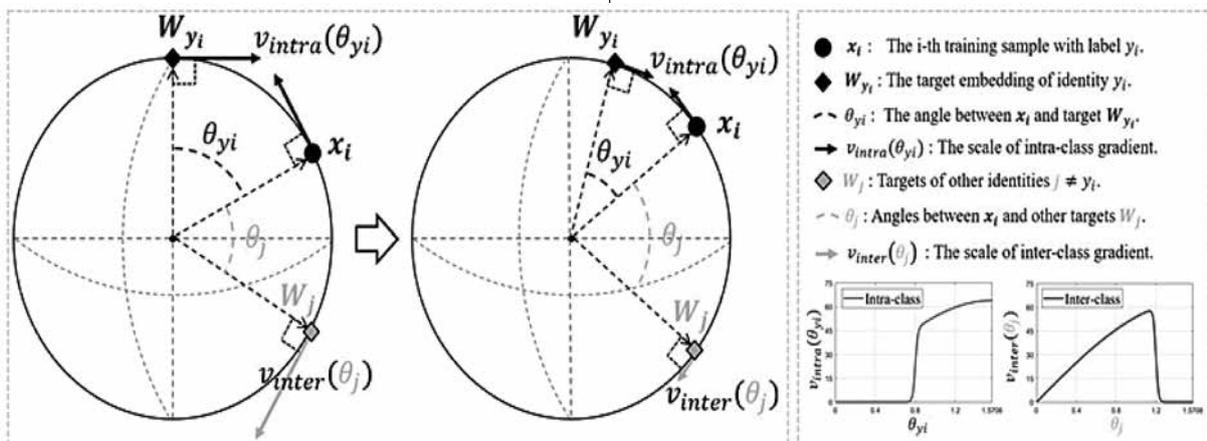


Рис. 5. Схематичное представление функции потерь SFace

Методы ускорения алгоритма распознавания лиц для видеопотока

Исходя из полученных в ходе эксперимента результатов, можно сделать вывод, что главным недостатком данной системы распознавания является высокая асимптотическая сложность алгоритма идентификации лиц, которая имеет вид $m \times n$, где m — количество распознаваемых лиц, n — количество записей в базе данных. Иными словами — время работы алгоритма зависит одновременно и от количества лиц в текущем кадре, которые необходимо распознать, и от размера хранилища, в котором производится поиск соответствий. Для решения данной проблемы и, как следствие, повышения производительности системы можно использовать несколько способов.

В первую очередь для ускорения процесса распознавания необходимо обратить внимание на оптимизацию хранения данных. В текущей реализации системы вычисление векторов признаков на каждой итерации выполняется не только для лиц, идентификацию которых необходимо осуществить, но и для всех лиц, хранящихся в базе данных, в результате чего общее время, затрачиваемое на распознавание, кратно возрастает. Оптимальная стратегия, при которой удастся избежать данной проблемы, заключается в хранении данных в виде заранее посчитанных при помощи модели значений, которые в дальнейшем будут использоваться напрямую при решении задачи поиска соответствий.

Еще один способ предполагает уменьшение объема обрабатываемых данных. В настоящий момент система

распознавания осуществляет поиск соответствий во всей базе данных для каждого лица на изображении. В худшем случае эталоны людей, находящихся в кадре, могут располагаться в самом конце хранилища, из-за чего при каждом распознавании система будет производить количество сравнений равное общему числу записей в базе данных. Очень часто в кадре могут находиться одни и те же люди продолжительный период времени, в связи с чем нет необходимости искать их эталоны во всем хранилище. Наиболее корректным способом в такой ситуации будет хранение эталонов недавно идентифицированных людей в промежуточном хранилище меньшего размера. В качестве такого хранилища может выступать LRU (Least Recently Used) [18] кэш. Принцип работы данного алгоритма кэширования (рис. 6) довольно прост. Временное хранилище ограничивается размером, зависящим от размера базы данных (к примеру, при нахождении в базе данных 1000 человек, размер временного хранилища может составлять 50). При его заполнении происходит вытеснение того значения, которое дольше всего не использовалось (находится в конце). Таким образом, эталон вновь распознанного лица будет заноситься в кэш, после чего при следующем распознавании система сперва осуществит поиск в нем, а дальше в случае отсутствия совпадений обратиться к основному хранилищу.

Третий способ, при помощи которого можно добиться кратного повышения быстродействия — параллельные вычисления. При параллельных вычислениях крупная задача (в данном случае — распознавание всех лиц в кадре) может быть разбита на ряд более мелких задач (распознавание отдельного лица), которые впослед-

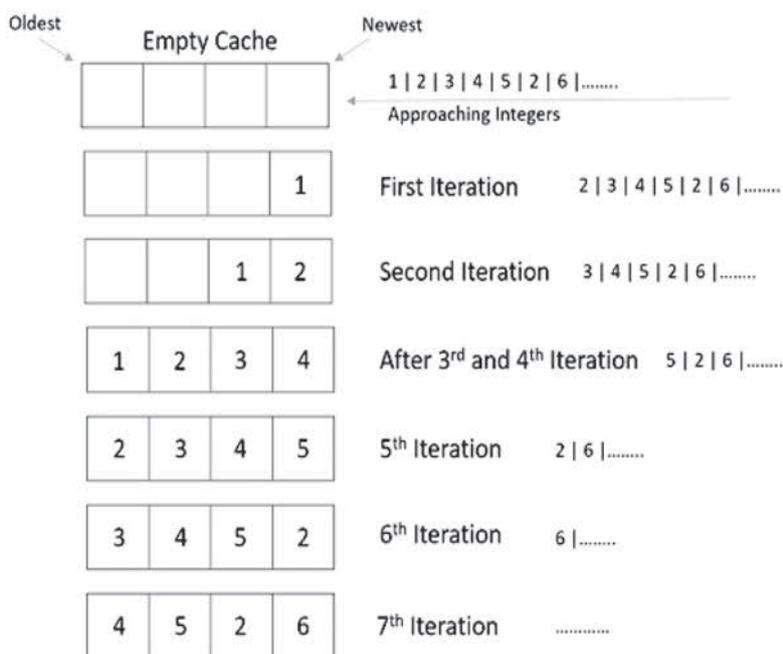


Рис. 6. Принцип работы алгоритма LRU

ствии будут выполнены одновременно. Архитектура современных процессоров позволяет производить параллельные вычисления на нескольких ядрах в составе одного процессора, благодаря чему нет необходимости в использовании многопроцессорных систем для решения данной задачи. Разумеется, вычислительные мощности каждого процессора ограничены, из чего не получится распознавать одновременно бесконечно большое число лиц, однако, даже при распознавании двух лиц одновременно, прирост производительности системы может составлять до 50 %.

Произведем повторную оценку скорости работы системы распознавания лиц (табл. 2). Условия, в которых будет осуществляться тестирование системы, полностью аналогичны тем, что использовались в ходе первой серии экспериментов. Единственным различием является использование ускоренной системы распознавания, в которой были использованы методики повышения производительности, описанные ранее.

Таблица 2.

Результаты тестирования производительности
улучшенной системы распознавания лиц

Количество распознаваемых лиц	Количество записей в БД	Время на первом кадре, ms	Время на последующих кадрах, ms
1	100	26.7093	25.9772
	1000	30.6895	25.9717
	10000	59.8840	25.8811
5	100	41.0536	34.0683
	1000	49.0056	34.1181
	10000	68.3924	34.3661
10	100	82.9249	65.0563
	1000	92.2416	65.5662
	10000	111.8773	65.2792

Из результатов второй серии экспериментов в первую очередь можно заметить, что появилась существенная разница между временем работы алгоритма распознавания на первом кадре и на последующих, достигнутая в результате кэширования векторов признаков идентифицированных лиц. Кроме того, наблюдается значительный прирост скорости распознавания во всех экспериментах, обусловленный параллельными вычислениями и оптимальной стратегией хранения эталонов. Применение методов ускорения позволило свести к минимуму зависимость времени работы алгоритма от количества распознаваемых лиц, благодаря чему данная система стала осуществлять идентификацию лиц в видеозвонках без значительных задержек. Отклонения в результатах тестирования системы обусловлены синхронизацией работы потоков, а также дополнительными накладными расходами на операции создания потоков и пр.

Заключение

Таким образом, современные методы распознавания лиц, представленные различными обученными моделями, демонстрируют высокую точность в условиях воздействия внешних факторов, обеспечивая в то же время сравнительно высокую производительность. В то же время скорости работы таких алгоритмов недостаточно для комфортного использования системы распознавания при обработке видеопотока в реальном времени. Поэтому были применены методы методики ускорения работы алгоритма, которые значительно уменьшили задержки, что в свою очередь позволило получить приемлемое время работы алгоритма, достаточное для распознавания лиц в ходе видеозвонка. Однако, несмотря на положительные результаты увеличения производительности алгоритма идентификации лиц, все еще существуют проблемы, связанные с недостаточным быстродействием, которые можно решить другими способами, например, посредством предварительного выбора идеального кадра из видеопотока, подходящего для распознавания. Поэтому дальнейшее совершенствование данной технологии является актуальной задачей.

ЛИТЕРАТУРА

1. «Компьютерное зрение: технологии, рынок, перспективы», отчет TAdviser, 2018.
2. Voulodimos A. et al. Deep learning for computer vision: A brief review // Computational intelligence and neuroscience. — 2018. — Т. 2018.
3. Khan A.A., Laghari A.A., Awan S.A. Machine learning in computer vision: a review // EAI Endorsed Transactions on Scalable Information Systems. — 2021. — Т. 8. — №. 32. — С. e4-e4.
4. Mainenti D. User perceptions of apple's face id // Information Science, Human Computer Interaction (DIS805). — 2017.
5. Osadchy M. et al. Scifi-a system for secure face identification // 2010 IEEE Symposium on Security and Privacy. — IEEE, 2010. — С. 239–254.
6. Samaria F., Young S. HMM-based architecture for face identification // Image and vision computing. — 1994. — Т. 12. — №. 8. — С. 537–543.
7. Кузнецов Р.А. и др. Технология «компьютерного зрения» в вопросе визуальной идентификации человека // Научная визуализация. — 2017. — Т. 9. — №. 1. — С. 124.
8. Лавров Д.Н., Назарова Т.Ю. Алгоритмы предварительной обработки изображения для поиска границ радужной оболочки глаза // Молодёжь третьего тысячелетия. — 2017. — С. 828–832.

9. Бекназарова С.С. и др. Алгоритмы эквализации гистограммы изображения // Finland International Scientific Journal of Education, Social Science & Humanities. — 2023. — Т. 11. — №. 3. — С. 93–97.
10. Ho J., Kriegman D. On the effect of illumination and face recognition // Face Processing: Advanced Modeling and Methods. — 2005.
11. Gros C., Straub J. Video Recordings of Male Face and Neck Movements for Facial Recognition and Other Purposes // Data. — 2019. — Т. 4. — №. 3. — С. 130.
12. Chen S. et al. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices // Biometric Recognition: 13th Chinese Conference, CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings 13. — Springer International Publishing, 2018. — С. 428–438.
13. G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, Tech. rep., Technical Report 07-49, University of Massachusetts, Amherst (2007).
14. Moschoglou S. et al. Agedb: the first manually collected, in-the-wild age database // proceedings of the IEEE conference on computer vision and pattern recognition workshops. — 2017. — С. 51–59.
15. Zhong Y. et al. SFace: Sigmoid-constrained hypersphere loss for robust face recognition // IEEE Transactions on Image Processing. — 2021. — Т. 30. — С. 2587–2598.
16. He K. et al. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2016. — С. 770–778.
17. Bradski, G. (2000). The OpenCV Library. Dr. Dobbs & Co. Journal of Software Tools.
18. Fricker C., Robert P., Roberts J. A versatile and accurate approximation for LRU cache performance // 2012 24th international teletraffic congress (ITC 24). — IEEE, 2012. — С. 1–8.

© Лукьянчиков Олег Игоревич (lukoilo@communicate.ru), Трушин Максим Михайлович (maxmtrushin@gmail.com);
Русаков Алексей Михайлович (rusal@bk.ru); Лукьянчиков Андрей Игоревич (dronluk@yandex.ru); Филатов Вячеслав Валерьевич (filv@mail.ru);
Долженков Сергей Сергеевич (dolzhenkov@mirea.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»