

## ОЦЕНКА МЕТОДОВ АНАЛИЗА ЛИЦЕВЫХ ХАРАКТЕРИСТИК ДЛЯ ПЕРСОНАЛИЗАЦИИ В СИСТЕМАХ МОНИТОРИНГА ВОДИТЕЛЯ

### EVALUATION OF FACIAL ANALYSIS METHODS FOR PERSONALIZATION IN DRIVER MONITORING SYSTEMS

*Pham T. A.  
M. Kashevnik  
A. Chechulin*

*Summary.* Recently, research and development on the topic of driver monitoring systems have become more and more popular. Since self-driving cars remain the long perspective, the research & development community in the field of intelligent transportation concentrated on driver monitoring in-vehicle cabin to reduce the number of traffic accidents. The paper presents an analysis of the relevant work and reviews modern methods for identifying potentially distracting driver situations based on the analysis of images received from the front camera of the smartphone. The study aims to answer the question: how to set up methods for analyzing facial images for different types of faces (for example, in Europe or Asia) to personalize the system and increase the accuracy of facial features detection. Progress in the field of deep learning is widely used in modern identification systems and can also be applied to the problem under consideration. The paper assesses the advantages and disadvantages of the existing machine learning methods in relation to the problem under consideration.

*Keywords:* Face analysis, personification, deep learning, artificial neural networks, convolutional neural networks.

**Фам Туан Ань**

Университет ИТМО

anhpt@itmo.ru

**Кашевник Алексей Михайлович**

СПИИРАН

alexey.kashevnik@iias.spb.su

**Чечулин Андрей Алексеевич**

СПИИРАН

chchulin@comsec.spb.ru

*Аннотация.* В последнее время исследования и разработки по тематике систем мониторинга водителей становятся все более и более популярными. Поскольку появление автономных транспортных средств на дорогах общего пользования невозможно в ближайшие годы, сообщество исследователей и разработчиков в области интеллектуальных транспортных систем сосредоточилось на мониторинге водителя в салоне транспортного средства для уменьшения количества дорожно-транспортных происшествий. В статье представлен анализ современных работ в области определения лицевых характеристик с использованием анализа изображений лица водителя. Целью исследования является ответ на вопрос, каким образом настроить методы анализа изображений для разных типов лиц (например, в европейских или азиатских), для того чтобы персонализировать систему и повысить точность определения черт лица. Прогресс в области глубокого обучения широко используется в современных системах идентификации и также может быть применен к рассматриваемой проблеме. В статье дается оценка преимуществ и недостатков существующих методов машинного обучения по отношению к рассматриваемой проблеме.

*Ключевые слова:* Анализ лицевых характеристик, персонификация, глубокое обучение, искусственные нейронные сети, сверточные нейронные сети.

### Введение

**И**сследования и разработки по тематике систем мониторинга водителей в последние годы становятся все более и более популярными. В ноябре 2019 года Европейский союз объявил об обязательном введении систем контроля водителя для каждого транспортного средства [1]. Многие компании сосредоточены на разработке аппаратного и программного обеспечения для определения сонливости или отвлечения внимания водителя. Одним из эффективных методов выявления сонливости и отвлечения внимания является анализа лицевых характеристик водителя, путем анализа непрерывной информации с фронтальной камеры смартфона для наблюдения за глазами водителя. При этом алгоритмы распознавания изображений используются для обнаружения характеристик лица [2].

Исследования анализу изображений лиц ведутся с 1960-х годов, однако в этой области осталось много нерешенных вопросов. В начале 1990-х и 2000-х годов исследования по анализу лиц стали популярными в основных инженерных направлениях, таких как линейное подпространство, разнообразие и разреженное представление. В начале 2010-х годов многие исследователи в области анализа лицевых характеристик представили такие методы, как локальные описания, локальные фильтры для сокращения программ и признание точнее распознавания. В 2012 году, когда компания AlexNet представила методику глубокого обучения, он оказал значительное влияние на область компьютерного зрения.

Для определения состояния небезопасного вождения с использованием камеры, направленной на лицо

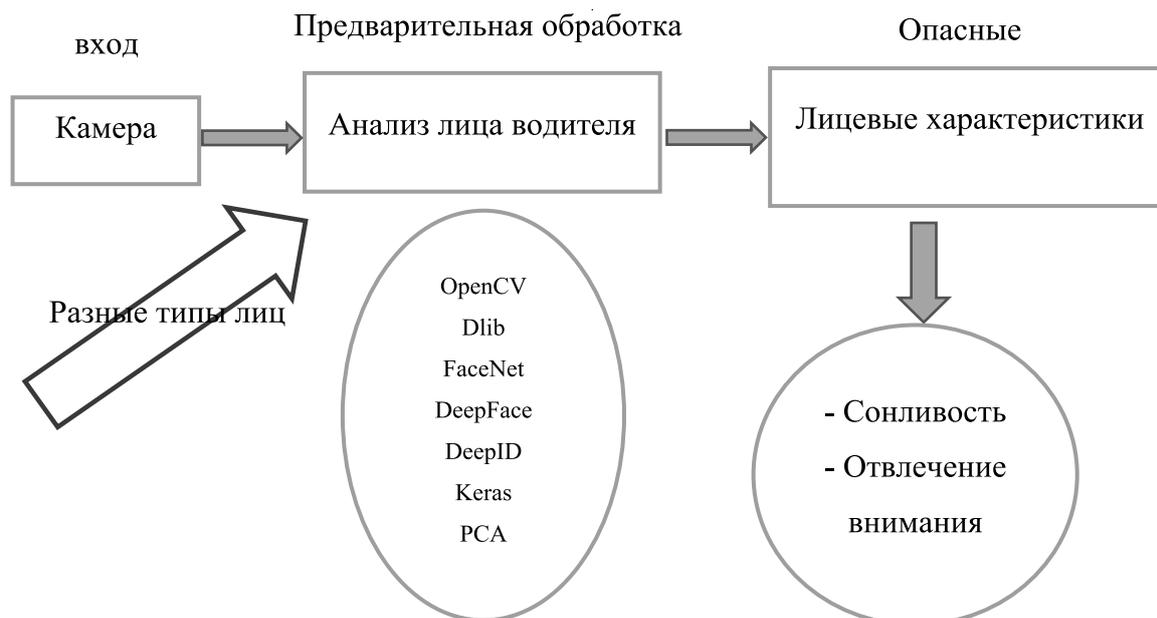


Рис. 1. Модель использования смартфона для выявления опасных состояний.

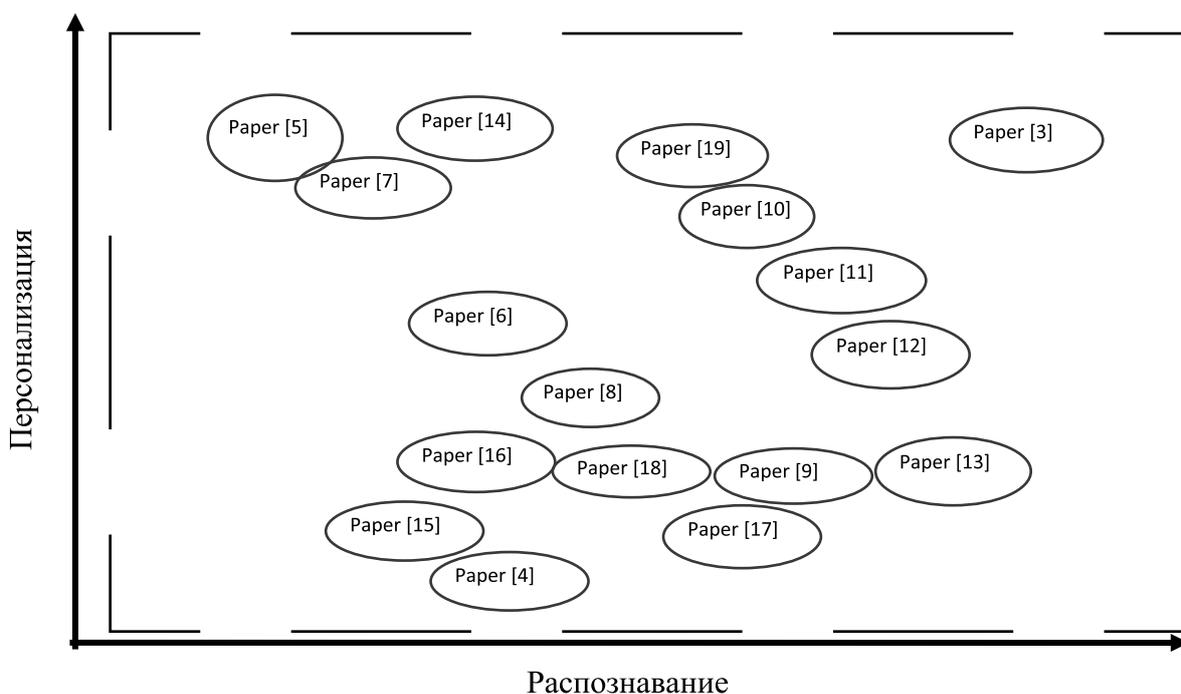


Рис. 2. Классификация документов по распознаванию объектов на лицах

водителя, в системе мониторинга необходимо проанализировать и идентифицировать лицевые характеристики водителя (глаза, нос, рот, уши) [2]. В данной статье анализируются методы идентификации лицевых характеристик по изображению с камеры применительно к разным типам лиц в разном контексте (Рисунок 1). В статье рассматриваются преимущества и недостатки популярных методов анализа лицевых характеристик,

а также влияния расовых признаков и черт лица на такое распознавание с целью его персонализации в дальнейшем.

Статья структурирована следующим образом. В разделе II описываются работы, связанные с темой анализа лиц, сравниваются преимущества и недостатки рассматриваемых методов, указываются и обсуждаются

технические вопросы, требования и спецификации для персонализации анализа лиц. В разделе III представлены результаты экспериментов с европейскими и азиатскими изображениями лиц с использованием рассмотренных методов. В заключение подведены итоги работы.

## Обзор современных исследований

Технология идентификации объектов на лице человека имеет достаточно долгую историю. Такие всемирно известные платформы открытый исходный код, как OpenCV и DLib, предоставляют комплексные инструменты для идентификации объектов на лице человека. Алгоритмы и методы идентификации делятся на большие группы: классические алгоритмы распознавания лиц; искусственные нейронные сети; вейвлеты Габора; методы, основанные на дескрипторе лица; 3D-распознавание лица. Проблемы, которые затрудняют идентификацию лиц, включают в себя: разные положения лиц, движения головы, разницу в освещенности.

Авторы исследовали и проанализировали научные статьи о методах распознавания объектов на лице водителя, которые были классифицированы с точки зрения качества распознавания и степени персонализации (см. рис. 2).

В статье [4] авторы представили исследование по распознаванию эскизов лица (нарисованных изображений). Предложено выделить 66 точек на лице, которые разделяют лицо на сегменты. Из полученных сегментов выделяются волосы, брови, глаза, нос и рот.

Статья [5] анализирует применение Hidden Markov model (HMM) для определения направления взгляда. Авторы показали, что во время процесса анализа лицевых характеристик, различные области лица получают разные уровни внимания. Принято считать, что глаза, нос и рот являются чертами лица, на которых чаще всего фокусируется внимание.

В статье [6] предложено делить методы распознавания лицевых характеристик на три категории: общие алгоритмы, 2D-методы и 3D-методы. Подход Eigenfaces представляется быстрым, простым и практичным методом, он стал наиболее широко используемым методом распознавания лицевых характеристик.

В статье [7] говорится о влиянии подмигивания на алгоритм распознавания лиц. Тестируемый алгоритм распознавания лиц представляет собой слияние анализа основных компонентов (Principal Component Analysis — PCA), извлекающего функции на основе локальных двоичных шаблонов (Local Binary Patterns — LBP) и клас-

сификации на основе поддержки (Support-Based Classification — SVM). Целью исследования является не улучшение работы алгоритма распознавания лиц, а тестирование производительности во время подмигивания.

Методы, используемые в настоящее время для распознавания лиц, обсуждаются в статье [8], важным используемым методом является метод использования инфракрасного спектра (Infrared Spectrum — IRS) для идентификации лиц. В статье [9] представлен метод анализа главных компонент (PCA), который состоит из 4 этапов: позиционирование лица и предварительная обработка изображения; стандартное разложение PCA; Local PCA (LPCA), сложная реализация PCA; Последний шаг — переоценка значений в векторе результатов, чтобы подчеркнуть уникальные признаки. Предложенные альтернативные преобразования включают в себя длины волн Габора, локальные двоичные структуры и корреляционные фильтры.

Авторы статьи [10] упомянули метод распознавания по локальным двоичным структурам LBP. В рамках этого метода изображение лица делится на набор областей. Локальные таблицы распределения двоичных объектов извлекаются из этих наборов. Затем таблицы объединяются в расширенный векторный элемент, который используется в качестве описания лица. Результаты ясно показывают, что изображения лица можно рассматривать как компонент микро-моделей, таких как плоскости, пятна, линии и края, которые могут быть хорошо описаны локальными двоичными структурами.

Статья [11] направлена на выявление черт лица (положение глаз, рта, центра глаза и угла рта) на основе геометрической информации с использованием векторных полей расстояния. Распознавание лица выполняется с использованием сегментации цвета кожи в цветовом пространстве YIQ — используется в видеосистемах, после чего следуют этапы точной настройки на основе генетических алгоритмов.

Авторы статьи [12] предложили последовательность алгоритмов с возможностью идентификации лица с более высокой точностью, включающих в себя комбинирование SVM, PCA, LBP и нейронных сетей различными методами. Из результатов становится ясно, что классификация на основе SVM для операторов LBP и извлечение признаков на основе PCA обеспечивают более высокую точность по сравнению с классификацией на основе прямой передачи с обратным распространением (Feed Forward Back Propagation Neural Network — FFBPNN), сохраняя при этом изменение характеристик извлечения модели. Кроме того, в случае SVM трехмерные изображения лица обеспечивают лучшую классификацию бла-

Таблица 1. Обзор методов распознавания лиц

№	Метод	Преимущества	Недостатки
1	Традиционные алгоритмы распознавания лиц [14], [15]	Алгоритмы ориентированы на локальную структуру многообразия. Они проецируют лицо на линейное подпространство, охватываемое изображениями собственного лица. Расстояние от пространства лица является ортогональным к плоскости среднего изображения, поэтому может быть легко преобразовано в расстояния Махаланобиса с вероятностной интерпретацией.	Методы не могут адекватно осуществлять распознавание объектов лиц, при увеличении шумов, например лохом освещении. Подпространственные подходы применяются в методах на основе Discriminant common vectors (DCV) и support vector machine (SVM). Сохранение локальной структуры между образцами является областью методов Neighbourhood Preserving Projection (NPP) и Orthogonal Neighborhood Preserving Projection (ONPP).
2	Искусственные нейронные сети (Artificial neural networks — ANN) [16], [17]	Этот метод представляет также другие подходы для упрощения процесса, касающегося собственной функции линеаризации ANN и ускорения вычислений. Он является идеальным решением, особенно для распознавания изображений лица с частичным искажением и окклюзией.	Основным недостатком данного подхода является требование большого количества обучающих образцов (вместо одного образца или их ограниченного количества).
3	Вейвлеты Габора [17], [18]	Вейвлеты Габора демонстрируют желательные характеристики захвата характерных визуальных свойств, таких как селективность ориентации пространственной локализации и пространственная частота.	Недостатком методов, основанных на Габоре, является значительно большая размерность пространства признаков Габора. Подход требует большого объема вычислений и нецелесообразен для приложений реального времени.
4	Методы, основанные на дескрипторе лица [3], [19]	Основная идея заключается в изучении наиболее различимых локальных особенностей, которые минимизируют разницу между изображениями одного и того же человека и максимизируют разницу между изображениями других людей. Эти методы являются дискриминационными и устойчивыми к изменениям освещения и выражения лица. Они предлагают компактный, легкий для извлечения и очень различительный дескриптор.	Подход требует значительных вычислительных ресурсов на этапе извлечения дескриптора, но поощряет простоту и производительность по сравнению с онлайн-приложениями.
5	3D распознавание лиц [20], [21]	Расширение традиционного процесса 2D захвата и имеет большой потенциал для точности. Информация о глубине не зависит от позы и освещения, что делает решение более надежным	3D-системы распознавания требуют калибровку и синхронизацию с существующими 2D-данными. Эта система обычно имеет время вычислений и большую задержку

годаря способности трехмерных изображений сохранять информацию.

В работе [13] проведены эксперименты по определению влияния пола и возраста на результаты идентификации. Результаты показывают, что точность распознавания лица у мужчин выше, чем у женщин, но анализ специфичности у мужчин ниже, чем у женщин. Причиной этого результата может являться то, что черты мужского лица значительно крупнее женских. А также, женщины наносят макияж на лицо, что тоже отрицательно сказывается на распознавании, поэтому степень распознавания мужского лица выше чем женского, но женские особенности лица выделяются сильнее чем мужские. При этом влияние возраста у мужчин намного выше, чем на женщин.

### Сравнение исследований

Предложено разделять методы распознавания объектов на лице на следующие основные группы: Традиционные алгоритмы распознавания лиц, искусственные нейронные сети, вейвлеты Габора, методы на основе дескрипторов лиц, 3D-распознавание лиц (Таблица 1).

### Технические требования

Из сравнений и анализа вышеупомянутых методов распознавания лицевых характеристик видно, что система глубокого обучения, использующая искусственные нейронные сети, является эффективным и передовым решением в области распознавания лиц [22]. Однако это решение по-прежнему сталкивается с проблемами, ко-

Таблица 2. Сравнение производительности

Лицо	Традиционные алгоритмы распознавание лица (время, сек)	Искусственная нейронная сеть (время, сек)	3D распознавание лиц (время, сек)
Азиатский анфас	0,157	0,195	2,684
Азиатский повернуто	0,113	0,188	2,491
Европейский анфас	0,264	0,383	2,567
Европейский повернуто	0,158	0,287	2,512

Таблица 3. Точность сравнения

Лицо	Традиционные алгоритмы распознавание лица (точность,%)	Искусственная нейронная сеть (точность,%)	3D распознавание лиц (точность,%)
Азиатский анфас	40	80	98
Азиатский повернуто	24	92	96
Европейский анфас	56	100	98
Европейский повернуто	40	92	95

торые приводят к снижению производительности распознавания. Поэтому, мы обсудим требования, которые необходимо учитывать при определении лицевых характеристик для решения этой проблемы:

- ◆ **Качество изображения:** Качество изображения влияет на эффективность работы алгоритмов распознавания лиц. В зависимости от задачи качество изображения может играть решающую роль при распознавании лицевых характеристик. Качество изображения при сканировании видео обычно ниже по сравнению с фотографией.
- ◆ **Размер изображения:** когда алгоритм обнаружения лица детектирует лицо на изображении, относительный размер этого лица по сравнению с зарегистрированным размером изображения влияет. Малый размер изображения означает, что обнаруженное лицо имеет низкое разрешение. Кроме того, необходимость сканировать изображение на предмет изменения размера лица является трудозатратной задачей с точки зрения ресурсов.
- ◆ **Угол лица:** относительный угол поворота и наклона лица сильно влияет на оценку распознавании лицевых характеристик. Сочетание решений 3D-идентификации и расширения хранилища данных достаточно ресурсозатратно, чтобы быть хорошим вариантом и хорошим решением для этой проблемы.
- ◆ **Обработка и хранение:** несмотря на то, что видео высокой четкости имеет довольно низкое разрешение по сравнению с изображениями с цифровых камер, оно все равно занимает значительные объемы дискового пространства. Обработка

каждого кадра видео является огромной задачей, поэтому обычно только часть (от 10 до 25 процентов) фактически проходит через систему распознавания. Чтобы минимизировать общее время обработки, могут быть использованы кластеры компьютеров. Однако добавление компьютеров включает в себя значительные затраты на передачу данных по сети, что также ограничивает скорость обработки лица.

- ◆ **Данные для обучения:** Deep neural network (DNN) зарекомендовала себя как доминирующий метод в машинном обучении. Combined neural networks (CNN) использовались почти во всех наиболее эффективных методах распознавания. После обучения, такие системы способны к глубокому анализу, решая самые сложные проблемы, которые все еще влекут за собой другие методы. Однако все эти методы подчеркивают необходимость в более широком и более сложном наборе общедоступных данных для оценки этих систем. Глубокие нейронные сети требуют больших объемов данных, предпочтительно десятков миллионов изображений. Поэтому часто в распоряжении разработчика просто не имеется достаточного объема данных для современных систем глубокого обучения.
- ◆ **Отсутствие разнообразия и плохое обобщение:** Существующие наборы данных, такие как база данных Yale Face, были получены в строго контролируемых лабораторных условиях. Наборы данных, такие как LFW и Cross-Age Celebrity Dataset (CACD), утверждают, что они получены «в дикой природе», но принимаются почти исключительно профессиональными фотографами

с цифровыми однообъективными зеркальными камерами, есть хороших освещенных условиях. Обучение таким наборам данных, вероятно, приведет к плохой ошибке обобщения, когда модели столкнутся с менее ограниченной операционной средой, такой как фотопоток с мобильного телефона или камеры низкого качества.

## Методы оценки

Чтобы ответить на вопрос: «Как адаптировать метод распознавания лицевых характеристик для разных типов лиц (например, европейские или азиатские) для обогащения персонализации в процессе распознавания лиц?» Авторы оценили классический алгоритм распознавания лиц, искусственную нейронную сеть и 3D распознавания лиц (см. (1) и (2) в таблице 1).

Авторы сформировали набор данных, который содержит 100 изображений, 50 азиатских лиц и 50 европейских, включая 25 изображений лиц в анфас и 25 изображений лиц в повернуто. Фотографии были взяты из доступных наборов данных. Были выполнены тесты, используя традиционные алгоритмы распознавания лиц, метод искусственной нейронной сети и 3D распознавания лиц. Измерена эффективность распознавания лиц (см. таблицу 2), а также точность распознавания (см. таблицу 3). Результаты исследований показывают, что традиционные алгоритмы имеют быструю работы, чем искусственные нейронные сети, однако, уровень точности намного ниже.

Для метода 3D-распознавания существует стабильная и высокая точность, но существует большая задержка. Точность распознавания европейских лиц часто немного выше, чем у азиатских лиц, в частности, результаты поиска с европейскими анфасами с искусственными нейронными сетями очень точны и стабильны.

## Заключение

В статье представлен анализ и спецификация требований к системам распознавания лицевых характеристик. Глубокое обучение было интегрировано в большинство современных моделей и методов распознавания лицевых характеристик. Это изменение привело к значительному повышению точности систем распознавания лицевых характеристик и стало действующим стандартом онлайн-стандартов распознавания лицевых характеристик. Модели глубокого обучения в сочетании с нейронами добились замечательных достижений, значительно повысив точность и скорость выполнения. Степень зависимости от людей, а также влияние человека на систему уменьшается. Системы постепенно становятся умными и достигают той же производительности, что и люди. В настоящее время существует множество наборов данных лиц различных организаций, как государственных, так и частных, которые можно использовать для исследования методов распознавания визуальных сигналов. Для обучения современных систем машинного обучения набор данных должен быть большого масштаба (десятки миллионов изображений).

## ЛИТЕРАТУРА

1. Regulation (EU) 2019/2144 of the European Parliament and of the Council of 27 November 2019.
2. Kashevnik, A., Lashkov, I., & Gurtov, A. (2019). Methodology and Mobile Application for Driver Behavior Analysis and Accident Prevention. *IEEE Transactions on Intelligent Transportation Systems*.
3. Abudarham, N., Shkiller, L., & Yovel, G. (2019). Critical features for face recognition. *Cognition*, 182, 73–83.
4. Liu, D., Li, J., Wang, N., Peng, C., & Gao, X. (2018). Composite components-based face sketch recognition. *Neurocomputing*, 302, 46–54.
5. Chuk, T., Crookes, K., Hayward, W. G., Chan, A. B., & Hsiao, J. H. (2017). Hidden Markov model analysis reveals the advantage of analytic eye movement patterns in face recognition across cultures. *Cognition*, 102–117.
6. Zhang, X., & Gao, Y. (2009). Face recognition across pose: A review. *Pattern Recognition*, 42(11), 2876–2896.
7. Shankar, S. (2015). Assessment of the effect of variations in Eye Blinks on a face recognition algorithm. In 2015 IEEE International Advance Computing Conference (pp. 371–376). IEEE.
8. Singh, S., & Prasad, S.V.A. V. (2018). Techniques and Challenges of Face Recognition: A Critical Review. *Procedia computer science*, 143, 536–543.
9. Givens, G. H., Beveridge, J. R., Phillips, P. J., Draper, B., & Bolme, D. (2013). Introduction to face recognition and evaluation of algorithm performance. *Computational Statistics & Data Analysis*, 67, 236–247.
10. Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12), 2037–2041.
11. Asteriadis, S., Nikolaidis, N., & Pitas, I. (2009). Facial feature detection using distance vector fields. *Pattern Recognition*, 42(7), 1388–1398.
12. Shankar, S., & Udipi, V. R. (2016). Recognition of Faces-An Optimized Algorithmic Chain. *Procedia Computer Science*, 89, 597–606.
13. Bowyer, K., & King, M. (2019). Why face recognition accuracy varies due to race. *Biometric Technology Today*, 2019(8), 8–11.
14. Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1), 71–86.
15. Bartlett, M.S., Movellan, J.R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on neural networks*, 13(6), 1450–1464.

16. Zhang, B. L., Zhang, H., & Ge, S. S. (2004). Face recognition by applying wavelet subband representation and kernel associative memory. *IEEE Transactions on neural networks*, 15(1), 166–177.
17. Trigueros, D. S., Meng, L., & Hartnett, M. (2018). Face recognition: from traditional to deep learning methods. *arXiv preprint arXiv:1811.00116*.
18. Serrano, Á., de Diego, I. M., Conde, C., & Cabello, E. (2010). Recent advances in face biometrics with Gabor wavelets: A review. *Pattern Recognition Letters*, 31(5), 372–381.
19. Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., & Gao, W. (2009). WLD: A robust local image descriptor. *IEEE transactions on pattern analysis and machine intelligence*, 32(9), 1705–1720.
20. Kemelmacher-Shlizerman, I., & Basri, R. (2010). 3D face reconstruction from a single image using a single reference face shape. *IEEE transactions on pattern analysis and machine intelligence*, 33(2), 394–405.
21. Bowyer, K. W., Chang, K., & Flynn, P. (2006). A survey of approaches and challenges in 3D and multi-modal 3D+ 2D face recognition. *Computer vision and image understanding*, 101(1), 1–15.
22. Balaban, S. (2015, May). Deep learning and face recognition: The state of the art. In *Biometric and Surveillance Technology for Human and Activity Identification XII* (Vol. 9457, p. 94570B).

© Фам Туан Ань ( anhpt@itmo.ru ), Кашевник Алексей Михайлович ( alexey.kashevnik@iiias.spb.su ),

Чечулин Андрей Алексеевич ( chechulin@comsec.spb.ru ).

Журнал «Современная наука: актуальные проблемы теории и практики»



НИУ ИТМО