

УДК 004.932.2

doi: 10.17586/2226-1494-2020-20-2-163-176

## АНАЛИТИЧЕСКИЙ ОБЗОР МЕТОДОВ РАСПОЗНАВАНИЯ ЭМОЦИЙ ПО ВЫРАЖЕНИЯМ ЛИЦА ЧЕЛОВЕКА

Е.В. Рюмина<sup>a,b</sup>, А.А. Карпов<sup>a,b</sup>

<sup>a</sup> Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, Российская Федерация

<sup>b</sup> Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация  
Адрес для переписки: [karpov@iias.spb.su](mailto:karpov@iias.spb.su)

### Информация о статье

Поступила в редакцию 20.01.20, принята к печати 19.02.20

Язык статьи — русский

**Ссылка для цитирования:** Рюмина Е.В., Карпов А.А. Аналитический обзор методов распознавания эмоций по выражениям лица человека // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 2. С. 163–176. doi: 10.17586/2226-1494-2020-20-2-163-176

### Аннотация

Распознавание эмоций человека по мимике его лица является важной научно-исследовательской проблемой, которая охватывает множество областей и дисциплин, такие как компьютерное зрение, искусственный интеллект, медицина, психология и безопасность. В статье приведен аналитический обзор баз видеоданных выражений лица и подходов к распознаванию эмоций по выражениям лица, которые включают в себя три основных этапа анализа изображений, такие как предварительная обработка, извлечение признаков и классификация. Представлены как традиционные подходы к распознаванию эмоций человека по визуальным признакам лица, так и подходы, основанные на глубоком обучении с использованием глубоких нейронных сетей. Приведены современные результаты некоторых существующих алгоритмов. При обзоре научно-технической литературы упор главным образом делался на источники, содержащие теоретическую и исследовательскую информацию рассматриваемых методов, а также сравнение традиционных методов и методов на основе глубоких нейронных сетей, которые подкреплены экспериментальными исследованиями. Анализ научно-технической литературы, описывающий методы и алгоритмы анализа и распознавания выражений лица, а также результаты мировых научных исследований показали, что традиционные методы классификации выражений лица уступают по скорости и точности искусственным нейронным сетям. Основные вклады данного обзора заключаются в следующем: предоставление общего понимания современных подходов к распознаванию выражений лиц, которое позволит новым исследователям понять основные компоненты и тенденции в области распознавания выражений лиц; сравнение результатов мировых научных исследований показало, что комбинирование традиционных подходов и подходов, основанных на глубоких нейронных сетях показывают лучшую точность классификации, однако лучшими методами классификации являются искусственные нейронные сети. Статья может быть полезна специалистам и исследователям в области компьютерного зрения.

### Ключевые слова

предварительная обработка изображений, классификация, распознавание выражений лица, извлечение признаков, глубокие нейронные сети, компьютерная паралингвистика

### Благодарности

Исследование выполнено при поддержке Российского научного фонда РНФ (проект № 18-11-00145).

doi: 10.17586/2226-1494-2020-20-2-163-176

## ANALYTICAL REVIEW OF METHODS FOR EMOTION RECOGNITION BY HUMAN FACE EXPRESSIONS

E.V. Ryumina<sup>a,b</sup>, A.A. Karpov<sup>a,b</sup>

<sup>a</sup> St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint Petersburg, 199178, Russian Federation

<sup>b</sup> ITMO University, Saint Petersburg, 197101, Russian Federation  
Corresponding author: [karpov@iias.spb.su](mailto:karpov@iias.spb.su)

### Article info

Received 20.01.20, accepted 19.02.20

Article in Russian

**For citation:** Ryumina E.V., Karpov A.A. Analytical review of methods for emotion recognition by human face expressions. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, vol. 20, no. 2, pp. 163–176 (in Russian). doi: 10.17586/2226-1494-2020-20-2-163-176

**Abstract**

Recognition of human emotions by facial expressions is an important research problem that covers many areas and disciplines, such as computer vision, artificial intelligence, medicine, psychology and security. This paper provides an analytical overview of video facial expression databases and approaches to recognition emotions by facial expressions, which include three main stages of image analysis, such as pre-processing, feature extraction and classification. The paper presents both traditional approaches to recognition of human emotions by visual facial features, and approaches based on deep learning using deep neural networks. We give the current results of some existing algorithms. In the review of scientific and technical literature we empathized mainly the sources containing theoretical and research information of the methods under consideration, as well as comparison of traditional methods and methods based on deep neural networks, which were supported by experimental studies. Analysis of scientific and technical literature describing methods and algorithms for study and recognition of facial expressions, as well as the results of world scientific research, have shown that traditional methods for classification of facial expressions are second in speed and accuracy to artificial neural networks. The main contribution of this review is providing a common understanding of modern approaches to recognition of facial expressions, which will enable new researchers to understand the main components and trends in the field of recognition of facial expressions. Moreover, comparison of world scientific findings has shown that a combination of traditional approaches and approaches based on deep neural networks achieves better classification accuracy, but artificial neural networks are the best classification methods. The paper may be useful to specialists and researchers in the field of computer vision.

**Keywords**

digital image processing, classification, facial expression recognition, feature extraction, deep neural networks, computational paralinguistics

**Acknowledgements**

This research was supported by the Russian Science Foundation (project No.18-11-00145).

**Введение**

На сегодняшний день существует большое количество алгоритмов способных автоматически распознавать эмоции человека по мимике лица [1–5]. Однако качество работы систем распознавания эмоций по выражениям лица ухудшается из-за следующих проблем:

- небольшое количество данных для обучения;
- этническая принадлежность, пол, возраст;
- наигранность эмоций;
- внутриклассовое различие и межклассовое сходство;
- окклюзия;
- различный угол поворота головы;
- освещенность;
- различия в пропорциях лица.

Современные системы распознавания выражений лица (РВЛ) включают следующие основные этапы [6–8]:

- предварительная обработка изображения, которая состоит из: нахождения области лица, обрезки и масштабирования найденной области, выравнивания лица и регулировки контрастности;
- извлечение визуальных признаков;
- классификация эмоций.

На рис. 1 представлена общая диаграмма метода анализа изображений для распознавания эмоций по выражениям лица человека.

Предварительная обработка позволяет справиться с описанными ранее проблемами [9, 10], на данном этапе выполняется локализация области лица, обрезка и масштабирование найденной области, изменение контрастности изображения. Извлечение признаков основано на геометрии и внешнем виде лица. Под геометрией подразумеваются такие компоненты лица (их форма и расположение на лице), как глаза, рот, нос и т. д., а под внешним видом лица — текстура кожи. Классификация признаков направлена на разработку



Рис. 1. Диаграмма метода анализа изображений для распознавания эмоций по выражениям лица человека

соответствующего алгоритма классификации для идентификации выражений лица.

Основной целью данного аналитического обзора является сравнение методов предварительной обработки изображений лиц, извлечения визуальных признаков и машинной классификации эмоций, что позволило определить дальнейшее направление исследования для создания новой автоматической системы распознавания эмоций человека по мимике лица.

В работе представлен краткий обзор баз данных выражений лица. Рассмотрены методы предварительной обработки, извлечения и классификации признаков, используемые для задачи РВЛ. Приведены современные результаты некоторых существующих алгоритмов, а также представлено сравнение методов РВЛ.

### Обзор баз данных выражений лиц

Базы данных эмоциональных выражений лиц подразделяются на содержащие статические и динамические изображения в виде последовательности кадров. Статические изображения фиксируют только пиковый уровень интенсивности переживаемой эмоции, в то время как динамические изображения фиксируют выражения лиц, изменяющиеся в динамике. Для создания систем РВЛ перспективным является использование баз данных, содержащих видеопоследовательности. В табл. 1 представлена краткая информация о некоторых существующих базах данных.

### Предварительная обработка изображений

Предобработка изображения позволяет справиться с такими проблемами, как недостаток данных выражений лица, внутриклассовые различия и межклассовые сходства, небольшие изменения внешнего вида лица, изменения позы головы, освещенности, а также повысить точность систем РВЛ. Предварительная обработка изображений может включать в себя следующие этапы: локализация области лица; обрезка и масштабирование найденной области; выравнивание лица и регулировка контрастности.

1. Локализация области лица позволяет определить размер и местонахождение лица на изображении. Наиболее часто используемые методы локализации:

- метод Виолы–Джонса (Viola–Jones object detection, VJ) [14];
- Single shot multibox detector (SSD) [15];
- гистограмма направленных градиентов (Histogram of oriented gradients, HOG) [16];
- Max margin object detection (MMOD) [17].

Например, использование методов Виолы–Джонса и SSD представлено в [18, 19].

2. Обрезка и масштабирование найденной области лица осуществляется согласно координатам, полученным методами локализации области лица. Так как найденные области лица имеют различный размер, то необходимо выполнять масштабирование

Таблица 1. Краткий обзор баз данных, содержащих эмоциональные выражения лиц

База данных	Количество примеров, ед.	Количество субъектов, ед.	Разрешение, пикс.	Цветность	Эмоции	Особенности
KDEF <sup>1</sup>	4900	70	562 × 562	RGB	7	Фиксирование лица в пяти ракурсах
JAFFE <sup>2</sup>	213	10	256 × 256	Серый	7	Азиатская внешность
CK+ <sup>3</sup>	72939	123	640 × 490, 640 × 480	RGB, серый	8	Лица разных национальностей
FER2013 <sup>4</sup>	35888	—	48 × 48	Серый	7	Изображения из интернета
SAVEE <sup>5</sup>	480 аудио-видео	4	320 × 256	RGB	7	Наличие 60-ти маркеров на лице
CREMA-D [11]	7442 аудио-видео	91	480 × 360	RGB	6	Возраст от 20 до 74 лет, разная национальность, 2443 аннотатора
RAMAS [12]	564 аудио-видео	10	1920 × 1080	RGB	7	Славянская внешность, датчики захвата движения
RAVDESS [13]	4904 аудио-видео	24	1280 × 720	RGB	7	Разная степень интенсивности эмоций, 247 аннотаторов

<sup>1</sup> The Karolinska Directed Emotional Faces [Электронный ресурс]. <http://www.emotionlab.se/resources/kdef/>, свободный. Яз. англ. (дата обращения 13.12.2019).

<sup>2</sup> The Japanese Female Facial Expression (JAFFE) Database [Электронный ресурс]. [zenodo.org/record/3451524#.XefRiegzZPY](http://zenodo.org/record/3451524#.XefRiegzZPY), свободный. Яз. англ. (дата обращения 13.12.2019).

<sup>3</sup> Cohn-Kanade AU-Coded Expression Database [Электронный ресурс]. [www.consortium.ri.cmu.edu/ckagree](http://www.consortium.ri.cmu.edu/ckagree), свободный. Яз. англ. (дата обращения 13.12.2019).

<sup>4</sup> Challenges in Representation Learning Facial Expression Recognition Challenge [Электронный ресурс]. <http://www.kaggle.com/challenges-in-representation-learning-facialexpression-recognition-challenge>, свободный. Яз. англ. (дата обращения 13.12.2019).

<sup>5</sup> Surrey audio-visual expressed emotion (SAVEE) database [Электронный ресурс]. [kahlan.eps.surrey.ac.uk/savee](http://kahlan.eps.surrey.ac.uk/savee), свободный. Яз. англ. (дата обращения 13.12.2019).

изображения, т. е. приведение всех изображений к одному разрешению. Для данных задач применимы:

- выборка Бесселя (Bessel's correction) [20];
- распределение Гаусса (Gaussian distribution).

Например, применение данных методов представлено в [21, 22].

3. Выравнивание лица позволяет уменьшить внутриклассовые различия. Так, например, для каждого выражения лица выбирается опорное изображение, которое разделяется по цветовым компонентам или наиболее информативным областям лица (например, лоб, глаза), остальные изображения выравниваются относительно опорных изображений. Для данной задачи применяются методы:

- масштабное инвариантное преобразование объектов (Scale-invariant feature transform, SIFT) [23];
- области интереса (Region of interest, ROI) [24].

Примеры использования данных методов представлены в [25, 26].

4. Регулировка контрастности позволяет сглаживать изображения, уменьшать шум, повышать контрастность изображения лица и улучшать насыщенность, что позволяет справиться, например, с проблемой освещенности. Методами регулировки контрастности являются:

- выравнивание гистограммы (Histogram equalization, HE) [27];
- линейное контрастное растяжение (Linear contrast stretching) [28].

Примеры использования данных методов представлены в [29].

Различные манипуляции с изображениями, такие как вращение или смещение, позволяют увеличить разноплановость изображений и расширить базы данных. Расширение базы данных вариативными изображениями полезно для методов на основе глубокого обучения. Тогда как для традиционных методов полезнее использовать модели на основе выравнивания лица, кото-

рые, наоборот, уменьшают вариативность, связанную с изменениями позы головы, что позволяет уменьшать внутриклассовые различия и увеличить межклассовые сходства, в этом случае для каждого класса подбирается свой эталон, по которому происходит выравнивание изображений каждого класса. Подбор правильных методов предварительной обработки занимает много времени, так как от этого зависит скорость и точность распознавания лиц.

### Извлечение визуальных признаков

Следующим этапом РВЛ является извлечение признаков. На данном этапе осуществляется нахождение элементов, которые являются наиболее информативными для дальнейшей обработки. В зависимости от выполняемых функций методы извлечения информативных визуальных признаков подразделяются на несколько основных типов (рис. 2). Детально каждый метод извлечения признаков описан далее.

1. Методы на основе геометрических объектов позволяют извлекать информацию о геометрических объектах, таких как рот, нос, брови и другие объекты, и определять их местоположение. Имея информацию о геометрических объектах и их местоположения, можно рассчитать расстояние между объектами; полученные расстояния и координаты положений объектов являются признаками для дальнейшей классификации эмоций. К методам на основе геометрических объектов относятся:

- дескриптор line edge map (LEM) [30] измеряет сходства выражений лица, основываясь на границах объектов, пример использования дескриптора представлен в [31];
- активная модель формы (Active shape model, ASM) [32] обнаруживает края объектов при помощи лицевых ориентиров, которые представляют собой цепочку последовательностей точек признаков;

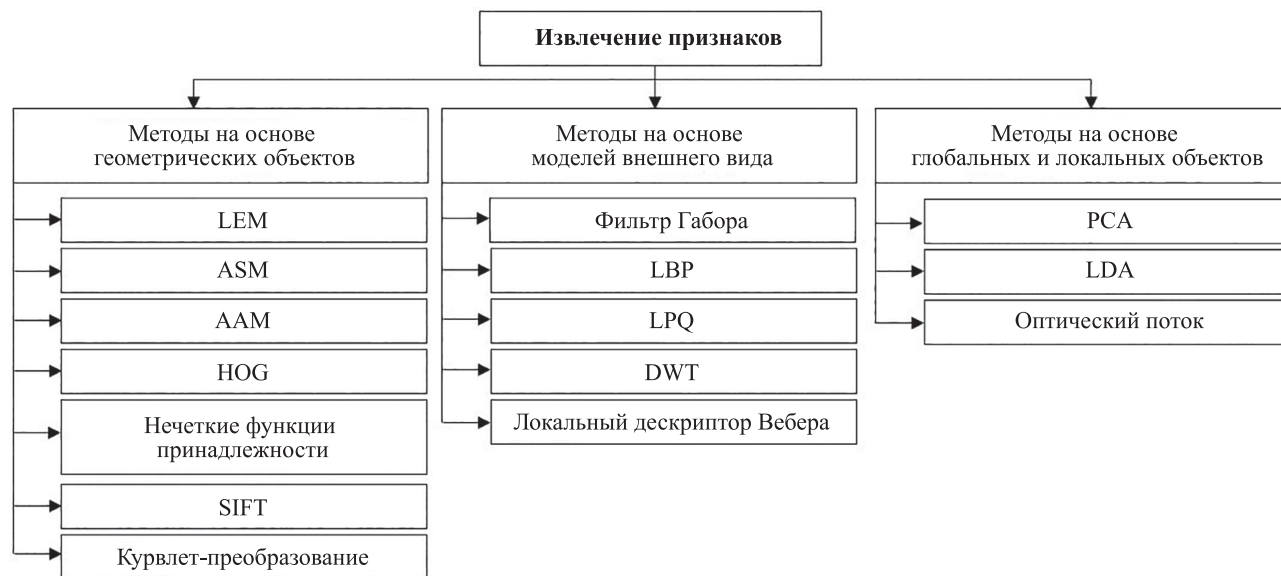


Рис. 2. Систематизация методов извлечения визуальных признаков



- активная модель внешнего вида (Active appearance model, AAM) [33] – расширенный вариант ASM, который также формирует текстурные особенности выражений лица. В [34] представлены примеры использования методов ASM и AAM;
  - HOG позволяет сравнивать выражения лица по направленности градиентов. Пример использования HOG представлен в [9];
  - нечеткие функции принадлежности (Fuzzy membership functions), предложенные в [35], извлекают объекты лица, представляя их в виде четырехугольных форм и сравнивая с четырехугольными геометрическими фигурами, степень неточности представляется в виде двенадцати нечетких функций;
  - дескриптор SIFT определяет потенциальные точки интереса на изображении посредством определения максимумов и минимумов для функции гауссовых фильтров, фильтры применяются в различных масштабах и поворотах. Пример использования дескриптора представлен в [35];
  - кривлет-преобразование (Curvelet Transform) [36] передает информацию о местоположении объекта и о пространственной частоте. Пример использования преобразования представлен в [37].
2. Методы на основе моделей внешнего вида позволяют извлекать информацию о текстурных особенностях лица. Например, разное количество морщин в области глаз, говорит о разных выражениях лица, поэтому применение методов на основе моделей внешнего вида является наиболее информативным для задач классификации эмоций. К методам на основе моделей внешнего вида относятся:
- фильтр Габора (Gabor filter) [38] является классическим методом выделения черт выражений лиц, который позволяет выделить различные модели деформации для каждой эмоции. Пример использования фильтра представлен в [35];
  - локальный бинарный шаблон (Local binary patterns, LBP) [39] позволяет представить окрестности пикселей изображения в двоичном коде, содержащих отличительные черты как локальных, так и глобальных (текстурных) областей выражений лица;
  - локальное фазовое квантование (Local phase quantisation, LPQ) устойчив к размытию изображений, основан на кратковременном преобразовании Фурье, которое позволяет выявить периодические компоненты в изображениях выражений лиц и оценить их вклад в формирование исходных данных. Примеры использования методов LBP и LPQ представлены в [40];
  - локальный дескриптор Вебера (Weber local descriptor) [41] извлекает признаки на два этапа, первый этап разделяет изображение на локальные участки (рот, нос и т. д.) и нормализует изображение, второй этап извлекает отличительные текстурные признаки, используя ориентацию градиента, описывающую выражения лица.
- Пример использования данного метода представлен в [42];
- дискретное вейвлет-преобразование (Discrete wavelet transform, DWT) [43] извлекает текстурные признаки, разбивая исходное изображение на участки низких и высоких частот. Например, авторы статьи [44] используют данный дескриптор.
3. Методами на основе глобальных и локальных объектов являются:
- метод главных компонент (Principal component analysis, PCA) [45] извлекает отличительные признаки выражений лица из ковариационной матрицы, уменьшая размерность векторов. Ковариационная матрица — матрица, элементами которой являются попарные линейные зависимости двух случайных величин;
  - линейный дискриминантный анализ (Linear discriminant analysis, LDA) [45] ищет векторы, которые имеют лучшие различия между классами и группирует признаки одного и того же класса. Примеры использования методов в PCA и LDA представлены в [1];
  - оптический поток (Optical flow, OF) [46] присваивает каждому пикселю вектор скорости, так извлекается информация о движении мышц лица из последовательности изображений, что позволяет учитывать деформацию лица в динамике. Например, авторы [47] используют данный метод.
- Из представленных методов извлечения признаков, методы на основе внешнего вида являются более полезными методами, так как они позволяют извлекать текстурные особенности внешнего вида, которые являются важными параметрами для РВЛ, однако они менее адаптированы к окклюзиям. Методы на основе геометрических объектов лучше адаптированы к окклюзиям, расстояние между лицевыми ориентирами больше характеризуют межклассовые различия. Использование гибридных методов для извлечения признаков выражений лица увеличивает точность классификации.

### Классификация эмоций

Классификация является последней стадией в РВЛ. На этой стадии осуществляется классификация извлеченных признаков на выражения лица: счастье, удивление, гнев, страх, отвращение, грусть и нейтральность.

Методы машинной классификации эмоций делятся на традиционные методы и искусственные нейронные сети. На рис. 3 можно увидеть методы машинной классификации эмоций.

Детально каждый метод машинной классификации эмоций описан далее.

1. Традиционные методы:

- расстояние Хаусдорфа (Hausdorff Distance) [48] позволяет измерять расстояние между областями интереса и сравнивать полученные расстояния между классами. Пример использования метода представлен в [10];



Рис. 3. Систематизация методов машинной классификации эмоций

- классификатор минимального расстояния (Minimum distance classifier, MDC) минимизирует расстояние между классами, чем меньше расстояние, тем больше сходство между классами. Пример использования метода представлен в [2];
  - алгоритм  $k$ -ближайших соседей ( $k$ -nearest neighbors algorithm, KNN) [49] классифицирует эмоции лица по первым  $k$ -элементам, расстояние которых минимально. Пример использования метода представлен в [3];
  - LDA представляет матричные данные в виде диаграммы рассеяния, что позволяет визуально разделять признаки классов. Например, авторы [50] используют метод LDA;
  - метод опорных векторов (Support vector machine, SVM) [51] строит гиперплоскость, разделяющую объекты выборки, чем больше расстояние между разделяющей гиперплоскостью и объектами разделяемых классов, тем меньше будет средняя ошибка классификатора. Например, авторы в [1, 3, 4] используют метод SVM;
  - скрытая марковская модель (Hidden Markov models, HMM) [52] сканирующим окном извлекает пиксели, преобразуя их в векторы наблюдений, полученные вектора классифицируются по выражениям лица. Например, авторы статьи [1] используют метод HMM;
  - дерево решений (Decision tree) [53] по заранее определенным правилам присваивает значение эмоции изображению в зависимости от того, какие признаки содержит данное изображение. Пример использования метода представлен в [4].
2. Искусственные нейронные сети:
- многослойный перцептрон (МП, Multilayer perceptron, MLP) [54] имеет три слоя, а именно: входной слой, скрытый слой и слой обработки,

- каждый слой содержит нейроны, которые имеют свое уникальное весовое значение. Данный вид перцептрона называется многослойным, потому что его скрытый (обучаемый) слой может состоять из нескольких слоев. Примеры использования MLP представлены авторами статей [3, 4];
- глубокие нейронные сети (Deep neural network, DNN):
  - многослойная прямая нейронная сеть (Multilayer feed forward neural network, MLFFNN) [55] представляет собой взаимосвязь перцептронов, где количество слоев в нейронной сети является количеством перцептронов. Пример использования MLFFNN представлен авторами в [5];
  - сверточная нейронная сеть (Convolutional neural network, CNN) — метод классификации, который имеет три слоя, а именно: слой свертки, слой подвыборки и полностью связанный слой. CNN улавливает текстуры на небольших участках изображений. CNN обучается на неподвижных изображениях. Примеры использования CNN представлены в [6–9];
  - рекуррентная нейронная сеть (Recurrent neural network, RNN) использует контекстную информацию о предыдущих изображениях. Так, один обучающий набор содержит последовательность изображений, классификация всего набора будет соответствовать классификации последнего изображения;
  - комбинация сверточной и рекуррентной нейронных сетей способна извлекать как локальные данные, так и использовать временную информацию для более точной классификации изображений. Например, авторы в [56]

используют комбинацию сверточной и рекуррентной нейронных сетей.

Имеются и другие DNN, которые являются модификациями представленных нейронных сетей.

Для анализа эффективности методов классификации используются метрики: полнота, точность, F-мера [57].

В настоящий момент при достаточном наборе обучающих данных лучшим методом классификации являются глубокие нейронные сети, так как они автоматически изучают и извлекают признаки из входных изображений, и обнаруживают эмоции на лице с более высокой точностью и скоростью в сравнении с другими методами.

### Сравнение алгоритмов распознавания выражений лиц

Поскольку точность алгоритмов РВЛ отчасти зависит от базы данных, которая используется для обучения и тестирования, правильным будет осуществлять сравнение использования различных методов на одинаковых наборах данных. Так, для сравнения методов РВЛ на статических изображениях исследовались полученные результаты на базе данных FER2013. Для оценки алгоритмов распознавания при динамическом изменении выражений лица использовалась база данных SAVEE. Для сравнения результатов исследований были выбраны широко используемые исследователями базы данных для данной задачи.

### Сравнение результатов на наборе данных FER2013

Авторы в [9] провели сравнение алгоритмов РВЛ с извлечением признаков при помощи методов HOG и CNN и выявили, что при небольшом расширении изображения  $48 \times 48$  и при невысоком качестве изображения результаты методов HOG и CNN не эффективны по сравнению с CNN.

Авторы в своей работе [8] экспериментировали с параметрами CNN. В качестве метода предварительной

обработки авторы использовали метод Виолы–Джонса, в результате локализованные области лица имели разрешение  $48 \times 48$  пикселей. На рис. 4 представлена архитектура CNN, предложенная авторами в [8].

Авторы статьи [56] предложили использовать многослойную функцию активации maxout (Multilayer maxout activation function, MMAF), которая позволяет справиться с такой проблемой при глубоком обучении как градиентный взрыв. Для извлечения визуальных признаков авторы использовали CNN и RNN с длинной кратковременной памятью (Long short-term memory network, LSTM). Комбинация моделей позволяет изучить информацию также о динамических особенностях выражения лица. Для классификации эмоций авторы использовали SVM. Все изображения случайным образом обрезались до  $24 \times 24$  пикселей.

Авторы в [58] использовали набор данных JAFFE в качестве набора обучающих данных метода *k*-средних (*k*-means), это было необходимо для того, чтобы получить набор центров кластеров. Метод *k*-средних — является методом кластеризации, используется для разделения *n* наблюдений на *k*-кластеров. Далее полученные значения центров кластера подавались как начальные значения ядра свертки в CNN. Классификацию в предложенном авторами алгоритме выполнил SVM. Для предварительной обработки авторы использовали метод Виолы–Джонса, в результате локализованные области лица имели разрешение  $48 \times 48$  пикселей.

Авторы в [59] для извлечения признаков использовали комбинацию функций ручной работы, которая включала в себя такие методы как eigenfaces, реализуемый через PCA, HOG, Dense-SIFT, комбинирование SIFT и ASM, CNN. Для классификации выражений лица авторы использовали классификатор SVM.

В табл. 2 приведена сводная информация об используемых методах РВЛ и точности классификации на наборе данных FER2013, представленная в [8, 9, 56, 58, 59].

По итоговой (табл. 2) можно заметить, что авторы применяли метод предварительной обработки Виолы–Джонса. Для извлечения признаков были использованы

Таблица 2. Результаты распознавания эмоций по лицу на базе данных FER2013

Авторы, год	Метод предобработки	Метод извлечения признаков	Классификатор	Точность, %
Karadeniz и др., 2019 [59]	—	PCA + HOG + SIFT + ASM	SVM	68,12
		CNN	SVM	68,34
		Все методы	SVM	69,54
Jumani и др., 2019 [9]	—	HOG	CNN	70,00
		—	CNN	72,00
Сао и др., 2019 [58]	VJ	CNN	KNN	72,86
		—	CNN	77,43
		CNN	SVM	78,86
		<i>k</i> -средних	CNN + SVM	80,29
An и др., 2019 [56]	—	MMAF	CNN	84,50
		MMAF + CNN + LSTM	SVM	86,60
Talegaonkar и др., 2019 [8]	VJ	—	CNN	89,78

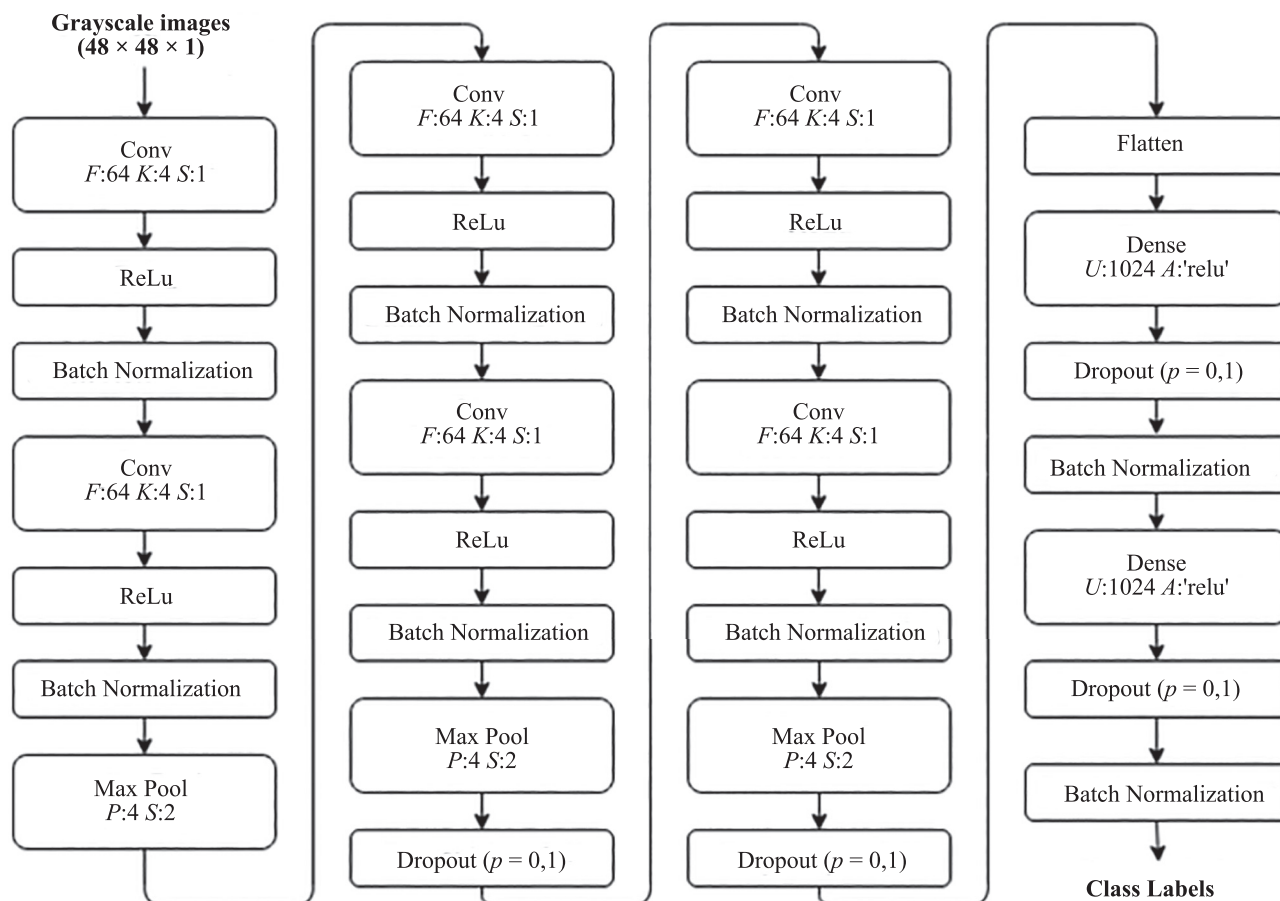


Рис. 4. Архитектура нейросети в системе распознавания эмоций [8]

grayscale images — полутоновые входные изображения размером  $48 \times 48$  пикселей; conv (Conv2D) — слой свертки с параметрами:  $F$  (filters) — количество выходных фильтров в свертке,  $K$  (kernel\_size) — длина окна свертки,  $S$  (strides) — длина шага свертки; ReLu — функция активации; Batch Normalization — нормализует слой активации, при котором среднее значение активации близко к 0, а стандартное отклонение активации близко к 1; Max Pool (MaxPooling2D) — слой подвыборки, где  $P$  (pool\_size) — уменьшает входные данные в  $n$  раз; Dropout — осуществляет случайное выпадение нейронов, для предотвращения переоснащения нейросети, где  $p = 0,1$ , обеспечивает 10 % выпадения нейронов; Flatten — преобразует входную матрицу в один массив; Dense — полносвязный слой, где  $U$  (units) — размерность выходного пространства; A:'relu' — функция активации ReLu; Class Labels — метки классов

методы на основе: геометрических объектов (ASM, HOG, SIFT), глобальных и локальных объектов (PCA). В том числе в некоторых представленных работах для извлечения признаков предпочтительными методами были CNN и комбинация CNN+LSTM. В качестве классификаторов авторы использовали SVM, KNN и CNN. Однако самый высокий результат показали авторы [8], которые использовали метод Виолы–Джонса и CNN, экспериментируя с параметрами CNN. Точность распознавания составила 89,78 %. Отсюда можно сделать вывод, что CNN справляется с задачей классификации точнее, чем традиционные методы классификации, такие как SVM, KNN.

### Сравнение результатов на наборе данных SAVEE

Как уже было сказано ранее, база данных FER2013 состоит из статических изображений, но поскольку выражения лица изменяются в динамике, необходимо произвести сравнение полученных результатов на базе данных, содержащих видеозаписи. Для сравне-

ния алгоритмов РВЛ будет использована база данных SAVEE.

Авторы в [60] используют новую гибридную структуру — тандемное моделирование (Tandem modelling, TM), которая состоит из двух иерархически объединенных нейронных сетей, имеющих прямую связь одного типа. Первой сетью является неполносвязная нейронная сеть (Not fully-connected neural net, NFCN), второй стандартная полносвязная нейронная сеть (Fully connected network, FCN), на обеих сетях применяется скрытый уровень соединения bottleneck, который позволяет объединять все выходы. В качестве предварительной обработки авторы использовали методы локализации Виолы–Джонса, в результате локализованные области лица имели разрешение  $128 \times 128$  пикселей, для извлечения признаков использовали LBP на трех ортогональных плоскостях (Local binary patterns on three orthogonal planes, LBP-TOP). Далее матрицы LBP каждого класса независимо распределялись через 5-слойную MLP и подавались на нейронную сеть.

В табл. 3 представлена сводная информация об используемых методах РВЛ и точности классификации



Таблица 3. Результаты распознавания эмоций по лицу на наборе данных SAVEE

Авторы, год	Метод предобработки	Метод извлечения признаков	Классификатор	Точность, %
Fan и др., 2018 [40]	CFAN	CNN	SVM	62,3
		Гист	SVM	62,7
		LPQ-TOP	SVM	64
		LBP	SVM	68,5
		LBP+LPQ-TOP	SVM	73,33
		CNN + Гист	SVM	75
		Все методы	SVM	85,83
Zhao и др., 2018 [47]	VJ, HE	OF	3D CNN	86,04
		AOF	3D CNN	90,83
		—	3D CNN	97,92
Kasraoui и др., 2019 [60]	VJ	LBP-TOP	MLP	95,08
		LBP-TOP	TM	100,00

на наборе данных SAVEE, представленная авторами в [40, 47, 60].

Авторы [40] использовали в своей работе следующий подход. За предварительную обработку отвечали грубые и точные сети с автоматическим кодировщиком (Coarse-to-fine auto-encoder networks, CFAN), которые позволили постепенно оптимизировать и выравнивать изображения выражений лица, в результате локализованные области лица имели разрешение  $64 \times 64$  пикселей. За извлечение глобальных признаков отвечали два набора функций: CNN и Гист, которая использует фильтры Габора. Локальные признаки извлекали LBP и LPQ-TOP. Для объединения локальных и глобальных признаков авторы применили дискриминирующий множественный канонический корреляционный анализ (Discriminative multiple canonical correlation analysis, DMCCA). Для уменьшения размерности функций использован алгоритм анализа компонент энтропии ядра (Kernel entropy component analysis, KECA). Классификатор SVM завершает работу алгоритма.

Авторы [47] для извлечения признаков использовали методы OF и накопительный OF (Accumulative optical flow, AOF). 3D CNN была выбрана авторами, так как она позволяет извлекать признаки из пространственных и временных измерений. В 3D CNN ядро движется в трех направлениях, входные и выходные данные в данной сети являются 4-х мерными. Для выбора подходящего набора гиперпараметров авторы выбрали байесовский оптимизатор. Методы Виолы–Джонса и выравнивание гистограммы (HE) были применены для предварительной обработки изображения, в результате локализованные области лица имели разрешение  $64 \times 64$  пикселей.

Итого, по результатам, полученным на базе данных SAVEE (табл. 3), можно сделать вывод. На этапе предварительной обработки изображения использованы методы: Виолы–Джонса, выравнивание гистограммы и CFAN. На этапе извлечения признаков использованы методы: LBP, LPQ, OF, CNN, Гист. Для классификаций выражений лица выбраны методы: SVM, MLP, TM, 3D CNN. Таким образом, самую высокую точность

получили авторы работы [60], точность распознавания составила 100 %. Авторы использовали метод Виолы–Джонса для предварительной обработки изображений. Метод локальный бинарный шаблон на трех ортогональных плоскостях был выбран на этапе извлечения признаков. Далее было использовано тандемное моделирование.

В результате сравнения алгоритмов РВЛ на базах данных SAVEE и FER2013 можно заключить:

- 1) на этапе предварительной обработки предпочтительным является метод локализации Виолы–Джонса;
- 2) на этапе извлечения признаков использованы методы:
  - на основе геометрических объектов (ASM, HOG, SIFT);
  - на основе внешнего вида (фильтр Габора, LBP);
  - на основе глобальных и локальных объектов (PCA, OF);
  - CNN;
- 3) на этапе классификации авторы использовали SVM, MLP, CNN и FCN.

Точность РВЛ является высокой и преимущественно достигается искусственными нейронными сетями, однако авторы статей не проводили экспериментов полученных ими обученных моделей с другими наборами данных, что может говорить об не универсальности обученных моделей.

### Заключение

Автоматическое распознавание выражения лица является важной составляющей многомодальных интерфейсов взаимодействия человека с компьютером и систем компьютерной паралингвистики. Для точного распознавания эмоций по мимике лица необходимо выбрать правильные методы предварительной обработки изображения, извлечения визуальных признаков выражений лица и классификации эмоций. На сегодняшний день традиционные методы классификации уступают по скорости и точности искусственным ней-

ронным сетям. Однако, несмотря на большое количество проведенных экспериментов, точность алгоритмов распознавания выражений лица пока недостаточно высока при различных входных параметрах, поэтому задача создания универсального алгоритма остается актуальной.

Для проведения дальнейших исследований авторами планируется объединить базы данных SAVEE, CREMA-D, RAMAS и RAVDESS, что позволит справиться с проблемой недостатка данных для обучения моделей. Кроме того, это позволит обучить классификатор автоматически распознавать эмоции вне зависимости от этнической принадлежности, возраста и пола. Для локализации области лица предполагается использовать активную модель формы, которая позволит выделить края объектов при помощи лицевых ориентиров, данный метод показывает высокую скорость и точность

обнаружения области лица. Масштабирование изображений будет осуществляться с разрешением  $48 \times 48$ ,  $64 \times 64$ ,  $128 \times 128$ ,  $224 \times 224$  пикселей, что позволит сделать вывод о влиянии разрешения на точность распознавания выражения лица. Также для регулирования контрастности будет использован метод выравнивания гистограммы, что позволит уменьшить шум и зависимость алгоритма распознавания выражений лица от освещенности. Использование активной модели формы также позволит извлечь признаки о расстоянии между ориентирами лица и центром масс, а также об угле поворота головы. Для классификации эмоций по полученным массивам пикселей с различным разрешением планируется использование CNN. Для классификации эмоций по извлеченным признакам (координаты ориентиров лица, расстояния до центра масс и угол поворота головы) планируется использовать методы FCN и SVM.

### Литература

1. Varma S., Shinde M., Chavan S.S. Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers // *Techno-Societal 2018. Proc. 2<sup>nd</sup> International Conference on Advanced Technologies for Societal Applications*. V. 1. 2020. P. 109–119. doi: 10.1007/978-3-030-16848-3\_11
2. Yin D.B.M., Mukhlas A.A., Chik R.Z.W., Othman A.T., Omar S. A proposed approach for biometric-based authentication using of face and facial expression recognition // *Proc. IEEE 3<sup>rd</sup> International Conference on Communication and Information Systems (ICCIS 2018)*, Singapore, 2018, P. 28–33. doi: 10.1109/ICOMIS.2018.8644974
3. Dino H.I., Abdulrazzaq M.B. Facial expression classification based on SVM, KNN and MLP classifiers // *Proc. International Conference on Advanced Science and Engineering (ICOASE 2019)*, Zakhoduhok, Iraq, 2019, P. 70–75. doi: 10.1109/ICOASE.2019.8723728
4. Tripathi A., Pandey S. Efficient facial expression recognition system based on geometric features using neural network // *Lecture Notes in Networks and Systems*. 2018. V. 10. P. 181–190. doi: 10.1007/978-981-10-3920-1\_18
5. Greche L., Es-Sbai N., Lavendelis E. Histogram of oriented gradient and multi layer feed forward neural network for facial expression identification // *Proc. International Conference on Control, Automation and Diagnosis (ICCAD 2017)*, Hammamet, Tunisia, 2017, P. 333–337. doi: 10.1109/CADIAG.2017.8075680
6. Целикова С.О., Горожанкин Я.П., Иванов А.О., Миронов А.А., Ахремчик Я.В. Использование нейросетевых технологий в задаче автоматического распознавания эмоций // *Молодой ученый*. 2019. № 26. С. 59–61 [Электронный ресурс]. URL: <https://moluch.ru/archive/264/61173/>, свободный. Яз. рус. (дата обращения: 12.12.2019).
7. Степанова О.А., Ивановский Л.И., Хрящев В.В. Использование глубокого обучения и сверточных нейронных сетей для анализа выражения лица // *DSPA: Вопросы применения цифровой обработки сигналов*. 2018. Т. 8. № 4. С. 170–173.
8. Talegaonkar I., Joshi K., Valunj S., Kohok R., Kulkarni A. Real time facial expression recognition using deep learning // *Proc. of International Conference on Communication and Information Processing (ICCIP)*, 2019 [Электронный ресурс]. URL: <https://ssrn.com/abstract=3421486>, свободный. Яз. англ. (дата обращения: 13.12.2019). doi: 10.2139/ssrn.3421486
9. Jumani S.Z., Ali F., Guriro S., Kandhro I.A., Khan A., Zaidi A. Facial expression recognition with histogram of oriented gradients using CNN // *Indian Journal of Science and Technology*. 2019. V. 12. N 24. P. 1–8. doi: 10.17485/ijst/2019/v12i24/145093
10. Babu D.R., Shankar R.S., Mahesh G., Murthy K.V.S.S. Facial expression recognition using bezier curves with hausdorff distance // *Proc. IEEE International Conference on IoT and Application (ICIOT 2017)*, Nagapattinam, India, 2017, P. 8073622. doi: 10.1109/ICIOTA.2017.8073622
11. Cao H., Cooper D.G., Keutmann M.K., Gur R.C., Nenkova A., Verma R. CREMA-D: Crowd-sourced emotional multimodal actors

### References

1. Varma S., Shinde M., Chavan S.S. Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers. *Techno-Societal 2018. Proc. 2<sup>nd</sup> International Conference on Advanced Technologies for Societal Applications*, vol. 1, 2020, pp. 109–119. doi: 10.1007/978-3-030-16848-3\_11
2. Yin D.B.M., Mukhlas A.A., Chik R.Z.W., Othman A.T., Omar S. A proposed approach for biometric-based authentication using of face and facial expression recognition. *Proc. IEEE 3<sup>rd</sup> International Conference on Communication and Information Systems (ICCIS 2018)*, Singapore, 2018, pp. 28–33. doi: 10.1109/ICOMIS.2018.8644974
3. Dino H.I., Abdulrazzaq M.B. Facial expression classification based on SVM, KNN and MLP classifiers. *Proc. International Conference on Advanced Science and Engineering (ICOASE 2019)*, Zakhoduhok, Iraq, 2019, pp. 70–75. doi: 10.1109/ICOASE.2019.8723728
4. Tripathi A., Pandey S. Efficient facial expression recognition system based on geometric features using neural network. *Lecture Notes in Networks and Systems*, 2018, vol. 10, pp. 181–190. doi: 10.1007/978-981-10-3920-1\_18
5. Greche L., Es-Sbai N., Lavendelis E. Histogram of oriented gradient and multi layer feed forward neural network for facial expression identification. *Proc. International Conference on Control, Automation and Diagnosis (ICCAD 2017)*, Hammamet, Tunisia, 2017, pp. 333–337. doi: 10.1109/CADIAG.2017.8075680
6. Tselikova S.O., Gorozhankin Ya.P., Ivanov A.O., Mironov A.A., Akhremchik Ya.V. Neural network technologies in automatic recognition of emotions. *Young Scientist*, 2019, no. 26, pp. 59–61. Available at: <https://moluch.ru/archive/264/61173/> (accessed: 12.12.2019). (in Russian)
7. Stepanova O., Ivanovsky L., Khryashchev V. Deep learning and convolutional neural networks for facial expression analysis. *DSPA*, 2018, vol. 8, no. 4, pp. 170–173. (in Russian).
8. Talegaonkar I., Joshi K., Valunj S., Kohok R., Kulkarni A. Real time facial expression recognition using deep learning. *Proc. of International Conference on Communication and Information Processing (ICCIP)*, 2019. Available at: <https://ssrn.com/abstract=3421486> (accessed: 13.12.2019). doi: 10.2139/ssrn.3421486
9. Jumani S.Z., Ali F., Guriro S., Kandhro I.A., Khan A., Zaidi A. Facial expression recognition with histogram of oriented gradients using CNN. *Indian Journal of Science and Technology*, 2019, vol. 12, no. 24, pp. 1–8. doi: 10.17485/ijst/2019/v12i24/145093
10. Babu D.R., Shankar R.S., Mahesh G., Murthy K.V.S.S. Facial expression recognition using bezier curves with hausdorff distance. *Proc. IEEE International Conference on IoT and Application (ICIOT 2017)*, Nagapattinam, India, 2017, pp. 8073622. doi: 10.1109/ICIOTA.2017.8073622
11. Cao H., Cooper D.G., Keutmann M.K., Gur R.C., Nenkova A., Verma R. CREMA-D: Crowd-sourced emotional multimodal actors dataset. *IEEE Transactions on Affective Computing*, 2014, vol. 5, no. 4, pp. 377–390. doi: 10.1109/TAFFC.2014.2336244

- dataset // IEEE Transactions on Affective Computing. 2014. V. 5. N 4. P. 377–390. doi: 10.1109/TAFFC.2014.2336244
12. Perepelkina O., Kazimirova E., Konstantinova M. RAMAS: Russian multimodal corpus of dyadic interaction for affective computing // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2018. V. 11096. P. 501–510. doi: 10.1007/978-3-319-99579-3\_52
  13. Livingstone S.R., Russo F.A. The Ryerson audio-visual database of emotional speech and song (RAVD ESS): A dynamic, multimodal set of facial and vocal expressions in North American English // PLoS ONE. 2018. V. 13. N 5. P. e0196391. doi: 10.1371/journal.pone.0196391
  14. Viola P., Jones M.J. Robust real-time face detection // International Journal of Computer Vision. 2004. V. 57. N 2. P. 137–154. doi: 10.1023/B:VISI.0000013087.49260.fb
  15. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A.C. SSD: single shot multibox detector // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2016. V. 9905. P. 21–37. doi: 10.1007/978-3-319-46448-0\_2
  16. Déniz O., Bueno G., Salido J., De la Torre F. Face recognition using histograms of oriented gradients // Pattern Recognition Letters. 2011. V. 32. N 12. P. 1598–1603. doi: 10.1016/j.patrec.2011.01.004
  17. King D.E. Max-margin object detection [Электронный ресурс]. URL: <https://arxiv.org/pdf/1502.00046.pdf>, свободный. Яз. англ. (дата обращения: 13.12.2019).
  18. Александров А.А., Кирпичников А.П., Ляшева С.А., Шлеймович М.П. Анализ эмоционального состояния человека на изображении // Вестник технологического университета. 2019. Т. 22. № 8. С. 120–123.
  19. Voronov V., Strelnikov V., Voronova L., Trunov A., Vovik A. Faces 2D-recognition and identification using the HOG descriptors method // Proc. 24<sup>th</sup> Conference of Open Innovations Association FRUCT. 2019. P. 783–789.
  20. Mohan P.G., Prakash C., Gangashetty S.V. Bessel transform for image resizing // Proc. 18<sup>th</sup> International Conference on Systems, Signals and Image Processing (IWSSIP 2011). Sarajevo, Bosnia-Herzegovina. 2011. P. 75–78.
  21. Owusu E., Abdulai J.-D., Zhan Y. Face detection based on multilayer feed-forward neural network and Haar features // Software: Practice and Experience. 2019. V. 49. N 1. P. 120–129. doi: 10.1002/spe.2646
  22. Su J., Gao L., Li W., Xia Y., Cao N., Wang R. Fast face tracking-by-detection algorithm for secure monitoring // Applied Sciences. 2019. V. 9. N 18. P. 3774. doi: 10.3390/app9183774
  23. Lowe D.G. Distinctive image features from scale-invariant keypoints // International Journal of Computer Vision. 2004. V. 60. N 2. P. 91–110. doi:10.1023/B:VISI.0000029664.99615.94
  24. Hernandez-Matamoros A., Bonarini A., Escamilla-Hernandez E., Nakano-Miyatake M., Perez-Meana H. A facial expression recognition with automatic segmentation of face regions // Communications in Computer and Information Science. 2015. V. 532. P. 529–540. doi: 10.1007/978-3-319-22689-7\_41
  25. Naz S., Ziauddin S., Shahid A.R. Driver fatigue detection using mean intensity, SVM, and SIFT // International Journal of Interactive Multimedia and Artificial Intelligence. 2019. V. 5. N 4. P. 86–93. doi: 10.9781/ijimai.2017.10.002
  26. Priya R.V. Emotion recognition from geometric fuzzy membership functions // Multimedia Tools and Applications. 2019. V. 78. N 13. P. 17847–17878. doi: 10.1007/s11042-018-6954-9
  27. Wang X., Chen L. Contrast enhancement using feature-preserving bi-histogram equalization // Signal Image and Video Processing. 2018. V. 12. N 4. P. 685–692. doi: 10.1007/s11760-017-1208-2
  28. Mustapha A., Oulefki A., Bengherabi M., Boutellaa E., Algaet M.A. Towards nonuniform illumination face enhancement via adaptive contrast stretching // Multimedia Tools and Applications. 2017. V. 76. N 21. P. 21961–21999. doi: 10.1007/s11042-017-4665-2
  29. Oloyede M., Hancke G., Myburgh H., Onumanyi A. A new evaluation function for face image enhancement in unconstrained environments using metaheuristic algorithms // EURASIP Journal on Image and Video Processing. 2019. N 1. P. 27. doi: 10.1186/s13640-019-0418-7
  30. Gao Y., Leung M.K.H. Face recognition using line edge map // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002. V. 24. N 6. P. 764–779. doi: 10.1109/TPAMI.2002.1008383
  31. Fawwad Hussain M., Wang H., Santosh K.C. Gray level face recognition using spatial features // Communications in Computer and Information Science. 2019. V. 1035. P. 216–229. doi: 10.1007/978-981-13-9181-1\_20
  12. Perepelkina O., Kazimirova E., Konstantinova M. RAMAS: Russian multimodal corpus of dyadic interaction for affective computing. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11096, pp. 501–510. doi: 10.1007/978-3-319-99579-3\_52
  13. Livingstone S.R., Russo F.A. The Ryerson audio-visual database of emotional speech and song (RAVD ESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE*, 2018, vol. 13, no. 5, pp. e0196391. doi: 10.1371/journal.pone.0196391
  14. Viola P., Jones M.J. Robust real-time face detection. *International Journal of Computer Vision*, 2004, vol. 57, no. 2, pp. 137–154. doi: 10.1023/B:VISI.0000013087.49260.fb
  15. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A.C. SSD: single shot multibox detector. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9905, pp. 21–37. doi: 10.1007/978-3-319-46448-0\_2
  16. Déniz O., Bueno G., Salido J., De la Torre F. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 2011, vol. 32, no. 12, pp. 1598–1603. doi: 10.1016/j.patrec.2011.01.004
  17. King D.E. *Max-margin object detection*. Available at: <https://arxiv.org/pdf/1502.00046.pdf> (accessed: 13.12.2019).
  18. Alexandrov A.A., Kirpichnikov A.P., Lyasheva S.A., Shleymovich M.P. Analyzing the emotional states of a person in an image. *Herald of Technological University*, 2019, vol. 22, no. 8, pp. 120–123. (in Russian)
  19. Voronov V., Strelnikov V., Voronova L., Trunov A., Vovik A. Faces 2D-recognition and identification using the HOG descriptors method. *Proc. 24<sup>th</sup> Conference of Open Innovations Association FRUCT*, 2019, pp. 783–789.
  20. Mohan P.G., Prakash C., Gangashetty S.V. Bessel transform for image resizing. *Proc. 18<sup>th</sup> International Conference on Systems, Signals and Image Processing (IWSSIP 2011)*, Sarajevo, Bosnia-Herzegovina, 2011, pp. 75–78.
  21. Owusu E., Abdulai J.-D., Zhan Y. Face detection based on multilayer feed-forward neural network and Haar features. *Software: Practice and Experience*, 2019, vol. 49, no. 1, pp. 120–129. doi: 10.1002/spe.2646
  22. Su J., Gao L., Li W., Xia Y., Cao N., Wang R. Fast face tracking-by-detection algorithm for secure monitoring. *Applied Sciences*, 2019, vol. 9, no. 18, pp. 3774. doi: 10.3390/app9183774
  23. Lowe D.G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, vol. 60, no. 2, pp. 91–110. doi:10.1023/B:VISI.0000029664.99615.94
  24. Hernandez-Matamoros A., Bonarini A., Escamilla-Hernandez E., Nakano-Miyatake M., Perez-Meana H. A facial expression recognition with automatic segmentation of face regions. *Communications in Computer and Information Science*, 2015, vol. 532, pp. 529–540. doi: 10.1007/978-3-319-22689-7\_41
  25. Naz S., Ziauddin S., Shahid A.R. Driver fatigue detection using mean intensity, SVM, and SIFT. *International Journal of Interactive Multimedia and Artificial Intelligence*, 2019, vol. 5, no. 4, pp. 86–93. doi: 10.9781/ijimai.2017.10.002
  26. Priya R.V. Emotion recognition from geometric fuzzy membership functions. *Multimedia Tools and Applications*, 2019, vol. 78, no. 13, pp. 17847–17878. doi: 10.1007/s11042-018-6954-9
  27. Wang X., Chen L. Contrast enhancement using feature-preserving bi-histogram equalization. *Signal Image and Video Processing*, 2018, vol. 12, no. 4, pp. 685–692. doi: 10.1007/s11760-017-1208-2
  28. Mustapha A., Oulefki A., Bengherabi M., Boutellaa E., Algaet M.A. Towards nonuniform illumination face enhancement via adaptive contrast stretching. *Multimedia Tools and Applications*, 2017, vol. 76, no. 21, pp. 21961–21999. doi: 10.1007/s11042-017-4665-2
  29. Oloyede M., Hancke G., Myburgh H., Onumanyi A. A new evaluation function for face image enhancement in unconstrained environments using metaheuristic algorithms. *EURASIP Journal on Image and Video Processing*, 2019, no. 1, pp. 27. doi: 10.1186/s13640-019-0418-7
  30. Gao Y., Leung M.K.H. Face recognition using line edge map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, vol. 24, no. 6, pp. 764–779. doi: 10.1109/TPAMI.2002.1008383
  31. Fawwad Hussain M., Wang H., Santosh K.C. Gray level face recognition using spatial features. *Communications in Computer and Information Science*, 2019, vol. 1035, pp. 216–229. doi: 10.1007/978-981-13-9181-1\_20
  32. Cootes T.F., Taylor C.J., Cooper D.H., Graham J. Active shape models-their training and application. *Computer Vision and Image*



32. Cootes T.F., Taylor C.J., Cooper D.H., Graham J. Active shape models—their training and application // *Computer Vision and Image Understanding*. 1995. V. 61. N 1. P. 38–59. doi: 10.1006/cviu.1995.1004
33. Cootes T.F., Edwards G.J., Taylor C.J. Active appearance models // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2001. V. 23. N 6. P. 681–685. doi: 10.1109/34.927467
34. Iqtait M., Mohamad F.S., Mamat M. Feature extraction for face recognition via active shape model (ASM) and active appearance model (AAM) // *IOP Conference Series: Materials Science and Engineering*. 2018. V. 332. N 1. P. 012032. doi:10.1088/1757-899X/332/1/012032
35. Бобе А.С., Коньшев Д.В., Воротников С.А. Система распознавания базовых эмоций на основе анализа двигательных единиц лица // *Инженерный журнал: наука и инновации*. 2016. № 9. С. 7. doi: 10.18698/2308-6033-2016-9-1530
36. Candès E., Demanet L., Donoho D., Ying L. Fast discrete curvelet transforms // *Multiscale Modeling & Simulation*. 2006. V. 5. N 3. P. 861–899. doi: 10.1137/05064182X
37. Fu X., Fu K., Zhang Y., Zhou Q., Fu X. Facial expression recognition based on Curvelet transform and sparse representation // *Proc. 14<sup>th</sup> International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNN-FSKD 2018)*, Huangshan, China, 2018. P. 257–263. doi: 10.1109/FSKD.2018.8686989
38. Ahsan T., Jabid T., Chong U.P. Facial expression recognition using local transitional pattern on Gabor filtered facial images // *IETE Technical Review*. 2013. V. 30. N 1. P. 47–52. doi: 10.4103/0256-4602.107339
39. Shan C., Gong S., McOwan P.W. Facial expression recognition based on Local Binary Patterns: A comprehensive study // *Image and Vision Computing*. 2009. V. 27. N 6. P. 803–816. doi: 10.1016/j.imavis.2008.08.005
40. Fan J., Tie Y., Qi L. Facial expression recognition based on multiple feature fusion in video // *Proc. International Conference on Computing and Pattern Recognition (ICCP 2018)*, Shenzhen, China, 2018. P. 86–92. doi: 10.1145/3232829.3232839
41. Li S., Gong D., Yuan Y. Face recognition using Weber local descriptors // *Neurocomputing*. 2013. V. 122. P. 272–283. doi: 10.1016/j.neucom.2013.05.038
42. Revina I.M., Emmanuel W.R.S. Face expression recognition using weber local descriptor and F-RBFNN // *Proc. 2<sup>nd</sup> International Conference on Intelligent Computing and Control Systems (ICICCS 2018)*, Madurai, India, 2018. P. 196–199. doi: 10.1109/ICCONS.2018.8662891
43. Addison P.S. *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. CRC Press, 2017. 464 p. doi: 10.1201/9781315372556
44. Nigam S., Singh R., Misra A.K. Efficient facial expression recognition using histogram of oriented gradients in wavelet domain // *Multimedia Tools and Applications*. 2018. V. 77. N 21. P. 28725–28747. doi: 10.1007/s11042-018-6040-3
45. Martínez A.M., Kak A.C. PCA versus LDA // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2001. V. 23. N 2. P. 228–233. doi: 10.1109/34.908974
46. Negahdaripour S. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1998. V. 20. N 9. P. 961–979. doi: 10.1109/34.713362
47. Zhao J., Mao X., Zhang J. Learning deep facial expression features from image and optical flow sequences using 3D CNN // *Visual Computer*. 2018. V. 34. N 10. P. 1461–1475. doi: 10.1007/s00371-018-1477-y
48. Guo B., Lam K.-M., Siu W.-C., Yang S. Human face recognition using a spatially weighted Hausdorff distance // *Proc. IEEE International Symposium on Circuits and Systems (ISCAS 2001)*, Hong Kong, China, 2001. V. 2. P. 145–148. doi: 10.1109/ISCAS.2001.921027
49. Meftah I.T., Le Thanh N., Amar C.B. Emotion recognition using KNN classification for user modeling and sharing of affect states // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2012. V. 7663. P. 234–242. doi: 10.1007/978-3-642-34475-6\_29
50. Greche L., Akil M., Kachouri R., Es-Sbai N. A new pipeline for the recognition of universal expressions of multiple faces in a video sequence // *Journal of Real-Time Image Processing*. 2019. in press. doi: 10.1007/s11554-019-00896-5
51. Abdulrahman M., Eleyan A. Facial expression recognition using support vector machines // *Proc. 23<sup>rd</sup> Signal Processing and Understanding*, 1995, vol. 61, no. 1, pp. 38–59. doi: 10.1006/cviu.1995.1004
33. Cootes T.F., Edwards G.J., Taylor C.J. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, vol. 23, no. 6, pp. 681–685. doi: 10.1109/34.927467
34. Iqtait M., Mohamad F.S., Mamat M. Feature extraction for face recognition via active shape model (ASM) and active appearance model (AAM). *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 332, no. 1, pp. 012032. doi: 10.1088/1757-899X/332/1/012032
35. Bobe A.S., Konyshov D.V., Vorotnikov S.A. Emotion recognition system based on the facial motor units' analysis. *Engineering Journal: Science and Innovation*, 2016, no. 9, pp. 7. doi: 10.18698/2308-6033-2016-9-1530. (in Russian)
36. Candès E., Demanet L., Donoho D., Ying L. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 2006, vol. 5, no. 3, pp. 861–899. doi: 10.1137/05064182X
37. Fu X., Fu K., Zhang Y., Zhou Q., Fu X. Facial expression recognition based on Curvelet transform and sparse representation. *Proc. 14<sup>th</sup> International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNN-FSKD 2018)*, Huangshan, China, 2018, pp. 257–263. doi: 10.1109/FSKD.2018.8686989
38. Ahsan T., Jabid T., Chong U.P. Facial expression recognition using local transitional pattern on Gabor filtered facial images. *IETE Technical Review*, 2013, vol. 30, no. 1, pp. 47–52. doi: 10.4103/0256-4602.107339
39. Shan C., Gong S., McOwan P.W. Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*, 2009, vol. 27, no. 6, pp. 803–816. doi: 10.1016/j.imavis.2008.08.005
40. Fan J., Tie Y., Qi L. Facial expression recognition based on multiple feature fusion in video. *Proc. International Conference on Computing and Pattern Recognition (ICCP 2018)*, Shenzhen, China, 2018, pp. 86–92. doi: 10.1145/3232829.3232839
41. Li S., Gong D., Yuan Y. Face recognition using Weber local descriptors. *Neurocomputing*, 2013, vol. 122, pp. 272–283. doi: 10.1016/j.neucom.2013.05.038
42. Revina I.M., Emmanuel W.R.S. Face expression recognition using weber local descriptor and F-RBFNN. *Proc. 2<sup>nd</sup> International Conference on Intelligent Computing and Control Systems (ICICCS 2018)*, Madurai, India, 2018, pp. 196–199. doi: 10.1109/ICCONS.2018.8662891
43. Addison P.S. *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. CRC Press, 2017, 464 p. doi: 10.1201/9781315372556
44. Nigam S., Singh R., Misra A.K. Efficient facial expression recognition using histogram of oriented gradients in wavelet domain. *Multimedia Tools and Applications*, 2018, vol. 77, no. 21, pp. 28725–28747. doi: 10.1007/s11042-018-6040-3
45. Martínez A.M., Kak A.C. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, vol. 23, no. 2, pp. 228–233. doi: 10.1109/34.908974
46. Negahdaripour S. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, vol. 20, no. 9, pp. 961–979. doi: 10.1109/34.713362
47. Zhao J., Mao X., Zhang J. Learning deep facial expression features from image and optical flow sequences using 3D CNN. *Visual Computer*, 2018, vol. 34, no. 10, pp. 1461–1475. doi: 10.1007/s00371-018-1477-y
48. Guo B., Lam K.-M., Siu W.-C., Yang S. Human face recognition using a spatially weighted Hausdorff distance. *Proc. IEEE International Symposium on Circuits and Systems (ISCAS 2001)*, Hong Kong, China, 2001, vol. 2, pp. 145–148. doi: 10.1109/ISCAS.2001.921027
49. Meftah I.T., Le Thanh N., Amar C.B. Emotion recognition using KNN classification for user modeling and sharing of affect states. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7663, pp. 234–242. doi: 10.1007/978-3-642-34475-6\_29
50. Greche L., Akil M., Kachouri R., Es-Sbai N. A new pipeline for the recognition of universal expressions of multiple faces in a video sequence. *Journal of Real-Time Image Processing*, 2019, in press. doi: 10.1007/s11554-019-00896-5
51. Abdulrahman M., Eleyan A. Facial expression recognition using support vector machines. *Proc. 23<sup>rd</sup> Signal Processing and*



- Communications Applications Conference (SIU 2015), Malatya, Turkey, 2015. P. 276–279. doi: 10.1109/SIU.2015.7129813
52. Aleksic P.S., Katsaggelos A.K. Automatic facial expression recognition using facial animation parameters and multistream HMMs // *IEEE Transactions on Information Forensics and Security*. 2006. V. 1. N 1. P. 3–11. doi: 10.1109/TIFS.2005.863510
  53. Safavian S.R., Landgrebe D. A survey of decision tree classifier methodology // *IEEE Transactions on Systems, Man and Cybernetics*. 1991. V. 21. N 3. P. 660–674. doi: 10.1109/21.97458
  54. Burkert P., Trier F., Afzal M.Z., Dengel A., Liwicki M. Dexpression: Deep convolutional neural network for expression recognition [Электронный ресурс]. URL: <https://arxiv.org/abs/1509.05371.pdf>, свободный. Яз. англ. (дата обращения: 13.12.2019).
  55. Svozil D., Kvasnička V., Pospichal J. Introduction to multi-layer feed-forward neural networks // *Chemometrics and Intelligent Laboratory Systems*. 1997. V. 39. N 1. P. 43–62. doi: 10.1016/S0169-7439(97)00061-0
  56. An F., Liu Z. Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM // *Visual Computer*. 2020. V. 36. N 3. P. 483–498. doi:10.1007/s00371-019-01635-4
  57. Ахремчик Я.В., Горожанкин Я.П., Иванов А.О., Миронов А.А., Целикова С.О. Распознавание и извлечение 3D-моделей по двумерным изображениям // *Молодой ученый*. 2019. № 26. С. 25–28 [Электронный ресурс]. URL: <https://moluch.ru/archive/264/61169/>, свободный. Яз. англ. (дата обращения: 02.01.2020).
  58. Cao T., Li M. Facial Expression Recognition Algorithm Based on the Combination of CNN and K-Means // *Proc. 11<sup>th</sup> International Conference on Machine Learning and Computing (ICMLC 2019)*. 2019. P. 400–404. doi: 10.1145/3318299.3318344
  59. Karadeniz A.S., Karadeniz M.F., Weber G.W., Husein I. Improving CNN features for facial expression recognition // *Zero: Jurnal Sains, Matematika dan Terapan*. 2019. V. 3. N 1. P. 1–11. doi: 10.30829/zero.v3i1.5881
  60. Kasraoui S., Lachiri Z., Madani K. Tandem modelling based emotion recognition in videos // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2019. V. 11507. P. 325–336. doi: 10.1007/978-3-030-20518-8\_28

#### Авторы

**Рюмина Елена Витальевна** — программист, Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, Российская Федерация; студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0002-4135-6949, [ryumina\\_ev@mail.ru](mailto:ryumina_ev@mail.ru)  
**Карпов Алексей Анатольевич** — доктор технических наук, доцент, руководитель лаборатории, Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, Российская Федерация; профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, Scopus ID: 57195330987, ORCID ID: 0000-0003-3424-652X, [karpov@iias.spb.su](mailto:karpov@iias.spb.su)

#### Authors

**Elena V. Ryumina** — Software Engineer, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint Petersburg, 199178, Russian Federation; Student, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0002-4135-6949, [ryumina\\_ev@mail.ru](mailto:ryumina_ev@mail.ru)  
**Alexey A. Karpov** — D.Sc., Associate Professor, Laboratory Head, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint Petersburg, 199178, Russian Federation; Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, Scopus ID: 57195330987, ORCID ID: 0000-0003-3424-652X, [karpov@iias.spb.su](mailto:karpov@iias.spb.su)



**Рюмина Елена Витальевна** родилась в 1991 г. в г. Кустанае (Республика Казахстан). В 2019 году окончила с отличием Челябинский государственный университет, экономический факультет. В 2019 году поступила в магистратуру Университета ИТМО. В настоящее время работает программистом в Санкт-Петербургском институте информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов — аффективные вычисления, искусственный интеллект, машинное обучение, нейронные сети, биометрические системы.

**Elena V. Ryumina** was born in 1991 in the city of Kustanai (Republic of Kazakhstan). In 2019 she graduated with honors from Chelyabinsk State University, the Department of Economics. In 2019 she undertook a master's degree at ITMO University. Currently, she is a Software Engineer at St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences. Her scientific research interests include affective computing, artificial intelligence, machine learning, neural networks and biometric systems.



**Карпов Алексей Анатольевич** — руководитель лаборатории речевых и многомодальных интерфейсов Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН), профессор Университета ИТМО, доктор технических наук (2014 г.), доцент (2012 г.). В 2002 г. окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП). С 2002 г. по настоящее время работает в СПИИРАН в лаборатории речевых и многомодальных интерфейсов, с 2015 г. возглавляет данную лабораторию. Является трехкратным победителем международных соревнований по компьютерной паралингвистике Computational Paralinguistics Challenges (ComParE) в рамках международных конференций INTERSPEECH в 2015 г. (Германия), 2016 г. (США) и 2017 г. (Швеция). Член редколлегии научных журналов «Труды СПИИРАН», «Речевые технологии» (Москва) и «Информатика» (Минск), приглашенный редактор международных журналов «Journal on Multimodal User Interfaces», «Speech Communication» и «Journal of Electrical and Computer Engineering». Генеральный председатель (сопредседатель) серии международных конференций «Speech and Computer» SPECOM. Автор (соавтор) более 250 научных статей, опубликованных в

зарубежных и отечественных научных журналах и трудах международных конференций, автор трех монографий. Признанный эксперт в области речевых технологий и многомодальных пользовательских интерфейсов. Области научных интересов – речевые технологии, автоматическое распознавание речи, обработка аудиовизуальной речи, многомодальные человеко-машинные интерфейсы, компьютерная паралингвистика и другие.

**Alexey A. Karpov** is a Head of the Speech and Multimodal Interfaces Laboratory at St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), and Professor (part-time) at ITMO University, D.Sc. (2014), Associate Professor (2012). He graduated from St. Petersburg State University of Aerospace Instrumentation (SUAI) in 2002. He has been working in the Speech and Multimodal Interfaces Laboratory of SPIIRAS since 2002, leading this laboratory since 2015. He is a three-time winner of the Computational Paralinguistics Challenge (ComParE) awards at the INTERSPEECH international conferences in 2015 (Germany), 2016 (USA), and 2017 (Sweden). Currently, he is an editorial board member of the following journals: “SPIIRAS Proceedings” (St. Petersburg), “Speech Technologies” (Moscow), and “Informatics” (Minsk); he was also a guest editor for such special issues of international journals as “Journal on Multimodal User Interfaces”, “Speech Communication” and “Journal of Electrical and Computer Engineering”. He is a General (co-)chairman of the “Speech and Computer” SPECOM series of international conferences. He has published more than 250 papers in international and Russian scientific journals and proceedings of international conferences, including 3 monographs. He is an expert in the scientific domains of speech technology and multimodal (audio-visual) user interfaces. His research interests include speech technology, automatic speech recognition, audio-visual speech processing, multimodal human-computer interfaces, and computational paralinguistics.