

Ван Лянпэн

**ПРОГРАММНЫЕ СРЕДСТВА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ И
АЛГОРИТМОВ КРАТНОМАСШТАБНОГО ВЕЙВЛЕТ АНАЛИЗА ДЛЯ
ОБРАБОТКИ И ПОИСКА ИЗОБРАЖЕНИЙ**

Специальность 05.13.11

Математическое и программное обеспечение вычислительных машин,
комплексов и компьютерных сетей (технические науки)

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата технических наук

Москва – 2019

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана) на кафедре информационных систем и телекоммуникаций.

Научный руководитель: **Петросян Олег Гарегинович**
кандидат технических наук, МГТУ им. Н.Э. Баумана

Официальные оппоненты: **Хачумов Вячеслав Михайлович**
доктор технических наук, Федеральный исследовательский центр «Информатика и управление» Российской академии наук
главный научный сотрудник

Виноградов Андрей Николаевич
кандидат физико-математических наук
Федеральное государственное бюджетное учреждение науки Институт программных систем им. А. К. Айламазяна Российской академии наук
исполняющий обязанности руководителя исследовательского центра искусственного интеллекта

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В.А. Трапезникова Российской академии наук

Защита диссертации состоится «12» февраля 2020 г. в 15 часов 00 минут на заседании диссертационного совета Д 999.216.02. при МАИ и МГТУ им. Н.Э. Баумана по адресу: 105005, г. Москва, 2-я Бауманская ул., д. 5, стр. 1.

С диссертацией можно ознакомиться в библиотеке МГТУ им. Н.Э. Баумана и на сайте <http://www.bmstu.ru>.

Отзывы на автореферат в двух экземплярах, заверенные печатью учреждения, просьба направлять по адресу: 105005, г. Москва, 2-я Бауманская ул., д. 5, стр. 1, МГТУ им. Н.Э. Баумана, ученому секретарю диссертационного совета Д 999.216.02.

Автореферат разослан « » 2019 г.

Ученый секретарь
диссертационного совета

А.Н. Алфимцев

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Необходимость обработки и поиска изображения на основе результатов наблюдений за объектами и явлениями окружающего мира возникла значительно ранее появления современных компьютерных систем и технологий. Первые методы обработки и поиска изображения разрабатывались для электронных аналоговых систем и рассматривались в рамках теории обработки сигналов. В процессе развития вычислительной техники и информационных технологий наука обработки и поиска изображения, как и само понятие образа, претерпела значительные изменения и продолжает интенсивно развиваться. В связи с этим актуальными являются следующие задачи.

1. Существующие алгоритмы для извлечения локальных или глобальных признаков изображений сложны, а размерность векторов признаков слишком высока. Поэтому возникает необходимость в разработке алгоритма для извлечения признаков изображений, позволяющего:

- а) сократить время машинной работы.
- б) минимизировать размерности векторов признаков и сокращать время обучения модели.
- в) повысить точность классификации изображений.

2. Обнаружение границ объектов является составной частью алгоритмов обработки изображений, где чаще всего используется алгоритм первой производной, основным недостатком которого является чувствительность к шуму. Поэтому разработка метода и алгоритма, уменьшающего эту чувствительность, продолжает оставаться актуальной.

3. Поиск изображений в базе данных в основном реализуется с помощью перцептивного хеш-алгоритма, скорость работы которого высока, но не имеет достаточной точности. В связи с этим актуальным является увеличение точности упомянутого поиска похожих изображений.

4. Существующие методы LBP, SIFT, ORB не позволяют явным образом выделить один из признаков, с наилучшими результатами. Поэтому актуальной является задача выявления более информативных признаков лиц и разработка более эффективных и точных методов определения, с более гибкой структурой, с сокращением временем обучения и с высокой устойчивостью к переобучению (overfitting). Здесь необходимо выделить работы ряда ученых (Paul V., Rainer L., Navneet D., Ojala T., Pietikäinen M., Harwood D., Guo Z., Zhang L., Lienhart R., David L., Bay H., Ess A., Tuytelaars T., Ethan R., Vincent R., Kurt K., Janusz B., Gaviphat L., Mohamed O., Yassine K., LeCun Y., Travis W., Robert L.), в которых систематизируются знания о применении различных признаков в задачах классификации.

Цель диссертационной работы является разработка и исследование методов обработки и поиска изображений на основе кратномасштабного вейвлет анализа, позволяющего увеличить точность распознавания практически до 100% и улучшить обобщающую способность метода.

Для достижения поставленной цели сформулированы и решены следующие основные задачи:

1. Разработка алгоритма для извлечения признаков изображений на основе дерева коэффициентов для кратномасштабного вейвлет-преобразования.

2. Разработка метода для повышения чёткости обнаруженных границ объектов на зашумленном изображении и улучшения помехоустойчивости метода.

3. Разработка метода для поиска в базе данных максимального количества правильных изображений, похожих на исходное изображение.

4. Разработка более гибких методов для распознавания лиц, при котором сокращается время обучения модели и повышается точность распознавания.

Основные положения, выносимые на защиту:

1. Алгоритм для извлечения признаков изображений с помощью кратномасштабного вейвлет-преобразования.

2. Метод обнаружения границ объектов на зашумлённом изображении.

3. Метод поиска похожих изображений на основе вейвлет анализа и Марковской цепи.

4. Методы решения задач распознавания лиц с применением СММ (Скрытая Марковская Модель, СММ), вейвлет-нейронных сетей, и алгоритма Левенберга-Марквардта.

Научная новизна

Научная новизна работы состоит в следующем:

1. Разработан алгоритм построения признаков изображений с помощью дерева коэффициентов приближений в различных масштабах вейвлет-преобразования, позволяющий сократить размерность векторов признаков и увеличить точность классификации изображений.

2. Использован метод сингулярного разложения (Singular Value Decomposition, SVD) матрицы изображения, который совместно с вейвлет анализом, позволяет эффективно устранить негативное влияние шумов, и получить более чёткие границы объектов на зашумлённом изображении.

3. Разработан метод поиска похожих изображений, основанный на вейвлет-преобразовании и Марковской цепи, позволяющий выделить большее количество похожих изображений, по сравнению с классическими алгоритмами, и одновременно обеспечить полноту и более высокую схожесть выделенных изображений по сравнению с перцептивным хеш-алгоритмом.

4. Разработан метод распознавания лиц на основе вейвлет-признаков и скрытой Марковской модели, который позволяет сократить время обучения, а при добавлении новых образцов не требуется повторного обучения существующих.

5. Разработан метод распознавания лиц на основе вейвлет-признаков и вейвлет-нейронных сетей, позволяющий сократить время сходимости сети, увеличить точность распознавания лиц и обобщающую способность к переобучению при использовании алгоритма Левенберга-Марквардта для обучения вейвлет-нейронных сетей.

Тематика работы соответствует:

п. 7. «Человеко-машинные интерфейсы; модели, методы, алгоритмы и программные средства машинной графики, визуализации, обработки изображений, систем виртуальной реальности, мультимедийного общения».

п. 9. «Модели, методы, алгоритмы и программная инфраструктура для организации глобально распределенной обработки данных» паспорта специальности 05.13.11 Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей (технические науки).

Методы исследования. При решении задач, рассматриваемых в диссертации, были использованы: математический аппарат линейной алгебры, кратномасштабный вейвлет анализ, перцептивный хэш-алгоритм, Марковская цепь, скрытая Марковская модель, вейвлет-нейронные сети и алгоритм Левенберга-Марквардта для обучения сети. В процессе математического моделирования применялись: среда моделирования MATLAB, язык программирования Python 3.5, виртуальные серверы Amazon Elastic Compute Cloud (Amazon EC2).

Достоверность полученных результатов базируется на корректной математической постановке задач, сопоставлении предлагаемых теоретических решений с результатами экспериментов, с применениями современных методик моделирования и программных вычислительных средств. Правильность полученных результатов определяется корректностью выкладок и сопоставлением с известными результатами других авторов.

Практическая значимость работы. На основе предложенных методов разработано программное обеспечение, предназначенное для решения задач обнаружения границ зашумленного изображения, поиска похожих изображений в базах данных. Экспериментально показано, что совместное применение кратномасштабного вейвлет анализа и вейвлет-нейронных сетей позволяет избегать переобучения на небольших объемах обучающих изображений лиц, ускорить конвергенцию сети и улучшить обобщающую способность сетей.

Апробация работы. Основные положения диссертационной работы были доложены и обсуждены на следующих конференциях и научных семинарах:

1. Научный семинар кафедры информационных систем и телекоммуникаций МГТУ им. Н.Э. Баумана (г. Москва, 2017 г., 2018 г.).

2. Международная научная конференция «Наука и инженерное образование. SEE-2016» (г. Москва, 2016 г.).

Публикации. По теме диссертации опубликовано 6 научных работ, из них 4 статьи в рецензируемых журналах и изданиях, рекомендуемых ВАК РФ и 1 статья из Перечня международных научных изданий включенных в базу данных Scopus и EI Compendex. Общий объем публикаций составляет 1,89 п.л.

Структура и объём диссертации. Диссертация включает в себя введение, четыре главы, общие выводы и заключение, список литературы, изложена на 138 страницах машинописного текста, в том числе 72 рисунка и 2 таблицы. Список литературы насчитывает 124 наименования.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертации в различных областях обработки и поиска изображений. Указаны основные цели работы, кратко изложена структура диссертации, охарактеризована её научная новизна, и практическая значимость, сформулированы основные положения, выносимые на защиту, а также представление результатов её различных частей и всей работы в целом научной общественности на семинарах кафедры информационных систем и телекоммуникаций МГТУ им. Н.Э. Баумана.

В первой главе описана математическая теория вейвлет-преобразования и его преимущества по сравнению с преобразованием Фурье, и проанализирован процесс построения вейвлет-функций и их соответственных масштабных функций для кратномасштабного анализа. Представлены процессы разложения и восстановления одномерного сигнала и двухмерного изображения с помощью одномерного быстрого вейвлет-преобразования и двумерного быстрого вейвлет-преобразования и роли различных коэффициентов (коэффициенты приближений и коэффициенты деталей) в восстановлении исходного сигнала или изображения.

Использованы коэффициенты деталей для обнаружения границ объектов на изображении и выделения высокочастотной информационной составляющей изображения в поиске похожих изображений.

Для преодоления недостатков существующих глобальных и локальных признаков разработан алгоритм для извлечения признаков изображений, основанный на дереве коэффициентов для трёхмасштабного вейвлет-преобразования (в 4-ом, 3-ом и 2-ом масштабах). В соответствии с теорией быстрого вейвлет-преобразования высокомасштабные коэффициенты приближения получают путём низкочастотной фильтрации и прореживающей выборки коэффициентов приближения в соседнем низком масштабе.

Для каждого коэффициента приближения в четвёртом масштабе, был создан 21-мерный вектор с помощью структуры дерева коэффициентов для трёхмасштабного вейвлет-преобразования, и получен 21-мерный тензор признаков, соответствующий коэффициентами приближений в 4-ом масштабе на Рис.1.

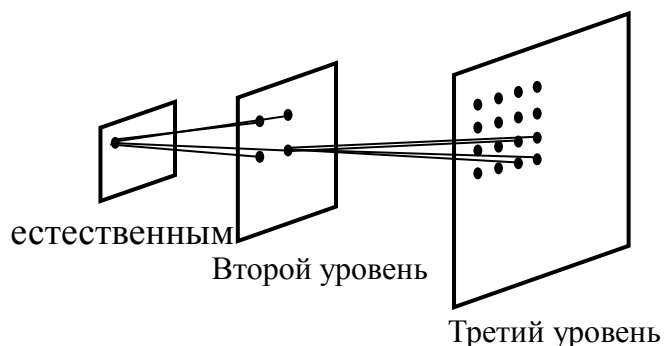


Рис.1.

Структура дерева коэффициентов для
трёхмасштабного вейвлет-преобразования

В четвертой главе для классификации изображений лиц были также

использованы эти признаки.

Во второй главе проведен обзор существующих методов для обнаружения границ объектов на изображении, таких как метода первой производной, кратномасштабного вейвлет анализа, и фрактального метода.

Для улучшения помехоустойчивости разработан метод сингулярного разложения матрицы изображения совместно с вейвлет анализом. Использовано сингулярное разложение матрицы для предварительной обработки зашумлённого изображения на Рис. 2.

Матрица $A \in R^{p \times q}$, состоящая из вещественных чисел, имеет p строк и q столбцов, причём $q > p$, представлена в виде SVD разложения:

$$A = \sum_{i=1}^m \lambda_i \mathbf{u}_i \mathbf{v}_i^T = U \Lambda V^T, \quad (1)$$

где \mathbf{u}_i и \mathbf{v}_i – левый и правый сингулярные векторы, $\lambda_i \geq 0$ – сингулярные числа матрицы A . Столбцы матриц U (ортогональная матрица с размером $p \times p$) и V (ортогональная матрица с размером $q \times q$) называются соответственно левыми и правыми сингулярными векторами, и Λ – диагональная матрица с размером $p \times q$.

Распределения сингулярных значений исходного изображения и зашумлённого изображения, полученные расчётным путём с помощью встроенных функции в среде MATLAB, показано на Рис. 3.

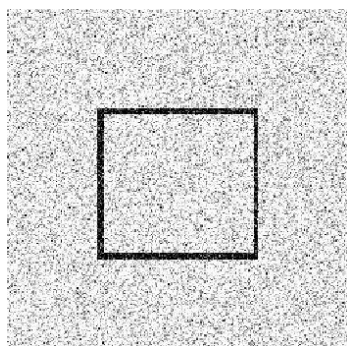


Рис. 2.

Зашумлённое тестовое
изображение

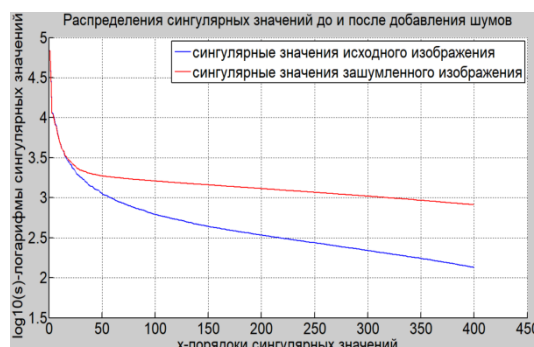


Рис. 3.

Изменение распределения
сингулярных значений

Ось x на Рис.3 характеризует порядок сингулярных значений, ось y представляет собой логарифмы сингулярных значений по основанию 10. Синяя линия представляет собой распределение сингулярных значений исходного изображения, красная линия – распределение сингулярных значений зашумлённого изображения. Логарифмирование сингулярных значений наглядно представляет изменение распределения сингулярных значений до и после добавления шума.

Из Рис.3 видно, что интенсивность шума доминирует в области сингулярных значений высших порядков. В то же время сингулярные значения низших порядков менее подвержены влиянию шумов и могут эффективно использоваться для восстановления изображения. Выбирая количество сингулярных значений $k = 50$, и соответствующие им левые, правые сингулярные векторы \mathbf{u}_i и \mathbf{v}_i в низших порядках, восстанавливаем изображение, которое содержит информацию границы, и менее подвержено воздействию шума.

Коэффициенты вейвлет-преобразования изображения коррелируют во всех масштабах, и эти коэффициенты имеют большие числовые значения. Но у вейвлет-преобразования шума корреляция отсутствует, и при этом числовые коэффициенты вейвлет-преобразования в каждом масштабе достаточно малы. Это означает, что можно уменьшить влияние шума при определении границы изображения, установлением порога для каждого масштаба:

$$T_j = \bar{\sigma} \cdot \sqrt{2 \ln(N_j)}, \quad (2)$$

где T_j – порог в масштабе j , N_j – количество коэффициентов вейвлет-преобразования в масштабе j , $\bar{\sigma}$ – оценка среднеквадратического отклонения шума изображения, может быть выражена как,

$$\bar{\sigma} = \frac{\text{Median}(|W_{\psi}^i(j, m, n)|)}{0,6745}, \quad (3)$$

где $W_{\psi}^i(j, m, n)$ – коэффициенты деталей вейвлет-преобразования, $i = \{H, V, D\}$ (H – коэффициенты горизонтальных, V – вертикальных и D – диагональных деталей в j -м масштабе).

Для оценки результатов методов используются среднеквадратичные ошибки (Mean Squared Error, MSE) между полученными и фактическими границами.

Среднеквадратичные ошибки при использовании метода первой производной – $8,28 \times 10^4$, кратномасштабного вейвлет анализа – $9,91 \times 10^3$, фрактального метода – $6,87 \times 10^3$, и предложенного метода – $4,62 \times 10^3$.

С помощью оставшихся коэффициентов вейвлет-преобразования в различных масштабах и обратного вейвлет-преобразования получим границы объектов.

Разработанный метод может быть успешно использован для выделения границ объекта на изображении со всеми подробностями за счёт значительного уменьшения влияния шумов.

Для представления преимущества метода, предложенного в диссертации, был выбрано изображение с естественным фоном с разрешением 640×1024 , с добавленным гауссовым шумом со средним значением $\mu = 0$, среднеквадратическим отклонением $\sigma = 40$, и пиковым отношением сигнала к шуму $PSNR = 16,97 \text{ dB}$ (Рис.4, Рис.5).



Рис. 4.
Изображение с



Рис. 5.
Границы объектов на
изображении

Сравнивая полученные результаты, можно утверждать, что предложенный в диссертации метод обнаружения границ изображения при одинаковом для всех

методов уровне шума на входе, превосходит рассмотренные в диссертации существующие методы на визуальном уровне по чёткости грани.

В третьей главе проведен обзор существующих методов для поиска похожих изображений, таких как алгоритма SIFT (Scale-Invariant Feature Transform, SIFT), алгоритма ORB (Oriented FAST and Rotated BRIEF, ORB) и перцептивного хеш-алгоритма (Perceptual Hash Algorithm, PHA). Для повышения относительной точность выделенных похожих изображений разработан метод, основанный на высокочастотной информационной составляющей изображения и Марковской цепи.

Исходная функция $f(x, y)$ может быть восстановлена по заданным коэффициентам W_φ и W_ψ^i при помощи обратного дискретного вейвлет-преобразования,

$$f(x, y) = f_\varphi(j_0, x, y) + f_\psi(J, x, y), \quad (4)$$

$$f_\varphi(j_0, x, y) = \frac{1}{\sqrt{MN}} \sum_m \sum_n W_\varphi(j_0, m, n) \varphi_{j_0, m, n}(x, y), \quad (5)$$

$$f_\psi(J, x, y) = \frac{1}{\sqrt{MN}} \sum_{i=H, V, D} \sum_{j=j_0}^{J-1} \sum_m \sum_n W_\psi^i(j, m, n) \psi_{j, m, n}^i(x, y), \quad (6)$$

где j_0 – начальный масштаб вейвлет-преобразования, J – самый крупный масштаб, $f_\varphi(j_0, x, y)$ – компонент исходной функции, соответствующий коэффициентам приближений W_φ , $f_\psi(J, x, y)$ – компонент исходной функции, соответствующий коэффициентам деталей W_ψ^i , $i = \{H, V, D\}$, $\varphi(x, y)$ – двухмерная масштабирующая функция, $\psi^i(x, y)$, $i = \{H, V, D\}$ – двухмерные вейвлет-функции, M , N – общее число пикселей изображения в вертикальном и горизонтальном направлении.

В процессе выделения высокочастотной информационной составляющей изображения коэффициенты приближения самого мелкого масштаба дискретного вейвлет-преобразования исходного изображения (Рис.6), удалены путём обращения в нуль её значений ($W_\varphi(j_0, m, n) = 0$). Высокочастотная информационная составляющая изображения извлечена при вычислении обратного преобразования с использованием этих изменённых коэффициентов.

В соответствии со свойством дискретного косинусного преобразования (Discrete Cosine Transform, DCT) выбираем в квадрате 8×8 коэффициенты в верхнем левом углу матрицы ДКП, и используем хеш-алгоритм для выделения признаков изображения (64-бит хеш-последовательность).

Для сопоставления признаков, вместо используемых в литературе кратчайшего евклидова расстояния и алгоритма ближайшего соседа, предложен алгоритм рейтинг подобия, который состоит из двух этапов:

1. Создание стохастической матрицы на основе матрицы расстояния Хэмминга между любыми двумя изображениями, образованной с помощью Марковской цепи.

2. Введения весовой функции для определения вероятности перехода из одного состояния в другое, что позволит повысить относительную точность.

Вычислим расстояние Хэмминга между любыми двумя изображениями в наборе ImageNet, в котором находятся по 500 изображений пяти различных объектов.

Допустим, что расстояние изображений между собой равно 1. Если расстояние Хэмминга между двумя изображениями больше порога и два изображения не связаны друг с другом, то расстояние Хэмминга полагаем равным бесконечности. В этом случае диаграмма расстояний Хэмминга всех изображений в наборе представлена на Рис. 7.



Рис. 6.

Исходное изображение

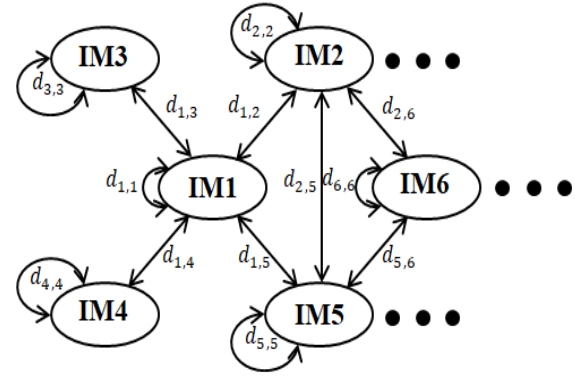


Рис. 7.

Фрагмент диаграммы расстояний Хэмминга изображений для анализа рейтинга подобия

Из приведённой выше диаграммы, можем получить реальную симметричную матрицу, составленную расстояниями Хэмминга между изображениями в наборе.

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,N-1} & d_{1,N} \\ d_{2,1} & d_{2,2} & d_{2,j} & d_{2,N-1} & d_{2,N} \\ \vdots & d_{i,2} & \ddots & d_{i,N-1} & \vdots \\ d_{N-1,1} & d_{N-1,2} & d_{N-1,j} & d_{N-1,N-1} & d_{N-1,N} \\ d_{N,1} & d_{N,2} & \cdots & d_{N,N-1} & d_{N,N} \end{bmatrix}, \quad (7)$$

где N – общее количество изображений в наборе, D – матрица расстояний Хэмминга, $d_{i,j}$ – расстояния Хэмминга между изображениями i и j , T – порог расстояния Хэмминга. Матрица D должна обладать следующими свойствами:

$$\begin{cases} d_{i,j} = d_{j,i} \\ d_{i,j} = 1 \text{ если } i = j \\ d_{i,j} = \infty \text{ если } d_{i,j} > T. \end{cases} \quad (8)$$

Рассмотрим расстояние Хэмминга между изображениями как вероятность перехода между двумя состояниями. Это означает, что чем короче расстояния Хэмминга, тем больше вероятность перехода из одного состояния в другое. Преобразуем матрицу расстояний Хэмминга в матрицу вероятности перехода состояний.

Введём понятие подобных слоёв (на Рис. 8). $S1$ – это состояние, которое находится в первом слое, соответствующее исходному изображению. Для остальных 499 состояний полагаем расстояние Хэмминга от состояния $S1$ меньше 6, они будут размещены во втором подобном слое. Расстояние Хэмминга между другими остальными состояниями и состояниями во втором слое меньше 4, они

размещаются в третьем подобном слое. В настоящей работе создано 4 слоя.

Чем ниже слои, состояния, тем соответственные изображения более похожи на исходное изображение. После необходимых вычислений, получим в общей сложности 22 состояния (1 состояние в первом слое, 11 состояний во втором слое, 9 состояний в третьем слое и 1 состояние в четвёртом слое). Эти похожие 22 изображения представлены на Рис. 9.

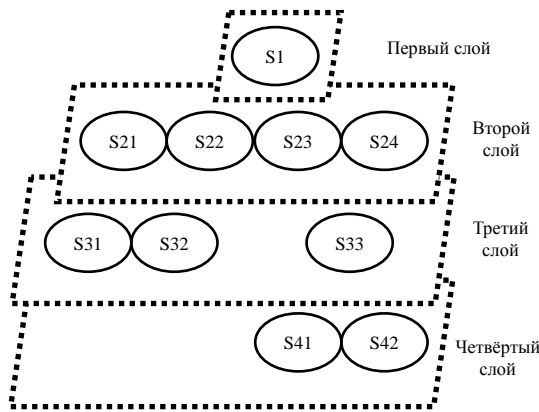


Рис. 8.

Схематическая диаграмма
подобных слоёв



Рис. 9.

Соответственные изображения
в подобных слоях

Для столбцов j матрицы, соответствующим этим 22 состояниям, весовая функция определяет как:

$$\begin{cases} o_{i,j} = d_j^{max} / d_{i,j} \\ o_{i,j} = 0 \text{ если } d_{i,j} = \infty \end{cases} \quad i, j = 1, 2, 3 \dots, N, \quad (9)$$

для остальных столбцов j :

$$\begin{cases} o_{i,j} = 1/N_j \\ o_{i,j} = 0 \text{ если } d_{i,j} = \infty \end{cases} \quad i, j = 1, 2, 3 \dots, N, \quad (10)$$

где $o_{i,j}$ – весовой коэффициент вероятности перехода между состояниями i и j , d_j^{max} – максимальное значение в j -й столбце, N_j – общее количество ненулевых значений в j -й столбце.

В соответствии с фундаментальной теоремой цепей Маркова матрица расстояний Хэмминга D , преобразованная в стохастическую матрицу слева B путём нормирования значений каждого столбца, можно утверждать, что этот процесс сходится.

Когда порог расстояния Хэмминга становится достаточно малым, т.е. несколько состояний не могут быть преобразованы в другое состояние, то это приводит к бесконечному циклу, и вероятности этих состояний можно представить как 1 или 0. Для этого необходимо ввести коэффициент затухания α и минимальную вероятность перехода $(1 - \alpha)/N$ в формулу (11) для регулировки стохастичности. Конечный рейтинг вероятности состояний P_{n+1} вычисляется с помощью выражения (11).

$$\begin{cases} \hat{B} = \alpha B + \frac{(1-\alpha)}{N} \mathbf{e} \mathbf{e}^T, \\ P_{n+1} = \hat{B} * P_n \end{cases}, \quad (11)$$

где \hat{B} – матрица вероятности перехода состояний после регулировки стохастичности, \mathbf{e} – вектор, все значения которого равны единицы, \mathbf{e}^T – транспонированная матрица, P_{n+1} – конечный рейтинг вероятности состояний.

Состояние, соответствующее изображению, которое аналогично исходному изображению, имеет более высокий рейтинг вероятности состояния. По мере изменения значения порога расстояния Хэмминга, точность и полнота алгоритма поиска изображения изменяются, как показано на Рис. 11. На Рис. 10 показывают изображения, соответствующие верхним 15 состояниям в рейтинге.



Рис. 10.
15 похожих изображений с
высоким рейтингом

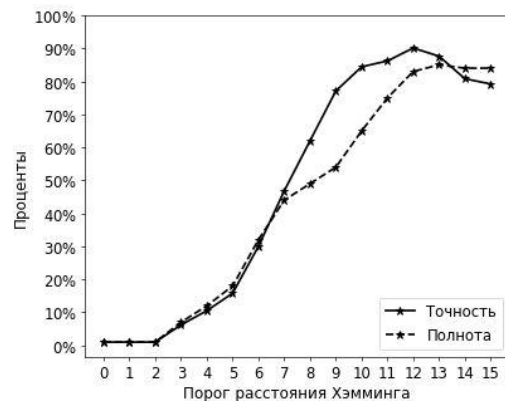


Рис. 11.
Точность и полнота после
рейтинга подобия при изменении
порога расстояния Хэмминга

В четвертой главе разработаны методы для распознавания лиц при классификации вейвлет-признаков изображений лиц, выделенных в первой главе.

Для коэффициентов приближений изображений лиц в четвёртом масштабе, создан 21-мерный тензор признаков на Рис. 12.

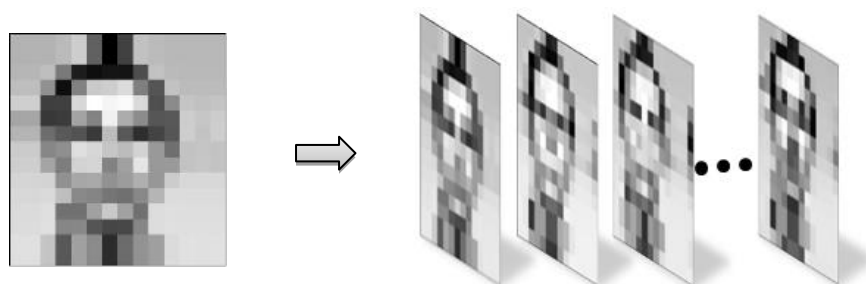


Рис. 12.
21-мерный тензор признаков, соответствующий
коэффициентами приближений в 4-ом масштабе

Набор данных FaceWarehouse: <http://kunzhou.net/zjugaps/facewarehouse/>, 150 людей, для каждого человека 24 изображения в естественном фоне, среди которых 10 изображений для обучения модели этого человека, остальные 14 изображений для распознавания лиц.

Использование скрытой Марковской модели (Hidden Markov Model, НММ) для классификации вейвлет-признаков изображений лиц.

Процесс обучения модели включает в себя следующие шаги:

Шаг 1. Выделение вейвлет-признаков изображений лиц – наблюдаемая последовательность O .

Шаг 2. Декорреляция данных признаков с помощью метода главных компонент (Principal Component Analysis, PCA).

Шаг 3. Создаём модель $\lambda = (A, B, \Pi)$, определяем число скрытых состояний N , для вероятностей начальных состояний $\Pi = \{\pi_i\}$. Определяем из условия, что $\pi_1 = 1$, $\pi_i = 0$ ($i \neq 1$), матрицу вероятностей переходов $A = \{a_{ij}\}$, определяем из условий $a_{ij} = 1$, ($i = j$), $a_{ij} = 0$ ($i \neq j$), матрицу вероятностей наблюдаемых последовательностей $B = \{b_j(o_t)\}$:

$$b_j(o_t) = \sum_{k=1}^M c_{jk} (2\pi)^{-N/2} |\Sigma_{jk}|^{-1/2} \exp[-(o_t - \mu_{jk})^T \Sigma_{jk}^{-1} (o_t - \mu_{jk})]/2, \quad (12)$$

где M – это размерность алфавита наблюдений, c_{jk} – это k -й весовой коэффициент смеси нормальных распределений в j -м скрытом состоянии, μ_{jk} и Σ_{jk} являются средним значением и ковариационной матрицей k -го компонента смеси в j -м скрытом состоянии, $j = 1, \dots, N, k = 1, \dots, M$:

$$\mu_{jk} = (\sum_{t=1}^{E_{jk}} o_t^{(j,k)}) / E_{jk}, \quad (13)$$

$$\Sigma_{jk} = [\sum_{t=1}^{E_{jk}} (o_t^{(j,k)} - \mu_{jk})(o_t^{(j,k)} - \mu_{jk})^T] / E_{jk}, \quad (14)$$

где E_{jk} – длина наблюдаемой последовательности, соответственно k -му компоненту смеси нормальных распределений в j -м скрытом состоянии, $o_t^{(j,k)}$ – соответственно наблюдаемые признаки изображений лиц.

На Рис. 13 (а) показан результат уменьшения размерности 21-мерного тензора до 3-мерного. Разделим каждое изображение в тензоре признаков на несколько частей в соответствии с числом скрытых состояний ($N = 5$). На Рис.13 (б) область, обведённую пунктиром, соответствует наблюдаемой последовательности первого скрытого состояния,

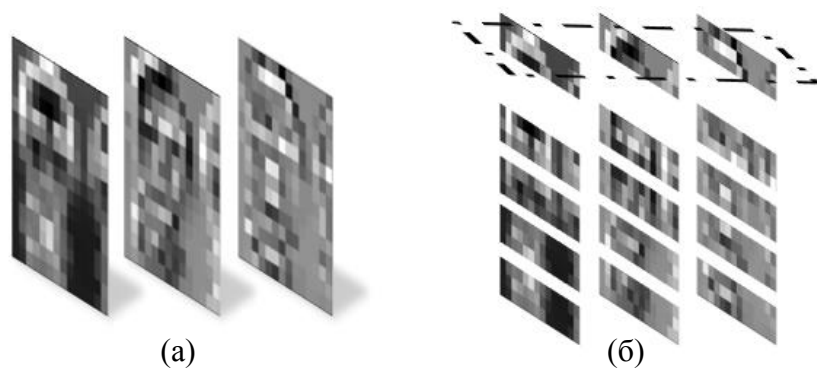


Рис. 13.

(а) Трёхмерный тензор признаков, уменьшены размерности методом PCA;

(б) Сегментация признаков по числу скрытых состояний ($N = 5$)

Шаг 4. Обучение модели с помощью алгоритма Баума-Велша. Находим модель $\lambda_1 = (A_1, B_1, \Pi_1)$, соответствующую первому человеку.

Шаг 5. Повторяем шаги 2 – 4, и переходим к обучению модели для всех лиц.

Шаг 6. Выделяем признаки изображения лица O_k в базе тестирования с помощью дерева коэффициентов для трёхмасштабного вейвлет-преобразования и методом PCA. Вычисляем вероятность $P(O_k|\lambda_i)$ того, что модель λ_i порождает O_k . Если n -я модель λ_n имеет наибольшую вероятность генерирования последовательности O_k , то это соответствует изображению n -му человеку. Блок-схема распознавания лиц с помощью скрытой Марковской модели представлена на Рис. 14.

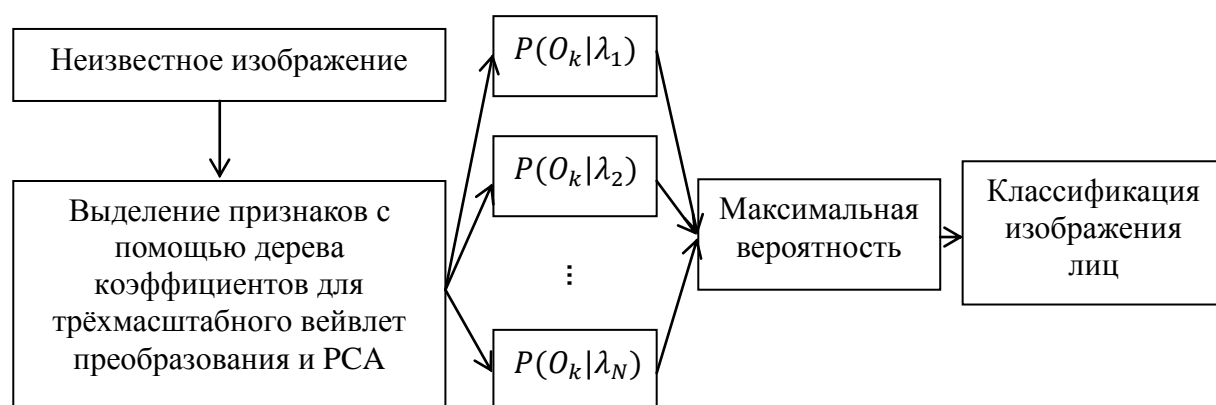


Рис. 14.

Блок-схема распознавания лиц скрытой Марковской модели

При условиях, что число скрытых состояний равно пяти, и размерность тензора признаков равна трём, точность распознавания лиц имеет максимальное значение 95,71%.

Для повышения точности распознавания лиц введены весовые коэффициенты вклада, которые определяются функцией двумерного нормального распределения с параметрами: коэффициент корреляции $\rho_{xy} = 0$, математические ожидания $\mu_x = \mu_y = 0$, дисперсии $\sigma_x = \sigma_y = 0,67$.

$$f(x, y) = \frac{(1-\rho_{xy}^2)^{-\frac{1}{2}}}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2(1-\rho_{xy}^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - \rho_{xy}\frac{2(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}. \quad (15)$$

При использовании взвешенных признаков для обучения моделей, точность распознавания лиц повысится. На Рис.15 приведены графики, характеризующие точность распознавания при изменении числа скрытых состояний S и количества обучающих изображений K .

На Рис.15 по оси x отложено число обучающих лиц K , ось y – проценты распознавания. Линия $\cdots*$ характеризует точность распознавания лиц, когда число скрытых состояний S равно 3, линия $-o$ соответствует случаю, когда число скрытых состояний равно 4, линия $-\triangleright$ – случаю, когда число скрытых состояний равно 5, линия $-\blacksquare$ – случаю, когда число скрытых состояний равно 6. Видно, что точность распознавания будет возрастать с увеличением числа обучающих изображений лиц, и когда $K = 10$ и $S = 5$, точность распознавания лиц имеет максимальное значение 98,57%.

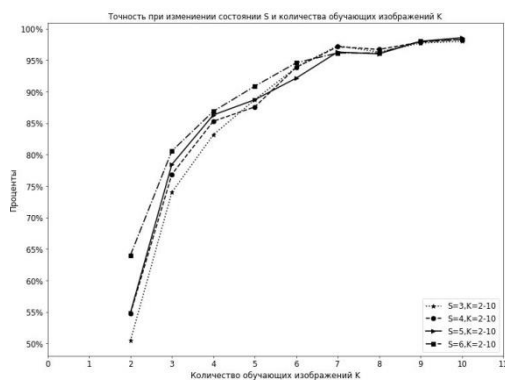


Рис. 15.

Точность распознавания лиц при изменении скрытых состояний и количества обучающих изображений

По сравнению с другими методами в машинном обучении и глубоком обучении, метод, основанный на классификации вейвлет-признаков с помощью скрытой Марковской модели, имеет следующие преимущества:

1. Сократилась размерность признаков изображений.
2. При добавлении новых образцов в модель, обучаются индивидуально только новые образцы.
3. Возрастает скорость обучения.
4. Обеспечивается более высокая точность распознавания лиц.

Использование вейвлет-нейронных сетей (Wavelet Neural Networks, WNN) для классификации вейвлет-признаков изображений лиц.

Вейвлет-нейронные сети имеют следующие преимущества по сравнению с искусственными нейронными сетями:

1. Из-за низкой корреляции нейронов в скрытом слое, вейвлет-нейронные сети сходятся быстрее, чем искусственные нейронные сети.
2. Переменные характеристики вейвлет-функции с помощью сдвига и сжатия значительно повышают аппроксимирующую способность сетей и преодолевают проблему локального минимума.
3. В тех же условиях вейвлет-нейронные сети имеют более высокую точность классификации, чем искусственные нейронные сети.

Алгоритм обратного распространения ошибки, который чаще всего используется для обучения многослойных искусственных нейронных сетей, имеет некоторые недостатки.

Во-первых, следует отметить его невысокую производительность и, во-вторых, склонность к попаданию в локальный минимум ошибки. В связи с этими была создана специализированная архитектура нейронных сетей, предназначенных для преодоления указанных недостатков. Так, например, для аппроксимации сигналов была предложена архитектура нейронной сети на основе вейвлетов — вейвлет-нейронные сети, обладающие хорошими показателями скорости и качеством обучения при использовании алгоритма Левенберга-Марквардта для обучения сети.

Блок-схема вейвлет-нейронных сетей приведена на Рис. 16.

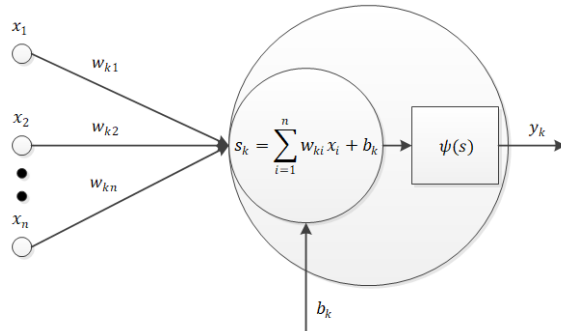


Рис. 16.

Нейрон вейвлет-нейронных сетей

где x_i – входные сигналы, совокупность всех входных сигналов нейрона образует вектор \mathbf{x} , w_{ki} – весовые коэффициенты, совокупность весовых коэффициентов образует вектор весов \mathbf{w} , b_k – постоянная величина сдвига, s_k – взвешенная сумма k -го нейрона входных сигналов, значение s_k передаётся на нелинейный элемент.

Активационная функция $\psi(x)$ формализуется в виде непрерывных вейвлет-функций, таких как POLYWOG (POLYnomials WindOwed with Gaussians) вейвлет-функции, которые определяются следующим образом:

POLYWOG1:

$$\psi_{POLYWOG1}(x) = k_1 x e^{-\frac{x^2}{2}}, \quad k_1 = \sqrt{e}; \quad (16)$$

POLYWOG2:

$$\psi_{POLYWOG2}(x) = k_2 (x^3 - 3x) e^{-\frac{x^2}{2}}, \quad k_2 = 0,7246; \quad (17)$$

POLYWOG3:

$$\psi_{POLYWOG3}(x) = k_3 (x^4 - 6x^2 + 3) e^{-\frac{x^2}{2}}, \quad k_3 = \frac{1}{3}; \quad (18)$$

POLYWOG4:

$$\psi_{POLYWOG4}(x) = (1 - x^2) e^{-\frac{x^2}{2}}. \quad (19)$$

Выходной сигнал k -го нейрона вейвлет-нейронных сетей можно смоделированы с помощью следующего выражения (20).

$$y_k = \psi\left(\frac{s_k - \hat{b}_k}{a_k}\right), \quad (20)$$

где y_k – выходной сигнал k -го нейрона, $\psi(x)$ – непрерывная вейвлет-функция, s_k – взвешенная сумма k -го нейрона входных сигналов \mathbf{x} , \hat{b}_k – коэффициент сдвига (translations) вейвлет-функции, a_k – коэффициент сжатия (dilations).

Большинство опубликованных научных исследований опираются на наиболее известный алгоритм обучения многослойных перцептронов – алгоритм градиентного спуска с обратным распространением ошибки. Однако в ряде публикации [Hinton, 1986] убедительно доказано, что этот достаточно универсальный алгоритм не обеспечивает наилучшей сходимости, требует достаточно много времени для обучения нейронных сетей, а также требует эмпирического или алгоритмического определения оптимальной скорости обучения, которая значительно влияет на результативность алгоритма.

Алгоритм Левенберга-Марквардта (Levenberg-Marquardt Algorithm, LMA) является наиболее распространённым алгоритмом для минимизации квадратичных отклонений. Его преимуществом, по сравнению с методом градиентного спуска, являются большая скорость счета и обеспечение сходимости.

В частности, он используется для оптимизации параметров нелинейных регрессионных моделей в предположении, что в качестве критерия оптимизации используется среднеквадратичная ошибка модели на обучающей выборке.

$$E(\mathbf{w}) = \frac{1}{2} \sum_{k=1}^M [e_k(\mathbf{w})]^2, \quad (21)$$

где $E(\mathbf{w})$ – среднеквадратичные ошибки (Mean Squared Error, MSE) между выходными данными и фактическими данными, M – число нейронов, ошибка $e_k(\mathbf{w})$ описывается формулой (22).

$$e_k = \hat{y}_k - y_k(\mathbf{w}), \quad (22)$$

где \hat{y}_k – фактические данные, y_k – выходные данные.

Алгоритм решения задачи минимизации среднеквадратичных ошибок получается в виде следующей итерации:

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \Delta \mathbf{w}^{(k)}, \quad (23)$$

где $\mathbf{w}^{(k+1)}$ – вектор весовых коэффициентов в $(k + 1)$ -й итерации, $\Delta \mathbf{w}^{(k)}$ – решение системы нелинейных уравнений (24).

$$[J^T(\mathbf{w})J(\mathbf{w}) - \mu I]\Delta \mathbf{w} = -J^T(\mathbf{w})\mathbf{e}(\mathbf{w}), \quad (24)$$

где $\mathbf{e}(\mathbf{w})$ – ошибки между выходными данными и фактическими данными, $J(\mathbf{w})$ – матрица Якоби, $J^T(\mathbf{w})$ – транспонированная матрица, μ – коэффициент регулирования, I – единичная матрица.

Ошибки $\mathbf{e}(\mathbf{w})$ определяются следующим образом:

$$\mathbf{e}(\mathbf{w}) = \begin{bmatrix} e_1(\mathbf{w}) \\ e_2(\mathbf{w}) \\ \vdots \\ e_M(\mathbf{w}) \end{bmatrix}. \quad (25)$$

Матрица Якоби $J(\mathbf{w})$ определяется в виде (26).

$$J(\mathbf{w}) = \begin{bmatrix} \frac{\partial e_1}{\partial w_1} & \frac{\partial e_1}{\partial w_2} & \dots & \frac{\partial e_1}{\partial w_N} \\ \frac{\partial e_2}{\partial w_1} & \frac{\partial e_2}{\partial w_2} & \dots & \frac{\partial e_2}{\partial w_N} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial e_M}{\partial w_1} & \frac{\partial e_M}{\partial w_2} & \dots & \frac{\partial e_M}{\partial w_N} \end{bmatrix}, \quad (26)$$

где N – число весовых коэффициентов w .

В экспериментах все сети имеют одинаковую структуру (1 скрытой слой, 180 нейронов в скрытом слое, learning rate $\lambda = 0,02$ для метода обратного распространения ошибки на основе алгоритма градиентного спуска, коэффициент регулирования $\mu = 0,05$ для алгоритма Левенберга-Марквардта) и используется алгоритм оптимизации Mini-batch стохастического градиентного спуска (Mini-batch stochastic gradient descent, Mini-batch SGD).

На Рис. 17 показаны связи между среднеквадратичными ошибками модели

(mse_loss), точностями распознавания в обучающих данных (accuracy on the training data) и числом итераций (epochs) в искусственных нейронных сетях, которые обучаются методом обратного распространения ошибки на основе алгоритма градиентного спуска.

Число нейронов в скрытом слое определяется формулой (27):

$$k = \sqrt{m \times n}, \quad (27)$$

где k – число нейронов в скрытом слое, m – число нейронов во входном слое, n – число нейронов в выходном слое.

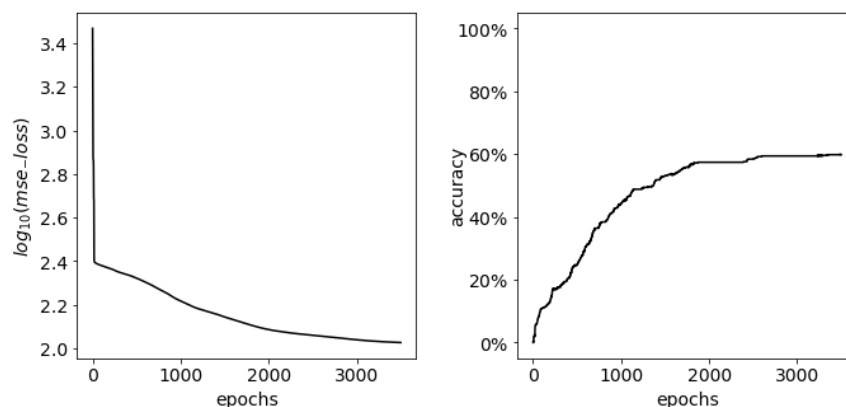


Рис. 17.

Среднеквадратичные ошибки и точности распознавания в обучающих данных модели ИНС

Ось x на Рис.17 характеризует число итераций, ось y в левом графике представляет собой логарифмы среднеквадратичных ошибок модели по основанию 10. Логарифмирование среднеквадратичных ошибок наглядно представлять скорость сходимости этой модели с увеличением числа итераций. Ось y в правом графике представляет собой точность распознавания лиц в обучающих данных. Видно, что скорость сходимости искусственных нейронных сетей является медленной, после 3000 итераций значение среднеквадратичных ошибок остается на уровне около 100, а точность распознавания в обучающих данных составляет 59,40%.

Используем вейвлет-нейронные сети с той же структурой сети, которые обучаются методом обратного распространения ошибки на основе алгоритма градиентного спуска.

Вейвлет-нейронные сети с 4 различными функциями активации имеют более высокую скорость сходимости и точность распознавания, чем искусственные нейронные сети. Лучшими сетями являются вейвлет-нейронные сети с функцией активации POLYWOG3, значение среднеквадратичных ошибок упало до 56 после 1500 итераций, а точность распознавания в обучающих данных достигла 82,00%.

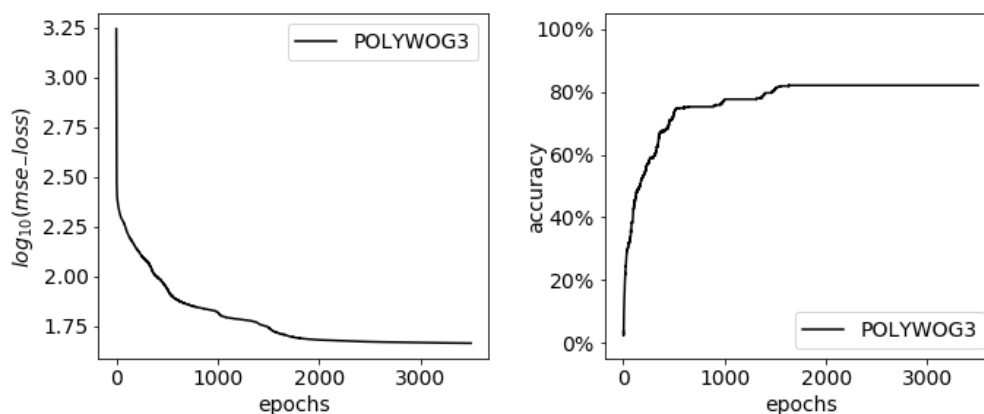


Рис. 18.

Среднеквадратичные ошибки и точности распознавания в обучающих данных модели ВНС при использовании алгоритма обратного распространения ошибки для обучения (функция активации POLYWOG3)

Затем используя алгоритм Левенберга-Марквардта для обучения тех же четырёх типов вейвлет-нейронных сетей, и лучшими сетями являются вейвлет-нейронные сети с функцией активации POLYWOG2, значение среднеквадратичных ошибок упало до 1,77 после 3000 итераций, и точность распознавания в обучающих данных осталась на уровне 100% после 800 итераций. Точность распознавания лиц вейвлет-нейронных сетей в данных тестирования составляют 99,20% (Рис. 19).

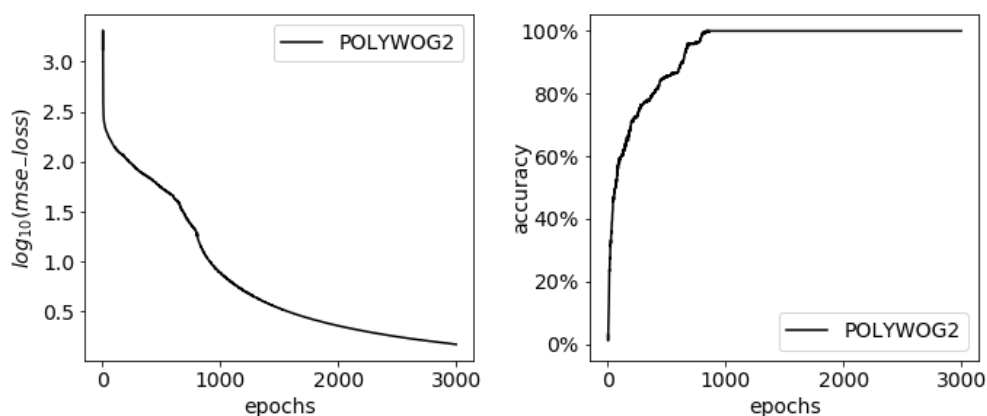


Рис. 19.

Среднеквадратичные ошибки и точности распознавания в обучающих данных модели ВНС при использовании алгоритма Левенберга-Марквардта для обучения (функция активации POLYWOG2)

В заключении подводятся итоги диссертационного исследования, излагаются его основные выводы и обобщающие результаты.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Разработан алгоритм для извлечения признаков изображений на основе дерева коэффициентов для кратномасштабного вейвлет-преобразования.
2. Разработан метод повышения чёткости обнаруженных границ объектов на изображении и повышена помехоустойчивость метода на основе сингулярного разложения матрицы с использованием вейвлет анализа.
3. Разработан метод для поиска похожих изображений, который выделяет в процентном отношении не только большее количество похожих изображений, но одновременно их полноту, с сохранением относительной точности и схожести выделенных изображений на основе высокочастотной информационной составляющей изображения, с помощью вейвлет анализа, и Марковской цепи.
4. Разработаны методы для распознавания лиц для сокращения времени обучения модели и повышения точность распознавания при классификации вейвлет-признаков. Для классификации вейвлет-признаков использованы скрытая Марковская модель и вейвлет-нейронные сети, обученные алгоритмом Левенберга-Марквардта.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Wang Liangpeng, Petrosyan O. G., Du Jiangming. Recognition of faces based on the coefficients tree of three scale wavelet transform // Automatic Control and Computer Sciences. 2019. Vol. 53, No.8. (принято к публикации).
2. Ван Лянпэн, Петросян О. Г. Использование новейших информационных технологий для обнаружения границ зашумлённого изображения // Информатизация образования и науки. 2018. №1. С. 58-67. (0,62п.л. / 0,31п.л.)
3. Ван Лянпэн, Петросян О. Г. Использование высокочастотных информационных составляющих изображений для поиска их похожих на основе системы CBIR // Информатизация образования и науки. 2018. №3. С. 106-115. (0,62п.л. / 0,31п.л.)
4. Ван Лянпэн, Петросян О. Г. Распознавание лиц на основе классификации вейвлет-признаков путём вейвлет-нейронных сетей // Информатизация образования и науки. 2018. №4. С. 129-139. (0,68п.л. / 0,34п.л.)
5. Ван Лянпэн, Петросян О. Г., Ду Цзяньмин. Распознавание лиц на основе дерева коэффициентов для трехмасштабного вейвлет-преобразования // Проблемы информационной безопасности. Компьютерные системы. 2018. №3. С. 126-136. (0,68п.л. / 0,34п.л.)
6. Ван Лянпэн, Петросян О. Г. Восстановление расплывчатых изображений в результате движения в среде LabVIEW // Инженерный вестник. 2015. №10. С. 523-530. (0,5п.л. / 0,25п.л.)