

ПРИМЕНЕНИЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ ХААРА, МЕТОДА ГЛАВНЫХ КОМПОНЕНТ И НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ СИМВОЛОВ НА ИЗОБРАЖЕНИЯХ В ПРИСУТСТВИИ ИМПУЛЬСНОГО ШУМА

В.Г. Спицын¹, Ю.А. Болотова¹, Н.Х. Фан², Т.Т.Ч. Буй²

¹Томский политехнический университет,

(национальный исследовательский университет) (ТПУ), Томск, Россия,

²Университет Ба Риа-Вунг, Вьетнам

Аннотация

В работе предложен новый алгоритм распознавания символов в условиях импульсного шума на основе применения вейвлет-преобразования, метода главных компонент и нейронных сетей. В предлагаемом алгоритме для устранения шума и извлечения признаков из исходного изображения и его отдельных частей с помощью вейвлет-преобразования Хаара выделяются низкочастотные компоненты. Сокращение размерности извлечённых признаков осуществляется методом главных компонент. В качестве классификатора используется многослойная нейронная сеть, на входы которой подаётся сокращённый набор признаков. Одной из ключевых особенностей предлагаемого подхода является создание отдельной нейронной сети для каждого типа символа. Результаты экспериментов показывают, что разработанный алгоритм сравним по точности распознавания с системами ABBYY FineReader и Tesseract в условиях импульсного шума.

Ключевые слова: распознавание символов, вейвлет-преобразование, метод главных компонент, нейронные сети.

Цитирование: Спицын, В.Г. Применение вейвлет-преобразования Хаара, метода главных компонент и нейронных сетей для оптического распознавания символов на изображениях в присутствии импульсного шума / В.Г. Спицын, Ю.А. Болотова, Н.Х. Фан, Т.Т.Ч. Буй // Компьютерная оптика. – 2016. – Т. 40, № 2. – С. 249-257. – DOI: 10.18287/2412-6179-2016-40-2-249-257.

Введение

Распознавание объектов на изображениях востребовано в различных сферах деятельности современного общества. Распознавание символов является одной из актуальных задач направления распознавания образов. Решение этой задачи может применяться в системах распознавания текстов, при аннотировании изображений, автоматическом переводе, для распознавания автомобильных номеров, регистрации железнодорожных составов и т.п. [1–7].

Основной трудностью при распознавании символов на изображениях, полученных путём сканирования, фотографирования или видеосъёмки, является наличие искажений и шумов. Точность распознавания даже признанными в данной области лидерами, такими как ABBYY FineReader или Tesseract OCR, значительно снижается в присутствии шума и искажений [6, 7].

К настоящему времени для решения задачи оптического распознавания символов предложено множество алгоритмов. Исторически первыми появились методы сопоставления с шаблонами, затем методы, анализирующие структуру символа [8]. Например, процесс распознавания в системе Tesseract основывается на полигональной аппроксимации символа и последующем сопоставлении полученной аппроксимации с сохранёнными на этапе обучения шаблонами [9]. В OCR-решениях ABBYY задействованы шесть классификаторов: растровый, признаковый, признаково-дифференциальный, контурный, структурный и структурно-дифференциальный, применяющиеся в зависимости от контекста документа, вход-

ных параметров изображения и задач распознавания [10]. В последнее время популярными стали методы, не требующие предварительной сегментации, такие как иерархические скрытые модели Маркова и свёрточные нейронные сети [18, 19]. Ввиду возникновения новой волны популярности нейросетевых классификаторов, в последнее время они часто используются в исследовательских работах по оптическому распознаванию символов [11].

Вейвлет-преобразование является эффективным методом выделения признаков объектов при решении задач распознавания. При разложении изображения вейвлет-преобразование разделяет его на субдиапазоны в низких и высоких частотных областях различного разрешения. Отбрасывание высокочастотных компонент, полученных при вейвлет-преобразовании, позволяет снизить размерность исходного изображения, а также избавиться от высокочастотного шума. К преимуществам вейвлет-преобразования относится и его быстрая вычислимость. В ряде работ показано, что при использовании вейвлет-преобразований для выделения признаков изображения точность классификации изображений составляет 76–99,7 % [12]. Кроме того, вейвлет-преобразование успешно используется при решении задач распознавания образов, в особенности задачи распознавания лиц [13, 14]. Точность распознавания лиц при этом составляет 90–98,5 %.

В данной работе предлагается новая технология оптического распознавания зашумлённых символов, основанная на комбинации алгоритма вейвлет-преобразования, метода главных компонент и нейронных сетей.

Предлагаемый алгоритм

Предлагаемый алгоритм для распознавания символов состоит из двух основных этапов: обучения и распознавания.

Этап 1. Обучение.

1.1. Выделение низкочастотных компонент изображения символа обучающей выборки с помощью вейвлет-преобразования Хаара.

1.2. Уменьшение размерности векторов выделенных признаков методом главных компонент.

1.3. Обучение нейронных сетей на полученных векторах признаков.

Этап 2. Распознавание символа.

2.1. Выделение низкочастотных компонент распознаваемого символа тестовой выборки на основе применения вейвлет-преобразования.

2.2. Уменьшение размерности вектора выделенных признаков с помощью метода главных компонент.

2.3. Распознавание символа обученными нейронными сетями.

Выделение признаков изображения символа с помощью преобразования Хаара

Выделение характерных признаков символов осуществляется с помощью вейвлет-преобразования Хаара [13]. Преобразование Хаара является базисным вейвлет-преобразованием. Пусть имеется одномерный дискретный сигнал $f(f_1, f_2, \dots, f_N)$, преобразование Хаара разлагает каждый сигнал на две составляющие равного размера. Первая составляющая называется средней, а вторая – детализирующей.

Средняя составляющая сигнала $a(a_1, a_2, \dots, a_{N/2})$ определяется как:

$$a_n = \frac{f_{2n-1} + f_{2n}}{\sqrt{2}}, n = 1, 2, 3, \dots, N/2. \tag{1}$$

Детализирующая составляющая $d(d_1, d_2, \dots, d_{N/2})$ описывается формулой (2):

$$d_n = \frac{f_{2n-1} - f_{2n}}{\sqrt{2}}, n = 1, 2, 3, \dots, N/2. \tag{2}$$

В данной работе преобразование Хаара применяется для выделения глобальных и локальных признаков.

Выделение глобальных признаков происходит следующим образом: область, содержащая символ, приводится к размеру 64×64 пикселя, затем к полученному изображению применяется вейвлет-преобразование, извлекаются низкочастотные вейвлет-коэффициенты, в результате чего получается матрица, состоящая из 32×32 низкочастотных вейвлет-коэффициентов.

Для выделения локальных характерных признаков символа его изображение делится на 12 частей одинакового размера – 32×32 пикселя согласно рис. 1, затем к каждой части применяется вейвлет-преобразование и извлекаются низкочастотные вейвлет-коэффициенты.

Исходное изображение I

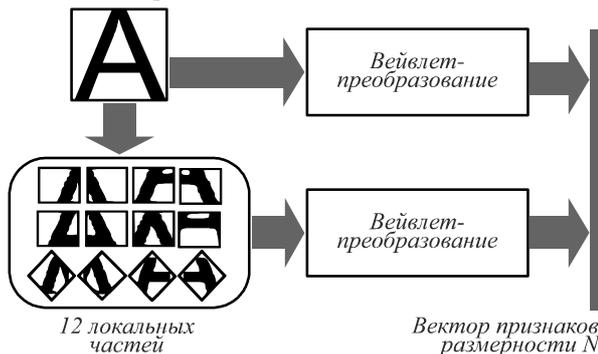


Рис. 1. Выделение признаков символа «А»

В результате получаются 12 матриц, каждая из которых состоит из 16×16 низкочастотных вейвлет-коэффициентов.

После этого формируется вектор характерных признаков символа, элементами которого являются все низкочастотные вейвлет-коэффициенты, полученные на предыдущих шагах. В результате формируется вектор характерных признаков символа, состоящий из $32 \times 32 + 12 \times 16 \times 16 = 4096$ элементов.

Снижение размерности вектора признаков

Перед подачей на входы нейронных сетей размерность вектора признаков необходимо уменьшить. Для решения этой задачи предлагается использовать метод главных компонент [14].

Для использования данного метода требуется его предварительное обучение.

Сначала с использованием множества M обучающих изображений символов создается пространство «собственных символов». При этом $M \ll N$, где $N = 4096$. Создание пространства собственных символов выполняется следующим образом.

К каждому из M изображений применяется разработанный способ выделения характерных признаков. В результате получается набор векторов признаков $\vec{I}_1, \dots, \vec{I}_M$, размерность каждого из которых равна N . Затем составляется средний вектор, значение каждого элемента которого по всем M векторам признаков вычисляется по формуле:

$$\vec{I}_{avg} = \frac{1}{M} \sum_{n=1}^M \vec{I}_n. \tag{3}$$

Далее из обучающих векторов формируется матрица A размерностью $N \times M$. Столбцами матрицы A является разность вектора признаков обучающих изображений и среднего вектора:

$$\vec{\Phi}_n = \vec{I}_n - \vec{I}_{avg}, n = 1, \dots, M, \tag{4}$$

$$A = [\vec{\Phi}_1, \vec{\Phi}_2, \dots, \vec{\Phi}_M]. \tag{5}$$

Вычисляются собственные векторы ковариационной матрицы C размера $N \times N$, определяемой соотношением (4):

$$C = AA^T. \tag{6}$$

Но ввиду её большой размерности можно рассмотреть матрицу $L [M \times M]$, определяемую:

$$L = A^T A. \tag{7}$$

В книге А.А. Лукьяница [15] доказано, что собственные векторы $U [N \times M]$ матрицы C могут быть получены путём линейной комбинации A и собственных векторов V матрицы L :

$$U = AV. \tag{8}$$

После этого создаётся пространство, состоящее из K собственных векторов \vec{u}_k , имеющих наибольшие собственные значения. При этом каждый k -й вектор \vec{u}_k удовлетворяет условию максимизации следующего выражения (9):

$$\lambda_k = \frac{1}{M} \sum_{n=1}^M (\vec{u}_k^T \vec{\Phi}_n)^2 \tag{9}$$

и условию ортогональности (10):

$$\vec{u}_l^T \vec{u}_k = \begin{cases} 1, & l = k \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Векторы \vec{u}_k и величины λ_k представляют собственные векторы и собственные значения ковариационной матрицы C . Пространством «собственных символов» является набор выбранных K собственных векторов (рис. 2).

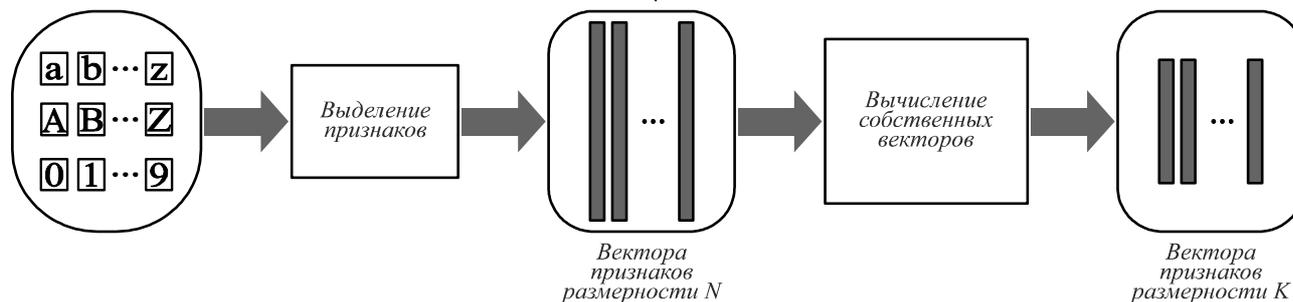


Рис. 2. Создание пространства собственных символов

После того как создано пространство «собственных символов», уменьшение размерности вектора характерных признаков символа \vec{I}_{in} осуществляется следующим образом [14].

Сначала вектор признаков символа разлагается по K имеющимся «собственным символам» \vec{u}_i и вычисляются соответствующие коэффициенты разложения, определяющиеся по формуле (11):

$$w_i = \vec{u}_i^T (\vec{I}_{in} - \vec{I}_{avg}), i = 1, \dots, K. \tag{11}$$

Затем составляется вектор, описывающий вклад каждого собственного символа в представление входного вектора признаков символа:

$$\vec{\Omega}^T = \{w_1, \dots, w_K\}. \tag{12}$$

В результате уменьшения размерности получается новый вектор признаков символа $\vec{\Omega}$, состоящий из K элементов. При этом $K \ll N$ (рис. 3).

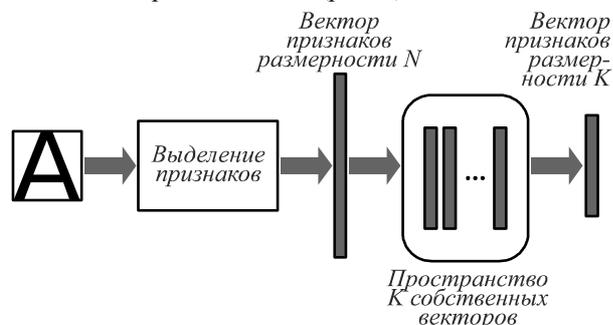


Рис. 3. Уменьшение размерности вектора признаков символа

Распознавание символов набором нейронных сетей

В качестве классификатора для распознавания символов предлагается использование многослойных нейронных сетей. Особенностью данного подхода является создание для каждого символа специальной нейронной сети, обучаемой алгоритмом обратного распространения ошибки [16].

Структура нейронных сетей идентична и представлена 1 входным, 1 скрытым и 1 выходным слоем. Входом каждой нейронной сети является вектор характерных признаков $\vec{\Omega}$ размерности K . Количество нейронов в скрытом слое равно 70% от количества входных нейронов. Выходной слой имеет только один нейрон, возвращающий значение в пределах от 0 до 1.

Сети обучаются алгоритмом обратного распространения ошибки, состоящим из следующих этапов:

1. Инициализация. Случайная генерация синаптических весов с помощью датчика равномерно распределённых чисел со средним значением 0.
2. Предъявление примеров обучения. На входы сети подаётся вектор значений текущего экземпляра n обучающей выборки $(x(n), d(n))$, где $x(n)$ – входной вектор, $d(n)$ – желаемый отклик сети на выходе.
3. Прямой проход. Последовательно для каждого слоя l , начиная с первого, вычисляется индуцированное поле каждого нейрона j по формуле (13):

$$v_j^{(l)}(n) = \sum_{i=0}^{m_0} w_{ji}^{(l)}(n) y_i^{(l-1)}(n), \tag{13}$$

где $y_i^{(l-1)}(n)$ – выходной сигнал нейрона i предыдущего слоя $(l-1)$, $w_{ji}^{(l)}(n)$ – синаптический вес связи нейрона j слоя l с нейроном i слоя $l-1$.

Выходной сигнал нейрона вычисляется по формуле (14):

$$y_j^{(l)}(n) = \varphi_j(v_j(n)), \tag{14}$$

где $\varphi_j(v_j(n))$ – активационная функция.

Нейроны первого скрытого слоя ($l = 1$) транслируют входные данные нейронам следующего слоя:

$$y_j^{(0)}(n) = x_j(n). \tag{15}$$

Для нейронов выходного слоя ($l = L$):

$$\delta_j^{(l)}(n) = \begin{cases} e_j^{(l)}(n) \varphi_j'(v_j^{(l)}(n)) \text{ для нейрона } j \text{ выходного слоя } L \\ \varphi_j'(v_j^{(l)}(n)) \sum_k \delta_k^{(l+1)}(n) \text{ для нейрона } j \text{ скрытого слоя } l \end{cases}. \tag{18}$$

Изменение синаптических весов осуществляется по формуле (19):

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n), \tag{19}$$

где η – параметр скорости обучения.

5. Итерации. Повторение шага 2 для следующего экземпляра обучающей выборки. При обработке всей обучающей выборки считается, что пройдена 1 эпоха обучения. Переход на следующую эпоху обучения осуществляется в случае, если критерий останова не был достигнут [17].

Использование отдельной нейронной сети для каждого символа позволяет ускорить процесс обучения за счёт параллельного обучения сетей. Каждая

$$y_j^{(L)}(n) = o_j(n). \tag{16}$$

После осуществления прямого прохода вычисляется ошибка для нейрона выходного слоя:

$$e_j(n) = d_j(n) - o_j(n). \tag{17}$$

4. Обратный проход. Локальные градиенты узлов сети вычисляются по формуле (18), где $\varphi_j'(\cdot)$ – дифференцирование по аргументу:

нейронная сеть определяет степень близости распознаваемого символа только к одному из символов обучающей выборки.

Распознавание входного символа нейронными сетями происходит следующим образом. Вначале извлекается вектор признаков символа и уменьшается его размерность. Затем полученный вектор признаков подаётся на входы всех обученных нейронных сетей, осуществляется распознавание с помощью прямого прохода по нейронной сети. Входной символ распознаётся как символ обучающей выборки, нейронная сеть которого возвращает наибольшее значение (рис. 4).

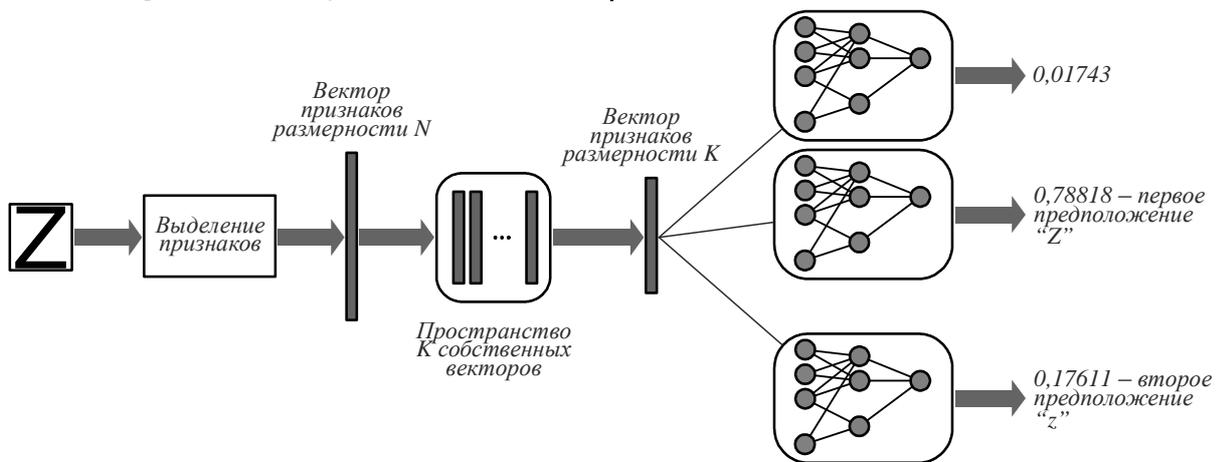


Рис. 4. Распознавание символа нейронными сетями

Оценка сложности алгоритма

Вейвлет-преобразование, применяемое в данной работе, является однократным и имеет квадратичную сложность $O(N^2)$.

Метод главных компонент требует вычисления собственных векторов и собственных значений матрицы ковариации размерности $N \times N$, сложность этой операции составляет $O(N^3)$. При $N \gg M$ в работе используется модификация [15], сложность составляет $O(M^3)$.

При создании многослойного персептрона строится нелинейная модель физического явления, обеспечивающая определённый выходной сигнал в ответ на входную информацию. Применяемый в данной работе алгоритм обучения обратного рас-

пространения ошибки имеет линейную сложность $O(w)$ по синаптическим весам w . При прямом проходе синаптические веса задействованы для вычисления выходных значений отдельных нейронов сети путём линейной комбинации со входными значениями. При обратном проходе синаптические веса используются для вычисления локальных градиентов скрытых нейронов и последующего изменения самих синаптических весов.

Заметим, что метод главных компонент обладает наибольшей сложностью среди всех этапов предлагаемого алгоритма, однако вычисление главных компонент требуется только во время обучения. На этапе распознавания сложность метода определяется как $O(N \times K)$.

Эксперименты по распознаванию рукописных и печатных символов

В первом эксперименте предложенный алгоритм был применён для распознавания рукописных цифр базы MNIST, созданной Яном Лекуном [18]. База состоит из 60000 изображений для обучения и 10000 тестовых изображений. Цифры центрированы на изображении размером 28×28 пикселей.

В этом эксперименте авторами статьи была создана дополнительная тестовая выборка путём добавления к исходным изображениям импульсного шума с вероятностью 5, 10, 15, 20, 25 и 30 %. Пример зашумлённого изображения приведён на рис. 5 (слева направо: изображение рукописных цифр с шумом 0, 10, 20 и 30 % соответственно).

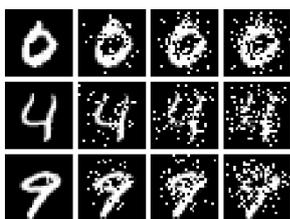


Рис. 5. Пример тестовых изображений рукописных цифр базы MNIST с добавлением шума

Исследования влияния размерности вектора признаков на точность, достигнутую разработанным алгоритмом при распознавании рукописных цифр базы MNIST, показали более устойчивую работу алгоритма при увеличении количества использованных признаков. При использовании вектора из 5 признаков точность распознавания составляет только 70,3%. При количестве использованных признаков, равном 37, точность распознавания составляет 97%. При количестве признаков больше 37 точность распознавания находится в пределах от 97% до 97,5%. При использовании вектора из 49 признаков точность распознавания достигает максимального значения 97,5%. Таким образом, для проведения дальнейших численных экспериментов по распознаванию рукописных цифр на зашумлённых выборках был выбран вектор, состоящий из 49 признаков.

Результаты распознавания предложенного алгоритма (табл. 1) оказались сопоставимыми с лучшими алгоритмами, представленными другими авторами [18–20].

Табл. 1. Результаты распознавания рукописных цифр MNIST различными алгоритмами

Алгоритм	Мин. ошибка, %	Макс. ошибка, %
Линейный классификатор [19]	7,6	12
Нелинейный классификатор [19]	3,3	3,6
Модель иерархической временной памяти [20]	1,5	1,5
Boosted Stumps [19]	0,87	7,7
К-ближайших соседей [19]	0,63	5
Машины опорных векторов [19]	0,56	1,4
Нейронные сети [19]	0,35	4,7
Свёрточные нейронные сети [19]	0,23	1,7
Предложенный алгоритм	2,5	3

Анализ табл. 1 показывает, что предложенный алгоритм является сопоставимым с нелинейным классификатором и нейронными сетями, превосходит по

точности линейный классификатор, boosted stumps, к-ближайших соседей, уступает модели иерархической временной памяти, машинам опорных векторов и свёрточным нейронным сетям, однако данные алгоритмы имеют большую вычислительную сложность по сравнению с предложенным.

При тестировании разработанного алгоритма на зашумлённой выборке рукописных цифр базы MNIST также использовался вектор из 49 признаков. Результаты распознавания рукописных цифр разработанным алгоритмом на зашумлённых выборках представлены на рис. 6.

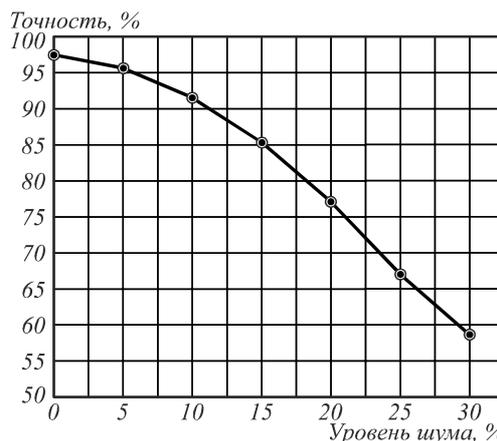


Рис. 6. Результаты распознавания рукописных цифр в зависимости от уровня шума

Следует отметить, что разработанный алгоритм обеспечивает возможность распознавания рукописных цифр в присутствии шума на изображениях. При этом точность распознавания существенно уменьшается, когда уровень шума превышает 20%.

Эксперименты по распознаванию печатных символов

Проводились эксперименты по распознаванию печатных символов латинского алфавита. Обучающая выборка состоит из 1488 изображений 10 цифр (0–9) и 52 букв латинского алфавита (a–z, A–Z). Каждый символ представлен 2 типами шрифтов: Times New Roman и Arial с жирным и нормальным начертанием, 16, 18, 20, 22, 24 и 26 кегля. Таким образом, в обучающей выборке каждый символ представлен 24 изображениями.

Тестирование проводилось на символах 8 популярных шрифтов: 4 шрифта с засечками – Times New Roman, Garamond, Courier New и Bookman Old Style, 4 шрифта без засечек – Arial, Lucida Sans, Tahoma и Verdana. Тестовая выборка состоит из 2480 изображений. Каждый символ представлен нормальным и жирным начертанием 12, 14, 16, 18, 20, 22, 24, 26, 28 и 36 кегля.

В случае распознавания символов проводились эксперименты по учёту второго предположения. Второе предположение в данном случае – это решение нейронной сети, возвратившей второе по величине значение. Учёт второго предположения применяется при распознавании символов, имеющих «похожее» написание заглавных и строчных букв, таких как {c, C}, {o, O}, {p, P}, {s, S}, {u, U}, {v, V}, {w, W}, {x, X} и {z, Z}.

Суть использования второго предположения заключается в том, что если в результате распознавания текущего символа первое и второе предположения представляют собой одну из вышеперечисленных пар, то вводится дополнительная проверка, при которой распознаваемый символ подаётся на вход специальной нейронной сети, предварительно обученной разделять символы вышеперечисленных пар строчных и прописных букв. Таким образом, для этой дополнительной проверки на этапе обучения создаётся ещё 9 дополнительных нейронных сетей (по одной на каждую пару «похожих» символов).

Результаты экспериментов показали (табл. 2), что предложенный алгоритм, обученный на 2 типах шрифтов, способен успешно распознавать символы других шрифтов.

Табл. 2. Результаты распознавания печатных символов разных шрифтов

Шрифт	Точность распознавания, %	
	Второе предположение	Первое предположение
<i>Times New Roman</i>	96,5	91,5
<i>Courier New</i>	77,4	63,5
<i>Garamond</i>	94,0	84,3
<i>Bookman Old Style</i>	88,5	78,6
<i>Arial</i>	97,1	92,4
<i>Lucida Sans</i>	90,6	85,4
<i>Tahoma</i>	91,5	80,4
<i>Verdana</i>	91,6	80,2

Результаты, приведённые в табл. 2, показывают, что точность распознавания предложенного алгоритма для всех тестируемых шрифтов является приемлемой. Точность распознавания на шрифтах без засечек выше и стабильнее, чем на шрифтах с засечками. Наилучшие результаты достигнуты при использовании 27 значимых признаков, полученных методом главных компонент. Вектор такого размера был использован для тестирования печатных символов с шумом.

Тестовая выборка была увеличена добавлением к исходным изображениям импульсного шума 5, 10, 15, 20, 25 и 30 %. Пример чистых и зашумлённых изображений приведен на рис. 7 (процент зашумления – 0, 10, 20 и 30 % слева – направо).



Рис. 7. Пример изображений «чистых» и «зашумлённых» печатных символов

Сопоставление точности распознавания систем *ABBYY FineReader 11* и *Tesseract OCR* на зашумлённых печатных символах шрифтов *Times New Roman* и *Arial* проводилось на тех же выборках, результаты распознавания приведены в табл. 3 и 4.

Показано, что при увеличении шума точность распознавания системами *ABBYY FineReader 11* and *Tesseract OCR* значительно снижается при зашумлении от 10 до 30 %, точность предложенного алгоритма снижа-

ется медленнее. Таким образом, можно сделать вывод, что предложенный алгоритм более эффективен при распознавании зашумлённых изображений символов, чем *ABBYY FineReader 11* и *Tesseract OCR*. При уровне шума более 15 % разница в точности распознавания становится более заметной (табл. 3, 4).

Табл. 3. Сравнение точности распознавания зашумлённых символов шрифта *Times New Roman* различными алгоритмами

Алгоритм распознавания	Уровень шума, %						
	0	5	10	15	20	25	30
<i>FineReader</i>	99,9	86,3	79	65,8	59,8	54,8	43,6
<i>Tesseract</i>	99,8	86,5	84,8	77,3	69,4	64	53
Предложенный алгоритм	91,5	89,8	89,6	86,8	83,2	81,5	78,1
Предложенный алгоритм с учётом второго предположения	96,5	95,9	95	93,2	90,8	89,8	86

Табл. 4. Сравнение точности распознавания зашумлённых символов шрифта *Arial* различными алгоритмами

Алгоритм распознавания	Уровень шума, %						
	0	5	10	15	20	25	30
<i>FineReader</i>	99,8	90,6	83,7	76,8	67,8	59	55,2
<i>Tesseract</i>	99,4	86,5	83,6	75,7	69,6	62,7	52,2
Предложенный алгоритм	92,4	88,5	87,8	85,6	85,4	81,5	78,3
Предложенный алгоритм с учётом второго предположения	97,1	96,3	95,2	93,5	92,4	90,2	88

Заключение

В данной работе представлен эффективный алгоритм распознавания рукописных и печатных символов, основанный на совместном применении вейвлет-преобразования, метода главных компонент и нейронных сетей. Преимуществом данного алгоритма над существующими решениями является его способность более качественно распознавать символы в присутствии импульсного шума. Этого удалось достичь благодаря применению вейвлет-преобразования Хаара и нейронных сетей. Заметим, что точность распознавания увеличивается в результате использования второго предположения. В работе показано, что предложенный алгоритм позволяет эффективно распознавать зашумлённые изображения рукописных и печатных символов.

Литература

- Bolotova, Yu.A.** License plate recognition with hierarchical temporal memory model / Yu.A. Bolotova, A.A. Druki, V.G. Spitsyn. – In: Proceedings of 9th International Forum on Strategic Technology (IFOST-2014), Chittagong, October 21-23, 2014. – Chittagong: CUET, 2014. – P. 121-124.
- Болотова, Ю.А.** Распознавание автомобильных номеров на основе метода связанных компонент и иерархической временной сети / Ю.А. Болотова, В.Г. Спицын, М.Н. Рудомёткина // Компьютерная оптика. – 2015. – Т. 39, № 2. – С. 275-280.
- Болотова, Ю.А.** Распознавание символов на цветном фоне на основе иерархической временной модели с предобработкой фильтрами Габора / Ю.А. Болотова, А.К. Кермани, В.Г. Спицын // Электромагнитные волны и электронные системы. – 2012. – Т. 16, № 8. – С. 14-19.

4. **Казанский, Н.Л.** Распределённая система технического зрения регистрации железнодорожных составов / Н.Л. Казанский, С.Б. Попов // Компьютерная оптика. – 2012. – Т. 36, № 3. – С. 419-428.
5. **Изотов, П.Ю.** Технология реализации нейросетевого алгоритма в среде CUDA на примере распознавания рукописных цифр / П.Ю. Изотов, С.В. Суханов, Д.Л. Головашкин // Компьютерная оптика. – 2010. – Т. 34, № 2. – С. 243-251.
6. **Breuel, T.M.** High-Performance OCR for Printed English and Fraktur using LSTM Networks / T.M. Breuel, A. Ull-Hasan, M.A.I. Azawi, F. Shafait // Proceedings of 12th International Conference on Document Analysis and Recognition. – 2013. – P. 683-687.
7. **Heliński, M.** Report on the comparison of Tesseract and ABBYY FineReader OCR engines / M. Heliński, M. Kmiecik, T. Parkoła. – Technical report. – Poznań, Poland: Poznań Supercomputing and Networking Center, 2012. – 24 p.
8. **Mori, Sh.** Historical review of OCR research and development / Sh. Mori, Ch.Y. Suen, K. Yamamoto // Proceedings of the IEEE. – 1992. – Vol. 80, Issue 7. – P. 1029-1058.
9. **Smith, R.** An overview of the Tesseract OCR Engine / R. Smith // Proceedings of 9th International Conference on Document Analysis and Recognition (ICDAR2007). – 2007. – Volume II. – P. 629-633.
10. **Крупин, А.** ABBYFineReader: взгляд изнутри [Электронный ресурс] / А. Крупин. – URL: <http://www.3dnews.ru/software/632560/>.
11. **Breuel, T.M.** The OCRopus Open Source OCR System Tech / T.M. Breuel // Proceedings of SPIE. – 2008. – Vol. 6815. – 68150F. – DOI: 10.1117/12.783598.
12. **Mehdi, L.** Combining wavelet transforms and neural networks for image classification / L. Mehdi, A. Solimani, A. Dargazany. – In: 41st Southeastern Symposium on System Theory, Tullahoma, TN, USA. – Tullahoma, TN, USA: 2009. – P. 44-48.
13. **Буй, Т.Т.Ч.** Распознавание лиц на основе применения метода Виолы-Джонса, вейвлет-преобразования и метода главных компонент / Т.Т.Ч. Буй, Н.Х. Фан, В.Г. Спицын // Известия Томского политехнического университета. – 2011. – Т. 320, № 5. – С. 54-59.
14. **Turk, M.A.** Face Recognition Using Eigenfaces / M.A. Turk, A.P. Pentland // Proceedings of the IEEE. – 1991. – P. 586-591.
15. **Лукьянича, А.А.** Цифровая обработка видеоизображений / А.А. Лукьянича, А.Г. Шишкин. – М.: Ай-Эс-Эс Пресс, 2009. – 518 с.
16. **Спицын, В.Г.** Интеллектуальные системы: учебное пособие / В.Г. Спицын, Ю.П. Цой. – Томск: Изд-во Томского политехнического университета, 2012. – 176 с.
17. **Хайкин, С.** Нейронные сети. Полный курс: / С. Хайкин; пер. с англ. – 2-е изд. – М.: Издательский дом «Вильямс», 2006. – 1104 с.
18. **LeCun, Y.** Gradient based learning applied to document recognition / Y. LeCun, L. Bottou, Y. Bengio, P. Haffner // Proceedings of the IEEE. – 1998. – Vol. 86, Issue 11. – P. 2278-2324.
19. **LeCun, Y.** Convolutional networks and applications in vision / Y. LeCun, K. Kavukcuoglu, C. Farabet. // International Symposium on Circuits and Systems. – Paris: 2010. – P. 253-256.
20. **Bolotova, Yu.A.** Analysis of hierarchically-temporal dependencies for handwritten symbols and gesture recognition / Yu.A. Bolotova, V.G. Spitsyn // 7th International Forum on Strategic Technology (IFOST-2012), Tomsk. – 2012. – Vol. 1. – P. 596-601.

Сведения об авторах

Спицын Владимир Григорьевич, 1948 года рождения, в 1970 году окончил Томский государственный университет по специальности «Радиофизика и электроника», профессор, д.т.н., профессор Национального исследовательского Томского политехнического университета. Область научных интересов: нейронные сети, обработка изображений, распространение электромагнитных волн в случайно-неоднородных средах.

E-mail: spvg@tpu.ru.

Болотова Юлия Александровна, 1986 года рождения, в 2009 году окончила Томский политехнический университет по специальности «Программное обеспечение вычислительной техники и автоматизированных систем», к.т.н., доцент кафедры вычислительной техники Томского политехнического университета. Область научных интересов: обработка изображений, распознавание образов, биологически-подобные алгоритмы.

E-mail: julya21@hotmail.ru.

Фан Нгок Хоанг, 1983 года рождения, в 2008 году окончил государственное образовательное учреждение высшего профессионального образования «Иркутский государственный технический университет», к.т.н., зам. декана факультета информационных технологий, Университет Ба Риа-Вунг Тау, Вьетнам. Область научных интересов: обработка изображений, вейвлет-преобразования, нейронные сети, распознавание образов.

E-mail: hoangpn285@gmail.com.

Буй Тхи Тху Чанг, 1984 года рождения, в 2009 году окончила государственное образовательное учреждение высшего профессионального образования «Томский политехнический университет», к.т.н., зав. кафедрой вычислительной техники факультета информационных технологий, Университет Ба Риа-Вунг Тау, Вьетнам. Область научных интересов: обработка изображений, вейвлет-преобразования, нейронные сети, распознавание образов. E-mail: trangbt.084@gmail.com.

Поступила в редакцию 3 февраля 2016 г. Окончательный вариант – 16 марта 2016 г.

USING A HAAR WAVELET TRANSFORM, PRINCIPAL COMPONENT ANALYSIS AND NEURAL NETWORKS FOR OCR IN THE PRESENCE OF IMPULSE NOISE

V.G. Spitsyn¹, Yu.A. Bolotova¹, N.H. Phan², T.T.T. Bui²

¹ Tomsk Polytechnic University, Tomsk, Russia,

² Ba Ria-Vung Tau University, Vietnam

Abstract

In this paper we propose a novel algorithm for optical character recognition in the presence of impulse noise by applying a wavelet transform, principal component analysis, and neural networks. In the proposed algorithm, the Haar wavelet transform is used for low frequency components allocation, noise elimination and feature extraction. The principal component analysis is used to reduce the dimension of the extracted features. We use a set of different multi-layer neural networks as classifiers for each character; the inputs are represented by a reduced set of features. One of the key features of the proposed approach is creating a separate neural network for each type of character. The experimental results show that the proposed algorithm can effectively recognize the characters in images in the presence of impulse noise; the results are comparable with ABBYY FineReader and Tesseract OCR.

Keywords: optical character recognition; wavelet transform; principal component analysis; neural networks.

Citation: Spitsyn VG, Bolotova YuA, Phan NH, Bui TTT. Using a Haar wavelet transform, principal component analysis and neural networks for OCR in the presence of impulse noise. *Computer Optics* 2016; 40(2): 249-257. DOI: 10.18287/2412-6179-2016-40-2-249-257.

References

- [1] Bolotova YuA, Druki AA, Spitsyn VG. License plate recognition with hierarchical temporal memory model. Proceedings of 9th International Forum on Strategic Technology (IFOST-2014). Chittagong: CUET; 2014: 121-124.
- [2] Bolotova YuA, Spitsyn VG, Rudometkina MN. License plate recognition algorithm on the base of connected components method and hierarchical temporal memory model. *Computer Optics* 2015; 39(2): 275-280.
- [3] Bolotova YuA, Kermani AK, Spitsyn VG. Colored background symbols recognition on the base of hierarchical temporal memory model with Gabor filter preprocessing [In Russian]. *Electromagnetic Waves and Electronic Systems* 2012; 17(1): 14-19.
- [4] Kazanskiy NL, Popov SB. The distributed vision system of the registration of the railway train. *Computer Optics* 2012; 36(3): 419-428.
- [5] Izotov PY, Sukhanov SV, Golovashkin DL. Technology of implementation of neural network algorithm in cuda environment at the example of handwritten digits recognition. *Computer Optics* 2010; 34(2): 243-251.
- [6] Breuel TM, Ul-Hasan A, Azawi MAI, Shafait F. High-Performance OCR for Printed English and Fraktur using LSTM Networks. Proc. of 12th International Conference on Document Analysis and Recognition 2013: 683-687.
- [7] Heliński M, Kmiecik M, Parkoła T. Report on the comparison of Tesseract and ABBYY FineReader OCR engines. Technical report. Poznań, Poland: Poznań Supercomputing and Networking Center; 2012.
- [8] Mori Sh, Suen ChY, Yamamoto K. Historical review of OCR research and development. *Proceedings of the IEEE* 1992; 80(7): 1029-1058.
- [9] Smith R. An overview of the Tesseract OCR Engine. *Proceedings of 9th International Conference on Document Analysis and Recognition (ICDAR2007)* 2007; II: 629-633.
- [10] Krupin A. ABBYFineReader: View from Inside [In Russian]. Source: (<http://www.3dnews.ru/software/632560/>).
- [11] Breuel TM. The OCRopus Open Source OCR System Tech. *Proceedings of SPIE* 2008; 6815: 68150F.
- [12] Mehdi L, Solimani A, Dargazany A. Combining wavelet transforms and neural networks for image classification. In: 41st Southeastern Symposium on System Theory. Tullahoma, TN, USA; 2009: 44-48.
- [13] Bui TTT, Phan NH, Spitsyn VG. Face recognition using Viola-Jones method, wavelet transforms and principal component analysis [In Russian]. *Bulletin of the Tomsk Polytechnic University* 2011; 320(5): 54-59.
- [14] Turk MA, Pentland AP. Face Recognition Using Eigenfaces. *Proc IEEE* 1991: 586-591.
- [15] Luk'janica AA, Shishkin AG. Digital video processing [In Russian]. Moscow: ISS Press; 2009.
- [16] Spitsyn VG, Tsoy YuR. Intelligent systems. Tomsk: Tomsk Polytechnic University Press; 2012.
- [17] Haykin S. Neural Networks: A Comprehensive Foundation Second Edition [In Russian]. Moscow: Williams Publishing; 2006.
- [18] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient based learning applied to document recognition. *Proceedings of the IEEE* 1998; 86(11): 2278-2324.
- [19] LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision. *International Symposium on Circuits and Systems*. Paris; 2010: 253-256.
- [20] Bolotova YuA, Spitsyn VG. Analysis of hierarchically-temporal dependencies for handwritten symbols and gesture recognition. *Proceedings of the 7th International Forum on Strategic Technology (IFOST-2012) (Tomsk)* 2012; 1: 596-601.

Authors' information

Vladimir Grigorievich Spitsyn (b. 1948) graduated from Tomsk State University in 1970, Radio-Physics department. He works as the Professor of Tomsk Polytechnic University. His research interests are currently focused on neural networks, image processing, and electromagnetic wave propagation in random discrete media. E-mail: spvg@tpu.ru.

Yuliya Alexandrovna Bolotova (b. 1986) graduated from Tomsk Polytechnic University in 2009, PhD, associate professor of Computer Science department of Tomsk Polytechnic University. Her research interests are image processing, object recognition, biologically-inspired models. E-mail: julya21@hotmail.ru.

Ngoc Hoang Phan (b. 1983) graduated from Irkutsk State University in 2008, PhD, Deputy Dean of Information Technology faculty, Ba Ria-Vung Tau University, Vietnam. His research interests are currently focused on image processing, wavelet-analysis, neural networks, image recognition. E-mail: hoangpnn285@gmail.com.

Thi Thu Trang Bui (b. 1984) graduated from Tomsk Polytechnic University in 2009, PhD, head of the Computer Science department, Information Technology faculty, Ba Ria-Vung Tau University, Vietnam. Her research interests are currently focused on image processing, wavelet-analysis, neural networks, image recognition. E-mail: trangbt.084@gmail.com.

Received February 3, 2016. The final version – March 16, 2016.
