

6. Литвиненко В.И., Бидюк П.И., Фелелов А.А., Баклан И.В. Гибридная иммунная сеть для решения задач структурной идентификации // *Нейронные сети*. – 2006. – № 9. – С. 143–155.
7. Аткина В.С. Применение иммунной сети для анализа катастрофоустойчивости информационных систем // *Известия ЮФУ. Технические науки. Информационная безопасность*. – 2011. – № 12. – С. 203–210.
8. Фокин В.А. Статистическое моделирование данных при оценке состояния биологических систем // *Известия Томского политехнического университета*. – 2007. – Т. 311. – № 5. – С. 132–135.
9. Хаитов Р.М., Игнатъева Г.Л., Сидорович И.Г. *Иммунология*. – М.: Медицина, 2000. – 432 с.
10. Аткина В.С. Использование программного комплекса для исследования катастрофоустойчивости информационных систем // *Вестник Волгоградского государственного университета. Серия 10. Инновационная деятельность*. – 2011. – Вып. 5. – С. 14–18.
11. Аткина В.С. Оценка эффективности катастрофоустойчивых решений // *Вестник Волгоградского государственного университета. Серия 10. Инновационная деятельность*. – 2012. – Вып. 6. – С. 89–93.
12. Погребной В.К. О построении активных моделей распределенных систем реального времени // *Известия Томского политехнического университета*. – 2008. – Т. 312. – № 5. – С. 78–84.
13. Погребной А.В., Погребной Д.В. Проектирование структуры локальной сети для распределенной вычислительной системы реального времени // *Известия Томского политехнического университета*. – 2007. – Т. 311. – № 5. – С. 91–96.
14. Вейбер В.В., Кудинов А.В., Марков Н.Г. Задача сбора и передачи технологической информации распределенного промышленного предприятия // *Известия Томского политехнического университета*. – 2011. – Т. 319. – № 5. – С. 69–74.

Поступила 23.09.2012 г.

УДК 004.931

РАСПОЗНАВАНИЕ СТРУКТУРИРОВАННЫХ СИМВОЛОВ НА ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ ГИСТОГРАММ СРЕДНЕЙ ИНТЕНСИВНОСТИ И СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ

А.А. Друки

Томский политехнический университет
E-mail: druki2008@yandex.ru

Разработаны и представлены алгоритм выделения области расположения символов на сложном фоне и алгоритм выделения символов на основе гистограмм средней интенсивности. Для решения задачи распознавания символов разработана и представлена сверточная нейронная сеть.

Ключевые слова:

Обработка изображений, распознавание символов, нейронные сети, гистограммы средней интенсивности.

Key words:

Image processing, character recognition, neural networks, histogram of average intensity.

Введение

Современные технологические, производственные и офисные системы в процессе своего функционирования используют информацию о маркировке объектов. Информация о маркировке грузов, вагонов, контейнеров, автомобильных номерных знаков позволяет рациональным образом организовать процесс технологической обработки информации, вести учет и контроль изделий, материалов, транспортных средств. В основе процессов использования маркировки (текстово-цифровых меток) лежит технология автоматизированного распознавания структурированных символов. Потребность в такой технологии вызвала необходимость создания методов, моделей и систем распознавания структурированных символов [1].

В настоящее время такие технологии реализуются тремя традиционными методами – структурным, признаковым и шаблонным. Каждый из этих методов ориентирован на свои условия применения, для которых они являются эффективными.

Вместе с тем всем этим методам присущи недостатки. Наиболее существенные из них – низкая устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям.

Эти недостатки особенно ярко проявились при масштабной эксплуатации программно-технологических систем, использующих в своей основе эти методы. Практически у всех систем распознавания структурированных символов точностные характеристики резко падают и становятся ниже технологически приемлемых при искажении аффинными и проекционными преобразованиями. Вместе с тем технологические условия получения информации о маркировке не позволяют полностью устранить эти искажения [2].

Анализ методов распознавания структурированных символов показал, что для решения данной задачи эффективно использовать искусственные нейронные сети, в связи с тем, что они являются слабо чувствительными к искажениям входного сигнала, а так же обеспечивают возможность полу-

чения классификатора, хорошо моделирующего сложную функцию распределения изображений символов, тем самым увеличивая точность решения по сравнению с остальными методами.

Целью работы является разработка методов, алгоритмов и программ распознавания автомобильных номерных знаков на изображениях, инвариантных к аффинным и проекционным искажениям и удовлетворяющих требованиям:

1. Программная система должна обеспечивать вероятность распознавания изображений, содержащих государственные регистрационные знаки транспортных средств на расстоянии 40 м, не менее 80 %.
2. Цикл обработки не должен превышать одной секунды.

Выделение области расположения символов на изображении

Алгоритм работы системы можно разделить на 3 этапа:

1. Выделение области расположения символов на изображении.
2. Выделение отдельных символов.
3. Распознавание символов.

Для анализа исходного изображения и поиска на нём объектов необходимо провести бинаризацию, которая представляет собой перевод полноцветного изображения в монохромное, где присутствуют только два типа пикселей (темные и светлые). Бинаризация изображения основывается на сравнении интенсивности каждого пикселя с пороговым значением интенсивности. Если значение интенсивности пикселя выше значения интенсивности порога, то данному пикселю присваивается значение 255, или в противном случае 0 [3].

Пороговое значение T вычисляется по следующей формуле:

$$T = \frac{I_{\max} + I_{\min}}{2},$$

где I_{\max} – максимальное значение интенсивности изображения, I_{\min} – минимальное значение интенсивности изображения.

Выделение области расположения символов на изображении начинается со сканирования входного изображения окном размером 3×3 :

P_0	P_1	P_2
P_7	$F(i,j)$	P_3
P_6	P_5	P_4

Внутри окна вычисляется количество точек с интенсивностью 255, затем вычисляется отношение количества данных точек и площади сканирующего окна и сравнивается с пороговым значением T . Если данное отношение превышает заданное пороговое значение, то данная область изображения определяется как область-кандидат на содержание пластины номерного знака автомобиля. Экспериментально было установлено, что данный порог равен 0,3.

Окно перемещается попиксельно по всему изображению и для точки, оказавшейся в центре окна, вычисляется новое значение интенсивности по формуле:

$$K_s(i, j) = \sqrt{X^2 + Y^2},$$

$$X = [P_2 + 2P_3 + P_4] - [P_0 + 2P_7 + P_6],$$

$$Y = [P_0 + 2P_1 + P_2] - [P_6 + 2P_5 + P_4],$$

где $K_s(j,k)$ – новое значение интенсивности точки с координатами (i,j) ; P_n – значение интенсивности n -го пикселя.

Затем на изображении производится исследование областей, имеющих наибольший контраст, и вычисляется количество их граней. Если количество граней находится в некотором диапазоне, данная область отмечается как кандидат на местоположение номерного знака. Результатом этого подхода является нахождение на изображении всех областей кандидатов, в которых может содержаться пластина номерного знака [4].

Основным достоинством данного метода является то, что он основан на использовании информации о символах и номерной пластине и не чувствителен к параметрам контрастности и цвета изображения.

Недостаток данного метода заключается в том, что результатом определения границ является большое количество линий как горизонтальных, так и вертикальных, следовательно, большое количество ложных областей изображения может быть отнесено к кандидатам на содержание номерного знака. Поэтому предлагается использовать выделение только вертикальных границ изображения, во избежание переизбытка количества границ и ложных областей кандидатов.

Выделение символов с помощью гистограмм средней интенсивности

Для выделения строки символов на изображении в каждой области-кандидате предлагается строить гистограммы средней интенсивности, каждая из которых строится под заданным углом.

Гистограмма средней интенсивности представляет собой график с осями координат x и y (рис. 1). Изображение сканируется слева направо, сверху вниз, при этом вычисляется средняя интенсивность в каждом столбце. В тех местах, где нет символа, средняя интенсивность будет значительно отличаться от интенсивности тех мест, где символы присутствуют. Далее, выполняя ту же операцию по строкам, получается набор отдельных символов, которые уже можно анализировать [1].

В результате работы этого подхода получается набор изображений объектов, в этот набор входят как символы, так и различные помехи, которые необходимо устранить. Это достигается путем проверки условий, которым должен удовлетворять символ как геометрический объект [5].

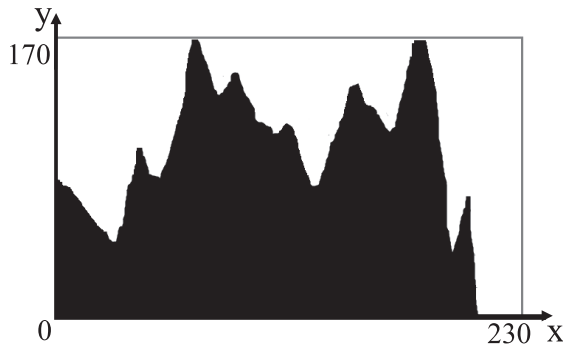


Рис. 1. Пример гистограмм средней интенсивности

Во-первых, проверяется выполнение условия

$$0,9 \leq \frac{b}{a} \leq 3,5, \text{ где } a \text{ и } b \text{ – соответственно ширина}$$

и высота объекта. Объекты, не удовлетворяющие данному соотношению, отбрасываются из рассмотрения.

Второе условие имеет целью удаление мелких областей, не представляющих интереса. Среди объектов, удовлетворяющих вышеописанным условиям, выбирается объект с максимальной площадью описываемого прямоугольника $S_{\max} = ab$. Затем среди оставшихся областей выбираются те, которые удовлетворяют условию $\frac{S}{S_{\max}} \geq 0,1$, где S –

площадь прямоугольника объекта-кандидата.

Для того чтобы отделить строку символов от всего изображения, для начала предлагается вычислить горизонтальные гистограммы средней интенсивности. Так как на изображении самой яркой областью является фон номерной пластины, то два наибольших максимума будут соответствовать областям 1 и 2 (рис. 2).



Рис. 2. Построение горизонтальных гистограмм. Линиям 1 и 2 соответствуют два наибольших максимума. x – номер строки изображения, y – средняя интенсивность строки изображения

Далее строятся вертикальные гистограммы под углом взаимно перпендикулярным к n , и выделяется уже около 10 максимумов в промежутках между символами. Таким образом, выделяются области расположения отдельных символов на номерной пластине, которые можно дальше анализировать (рис. 3). Ненужные мелкие объекты, не являющиеся символами, отбрасываются из рассмотрения путём сравнения с некоторыми условиями.

При регистрации изображение номерных пластин подвергается аффинным и проекционным искажениям, следовательно, линии, соответствующие областям 1 и 2, будут располагаться не по го-

ризонтالي, а под неизвестным углом. В связи с чем предлагается строить не одну, а n гистограмм средней интенсивности, каждая из которых строится не по горизонтали, а под заданным углом (рис. 4).

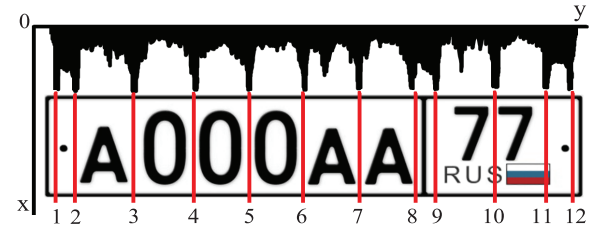


Рис. 3. Построение вертикальных гистограмм. Линиям от 1 до 12 соответствуют наибольшие максимумы. y – номер строки изображения, x – средняя интенсивность строки изображения

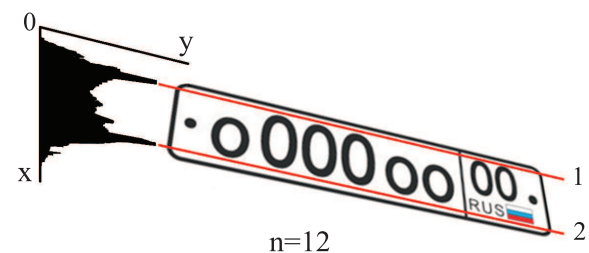


Рис. 4. Изображение номерной пластины, его гистограмма, построенная под углом 12° и выделенная область изображения, содержащая строку символов

Необходимое количество гистограмм средней интенсивности определяется из технических условий регистрации изображений, по данным условиям угол разворота изображения не превышает 15° по горизонтали как в правую, так и в левую сторону, следовательно, $n=31$. Из n построенных гистограмм выбирается та, которая содержит наибольшее значение по y , т. к. наибольшее значение будет соответствовать области 1 или 2 (рис. 4) [5, 6].

Распознавание символов с помощью сверточной нейронной сети

Анализ методов распознавания образов показал, что для решения данной задачи эффективно использовать искусственные нейронные сети. Преимущество использования нейронных сетей для обработки изображений – обучаемость системы для выделения ключевых характеристик символов из учебных наборов [7].

Наиболее часто в задачах распознавания и идентификации изображений используются классические нейросетевые архитектуры (многослойный персептрон, сети с радиально-базисной функцией и др.), но из анализа данных методов и экспериментальных исследований следует, что применение классических нейросетевых архитектур в данной задаче неэффективно по следующим причинам:

- изображения имеют большую размерность, соответственно возрастает размер нейронной сети;

- большое количество параметров увеличивает вместимость системы и соответственно требует большей тренировочной выборки, увеличивает время и вычислительную сложность процесса обучения;
- для повышения эффективности работы системы желательно применять несколько нейронных сетей (обученные с различными начальными значениями синаптических коэффициентов и порядком предъявления образов), но это увеличивает вычислительную сложность и время выполнения задачи;
- отсутствует инвариантность к изменениям масштаба изображения, ракурсов съёмки камеры и других геометрических искажений входного сигнала [7, 8].

Поэтому для распознавания символов на номерных знаках были выбраны сверточные нейронные сети, т. к. они обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям.

Архитектура сверточной нейронной сети состоит из многих слоёв. Слои бывают двух типов: сверточные и подвыборочные, они чередуются друг с другом.

Нейроны в пределах слоя организованы в плоскости. В каждом слое имеется набор из нескольких плоскостей, причём нейроны одной плоскости имеют одинаковые синаптические коэффициенты, ведущие ко всем локальным участкам предыдущего слоя. Каждый нейрон слоя получает входы от некоторой области предыдущего слоя (локальное рецептивное поле), т. е. входное изображение предыдущего слоя как бы сканируется небольшим окном и пропускается сквозь набор синаптических коэффициентов, а результат отображается на соответствующий нейрон текущего слоя. Набор плоскостей представляет собой карты характеристик, и каждая плоскость находит «свои» участки изображения в любом месте предыдущего слоя. Размер локального рецептивного поля выбирается самостоятельно в процессе разработки нейронной сети [9].

Подвыборочный слой уменьшает масштаб плоскостей путём локального усреднения значений

выходов нейронов. Таким образом достигается иерархическая организация. Последующие слои извлекают более общие характеристики, меньше зависящие от искажений изображения.

Постепенно нейронная сеть обучается выделять ключевые характеристики поступающих на вход изображений [9, 10].

Разработана нейронная сеть, состоящая из 3-х скрытых слоёв, представлена на рис. 5. Входными данными нейронной сети являются отмасштабированные изображения размером 29×29 пикселей.

Следом за входным слоем находится первый скрытый слой, который является свёрточным. Он состоит из шести карт признаков размером 13×13. Каждый элемент карты признаков соединен с рецептивным полем размером 5×5 на входном изображении и единичным смещением. Следовательно, каждый элемент карты имеет 26 обучаемых весовых коэффициентов. Значения всех элементов карты признаков вычисляются путем последовательного сканирования рецептивным полем входного слоя. Таким образом, в первом скрытом слое содержится 26364 синаптических связей и 156 обучаемых параметров.

Размер свёрточной плоскости определяется в соответствии со следующими выражениями:

$$w_c = \frac{(w_u - 3)}{2},$$

где w_c – ширина свёрточной плоскости; w_u – ширина плоскости предыдущего слоя.

$$h_c = \frac{(h_u - 3)}{2},$$

где h_c – высота свёрточной плоскости; h_u – высота плоскости предыдущего слоя.

Второй скрытый слой также является слоем свёртки. Он состоит из 50 карт признаков размером 5×5. В этом слое 5×5×50=1250 нейронов. Каждый элемент в карте признаков связан с шестью областями размером 5×5 шести карт предыдущего слоя. Таким образом, во втором скрытом слое содержится 7550 весов и 188750 связей.

Два первых свёрточных слоя можно рассматривать как слои для извлечения признаков из изобра-

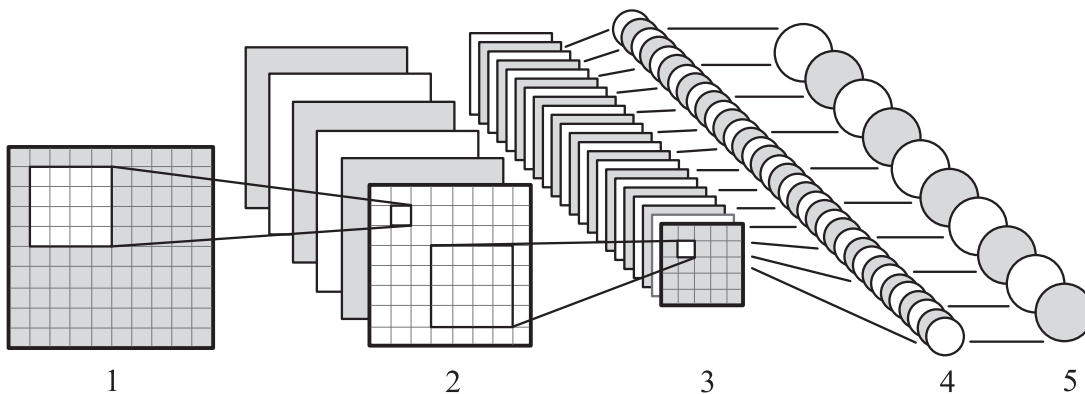


Рис. 5. Архитектура свёрточной нейронной сети: 1) вход; 2, 3) скрытые сверточные слои; 4) слой из обычных нейронов; 5) выходы сети

жения. Следующие два слоя являются слоями классификации. В этих слоях каждый нейрон соединён со всеми нейронами предыдущего слоя.

Третий скрытый слой состоит из 100 нейронов. В этом слое имеется 125100 связей и 125100 обучаемых параметров. Для десяти классов изображений, соответствующих десяти цифрам, 100 нейронов достаточно для хорошей обобщающей способности сети.

Данная нейронная сеть работает как классификатор, поэтому четвёртый выходной слой состоит из 21 нейрона, так как распознается 21 символ. Согласно ГОСТ Р 50577-93 на регистрационные знаки транспортных средств могут содержать следующие символы: А, В, Е, К, М, Н, О, Р, С, Т, Х, У и все цифры от 0 до 9. Следовательно система должна распознавать 21 символ.

В качестве активационной функции был выбран гиперболический тангенс:

$$f(a) = \text{Atanh}(Sa).$$

где A – амплитуда этой функции, S – определяет ее положение относительно начала отсчета. Функция f – нечетная, ее горизонтальные асимптоты равняются A и $-A$.

Эта функция имеет ряд преимуществ для решения задачи:

- симметричные активационные функции, типа гиперболического тангенса, обеспечивают более быструю сходимость, чем стандартная логистическая функция;
- имеет непрерывную первую производную;
- имеет простую производную, которая может быть вычислена через ее значение, что дает экономии вычислений [11].

Формула функционирования нейрона свёрточного слоя:

$$y_k^{(i,j)} = b_k + \sum_{s=1}^K \sum_{t=1}^K w_{k,s,t} x^{((i-1)+s, (j-1)+t)},$$

где $y_k^{(i,j)}$ – нейрон k -й плоскости свёрточного слоя; b_k – нейронной смещение k -й плоскости; K – размер рецептивной области нейрона; $w_{k,s,t}$ – матрица синаптических коэффициентов; x – выходы нейронов предыдущего слоя.

Используется стандартный для нейронных сетей алгоритм обратного распространения ошибки. Для измерения качества распознавания использовалась функция среднеквадратической ошибки:

$$E_p = \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2,$$

где E_p – величина функции ошибки для образа p ; t_{pj} – желаемый выход нейрона j для образа p ; o_{pj} – действительный выход нейрона j для образа p .

Окончательная коррекция синаптических коэффициентов происходит по формуле:

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_{pj} o_{pj},$$

где η – коэффициент пропорциональности, влияющий на скорость обучения.

На каждой итерации алгоритм обратного распространения ошибки рассчитывается для всего обучающего набора данных, чтобы вычислить средний или истинный градиент [12].

Когда ненастроенной сети предъявляется входной образ, она выдает некоторый случайный выход. Функция ошибки представляет собой разность между текущим выходом сети и идеальным выходом, который необходимо получить. Для успешного обучения сети требуется приблизить выход сети к желаемому выходу, т.е. последовательно уменьшать величину функции ошибки. Это достигается настройкой межнейронных связей. Каждый нейрон в сети имеет свои веса, которые настраиваются, чтобы уменьшить величину функции ошибки [13, 14].

Значения весовых коэффициентов были выбраны случайным образом из нормального распределения с нулевым средним и стандартным отклонением:

$$\delta_w = \sqrt{m},$$

где m – число связей входящих в нейрон.

Экспериментально доказано, что локальное рецептивное поле, равное 5×5 пикселей, позволяет избежать вычислительной избыточности и обеспечить надежное распознавание символов.

Для обучения сети использовалась база изображений, состоящая из 5000 символов. Значения исходных синаптических весов для всех слоёв сети с гиперболическим тангенсом выбираются на основе равномерного распределения с нулевым математическим ожиданием и дисперсией, обратной квадратному корню, из количества синаптических связей нейрона.

Нейронная сеть обеспечивает вероятность распознавания символов на уровне 95 % и скорость работы 0,5 секунды. В ходе экспериментов было установлено, что сеть ошибочно распознаёт символы при отклонении пластины номерного знака по горизонтали и вертикали более 45° и повороте пластины на плоскости до 15° . В этих случаях атрибуты символов наслаиваются друг на друга и не подаются распознаванию даже человеческим глазом.

При различных шумах на изображениях, таких как грязь, снег, дождь, которые визуально меняют структуру символов, вероятность распознавания составляет 70 %, это является хорошим результатом по сравнению с другими методами. Одним из достоинств свёрточных нейронных сетей является то, что они игнорируют мелкие дефекты и шумы на изображениях, тем самым обеспечивая высокую вероятность распознавания зашумлённых изображений.

Заключение

На основе представленных алгоритмов выделения и распознавания символов разработана программная система, которая обеспечивает точность распознавания символов на уровне 95 % и соответствует техническим требованиям. Разработана че-

тырёхслойная сверточная нейронная сеть без слоёв подвыборки, обеспечивающая инвариантность к изменениям условий съёмки и поворотам изображений. Несмотря на простую структуру сети, эксперименты показали высокие результаты распознавания символов. Преимущества разработанных

методов заключаются в быстродействии (0,5 секунд), инвариантности к искажениям входного сигнала и высокой вероятности распознавания, недостатки же заключаются в сложности настройки весовых коэффициентов сети и выборе её структуры.

СПИСОК ЛИТЕРАТУРЫ

1. Фаворская М.Н., Зотин А.С., Горошкин А.Н. Морфологическая обработка контурных изображений в системах распознавания текстовых символов // Вестник СибГАУ. – 2007. – № 1 (14). – С. 70–75.
2. Бредихин Р.Н. Об одном подходе к распознаванию оптических образов текстов // Вестник МЭИ. – 2005. – № 2. – С. 134–141.
3. Садыхов Р.Х., Ваткин М.Е. Модифицированный алгоритм обучения РБФ-сети для распознавания рукописных символов // Идентификация образов. – 2001. – Т. 1. – № 3. – С. 7–16.
4. Нгуен Тоан Тханг, Спицын В.Г. Алгоритмическое и программное обеспечение для распознавания формы руки в реальном времени с использованием SURF-дескрипторов и нейронной сети // Известия Томского политехнического университета. – 2012. – Т. 320. – № 5. – С. 48–54.
5. Буй Тхи Тху Чанг, Фан Нгок Хоанг, Спицын В.Г. Распознавание лиц на основе применения метода Виолы-Джонса, вейвлет-преобразования и метода главных компонент // Известия Томского политехнического университета. – 2012. – Т. 320. – № 5. – С. 54–59.
6. Болотова Ю.А., Спицын В.Г., Фомин А.Э. Применение модели иерархической временной памяти в распознавании изображений // Известия Томского политехнического университета. – 2011. – Т. 318. – № 5. – С. 60–63.
7. Кермани Коланкех А., Спицын В.Г., Хамкер Ф. Нахождение параметров и удаление постоянной составляющей фильтра Габор для обработки изображений // Известия Томского политехнического университета. – 2011. – Т. 318. – № 5. – С. 57–59.
8. Plamondon R., Srinari S. On-line and off-line handwriting recognition: A comprehensive survey // IEEE Transactions on pattern analysis and machine intelligence. – 2000. – V. 22. – P. 105–109.
9. Ronse C., Najman L., Decenciere E. Mathematical morphology: 40 years on // Proceedings of the VII International symposium of mathematical morphology. – Netherlands, 2005. – № 30. – P. 350–370.
10. Le Cun Y., Bengio Y. Convolutional networks for images, speech and time series // The handbook of brain theory and neural networks. – 1998. – V. 7. – № 1. – P. 255–258.
11. Rowley H.A., Baluja S., Kanade T. Neural network-based face detection // Pattern anal. mach. intell. – 2000. – V. 5. – P. 23–38.
12. Feraud R., Bernier O., Viallet J., Collobert M. A fast and accurate face detector based on neural networks // Transactions on pattern analysis and machine intelligence. – 2002. – V. 3. – № 23. – P. 42–53.
13. Yu N., Notkin B.S., Sedov V.A. Neuro-iterative algorithm of tomographic reconstruction of the distributed physical fields in the fibre-optic measuring systems // Computer optics. – 2009. – V. 33. – № 4. – P. 446–455.
14. Wilson D.R., Martinez T.R. The general inefficiency of batch training for gradient descent learning // Neural Networks. – 2004. – V. 16. – № 4. – P. 1429–1451.

Поступила 30.01.2013 г.