

РАСПОЗНАВАНИЕ РАСТРОВЫХ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ ДИНАМИЧЕСКОЙ НЕЙРОННОЙ СЕТИ, ЗАДАННОЙ В ПРОСТРАНСТВЕ КОМПЛЕКСНЫХ ЧИСЕЛ

© 2003 А.А. Юдашкин

Институт проблем управления сложными системами РАН, г. Самара

Представлен новый метод распознавания растровых изображений, для каждой точки которых введены две характеристики. Изображение переводится в вектор с комплексными элементами, отвечающими характеристикам точек исходного образа. Распознавание производится с помощью модели динамической нейронной сети с конкуренцией, в которой каждый искусственный нейрон характеризуется комплексным уровнем активности. В процессе распознавания ненулевым по абсолютной величине остается только уровень активности нейрона, отвечающего запомненному образу, наиболее похожему на предъявленный.

Введение

В задачах распознавания образов часто приходится иметь дело с оцифрованными фотографиями, полученными в условиях, отличных от идеальных, например, при наличии глубоких теней или засвеченных участков на изображениях. Оцифрованные изображения представляют собой наборы данных, состоящие из оптических характеристик каждого пикселя. В частности, при обработке полутоновых картин набор данных состоит из значений яркости для каждого пикселя. В результате любые изменения освещения, качества съемки или характеристик объекта съемки принципиально сказываются на качестве распознавания вследствие нелинейности подобных искажений, что невозможно учесть простыми способами. Подобные случаи, в частности, характерны для распознавания фотографий человеческих лиц, где искажения освещения и естественные изменения внешности становятся равноценно нежелательными для любой системы распознавания. Для снижения влияния таких факторов, как тени и засвеченные участки применяются различные методы предварительной обработки с помощью фильтров, одним из которых является простейший расчет перепадов яркости для двух соседних точек (пикселей). При этом перепад яркости Δq_{jk} точки на пересечении j -й строки и k -го столбца определяется следующим образом:

$$\Delta q_{jkx} = q_{jk} - q_{(j-1)k} \quad (1)$$

для расчета перепада вдоль горизонтальной оси X (вдоль строки матрицы, представляющей образ) или

$$\Delta q_{jky} = q_{jk} - q_{j(k-1)} \quad (2)$$

для расчета перепада вдоль вертикальной оси Y (вдоль столбца матрицы). При этом существующие алгоритмы позволяют учесть либо только одну из этих двух характеристик, либо обе последовательно, либо проводить расчеты с одним параметром, полученным в результате аппроксимации по двум приведенным, например, посредством их взвешенного суммирования. В результате неизбежно возникает проблема выбора наиболее значимой из характеристик (1) и (2) и соответствующая потеря качества распознавания при изменении факторов, влияющих на дефекты изображений. Поэтому представляется интересным учесть обе характеристики одновременно таким образом, чтобы они дополняли друг друга в некоторой интегральной зависимости по всем точкам изображения. Например, в случае изменения расположения источника освещения значения яркостей различных точек меняются по-разному, однако изменение параметров по фильтрам (1) и (2) обладает сходным характером для всего изображения. Подобная ситуация аналогична вращению некоторой геометрической фигуры на плоскости, когда существует единое преоб-

разование координат для всех элементов фигуры. Для решения задачи синтеза сложной фигуры с памятью, содержащей несколько желаемых конфигураций на плоскости, в работе [1] координаты точек фигуры были заданы в виде комплексных чисел, действительными и мнимыми частями которых служили соответственно x - и y -координаты. Кроме указанной постановки задачи большой интерес представляет распознавание изображений в общем случае многопараметрического описания точек образов. Это имеет место, например, при распознавании цветных изображений в палитре YCrCb, где компонент Y представляет собой интенсивность, а две другие составляющие определяют цветность. При этом каждая точка изображения характеризуется тремя числовыми параметрами. В ряде работ уже были показаны способы использования комплексных и гиперкомплексных чисел при построении систем распознавания образов для использования нескольких параметров для точек изображения. В частности, введение комплексных чисел в качестве характеристик изображений применялось в работе [2], где путем перехода в частотную область была достигнута инвариантность к сдвигам. В работе [3] показана возможность построения дискретной нейронной сети Хопфилда на основе комплексных значений активностей нейронов. Похожая проблема решалась в работах [4] и [5] для выполнения анализа основных компонентов изображения и синтеза многослойного перцептрона, работающего с комплексными значениями весов и входов, соответственно. Дальнейшее обобщение проблемы и переход к работе с кватернионами произведено в работе [6]. Тем не менее, ни в одной работе не приведено достаточно надежного и устойчивого метода, позволяющего использовать числа, более общие, чем действительные для характеристики изображения и выполнения распознавания в области непрерывного изменения переменных при отсутствии ложных устойчивых состояний.

В представленной работе предлагается новый подход к распознаванию образов, представляющих собой полутоновые оциф-

рованные плоские изображения, основанный, с одной стороны, на переходе к рассмотрению нескольких относительных характеристик точек изображения вместо одной абсолютной, и, с другой стороны, использующий модель самоорганизующейся динамической нейронной сети, представленной и исследованной в работах [7-9]. Указанная модель нейронной сети применялась ранее для распознавания обычных полутоновых изображений, где обрабатывались именно абсолютные значения яркости точек, причем был доказан факт отсутствия ложных устойчивых состояний.

Модель нейронной сети

Пусть рассматриваются полутоновые изображения, которым соответствуют векторы $\mathbf{v}^{(j)}$ ($j = 1, 2, \dots, M$). Каждый элемент вектора в общем случае представляет собой комплексное число $v_{jk} = \sigma_{jk} + i\omega_{jk}$ ($k = 1, 2, \dots, N$), где действительная и мнимая части являются двумя статистически независимыми характеристиками точки изображения. В данной работе принято, что характеристики σ и ω принадлежат множеству вещественных чисел, что не ограничивает общности. Пусть значениями σ и ω в каждой точке служат числа, получившиеся после применения фильтров (1) и (2) к исходным изображениям. После фильтрации все изображения, изначально представляющие собой прямоугольные матрицы яркостей пикселей, разворачиваются в векторы $\mathbf{v}^{(j)}$. Для распознавания предъявляется искаженная или неизвестная картина $\mathbf{q}(0)$, которая также представлена вектором длины N с комплексными координатами. Необходимо синтезировать структуру нейронной сети, которая данному представленному вектору сопоставляет один наиболее близкий к нему запомненный образ $\mathbf{v}^{(j)}$.

В работах [7-9] для распознавания образов, представленных векторами с действительными элементами, использовалась синергетическая модель нейронной сети, где производился переход от рассмотрения самих векторов с образами к конкуренции скалярных функций, соответствующих им. Тот же

подход предлагается в данной работе, но для комплексных элементов вектора. Пусть используется форма потенциала в виде формы четвертой степени в качестве потенциальной функции, минимумами которой по вектору являются запомненные векторы:

$$W(\mathbf{q}) = -\frac{1}{2} \bar{\mathbf{q}} \mathbf{J} \mathbf{q} + \frac{1}{4} (\bar{\mathbf{q}} \mathbf{q})^2 + \frac{1}{4} \sum_{k=1}^M \sum_{l \neq k} \bar{\mathbf{q}} \mathbf{v}^{(k)} \bar{\mathbf{u}}^{(k)} \mathbf{q} \bar{\mathbf{q}} \mathbf{v}^{(l)} \bar{\mathbf{u}}^{(l)} \mathbf{q} \quad (3)$$

где $\bar{\mathbf{q}}$ обозначает вектор, сопряженный с \mathbf{q} . Здесь векторы $\mathbf{u}^{(k)}$ образуют набор, дополнительный к набору векторов $\mathbf{v}^{(k)}$, который вводится для обеспечения взаимной ортонормированности по правилам

$$\bar{\mathbf{u}}^{(k)} \mathbf{v}^{(j)} = \delta_{kj}, \quad \mathbf{U} = \mathbf{V}(\bar{\mathbf{V}} \mathbf{V})^{-1}, \quad (4)$$

где δ_{kj} - символ Кронекера, столбцами матриц \mathbf{U} и \mathbf{V} являются векторы $\mathbf{v}^{(k)}$ и $\mathbf{u}^{(k)}$, и вводится матрица, определяющая память (минимума потенциала) сети:

$$\mathbf{J} = \sum_{k=1}^M \mathbf{v}^{(k)} \bar{\mathbf{u}}^{(k)}. \quad (5)$$

Минимизация потенциала (3) по \mathbf{q} может происходить согласно методу наискорейшего спуска, когда элементы вектора \mathbf{q} меняются во времени. Согласно подходу, изложенному в [7], принимается, что $\mathbf{q}(t)$ может быть представлен в виде следующей линейной комбинации:

$$\mathbf{q}(t) = \sum_{i=1}^M d_i(t) \mathbf{v}^{(i)} + \xi(t), \quad (6)$$

где $\xi(t)$ является затухающей во времени по модулю случайной добавкой, а $d_i(t)$ – скалярные функции времени. В новой постановке задачи векторы $\mathbf{v}^{(k)}$ состоят из комплексных элементов и, соответственно, $d_i(t)$ в общем случае также комплексные. Подстановка (6) в (3) дает с учетом (4) следующее выражение для потенциала, теперь зависящего от $d_i(t)$:

$$W(\mathbf{d}) = -\frac{1}{2} \bar{\mathbf{d}} \mathbf{G} \mathbf{d} + \frac{1}{4} (\bar{\mathbf{d}} \mathbf{G} \mathbf{d})^2 + \frac{1}{4} \sum_{k=1}^M \sum_{l \neq k} \bar{\mathbf{d}} \mathbf{g}^{(k)} d_k \bar{\mathbf{d}} \mathbf{g}^{(l)} d_l \quad (7)$$

Здесь

$$\mathbf{G} = \bar{\mathbf{V}} \mathbf{V},$$

а $\mathbf{g}^{(k)}$ – k -й столбец матрицы \mathbf{G} . После ввода новых переменных

$$\mathbf{d}_u = \mathbf{G} \mathbf{d}$$

и соответствующей процедуры построения уравнений движения для точки в поле потенциала (7) получаем уравнения, минимизирующие данный потенциал согласно методу наискорейшего спуска:

$$\dot{\mathbf{d}} = \mathbf{d} - 2\mathbf{d} \bar{\mathbf{d}} \mathbf{d} \mathbf{d}_u + \mathbf{D} \mathbf{d}, \quad (8)$$

где $\mathbf{D} = \text{diag}(d_k d_{uk})$. В уравнениях (8) фактически присутствует матрица \mathbf{G} , являющаяся, как нетрудно видеть, комплексной матрицей Грама для векторов $\mathbf{v}^{(k)}$, и, следовательно, эрмитовой матрицей. Это дает возможность предположить, что свойства системы (8) будут близки к полученным в [8], и решение этой системы будет иметь требуемый качественный характер. А именно, необходимо, чтобы при движении из начального состояния

$$\mathbf{d}(0) = \mathbf{G}^{-1} \bar{\mathbf{V}} \mathbf{q}(0) \quad (9)$$

все переменные состояния $d_i(t)$, кроме одной $d_l(t)$, релаксировали к нулю по абсолютной величине. Тогда оставшаяся переменная указывает на образ $\mathbf{v}^{(l)}$, наиболее близкий к $\mathbf{q}(0)$. Здесь каждая переменная состояния $d_i(t)$ характеризует уровень активности отдельного искусственного нейрона, а связь между отдельными нейронами определена матрицей \mathbf{G} и уравнениями (8). Так производится распознавание в данной модели динамической нейронной сети с конкуренцией. Как нетрудно видеть, если $\mathbf{q}(0)$ совпадает с $\mathbf{v}^{(l)}$, то в выражении (9) сразу получается вектор $\mathbf{d}(0)$, в котором $d_i(0)=0$ для $i \neq l$ и $d_l(0)=1$. В работах [8,9] показано, что для действительных векторов $\mathbf{v}^{(k)}$ динамика системы в фазовом пространстве \mathbf{d} определяется набором устойчивых узлов, заданных наборами координат

описанного вида, а также расположением седел, разделяющих фазовое пространство на области притяжения каждого узла. По-видимому, для векторов, состоящих из комплексных элементов, качественный характер соответствующей системы будет очень близким, но более сложным вследствие того, что теперь $d_i(t)$ могут быть в общем случае комплексными. Интерпретация факта, когда в конце распознавания остается одна переменная, равная 1 по модулю, но имеющая и мнимую, и действительную части, является предметом будущих исследований. В данной работе далее рассматриваются результаты численного моделирования распознавания набора полутоновых образов, заданных с помощью применения фильтров (1) и (2), на основе модели (8) с начальными условиями (9) и определениями (4).

Численные эксперименты

С помощью пакета MATLAB был произведен ряд экспериментов по исследованию свойств предложенной модели для задач распознавания полутоновых фотографий человеческих лиц. Образец подобной фотографии приведен на рис.1.

Применение фильтров (1) и (2) к представленному изображению приводит к тому, что два результирующих массива в расширенной палитре 32 градации серого могут иметь вид изображений, показанных на рис.2.

Все изображения состояли из 192x192 пикселей в целочисленной палитре 16 градаций серого. Образы, участвовавшие в распоз-

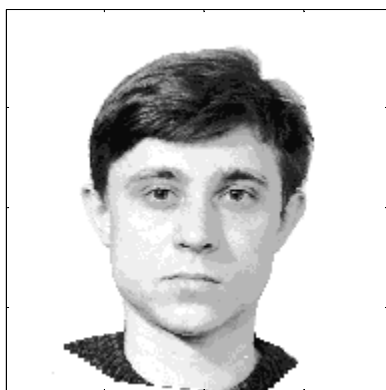


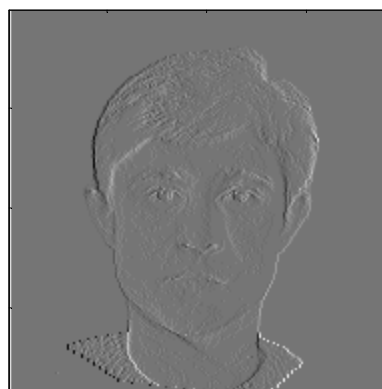
Рис.1. Полутоновый портрет в палитре 16 градаций серого

новании, были представлены векторами $\mathbf{v}^{(k)}$, каждый элемент которого был получен с помощью построчного преобразования исходной матрицы портрета в вектор и имел вид

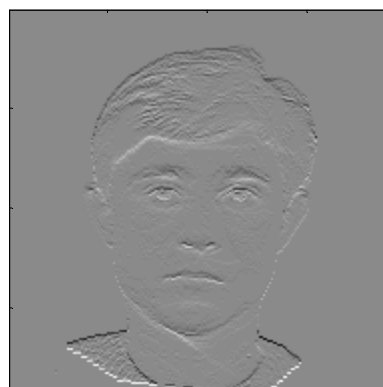
$$v_{jk} = \sigma_{jk} + i\bar{\omega}_{jk},$$

причем σ_{jk} и $\bar{\omega}_{jk}$ соответствовали результатам фильтрации по горизонтали по формуле (1) и вертикали по формуле (2) для пикселя исходной матрицы, отображенного в j -й элемент вектора $\mathbf{v}^{(k)}$. После данного преобразования все векторы были подвергнуты процедуре нормировки путем деления на свой модуль, что привело к уходу от целочисленных значений. В процедуре распознавания участвовало 16 образов. Численные эксперименты сводились к двум основным задачам:

- 1) исследовать влияние шума на качество распознавания в случае комплексных значений элементов векторов, представляющих образы;
- 2) исследовать качество алгоритма при



а)



б)

Рис.2. Результат применения фильтров взятия первой разности между значениями яркостей соседних точек: а) по горизонтали; б) по вертикали

изменении некоторых характеристик освещения.

Для изучения первого вопроса для некоторого запомненного образа генерировался шум η , заданный нормальным распределением. Шум накладывался на некоторый исходный портрет, после чего выполнялась фильтрация и последующее распознавание зашумленного образа во всей совокупности, как это было описано выше. Для отдельного прототипа проводилось распознавание с различным уровнем шума η . Уровень шума определялся как отношение дисперсии шума к дисперсии значений яркости точек исходного образа и менялся от 0 до 10. Процедура повторялась для каждого из прототипов. После этого было рассчитано качество распознавания Q_η для всей совокупности прототипов как отношение числа успешных распознаваний к общему их числу для каждого значения уровня шума. Как видно из графика, представленного на рис.3а, качество распознавания практически не менялось с ростом уровня шума и не опускалось ниже 0,98 в конце интервала увеличения искажений. Это согласуется с данными, полученными ранее для стандартного метода распознавания по яркостям точек, и говорит о том, что увеличение числа параметров, характеризующих точки изображения, по меньшей мере, не ухудшает характеристик алгоритма.

Для определения характеристик метода по отношению к изменениям освещения для

каждого из запомненных образов производилась процедура изменения средней яркости образа. Она моделировала ситуацию усиления или ослабления освещения от почти абсолютной темноты (средняя яркость 0,1 при минимуме 0) до самого яркого, практически ослепляющего (средняя яркость 14,9 при максимуме 15). Поскольку диапазон изменения градаций серого ограничен, то данная процедура эквивалентна нелинейным искажениям исходного образа. Кроме того, для фильтрации вида (1), (2) данное искажение является также существенно деструктивным, поскольку численно смещает уровни перепадов, сохраняя при этом их качественную структуру. Изменение яркости I производилось в полном диапазоне. Определялся показатель качества распознавания Q_I по тому же принципу, что и для случая аддитивного шума, но теперь по отношению к уровню яркости.

Снова предложенный метод показал высокую степень устойчивости к искажениям и качество Q_I не опускалось ниже 0,98, что видно из рис.3б. Срыв графика в ноль при значениях яркости 0 и 15 обусловлен тем, что при абсолютной темноте и абсолютно ярком освещении в данной модели изображение исчезает и превращается в сплошной черный или белый квадрат.

Примерная динамика изменения абсолютных величин переменных состояния для всей совокупности образов показана на рис.4а, где можно видеть, как из первоначаль-

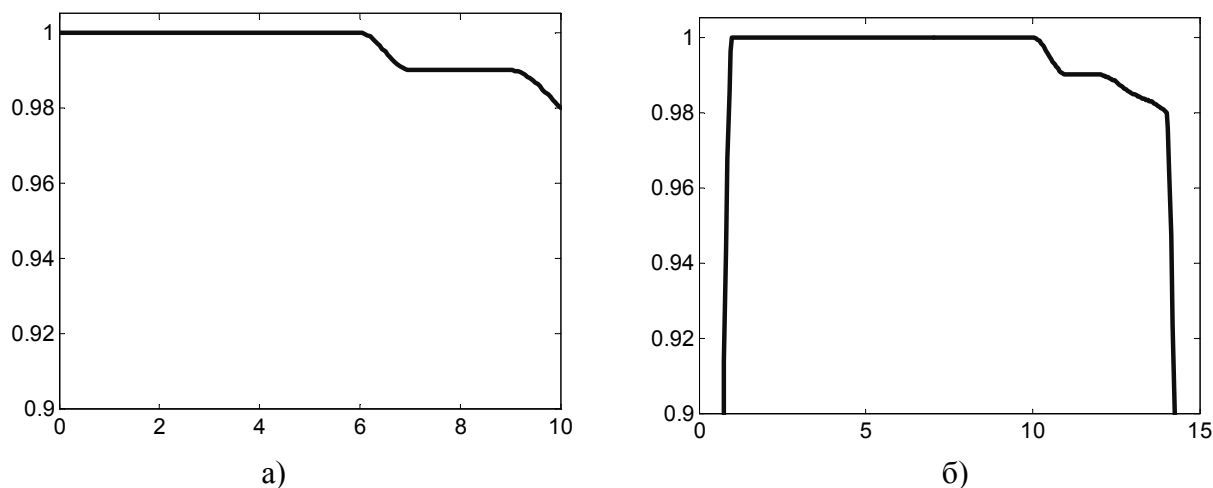


Рис.3. Диаграммы изменения качества распознавания в зависимости от: а) уровня шума; б) яркости освещения

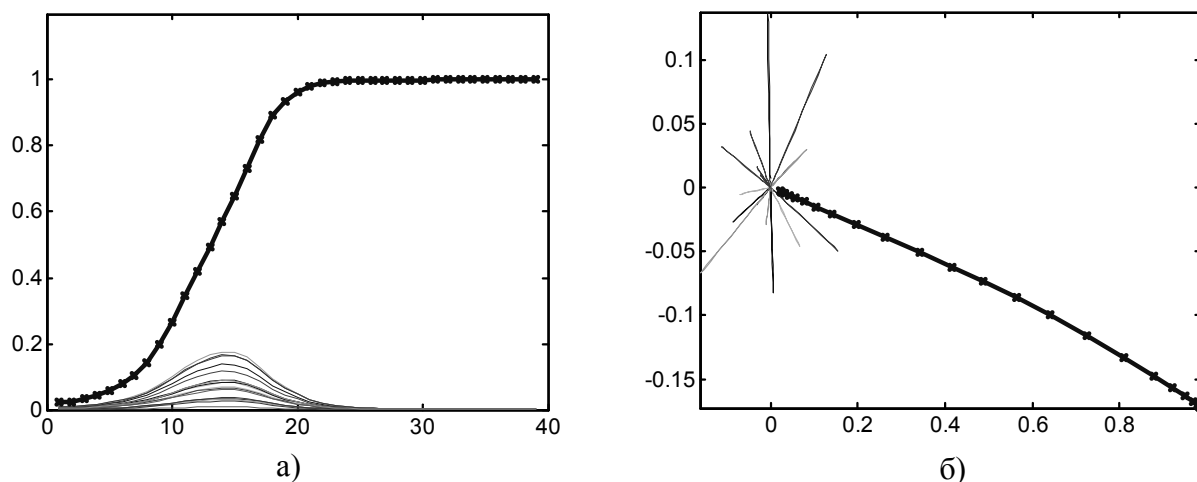


Рис.4. Динамика процесса распознавания, показывающая победившую переменную состояния: а) во времени; б) на комплексной плоскости

но очень близкого состояния почти все переменные состояния затухают, в то время как одна постепенно приближается к значению 1 по абсолютной величине.

Этому же случаю соответствует диаграмма изменения действительных и мнимых частей переменных состояния на комплексной плоскости, приведенная на рис.4б.

Выводы

В статье рассмотрен новый подход к синтезу алгоритмов распознавания сложных полутонных изображений, предполагающий работу с двумя независимыми параметрами, характеризующими каждый пиксель изображения. Для реализации алгоритма был произведен переход в комплексную область относительно значений параметров изображений, после чего в соответствии с ранее разработанным подходом была сформирована самоорганизующаяся нейронная сеть, в которой каждый нейрон во времени характеризовался переменной состояния, также являющейся комплексным числом, отвечающим отдельному запомненному прототипу. Используемая модель динамической нейронной сети не обладает ложными устойчивыми состояниями и работает по принципу “победитель забирает все”. Переход в комплексное пространство в данной статье осуществлялся посредством цифровой фильтрации в виде

определения перепадов яркости соседних пикселей, причем перепад по оси X образовал действительную, а перепад по оси Y – соответственно мнимую часть характеристики пикселя. Метод показал достаточно сильную устойчивость к помехам как в случае с аддитивным шумом высокого уровня, так и для нелинейных искажений, вызванных изменением уровня освещенности изображения. Качество распознавания для искаженных образов оставалось выше 0,98 для всех искажений. Представленная модель может быть достаточно эффективно использована как для распознавания изображений, предварительно подвергнутых цифровой фильтрации более чем одного вида, так и для работы с цветными изображениями и многопараметрическими наборами данных.

СПИСОК ЛИТЕРАТУРЫ

1. *Yudashkin A.A.* A Synthesis of Invariant Structures With Memory in Two-Dimensional Space. *Int. J. Mod. Phys. C.* 2000. V.11. №.5.
2. *De Castro M.C.F., De Castro F.C.C., Amaral J.N. and Franco P.R.G.* A complex valued Hebbian learning algorithm. *Proc. of IEEE-IJCNN.* 1998.
3. *Jankowski S., Lozowski A. and Zurada J.M.* Complex-Valued Multistate Neural Associative Memory. *IEEE Trans. on*

- Neural Networks, 1996. V.7. №6.
4. *Zhang Y. and Ma Y.* CGHA for Principal Component Extraction in the Complex Domain. IEEE Trans. on Neural Networks, 1997. V.8. №5.
 5. *Nitta T.* A Back-Propagation Algorithm for Complex Numbered Neural Networks. Proc. of IEEE/INNS International Joint Conference on Neural Networks, IJCNN'93-Nagoya, Oct. 25-29, 1993. V.2.
 6. *Arena P., Fortuna L., Muscato G. and Xibilia M.G.* Multilayer Perceptrons to Approximate Quaternion Valued Functions. Neural Networks, 1997. V.10. №2.
 7. *Haken H.* Synergetic computers and cognition: A top-down approach to neural nets. Berlin: Springer-Verlag, 1991.
 8. *Юдашкин А.А.* Бифуркации стационарных решений в синергетической нейронной сети и управление распознаванием образов//Автоматика и телемеханика. 1996. №11.
 9. *Юдашкин А.А.* Классификация образов и ее связь с топологией многообразий динамических систем//Известия Самарского научного центра РАН. 2001. №2.

BITMAP IMAGES RECOGNITION BY DYNAMICAL NEURAL NETWORK DEFINED IN THE SPACE OF COMPLEX NUMBERS

© 2003 A.A. Yudashkin

Institute for the Control of Complex Systems of Russian Academy of Sciences, Samara

A novel method of bitmap images recognition is considered for the case of introducing two parameters per each image pixel. The image is being mapped to the vector with complex elements corresponding to the pixel characteristics of the original image. A model of the dynamical competitive neural network performs the recognition where each artificial neuron is defined by the complex activity level. Only the activity level of the neuron corresponding to the most matching pattern to the offered one achieves non-zero value at the end of the recognition process.