

УДК 004.931

## ПРИМЕНЕНИЕ МОДЕЛИ ИЕРАРХИЧЕСКОЙ ВРЕМЕННОЙ ПАМЯТИ В РАСПОЗНАВАНИИ ИЗОБРАЖЕНИЙ

Ю.А. Болотова, В.Г. Спицын, А.Э. Фомин\*

Томский политехнический университет,  
\*Московский физико-технический институт  
E-mail: julya21@hotmail.ru; arseniy-fe@mail.ru

На основе модели иерархической временной памяти Дж. Хокинса и Дж. Дилипа разработана система для решения задачи распознавания образов. Описана структура системы, алгоритм ее обучения и работы. Представлены результаты тестирования системы на примере распознавания стилизованных цифр.

### Ключевые слова:

Модель иерархической временной памяти, распознавание образов, временная и иерархическая составляющая в процессе видения, эволюционный алгоритм.

### Key words:

Hierarchical temporal memory model, pattern recognition, time and hierarchical components in vision, evolutionary algorithm.

Задачи распознавания образов, чтения текстов, понимания речи и ориентации в пространстве, являющиеся элементарными для человека, до сих пор не были решены машиной. Американские ученые Дилип и Хокинс совместно с Редвудским Институтом нейробиологии выдвинули два ключевых отличия функционирования человеческого мозга от работы современных систем распознавания образов.

Первым отличием является способность мозга учитывать роль временной составляющей в процессе зрения. Это выражается в способности относить близкие по времени видимые объекты к одной и той же категории изображений [1, 2].

Вторым отличием является иерархичность. Мозг иерархичен по своей структуре. Нейроны, составляющие неокортекс, находятся в иерархической связи друг с другом и образуют несколько уровней иерархии. Чем выше уровень иерархии, тем с более абстрактной информацией он работает. Следовательно, в системе распознавания образов сложные объекты должны быть иерархически связаны с более простыми их составляющими [1–4].

Целью работы является реализация системы распознавания образов на основе модели иерархической временной памяти, а также оценка результатов работы системы в зависимости от ее структуры и количества распознаваемых категорий.

### Структура сети

Система иерархической временной памяти организована в виде иерархической сети, представляющей собой древовидную структуру. Каждый уровень состоит из набора узлов. Узлы одного уровня не связаны между собой. Узел верхнего уровня получает информацию с нескольких узлов предыдущего уровня. Самый верхний уровень состоит из одного узла и в терминах распознавания изображений оперирует образами объектов. На рис. 1 приведен пример сети, состоящей из 3-х уровней.

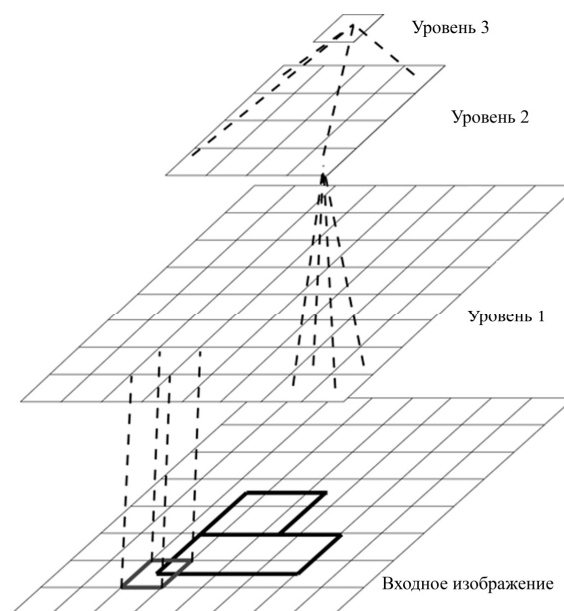


Рис. 1. Схема передачи информации между уровнями

На вход узлов первого уровня подается видеопоследовательность из кадров плавно перемещающихся во времени объектов. Так обеспечивается временная составляющая сигнала. Тестирование и обучение сети происходит последовательно. В первую очередь обучаются нижние уровни, потом, на основе их данных обучаются последующие уровни.

### Структура узла

Входной информацией для узла, независимо от уровня иерархии, является бинарный вектор. Для узлов первого уровня это вектор пикселей изображения. Каждый узел содержит набор групп. Задача узла – выбрать группу, которой больше всего соответствует входная последовательность. Выбор группы происходит в 2 этапа. На первом этапе сравнивается пространственное расположение эл-

элементов в векторе (пространственное объединение), в результате чего выбирается наиболее подходящий пространственный центр. На втором этапе выбирается временная группа, в которой объединены близкие по времени появления пространственные центры (временное объединение). На рис. 2 изображена структура узла первого уровня.

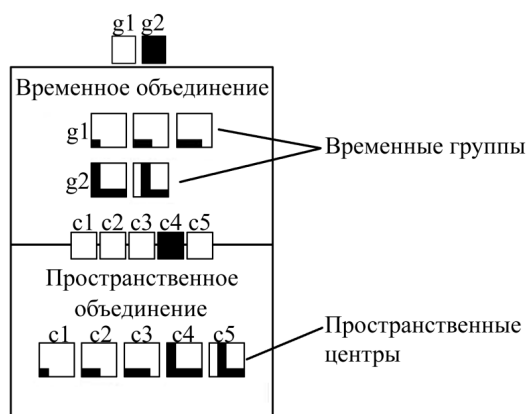


Рис. 2. Структура узла. Пример узла первого уровня:  $c1-c5$  – пространственные центры;  $g1, g2$  – временные группы

### Пространственное объединение

Пространственное объединение необходимо для фильтрации и компрессии входных данных и их первичного обобщения.

На этапе обучения в узле из входящих бинарных векторов формируются пространственные центры. Чтобы определить, был ли данный центр уже сохранен, рассчитывается евклидово расстояние  $d_i$  между входным вектором и существующими центрами. Если  $d_i < \varepsilon$ , где  $\varepsilon$  – заданная погрешность распознавания, то такой центр уже присутствует в узле, иначе, добавляем в узел новый центр. Обучение происходит до тех пор, пока скорость появления новых центров не снизится до некоторого малого значения.

Если узел не является узлом входного уровня, его пространственные центры представляют собой вектора из индексов временных групп предыдущего уровня.

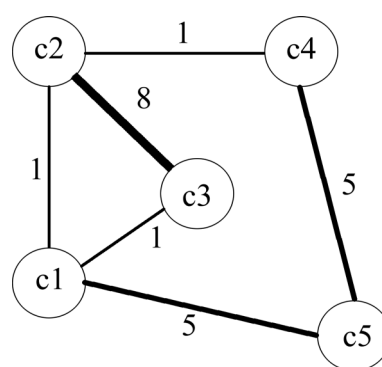
### Временное объединение

В процессе обучения сети, помимо запоминания пространственных центров, происходит их объединение во временные группы. В одну группу попадают пространственные центры, часто приходящие на вход узла друг за другом. Для этого составляется матрица смежности  $T$ , хранящая в себе количество следований одного центра за другим. По матрице  $T$  строится цепь Маркова, где номера пространственных центров являются вершинами, а частота их взаимных переходов – весом ребер. После того, как обучающий сигнал закончился, из вершин цепи Маркова формируются временные группы. В одну временную группу отбираются вершины, соединенные ребрами с наибольшими весами. На рис. 3 приведены примеры ма-

трицы смежности  $T$  (рис. 3, а) и цепи Маркова, состоящей из 5 пространственных центров. В результате выделились 2 временные группы:  $g1$  и  $g2$  (рис. 3, б).

	c1	c2	c3	c4	c5
c1	0	1	1	0	5
c2	0	0	8	1	0
c3	0	0	0	0	0
c4	0	0	0	0	5
c5	0	0	0	0	0

а



б

Рис. 3. Формирование временных групп: а) матрица смежности  $T$ ; б) цепь Маркова  $g1: c1, c4, c5; g2: c2, c3$

Ключевым элементом в работе сети, связывающим этапы пространственного и временного объединения, является матрица  $P(C|G)$ , строки которой соответствуют временным группам, а столбцы – пространственным центрам. Если какой-либо из центров не содержится в определенной временной группе, то на пересечении соответствующего столбца и строки матрицы  $P(C|G)$  стоит «0», иначе этот элемент матрицы содержит относительную нормированную вероятность, равную отношению частоты возникновения данного пространственного центра во входном сигнале к частоте возникновения всех пространственных центров данной временной группы.

Обучение узла верхнего уровня сети может производиться с учителем – элементом, который позволяет правильно разбивать пространственные центры узла по временным группам на основе ярлыков, присвоенных обучающим данным. В этом случае каждая временная группа верхнего узла однозначно соответствует определенному классу объектов во входном сигнале.

### Тестирование сети

В случае идеального тестирования (без помех) используется стратегия «победитель получает все»: находится пространственный центр, наиболее соответствующий входному сигналу (с наименьшим отклонением по евклидовой метрике), после

чего выбирается временная группа, в которую этот центр входит. Информацию о временной группе – «победителе» узел подает на вход узлу вышестоящего уровня.

В процессе основной работы сети входной сигнал, как правило, подается с помехами. Соответствующий пространственный центр находится по формуле Гаусса, с помощью которой рассчитываются вероятности совпадения входного вектора с существующими пространственными центрами:

$$p(c_j) = e^{-d_j^2/\sigma^2},$$

где  $d$  – степень отклонения входного вектора от текущего центра по евклидовой метрике;  $\sigma$  – настраиваемый нормировочный параметр, соответствующий степени предполагаемого отклонения (пропорционален степени зашумления входных данных).

Пример распределения вероятности совпадения входного вектора с пространственными центрами для узла первого уровня (рис. 4):  $p(c1)=0,1$ ;  $p(c2)=0,02$ ;  $p(c3)=0,7$ .

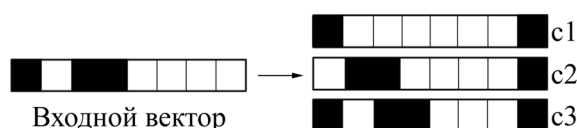


Рис. 4. Выбор соответствующего центра квантизации во время пространственного объединения. Этап тестирования с шумом

Таким образом, в ходе пространственного объединения строится вектор распределения вероятности по пространственным центрам  $y(c_j)$ . Полученный вектор умножается на матрицу  $P(C|G)$ , результатом является распределение вероятности по временным группам  $\lambda(g_i)$  для текущего узла, которое подается на вход узла следующего уровня:

$$\lambda(g_i) = \sum_{j=1}^{N_c} P(c_j | g_i) y(c_j).$$

Если узел находится на втором уровне и выше, вероятность совпадения входного сигнала с существующими в его памяти пространственными центрами находится перемножением вероятностей вхождения временных групп из нижестоящих узлов в соответствующий центр:

$$p(c_j) = \alpha \prod_{k=1}^M \lambda^k(r_j^k),$$

где  $M$  – число присоединенных узлов;  $r_j^k$  – индекс временной группы  $k$ -го узла, входящей в  $j$ -й пространственный центр;  $\alpha$  – нормировочный параметр.

Выходной информацией из узла на этапе тестирования является вектор распределения вероятности входного вектора по временным группам  $\lambda(g_i)$ . Выходом верхнего слоя является вектор распределения вероятности по классам объектов.

### Результаты численных экспериментов

Авторами данной статьи была реализована система распознавания образов на основе модели иерархической временной памяти. Система представляет собой сеть минимально возможной конфигурации – двухслойную, с 64-мя узлами на первом уровне и 1-м на втором. Входными данными служили бинарные изображения  $32 \times 32$  («0» – белый цвет – отсутствие сигнала, «1» – черный цвет – сигнал). На вход каждого узла первого уровня подавалась область изображения  $4 \times 4$  пикселя. В качестве обучающей и тестовой информации для классификации были подготовлены изображения цифр, состоящих только из горизонтальных и вертикальных линий. Минимальный размер цифры на изображении должен был быть не меньше области видимости двух узлов ( $8 \times 4$  пикселя), максимальный был ограничен зоной  $30 \times 30$  пикселей (для возможности создания обучающего видео). Для каждого класса цифр было создано 30 обучающих изображений, рис. 5.

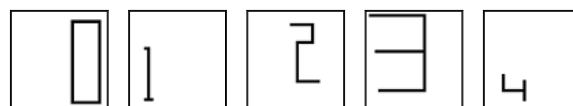


Рис. 5. Примеры обучающих изображений

Было поставлено три задачи классификации – разделение по 2, 4 и 10 классам объектов. В первой задаче для обучения использовались 60 изображений цифр «3» и «7», во второй 120 изображений цифр «3», «5», «7», «9», а в третьей 300 изображений всех десяти цифр.

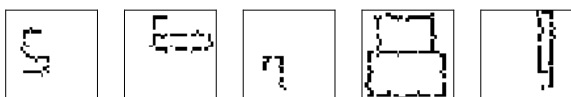
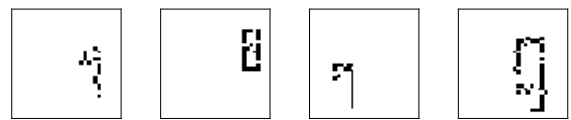
Одной из существенных трудностей было задание общих настроечных параметров, влияющих на функционирование сети, число которых было порядка десяти. Для их получения использовался эволюционный алгоритм, создающий популяции из сетей, обучающий их, тестирующий на обучающей выборке и отбирающий особи с лучшими результатами для кроссинговера и мутаций. Таким образом, изучалась чувствительность сети к настроечным параметрам и ее зависимость от количества классов, по которым требуется разделить входные данные. Для первой популяции сетей параметры задавались случайным образом в рамках допустимых значений. Когда эволюционный алгоритм сходился (средний коэффициент распознавания по популяции не увеличивался несколько поколений), из популяции отбирались наиболее удачные особи для прохождения набора тестов. Результаты тестов приведены в табл. 1.

Тестовые данные были созданы в специальной программе путем искажения обучающей выборки (рис. 6).

Как можно видеть из результатов, сеть обладает устойчивостью работы по отношению к значениям настроечных параметров, т. к. при их случайных значениях (I поколение) полученный результат распознавания выше случайного угадывания.

**Таблица 1.** Результаты тестирования на искаженных изображениях, % распознавания

Результат распознавания	Количество классов изображений		
	2	4	10
Лучшие особи сети (VII поколение эволюции)			
На обучающей выборке	100,0	100,0	99,0
На тестовой выборке	<b>96,7</b>	<b>95,8</b>	<b>95,0</b>
Средний результат распознавания на обучающей выборке			
I поколение (случайные значения)	85,8	66,8	51,3
III поколение	98,2	96,8	94,3
V поколение	99,0	98,2	97,8

**Рис. 6.** Пример тестовых данных**Рис. 7.** Примеры неправильно распознанных изображений

Сеть с 4 классами распознавания была протестирована на всех 300 тестовых изображениях для того, чтобы выяснить, не находит ли сеть зависимости, общие не для одного, а для нескольких классов цифр сразу. Результаты тестирования приведены в табл. 2.

Из табл. 2 можно заключить что, распределение «неизвестных» цифр по обученным классам происходит неравномерно (например, склонность к распознаванию «1» и «4» как «7»). Однако, погреш-

ность такого распознавания гораздо выше погрешности распознавания «знакомых» цифр. Следовательно, найденные сетью закономерности или свойства для каждого обученного класса не являются обобщающими для нескольких классов цифр.

**Таблица 2.** Распределение 10 классов изображений по 4, на которых производилось обучение сети

Распознаны как:	«3»	«5»	«7»	«9»
«0»	6	2	11	11
«1»	7	1	22	0
«2»	9	9	2	10
<b>«3»</b>	<b>29</b>	<b>1</b>	<b>0</b>	<b>0</b>
«4»	2	0	23	5
<b>«5»</b>	<b>2</b>	<b>28</b>	<b>0</b>	<b>0</b>
«6»	2	10	12	6
<b>«7»</b>	<b>1</b>	<b>0</b>	<b>29</b>	<b>0</b>
«8»	6	6	4	14
<b>«9»</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>29</b>
Всего по классам	64	57	104	75

### Выводы

1. Реализована система распознавания образов на основе модели иерархической временной памяти.
2. Предложен и разработан эволюционный метод подбора коэффициентов сети, что положительно сказалось на результатах тестирования.
3. Программа апробирована на примере распознавания стилизованных цифр. Лучшие результаты работы сети получились при коэффициентах, полученных на 7-ом поколении эволюции, и составляют 95...96 % правильно распознанных бинарных изображений цифр на тестовой выборке, искаженной на 20 % по сравнению с обучающей.
4. Исследована зависимость процента распознавания системы от количества категорий цифр.
9. Hierarchical Temporal Memory: Comparison with Existing Models // numenta.com: сайт компании Нумента 2007. URL: [http://numenta.com/htm-overview/education/HTM\\_Comparison.pdf](http://numenta.com/htm-overview/education/HTM_Comparison.pdf) (дата обращения: 14.09.2010).
10. Dileep G., Jaros B. The HTM Learning Algorithms // numenta.com: сайт компании Нумента. 2006. URL: [http://numenta.com/htm-overview/education/Numenta\\_HTM\\_Learning\\_Algos.pdf](http://numenta.com/htm-overview/education/Numenta_HTM_Learning_Algos.pdf) (дата обращения: 14.09.2010).
11. Problems that Fit HTM // numenta.com: сайт компании Нумента. 2007. URL: <http://www.numenta.com/htm-overview/education/ProblemsThatFitHTMs.pdf> (дата обращения: 14.09.2010).
12. Hawkins J. Hierarchical Temporal Memory: How a theory of neocortex may lead to truly intelligent machines, видеозапись лекции от 18.03.2010 г. 2010. URL: [http://www.youtube.com/watch?v=TDzr0\\_fbnVk](http://www.youtube.com/watch?v=TDzr0_fbnVk) (дата обращения: 24.09.2010).
13. Применение технологии HTM в коммерческих приложениях. 2010. URL: <http://numenta.com/about-numenta/customers.php> (дата обращения: 24.09.2010).
14. Wölpert D. The Lack of a Priori Distinctions Between Learning Algorithms // Neural Computation. – 1996. – V. 8. – № 7. – P. 1341–1390.
15. Hartung J., McCormack J. Support for the Use of Hierarchical Temporal Memory Systems in Automated Design Evaluation: a first experiment // www.uidaho.edu: сайт Университета Айдахо, США. 2010. URL: <http://www.webpages.uidaho.edu/~mccormack/papers/DETC2009-87702.pdf> (дата обращения: 24.09.2010).

### СПИСОК ЛИТЕРАТУРЫ

1. Hawkins J., Blakeslee S. On Intelligence. – N.Y.: Times Books, 2004. – 272 p.
2. Hawkins J., Dileep G. Temporal Memory Concepts, Theory, and Terminology // numenta.com: сайт компании Нумента. 2006. URL: [http://numenta.com/htm-overview/education/Numenta\\_HTM\\_Concepts.pdf](http://numenta.com/htm-overview/education/Numenta_HTM_Concepts.pdf) (дата обращения 04.08.2010).
3. Савельев С.В. Происхождение мозга. – М.: ВЕДИ, 2005. – 368 с.
4. Циммерман М., Ениг В. и др. Физиология человека. В 3 томах. Т. 2. – М.: Мир, 2005. – 320 с.
5. Vilis T. The Physiology of the Senses. Transformations for perception and action// The University of Western Ontario. 2008. URL: <http://www.physpharm.fmd.uwo.ca/undergrad/sensesweb/L1Eye/L1Eye.pdf> (дата обращения 04.06.2010).
6. Mather G. The Visual Cortex//www.lifesci.sussex.ac.uk: сайт университета Сассекс, Великобритания. URL: [http://www.lifesci.sussex.ac.uk/home/George\\_Mather/Linked%20Pages/Physiol/Cortex.html](http://www.lifesci.sussex.ac.uk/home/George_Mather/Linked%20Pages/Physiol/Cortex.html) (дата обращения 20.10.2010).
7. Dileep G., Jaros B. The HTM Learning Algorithms//numenta.com: сайт компании Нумента. 2007. URL: [http://numenta.com/htm-overview/education/Numenta\\_HTM\\_Learning\\_Algos.pdf](http://numenta.com/htm-overview/education/Numenta_HTM_Learning_Algos.pdf) (дата обращения 10.08.2010).
8. Dileep G. How the Brain Might Work: A Hierarchical and Temporal Model for Learning and Recognition // numenta.com: сайт компании Нумента. 2007. URL: <http://www.numenta.com/htm-overview/education/DileepThesis.pdf> (дата обращения: 14.09.2010).

Поступила: 30.11.2010 г.