

ИКОНИКА – НАУКА ОБ ИЗОБРАЖЕНИИ

УДК 612.84 + 519.72

ПРОБЛЕМЫ КОДИРОВАНИЯ СТЕРЕОИЗОБРАЖЕНИЙ В ПАМЯТИ ЧЕЛОВЕКА

© 2010 г. Г. Крумина*, Dr. phys; В. А. Ляховецкий**, канд. техн. наук

* University of Latvia, Riga, Latvia

** Институт физиологии РАН им. И.П. Павлова, Санкт-Петербург

E-mail: gkrumina@cfi.lu.lv, v_la2002@mail.ru

Изучено запоминание и воспроизведение человеком последовательности плоских или стереоскопических изображений, включающей 6 кадров, содержащих плоскую (8×8 положений стимула) или объемную полоску ($8 \times 4 \times 2$ положений). На этапе воспроизведения испытуемый в каждом кадре выбирал между стимулом и тремя дистракторами. Показано, что времена распознавания и воспроизведения меньше для объемных стимулов, а процент правильных ответов больше для плоских стимулов. Для объемных стимулов распределение ошибок зависит от диспаратности между целью и выбранным дистрактором. Модель, разработанная на основе гетероассоциативной нейросети, воспроизводит распределение ошибок для плоских, но не для объемных стимулов. Полученные данные свидетельствуют о том, что внутренние представления для плоских и трехмерных объектов существенно различаются: диспаратность существенно влияет на запоминание и узнавание трехмерных объектов.

Ключевые слова: стереоскопическое зрение, пространственная память, гетероассоциативные нейронные сети.

Коды OCIS: 330.1400

Поступила в редакцию 26.01.2010

Введение

Одним из важных признаков удаленности объектов является диспаратность, которая определяется как разность между координатами проекций объектов на сетчатки глаз. Одной диспаратности достаточно, чтобы человек воспринимал глубину объектов, расположенных на плоском изображении (например, в случайно-точечной стереограмме) [1]. Широко распространена точка зрения, согласно которой нейроны, ответ которых зависит от диспаратности, являются физиологической основой работы системы бинокулярного зрения [2, 3].

Диспаратность используется в моделях выделения глубины со временем работ Юлеша [1] и Марра [4]. В настоящее время существуют модели, способные фузировать части стереограммы, обладающие различным контрастом. Они также позволяют объяснить явления “притяжения” и “отталкивания” диспаратностей [5]. Сложные алгоритмы искусственного стереоско-

ического зрения широко применяются в робототехнике [6].

Отметим, что важность стереоскопической информации для восприятия глубины объектов реального мира была оспорена Гримсоном [7]. Он полагал, что эта информация используется лишь для отделения объектов от фона, так как для восстановления глубины объектов по диспаратности требуется точная калибровка зрительного аппарата, лежащая за пределами возможностей зрительной системы человека.

Представлению стереоскопической информации при восприятии посвящено множество работ, однако использование стереоскопических признаков при запоминании исследовано значительно менее полно. Существует две противоположные точки зрения на эту проблему. Согласно первой, в памяти хранится трехмерное представление объектов, инвариантное к точке обзора [8]. Согласно второй, при распознавании объектов реального мира человек опирается не на трехмерную карту диспаратностей, а на набор свя-

занных двумерных внутренних представлений объектов [9]. Такие представления используют в робототехнике в задачах распознавания образов [10, 11].

Однако литературные данные не позволяют сделать уверенный выбор между этими точками зрения. Известно, например, что стереоскопическая информация сохраняется в оперативной памяти, но не улучшает качества обработки сложных сцен [12]; что саккадическая система имеет доступ к стереоскопической информации из кратковременной памяти, однако в течение лишь малого промежутка времени [13].

Является ли диспаратность вспомогательной характеристикой, используемой низкоуровневым зрением, например, для вергентных движений глаз и для построения формы, или же значимым для высокоуровневого зрения параметром сцены? В последнем случае при запоминании 2D и 3D-объектов могут использоваться различные внутренние представления (схемы кодирования) пространства.

Методика исследования

Как правило, структуры внутренних представлений изучаются на основе анализа правильных ответов испытуемого при восприятии статических пространственных конфигураций [14]. Авторами настоящей статьи использован подход, основанный на анализе распределения ошибок при запоминании последовательностей [15–17]. Предполагается, что ошибки обуслов-

лены искажением элементов внутреннего представления, в котором запоминаются объекты. При этом более вероятно искажение небольшого числа элементов. Тогда при ошибочном воспроизведении испытуемым выбирается “близкий” к запомненному в выбранном внутреннем представлении объект.

Процедура эксперимента. Испытуемый должен был запомнить и немедленно воспроизвести последовательность из 6 случайных перемещений полоски по доске размером 8×8 клеток (схема А), по доске размером 8×4 клеток (схема Б); причем полоска могла быть расположена либо за плоскостью доски, либо перед ней (всего 8×4×2 возможных положения). В последнем случае для предъявления стимулов использовался метод анаглифов, диспаратность полоски составляла $\pm 0,15^\circ$.

На этапе воспроизведения в каждом опыте испытуемый должен был выбрать между правильной позицией (целью) и тремя дистракторами (рис. 1). Мерой близости между выбранным испытуемым дистрактором и целью является евклидово расстояние d , вычисляемое в клетках доски, в схеме (А) на плоскости или же в схеме (Б) в 3-х мерном пространстве.

Испытуемые. В опыте участвовали 22 человека в возрасте 20–30 лет. С каждым испытуемым было проведено 3 опыта по схеме (А), а затем 3 опыта по схеме (Б).

Математическое моделирование. Для моделирования была использована гетероассоциативная нейронная сеть, впервые предложенная

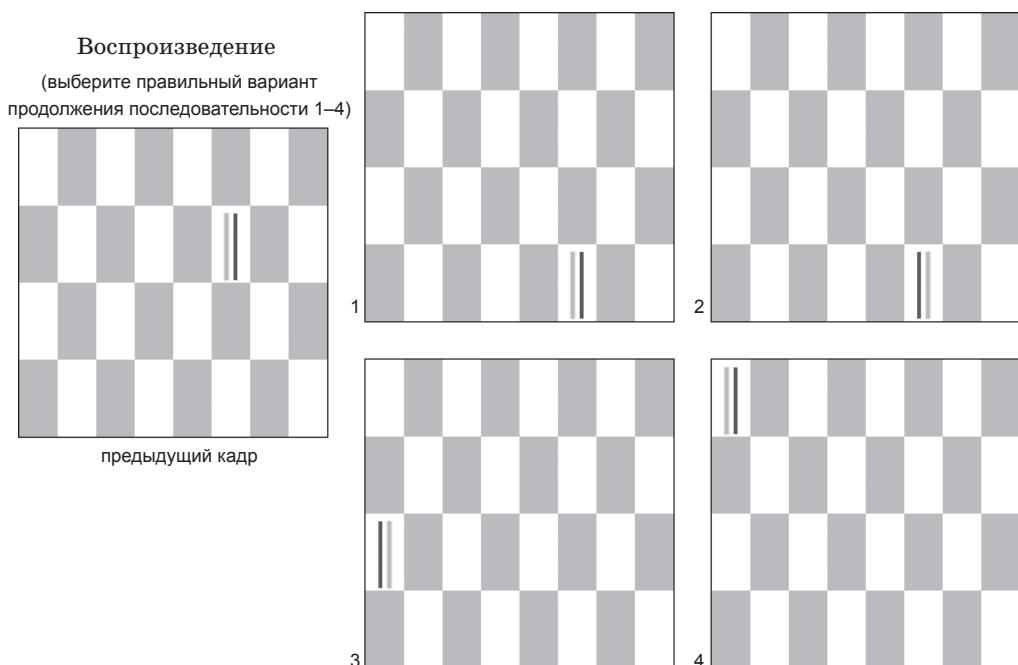


Рис. 1. Этап воспроизведения эксперимента с объемным стимулом.

Коско [18]. Она способна запомнить пары векторов, хранящих сведения об объектах, и восстановить вектор, когда ассоциированный с ним вектор предлагается ей в качестве подсказки. Важным свойством сети является способность “ошибаться”, не задаваемая искусственно при моделировании, но возникающая в результате особенностей распределенного хранения векторов.

В сети два слоя нейронов, соединенных двунаправленными связями. Нейроны работают с бинаризованными данными, поэтому элементы векторов могут принимать лишь значения “−1” и “1”. На этапе запоминания на основе цепочки пар векторов $Z_{i-1} \rightarrow Z_i$, $i = 1\text{--}6$ рассчитывается матрица весов W . Векторы Z хранят номер позиции i и декартовы координаты полоски на плоскости (А) или же в 3-х мерном пространстве (Б). При бинаризации эти компоненты вектора переводятся в двоичную систему счисления с последующей заменой “0” на “−1”.

На этапе воспроизведения для цели X_i и для дистракторов D_{ij} , $j = 1\text{--}3$, рассчитываются аддитивные составляющие матрицы весов $W_{X_i} = kZ_{i-1}^T X_i$ и $W_{D_{ij}} = kZ_{i-1}^T D_{ij}$, где $k = M \ln[(1 - r)/r]/[2N(1 - 2r)]$, M – количество запоминаемых пар векторов, N – количество бит в векторе, r – количество потенциально зашумленных бит в векторе [19]. Сеть пытается воспроизвести гипотетическую цель X'_i и дистракторы D'_{ij} на основе правильной предыдущей позиции Z_{i-1} и матриц $(W + W_{X_i})$ и $(W + W_{D_{ij}})$. В качестве ответа сеть выбирает тот, для которого минимально расстояние Хемминга $\|X_i - X'_i\|$ и $\|D_{ij} - D'_{ij}\|$.

С целью было проведено 10 000 опытов для схем (А) и (Б). Способность сети совершать собственные ошибки позволяет непосредственно сравнить особенности запоминания человека и модели. Поэтому результаты модельных опытов обрабатывались аналогично данным психофизиологических экспериментов.

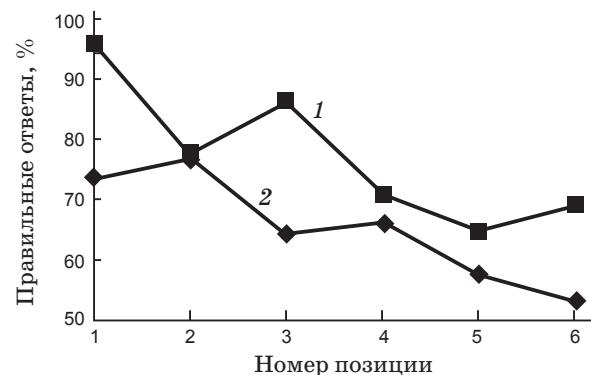


Рис. 2. Зависимость числа правильных ответов от номера позиции в последовательности. 1 – схема А, 2 – схема Б.

Результаты исследования

Психофизический эксперимент. Для стереостимулов среднее время запоминания и среднее время воспроизведения достоверно ($p < 0,05$) меньше, чем для плоских стимулов ($T_{A_{mem}} = 6,8$ с, $T_{B_{mem}} = 5,7$ с, $T_{A_{rec}} = 10,5$ с, $T_{B_{rec}} = 8,5$ с). Напротив, число правильно воспроизведенных позиций последовательности достоверно ($p < 0,05$, критерий знаков) выше для плоских стимулов ($N_A = 4,4$, $N_B = 3,6$) – для каждого номера позиции в последовательности процент правильных ответов ниже для объемных стимулов (рис. 2).

Распределение числа ошибок E от расстояния d между выбранным дистрактором и целью для схемы (А) имеет максимумы при $d = 2$ клеткам и 4 клеткам (рис. 3а). Распределения числа ошибок E от расстояния d между выбранным дистрактором и целью для схемы (Б) достоверно ($p < 0,05$, критерий Манна–Уитни) различны в зависимости от того, имеют ли цель и дистрактор одинаковую (eq) или же различную ($uneq$) диспаритетность. При $d = 3$ клеткам кривая $E_{B_{eq}}$ имеет минимум, а кривая $E_{B_{uneq}}$ – максимум (рис. 3б).

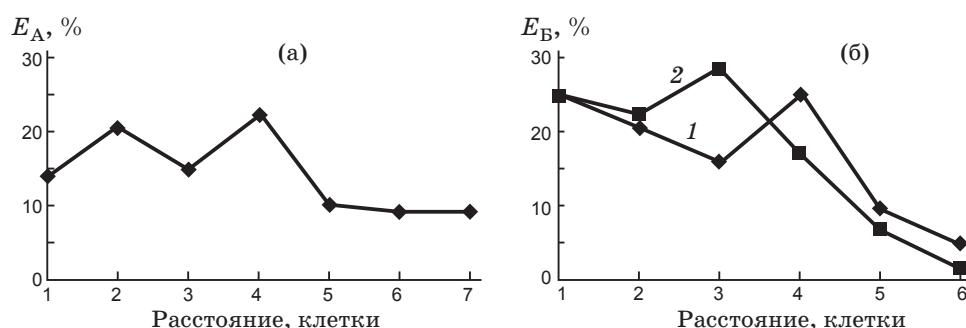


Рис. 3. Зависимости числа ошибок человека от расстояния для плоского (а) и объемного (б) стимулов. 1 – одинаковая диспаритетность, 2 – различная диспаритетность.

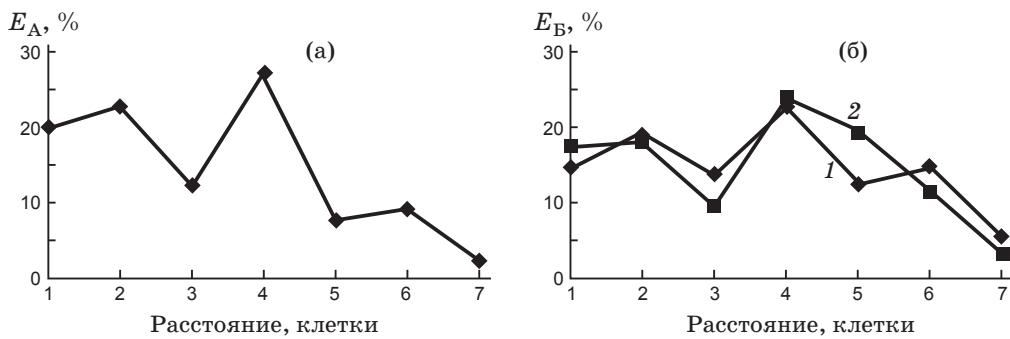


Рис. 4. Зависимости числа ошибок модели от расстояния для плоского (а) и объемного (б) стимулов. 1 – одинаковая диспаратность, 2 – различная диспаратность.

Математическое моделирование. Число правильно воспроизведенных моделью позиций составляет 85% ($N_A = N_B = 5,1$) независимо от номера позиции и от мерности стимула.

Распределение модельных ошибок E от расстояния d между выбранным дистрактором и целью для схемы (А) также, как и распределение ошибок человека, имеет максимумы при $d = 2$ клеткам и 4 клеткам (рис. 4а). Распределения модельных ошибок при условии (Б) не зависят от того, одинаковы ли диспаратности стимула и выбранного дистрактора (рис. 4б). Эти модельные распределения подобны распределению ошибок человека при условии, что стимул и выбранный дистрактор имеют одинаковую диспаратность.

Обсуждение

Следует отметить, что в данной экспериментальной схеме от испытуемого не требовалось специально запоминать глубину стереоскопического стимула. Задача выбора правильного ответа из четырех вариантов могла быть решена и при запоминании координат объекта на плоскости. При этом испытуемый должен был запомнить в каждом опыте лишь одно из $8 \times 4 = 32$ (а не 64) положений объекта. Тогда, очевидно, кривые распределения ошибок от расстояния между стимулом и выбранным дистрактором не зависели бы от диспаратности объекта. Однако это не так (рис. 3б).

Достоверные различия $T_{A_{mem}} > T_{B_{mem}}$, $T_{A_{rec}} > T_{B_{rec}}$ как таковые могут быть объяснены тем, что опыты с объемными стимулами проводились после опытов с плоскими стимулами, то есть тогда, когда у испытуемых уже появлялся навык выполнения тестового задания. А достоверное различие $N_A > N_B$ может быть объяснено

затруднением условий восприятия испытуемого из-за применения красно-зеленых очков. Однако сочетание этих достоверных различий свидетельствует о различных механизмах обработки плоских и объемных стимулов.

Полученные данные сопоставимы с результатами экспериментов по воспроизведению запомненных последовательных перемещений шахматной фигуры, слона, по шахматной доске, предъявляемой на экране [15–17]. Было показано, что распределение ошибок при воспроизведении шахматистами запомненной последовательности ходов слона (движение по диагоналям доски) имеет экспоненциальную форму. Такая форма кривой может быть объяснена тем, что тогда, когда фигура ходит по правилам, шахматисты запоминают именно смещения, а не текущее положение фигуры. Напротив, распределение ошибок при запоминании последовательности случайных перемещений слона имеет несколько максимумов и объяснимо в рамках представлений о кодировании абсолютных координат. Форма распределения E_A подобна форме распределений ошибок при запоминании последовательности случайных перемещений шахматной фигуры. Она может быть объяснена в рамках нейросетевой модели, использующей для кодирования объектов их абсолютные координаты. Распределения E_B , также как и E_A , имеют несколько максимумов, что позволяет предположить, что в памяти человека хранятся абсолютные координаты объемных объектов. Однако лишь одно из этих распределений ($E_{B_{eq}}$) можно воспроизвести с помощью разработанных моделей, путем простого увеличения мерности входного вектора данных Z_i .

Таким образом, внутреннее представление координат объемных объектов принципиально отличается от внутреннего представления ко-

ординат плоских объектов: диспаратность существенно влияет на запоминание и узнавание трехмерных объектов. Косвенным подтверждением этому тезису служат и данные о том, что вентральные области коры головного мозга, участвующие в обработке формы объемных объектов, не полностью пересекаются с областями, анализирующими форму плоских объектов [20]. По-видимому, вычислительной способности человеческого мозга достаточно для хранения сложных трехмерных внутренних представлений объектов окружающего мира. Вероятно, с повышением быстродействия аппаратной базы в робототехнике также будет уделено значительное внимание сложным трехмерным схемам кодирования формы объектов при обучении и распознавании.

ЛИТЕРАТУРА

1. Julesz B. Foundations of cyclopean perception. Chicago: The University of Chicago Press, 1971.
2. Cumming B.G., De Angelis G.C. The physiology of stereopsis // Ann. Rev. Neurosci. 2001. V. 24. P. 203–238.
3. Livingstone M.S., Hubel D.H. Psychophysical evidence for separate channels for the perception of form, color, movement and depth // Journal of Neuroscience. 1987. V. 7(11). P. 3416–3468.
4. Marr A., Poggio T. Cooperative computation of stereodisparity // Science. 1976. V. 194. P. 283–287.
5. Qian N., Zhu Y. Physiological computation of binocular disparity // Vision Research. 1997. V. 37. P. 1811–1827.
6. Pollard S.B., Pridmore T.P., Porritt J., Mayhew J.E.W., Frisby J.P. Geometrical modeling from multiple stereo views // The International J. of Robotics Research. 1989. V. 8. P. 3–32.
7. Grimson W.E. Why stereo vision is not always about 3D reconstruction // A.I. Memo. № 1435. 1993. MIT.
8. Biederman I., Gerhardstein P.C. Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance // J. of Experimental Psychology: Human Perception and Performance. 1993. V. 19. P. 1162–1182.
9. Bulthoff H.H., Edelman S.Y., Tarr M.J. How are three-dimensional objects represented in the brain? // Cerebral Cortex. 1995. V. 5. P. 247–260.
10. Saxena A., Driemeyer J., Ng A.Y. Robotic grasping of novel objects using vision // The International J. of Robotics Research. 2008. V. 27. P. 157–173.
11. Kragic D., Bjorkman M., Christensen H.I., Ekluendh J.-O. Vision for robotic object manipulation in domestic settings // Robotics and Autonomous Systems. 2005. V. 52. P. 85–100.
12. Busey T.A. The perception and memory of stereoscopic depth information in naturalistic objects // Technical Report #207. Indiana University Press. 1996.
13. Kapoula Z., Bucci M.P., Lavigne-Tomps F., Zamfirrescu F. Disconjugate memory-guided saccades to disparate targets: evidence for 3D sensitivity // Exp Brain Res. 1998. V. 122. P. 413–423.
14. Jiang Y., Olson I.R., Chun M.M. Organization of visual short-term memory // J. of Exp. Psychology: Learning, Memory and Cognition. 2000. V. 26. P. 683–702.
15. Ляховецкий В.А., Потапов А.С., Попечителев Е.П. Методика изучения и модель информационной структуры памяти человека // Известия ТРТУ. Тематический выпуск “Медицинские информационные системы”. 2006. № 11. С. 4–9.
16. Lyakhovetskii V., Potapov A., Ivanov S. Strategies for storing spatial transformations of chess positions // Perception Supplement. 2006. V. 35. P. 105.
17. Lyakhovetskii V.A., Bobrova E.V. The strategies of coding in spatial memory // Perception Supplement. 2007. V. 36. P. 51.
18. Kosko B. Bidirectional associative memories // IEEE Transactions on Systems, Man and Cybernetics. 1988. V. 18. P. 49–60.
19. Kothari R., Lotlikar R., Cahay M. State-dependent weights for neural associative memories // Neural Computation. 1998. V. 10. P. 59–71.
20. Kourtzi Z., Kanwisher N. Representation of perceived object shape by the human lateral occipital complex // Science. 2001. V. 293. P. 1506–1509.