

THE PROBLEMS OF 3-D SPACE CODING IN HUMAN MEMORY

G. Krumina,¹ V.A. Lyakhovetskii,²

¹ Department of Optometry and Vision Science, University of Latvia
Kengaraga st. 8, LV 1063 Riga, Latvia
E-mail: gkrumina@cfi.lu.lv

² Department of Biomedical Electronics, St. Petersburg Electrotechnical University
ul Prof. Popova 5, St. Petersburg, 197376, Russia

ABSTRACT

The objectives of this study were human memorization and recognition processes in respect of 2-D and 3-D objects sequences. In each trial we used six sequential slides, each of them containing a 2-D or 3-D bar located on the chessboard. This bar was positioned in one of 8x8 (for 2-D) or 8x4x2 (for 3-D) locations. At the recognition stage there were the target and three distracters on each slide, the subject to recognize the target.

It was shown that performance as well as memorization time and recognition time for the 3-D stimuli were significantly lower than for the 2-D ones. For the 3-D trial, the error distribution depended on whether the target and the chosen distracter had same or different binocular disparities.

The data obtained were simulated by creating a model developed on the basis of bidirectional associative memory network. The model successfully reproduced the error distribution for the 2-D sequences, but not for the 3-D ones if the target and the chosen distracter had different disparities.

Thus, we conclude that the internal representation of 3-D objects fundamentally differs from that of 2-D objects.

Keywords: spatial memory, stereovision, bidirectional associative memory network.

1. INTRODUCTION

One of the important attributes of object remoteness is the disparity, a difference between coordinates of object's projections on eyes retinas. Sole disparity is sufficient for perception of the 3-D forms from the 2-D image (for example, from the random-dot stereogram¹). It is widely believed that neurons responding to the disparity are a physiological basis of work of binocular vision system^{2, 3}. The disparity was used for depth selection simulation starting from the models by Julesz¹ and Marr⁴. Now there are models allowing for the fusion between images having different contrast. They also explain the phenomena of "attraction" and "repulsion" of disparities⁵. Complex artificial stereovision algorithms have practical use for many robotics applications⁶.

It should be noted that importance of the 3-D information for depth perception of the real world objects was challenged by Grimson⁷. He believed that this information was used only for separation of objects from their background, as restoration of objects' depth based on the disparity needs the exact calibration of the visual apparatus, which is beyond possibilities of the human visual system.

There are multiple studies on representation of the 3-D information at perception. At the same time little is known about significance of 3-D attributes at storing. There are two opposite

points of view on this problem. According to the first one, human memory stores a 3-D representation of objects that is invariant to a viewpoint⁸. According to the second one, recognition of the real world objects by humans is based not on a 3-D map of disparities, but on a set of interconnected 2-D viewpoint-specific representations of the objects⁹. The latter approach is also used in robotics in few works on objects recognition^{10,11}.

The literary data do not enable us to make the right choice between these approaches. E.g., it is known that the 3-D information is kept in working memory, but does not improve quality of complex scenes processing¹²; that saccadic system has access to the 3-D information retained in the short-term memory, however for a short time¹³. It is unclear whether the disparity is only auxiliary low level vision characteristic used for vergency eye movements and for the form reconstruction, or it is significant for high brain levels as an important scene parameter. In the latter case different internal representations of space can be used at perception and recognition of 2-D and 3-D objects.

2. METHODS

Usually the coding schemes are studied by analyzing the human performance of different static spatial configurations¹⁴. We use the different approach based on the analysis of error distribution at storing objects sequences¹⁵⁻¹⁷. It is supposed that errors are caused by the distortion of elements of internal representation in which objects are remembered, the distortion of a small number of elements being more probable. If the subject's recall is wrong, the subject chooses the answer the internal representation whereof is "close" to the correct object.

2.1. Psychophysical experiment

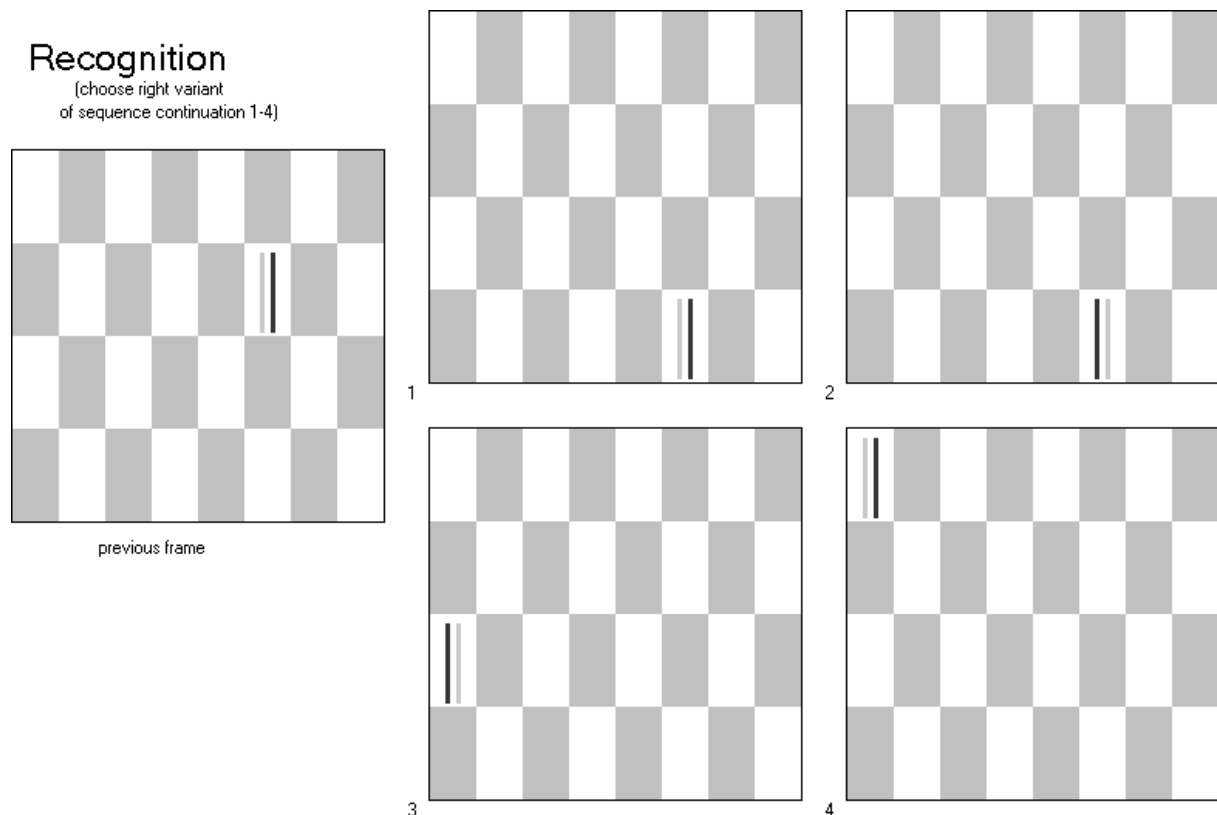


Fig. 1. The recognition stage of experiment with 3-D bar.

Within this approach, the stimuli enabling to estimate the quantitative closeness between the remembered objects are to be used. The subject should remember and after that immediately reproduce a sequence of six random displacements of a bar (a) on the 8 x 8 board or (b) on the 8 x 4 board, the bar located in front of or behind the board plane (8 x 4 x 2 bar positions total). In the latter case, the bar disparity equals to $\pm 0.15^\circ$, the anaglyph method is used for stereo presentation.

At the recognition stage in every trial the subject gets to see the correct previous position and should make a choice between the correct position (target) and three distracters (fig. 1). The closeness measure between the chosen distracter and the target is Euclidian distance d measured in board cells (a) in 2-D or (b) in 3-D space.

Twenty two subjects (age of 20-30 years) participated in experiments, for each there were carried out 3 series under (a) conditions and then 3 series under (b) conditions.

2.2. Model

Data obtained were simulated by creating the bidirectional associative memory network first proposed by Kosko¹⁸. At the memorization stage, the weight matrix W is calculated on the basis of a "chain" of vector pairs $Z_{i-1} \rightarrow Z_i$, $i = 1 \div 6$. Z_i hold in binary form the position's number i and bar coordinates (a) in 2-D or (b) in 3-D space. At the recognition stage the adaptive parts W_{X_i} and $W_{D_{ij}}$ of weight matrix¹⁹ are calculated for target X_i and for each distracter D_{ij} , $j = 1 \div 3$. The network recalls the hypothetical target X_i' and distracters D_{ij}' based on correct previous position Z_{i-1} and matrices $(W + W_{X_i})$ and $(W + W_{D_{ij}})$ respectively. As an answer the network chooses target or one of distracters for which the Hemming distance $\|X_i - X_i'\|$ and $\|D_{ij} - D_{ij}'\|$ is minimal.

10000 runs were carried out under (a) and (b) conditions. Due to ability of this network to make self-induced errors, it's possible to compare the human performance with the artificial one. Thus, the simulation results were processed in the same way as the psychophysical results.

3. RESULTS

3.1. Psychophysical experiment

It was shown that for 3-D stimuli the mean memorization time and the mean recognition time were significantly ($p < 0.05$) lower than for 2-D stimuli ($T_{a\ mem} = 6.8$ sec, $T_{b\ mem} = 5.7$ sec, $T_{a\ rec} = 10.5$ sec, $T_{b\ rec} = 8.5$ sec).

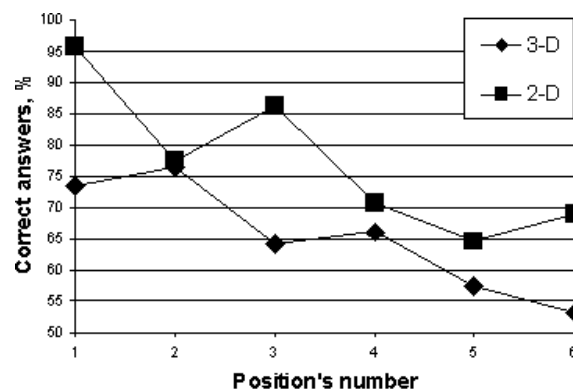


Fig. 2. The dependency of correct answers on position's number for 2-D and 3-D stimuli.

On the contrary, the amount of correctly reproduced positions is significantly ($p < 0.05$, sign test) higher for 2-D than for 3-D stimuli ($N_a = 4.4$, $N_b = 3.6$). This can be concluded from the

dependency of correct answers on position's number in the sequence. For all position's numbers the percentage of correct answers is lower for 3-D stimuli (fig. 2).

Distribution of human errors E from distance d between the chosen distracter and the target under (a) condition has maxima at $d = 2$ cells and $d = 4$ cells (fig. 3A). Distributions of human errors under (b) condition are significantly ($p < 0.05$, Mann-Whitney test) various, depending on whether the target and the distracter have the same (*eq*) or different (*uneq*) disparities. At $d = 3$ cells curve $E_{b\ eq}$ has a minimum, and curve $E_{b\ uneq}$ has a maximum (fig. 3B).

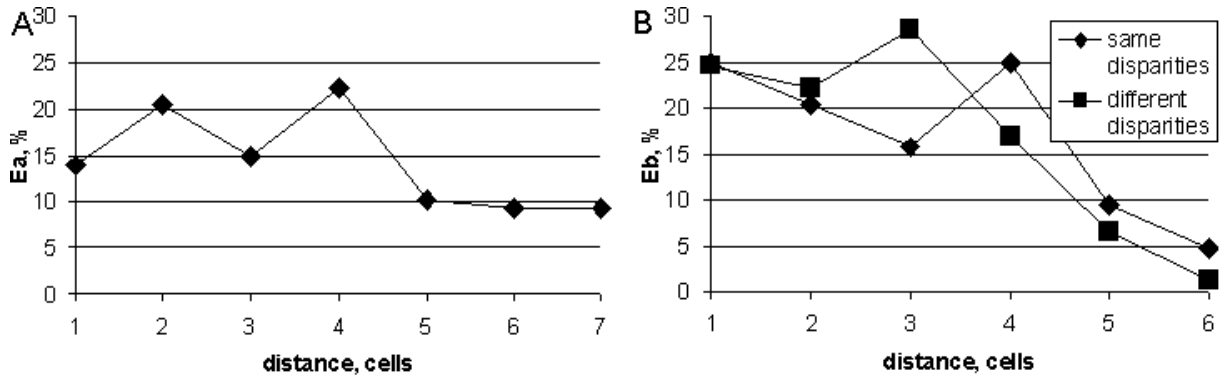


Fig. 3. The distributions of human errors from distance between the target and the chosen distracter (A) in 2-D space, (B) in 3-D space.

3.2. Simulation results

The amount of positions correctly reproduced by our network does not depend on space dimensionality and positions' number, it equals to 85%.

Distribution of model errors E from distance d between the chosen distracter and the target under (a) condition also (as a distribution of human errors) has maxima at $d = 2$ cells and $d = 4$ cells (fig. 4A). Distributions of model errors under (b) condition do not depend on whether the target and the distracter have the same or different disparities. At $d = 3$ cells both curves have a minimum, and at $d = 4$ cells they have a maximum (fig. 4B). These two model curves are similar to human error distribution when the target and the distracter have the same disparities.

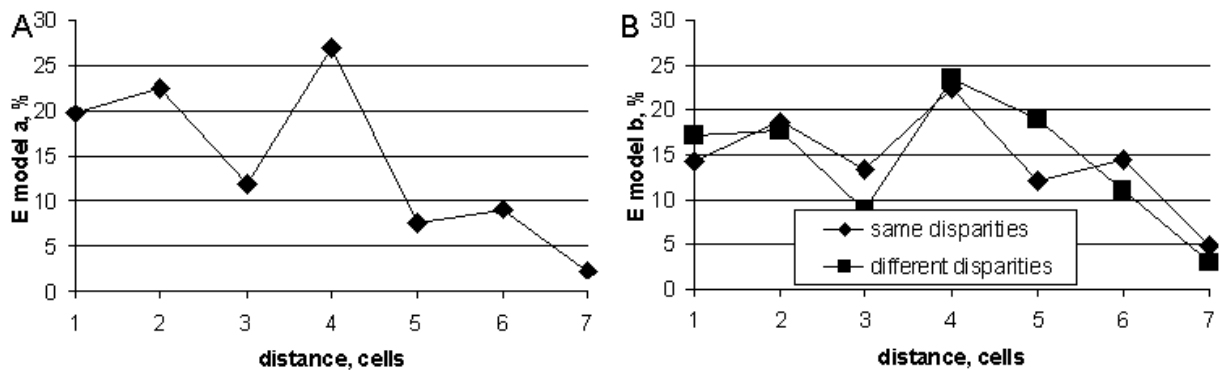


Fig. 4. The distributions of model errors from distance between the target and the chosen distracter (A) in 2-D space, (B) in 3-D space.

4. DISCUSSION

It should be noted that in the above experimental design it was not required to specially remember the depth. The problem of choosing the correct answer from four variants could be solved also by storing 2-D object coordinates. If it was so, in each 3-D slide the subject should remember one of 32 possible 2-D positions of the object. Thus, obviously, the distributions of

errors from the distance would not depend on the object disparity. However, this is not the case (fig. 3B).

The significant differences $T_{a\ mem} > T_{b\ mem}$, $T_{a\ rec} > T_{b\ rec}$ as such can be explained by the fact that experiments with 3-D stimuli were carried out after experiments with 2-D stimuli, i.e. when subjects were already familiar with this test task. The significant difference $N_a > N_b$ as such can be explained by difficulty in the subject perception due to the use of the red-green glasses. However the combination of these significant differences indicates that various mechanisms are involved in 2-D and 3-D stimuli processing.

The form of distribution E_a is similar to the form of error distributions at storing a sequence of random moves of a chess figure. It can be explained within the framework of the models using absolute 2-D coordinates¹⁵⁻¹⁷ for coding the objects (fig. 4A). Distributions E_b , as well as E_a , have more than one maximum that enables us to assume that absolute 3-D coordinates are stored in the human memory. However, only one of these two distributions can be reproduced by using the developed models, by increasing the vector Z_i length (fig. 4B).

Thus, we conclude that 3-D objects internal coordinates representation fundamentally differs from 2-D one. The indirect confirmation of this thesis is that the ventral regions of the cortex involved in the 3-D shape analysis do not perfectly overlap with the regions involved in the analysis of 2-D shape²⁰. Apparently, the computational power of the human brain is sufficient for storing complex 3-D representations of the real world objects. We expect that as soon as new high-speed hardware becomes available, the robotics will also pay attention to complex 3-D coding schemes in learning and recognition.

1. B. Julesz, Foundations of Cyclopean Perception, Chicago: The University of Chicago press. (1971).
2. B.G. Cumming, and G.C. De Angelis, "The physiology of stereopsis," *Ann. Rev. Neurosci.*, vol. 24, pp. 203-238 (2001).
3. M.S. Livingstone, and D.H. Hubel, "Psychophysical evidence for separate channels for the perception of form, color, movement and depth," *J. of Neuroscience*, vol. 7(11), pp. 3416-3468 (1987).
4. D. Marr, and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol.194, pp. 283-287 (1976).
5. N. Qian, and Y. Zhu, "Physiological computation of binocular disparity," *Vision Research*, vol. 37, pp. 1811-1827 (1997).
6. S.B. Pollard, T.P. Pridmore, J. Porrill, J.E.W. Mayhew, and J.P. Frisby, "Geometrical Modeling from Multiple Stereo Views," *The International J. of Robotics Research*, vol. 8, pp. 3-32 (1989).
7. W.E. Grimson, "Why stereo vision is not always about 3D reconstruction," *A.I. Memo No.1435*. MIT (1993).
8. I. Biederman, and P. C. Gerhardstein, "Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance," *J. of Experimental Psychology: Human Perception and Performance*, vol. 19, pp. 1162-1182 (1993).
9. H.H. Bulthoff, S.Y. Edelman, and M.J. Tarr, "How are three-dimensional objects represented in the brain?," *Cerebral Cortex*, May-Jun;5(3), pp. 247-260 (1995).
10. A. Saxena, J. Driemeyer, and A.Y. Ng, "Robotic Grasping of Novel Objects using Vision," *The International J. of Robotics Research*, vol. 27, pp. 157-173 (2008).
11. D. Kragic, M. Bjorkman, H.I. Christensen, and J.-O. Eklundh, "Vision for robotic object manipulation in domestic settings," *Robotics and Autonomous Systems*, vol. 52, pp. 85-100 (2005).

12. T.A. Busey, "The perception and memory of stereoscopic depth information in naturalistic objects," *Technical Report #207*, Indiana University Press (1996).
13. Z. Kapoula, M.P. Bucci, F. Lavigne-Tomps, and F. Zamfirescu, "Disconjugate memory-guided saccades to disparate targets: evidence for 3D sensitivity," *Experimental Brain Res.*, vol. 122(4), pp.413-423 (1998).
14. Y. Jiang, I.R. Olson, and M.M. Chun, "Organization of visual short-term memory", *J. of Exp. Psychology: Learning, Memory and Cognition*, vol. 26, pp.683-702 (2000).
15. V.A. Lyakhovetskii, A.S. Potapov, and E.P. Popechitelev, "The Principles of studying and the model of human memory informational structure," *Proceedings of Taganrog Radiotechnical University, Series "Medical Informational Systems"*, №11, pp.4-9 (2006).
16. V.A. Lyakhovetskii, A.S. Potapov, and S.V. Ivanov, "Strategies for storing spatial transformations of chess positions," *Perception Supplement*, vol. 35., p.105 (2006).
17. V.A. Lyakhovetskii, and E.V. Bobrova, "The strategies of coding in spatial memory," *Perception Supplement*, vol. 36, p.51 (2007).
18. B. Kosko, "Bidirectional associative memories," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 18, pp. 49-60 (1988).
19. R. Kothari, R. Lotlikar, and M. Cahay, "State-Dependent Weights For Neural Associative Memories," *Neural Computation*, vol. 10, pp. 59-71 (1998).
20. Z. Kourtzi, and N. Kanwisher, "Representation of perceived object shape by the human lateral occipital complex", *Science*, vol. 293, pp.1506-1509 (2001).