

A Computational Model of Spatial Representations That Explains Object-Centered Neglect in Parietal Patients

Rajesh P.N. Rao* and Dana H. Ballard

Department of Computer Science, University of Rochester,
Rochester, NY 14627-0226, USA
{rao,dana}@cs.rochester.edu

1 INTRODUCTION

Patients with parietal cortex lesions exhibit unusual visual deficits, typically involving the neglect of a part of their visual space. Without being consciously aware of it, they act as if that part of space is not visible. At first, it was thought that this neglected space coincided with the visual hemifield contralateral to the lesioned hemisphere, but more recently, an increasing number of experiments [1, 2, 5, 6, 7, 8, 11] have shown that in many cases, the neglected part of space is related to a reference object of immediate interest. For example, in a recent experiment by Behrmann et al. [3], a patient with a right parietal lobe lesion was asked to count the number of instances of the letter “A” in a field of letters on a TV screen (see Figure 1). The eye movements recorded from the subjects showed that they typically neglected to look at most of the “A”s in the left side of the TV screen. Note that the neglect cannot be explained as a visual hemifield neglect because, *as the patient makes eye movements, the letters that appear in each hemifield change*.

We know that the observed behavior pertains to the object of immediate interest owing to another experiment by Behrmann *et al.* [2]. In this case, the patient gazes at a dumbbell-shaped object consisting of two circles (one red and one blue) joined by a horizontal bar. The patient is asked to press a button when a target (a small white circle) appears within either the left circle or the right. As expected, the response to targets in the contralesional circle (opposite side of the lesion) is typically much slower. Now the patient sees the dumbbell slowly rotate 180 degrees clockwise about its midpoint so that the colored circles have now exchanged positions. When the same test is repeated using the rotated dumbbell, the patient is now much slower with respect the ipsilesional circle (same side as the lesion). This result is explained if the patient assigns an object centered frame to the dumbbell initially and maintains that frame throughout the experiment. The neglect is thus consistently in object-centered coordinates. Other experiments [6, 7] have demonstrated similar results.

While neglect in object-centered coordinates is easy to understand abstractly, it is much more difficult to explain how the brain’s object recognition machinery could be organized to explain such results. The purpose of this paper is to describe a systems-level model of spatial representations in the parietal cortex, and show by simulations that it can explain the above experimental data in a concise manner.

2 REFERENCE FRAMES

A central problem in object recognition is that of determining the *pose* of an object, where pose characterizes the transformation between an object-centered reference frame and the current view frame (retinal reference frame). For humans, the retinal frame is determined by the current fixation point. However, it is easy to demonstrate the usefulness of a third frame. In reading, the position of letters with respect to the retina is unimportant compared to their position in the encompassing word. In driving, the position of the car with respect to the fixation point is unimportant compared to its position with respect to the edge of the road. In both of these examples, the crucial information is contained in the transformation between an object-centered frame and a *scene* frame [9].

Figure 2 shows the relationships between the three reference frames for the image of the letter “A” (the object) depicted on a TV screen (the scene). The transformations between the frames are denoted by T_{os} (object-scene), T_{sr} (scene-retina), and T_{or} (object-retina). The position of remembered objects with respect to a scene, T_{os} , together with the scene-to-retina transformation T_{sr} , determines the current position of objects T_{or} in retinotopic space.

*To whom correspondence should be addressed.

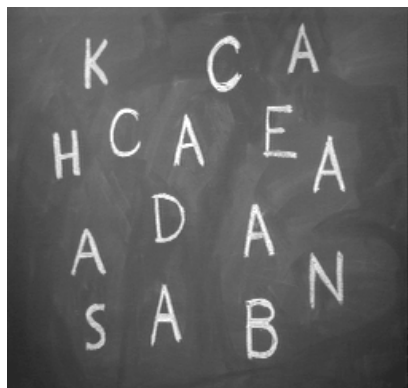


Figure 1: **The Visual Counting Task.** Subjects were asked to count the number of occurrences of the letter “A” on the current display screen.

3 THE MODEL

The reference frames discussed in the previous section provide a basis for constructing a systems-level model of spatial representations in the parietal cortex. The model uses “iconic representations” for recognizing and searching for targets in natural scenes. These iconic object representations are obtained by filtering the scene with a large number of oriented spatial filters at multiple scales [15]. This allows each location in the scene to be characterized by a vector of filter responses that serves as an effective signature of the photometric intensity variations in the region surrounding the given scene location. We refer the reader to [4, 12, 15, 19, 20] for more information regarding recognition methods based on such iconic representations.

In order to search for possible locations containing a target (for example, an “A”) in a given scene, the remembered filter response vector for the target is correlated with those for all scene locations. This results in a retinotopic *saliency map* (see Figure 3; brighter spots indicate higher correlations with the target, in this case, an “A”). Given a saliency map denoting possible target locations in the scene, the number of targets in the scene can be counted by approximately fixating on each candidate target location in succession. Unfortunately, this strategy changes the location of the targets in retinal coordinates. Thus, the central issue is *how to keep track of the already counted locations, which shift in a retinotopic frame after each eye movement*. One possible solution is to inhibit the counted locations in the retinal frame, but this requires elaborate circuitry to keep track of the inhibited locations across eye movements. A more elegant solution is to use a separate representation for the scene that describes the location of the scene with respect to retinal coordinates. In the model, the transform of the scene frame with respect to the retina is activated in a separate area T_{sr} in parietal cortex. The transform T_{sr} is continually updated using posture signals (“efference copies” or “corollary” discharges) as derived from eye, head, and body movements. The use of the scene frame allows the relations of the parts of the scene (for example, letters on a TV screen) to be depicted in a separate object-centered frame denoted by T_{os} , which is assumed to be represented bilaterally in parietal cortex, with each half in a separate hemisphere. In addition, the scene frame is assumed to be task-dependent. Once initialized (at the beginning of the task), it is maintained until the task is deemed completed.

Consider the situation where objects are being counted sequentially by approximately foveating them. After an eye movement is made to an object, its corresponding location in T_{os} is inhibited. Note that these locations do not change with eye movements, thereby avoiding the problem of shifting inhibitory markers. After an object has been foveated and counted, the saliency map is recomputed. The scene frame and the object-centered frame then combine to inhibit previously visited locations in the saliency map, yielding a new location for the next eye movement in retinal coordinates. Note that this location is represented in exactly the same coordinate system as that used by the oculomotor system, which allows a saccade to be executed to the desired target location ([14] suggests a possible method for learning such saccadic eye movements). Object centered neglect can be easily

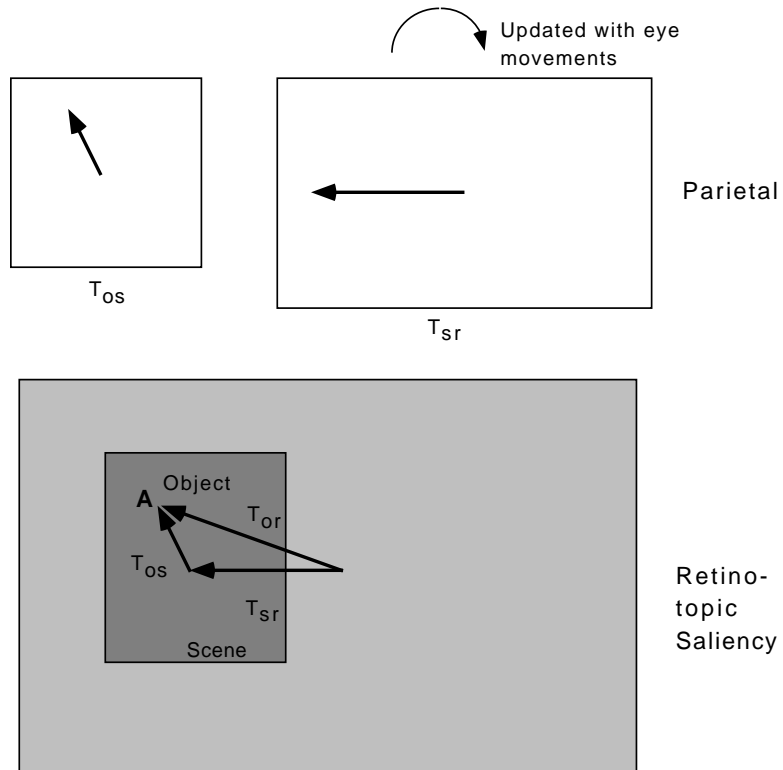


Figure 2: Three Fundamental Spatial Transformations. To represent the geometric relations of visual features, three transformations are fundamental. The first (T_{sr}) describes how a particular depiction of the world, or scene, is related to the retinal coordinate system with respect to the current fixation. The second (T_{os}) describes how objects can be related to the scene. The third (T_{or}), which is the composition of the other two, describes how objects are located with respect to the retina.

explained as unilateral damage to T_{os} . This damage does not allow parts of the scene on the damaged side to be represented. Thus, regardless of eye position, they can never be accessed.

4 EXPERIMENTAL RESULTS

The model was tested using the visual search task of Behrmann *et al.* [3] involving the counting of the number of occurrences of the letter “A” on a display. Figure 3 (left) shows the case without damage, where the search is successful for the display shown in Figure 1. The images depict the alternating sequence of saliency maps (in retinal coordinates) that are used to initiate the next saccade, followed by the resulting retinal image after the saccade (the fixation point is always the center of the image and the fovea is denoted by a circle). The correlation peaks in the saliency map (bright regions in (a)) depict possible “A” locations. An eye movement to the highest peak allows the corresponding “A” to be foveated and counted as shown in (b). The foveated location is then inhibited in object-centered coordinates. This inhibition shows up in retinal coordinates in the saliency map in (c) by virtue of the fact that the scene frame is continuously translating T_{os} locations to T_{or} locations. A subsequent eye movement is shown in (d), and the final state of the model at the end of the counting process is shown in (e).

Figure 3 (right) shows the model with right parietal lobe damage. The damage is assumed to lie in the area of parietal cortex responsible for the T_{os} transformation. Since T_{os} is required for computing the current saliency map, its damage prevents the appearance of any task-relevant saliency peaks in the contralesional side of the object-centered frame. As a result, any targets (such as the “A” locations) on the contralesional side of the current scene (the TV screen) fail to be noticed and are not attended to or foveated during the course of the task. Note that the “damage” (dark region) shows up in different parts of the retinal frame as a result of eye movements and the updating of the scene frame.

5 DISCUSSION

Observations from related experiments such as Behrmann *et al.*’s rotating dumbbell task [2] can be succinctly explained in the context of the present model by allowing the scene transformation to be sufficiently general so as to allow for rotation. In this case, the explanation for the neglect is the same. The targets are represented in T_{os} , the movement of the display is interpreted by the subject as a change in T_{sr} , and the neglect follows by the same mechanism.

Some object-centered neglect effects can be obtained by using only retinal frames, but such models cannot explain complicated effects such as those elicited by the dumbbell display. Previous models of unilateral neglect [10, 13] have either relied on highly abstract interpretations of experimental observations or have concentrated on deriving relatively low-level implementations that can explain a given set of experimental observations. The model presented herein attempts to bridge this gap by suggesting a systems-level mechanism for explaining a wide variety of neglect-related phenomena while simultaneously retaining the possibility of a neural implementation. Ongoing work involves integration of the present model with the Kalman filter-based neural model of the cortex [17, 18]. Preliminary results in this direction have been encouraging [16].

References

- [1] M. Behrmann and M. Moscovitch. Object-centered neglect in patients with unilateral neglect: Effects of left-right coordinates of objects. *Journal of Cognitive Neuroscience*, 6(1):1–16, 1994.
- [2] M. Behrmann and S.P. Tipper. Object-based attentional mechanisms: Evidence from patients with unilateral neglect. In C. Umiltà and M. Moscovitch, editors, *Attention and Performance XV: Conscious and Nonconscious Information Processing*. MIT Press, Cambridge, MA, 1994.

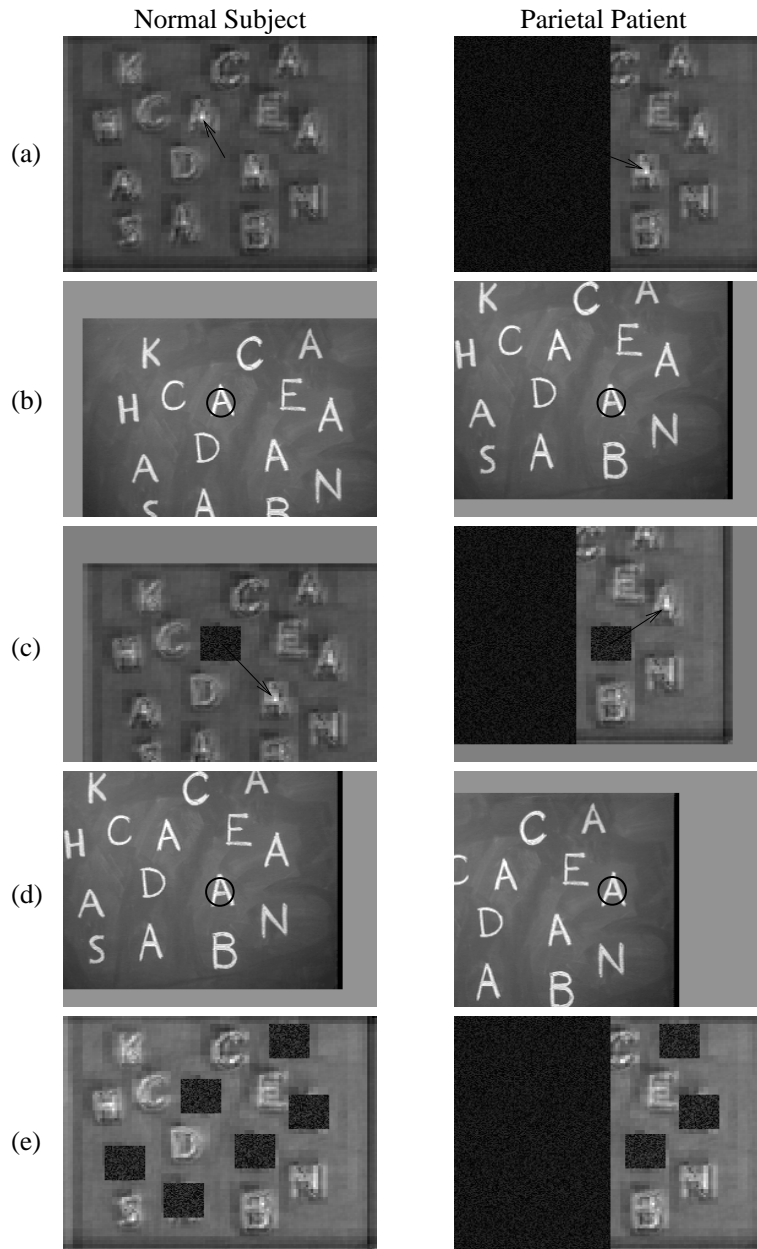


Figure 3: Simulation Results. The left side of the figure shows the model without a lesion (“normal subject”) counting the number of occurrences of the letter “A” in the display shown in Figure 1. The right side shows the model with a right hemispheric lesion (“parietal patient”). The panels in (a), (c), and (e) depict the retinotopic saliency map of Figure 2. (a) Correlation peaks in retinal coordinates. (b) The first “A” foveated with an eye movement and counted. (c) Inhibition of the location of the counted “A”. (d) The second “A” is foveated and counted. The process repeats until the correlation peaks fall below a preset threshold for detecting the presence of an “A”. (e) The final state of the model. Note that the model without a lesion (“normal subject”) is able to count all 6 occurrences of the letter “A” while the lesioned model (“parietal patient”) counts only 3.

- [3] M. Behrmann, S. Watt, S.E. Black, and J.J.S. Barton. Impaired visual search in patients with unilateral neglect: An oculographic analysis. Submitted, 1996.
- [4] J.M. Buhmann, M. Lades, and C.v.d. Malsburg. Size and distortion invariant object recognition by hierarchical graph matching. In *Proc. IEEE IJCNN, San Diego (Vol. II)*, pages 411–416, 1990.
- [5] R. Calvanio, P. N. Petrone, and D. Levine. Left visual spatial neglect is both environment-centered and body-centered. *Neurology*, 37:1179–1183, 1987.
- [6] A. Caramazza and A.E. Hillis. Spatial representation of words in the brain implied by studies of a unilateral neglect patient. *Nature*, 346:267–269, 1990.
- [7] M. J. Farah, J. L. Brunn, A. B. Wong, M. Wallace, and P. Carpenter. Frames of reference for the allocation of spatial attention: Evidence from the neglect syndrome. *Neuropsychologia*, 28:335–347, 1990.
- [8] M. Gazzaniga and E. Ladavas. Disturbances in spatial attention following lesion or disconnection of the right parietal lobe. In M. Jeannerod, editor, *Neurophysiological and Neuropsychological Aspects of Spatial Neglect*. North-Holland, Amsterdam, 1987.
- [9] G.E. Hinton. A parallel computation that assigns canonical object-based frames of reference. In *7th International Joint Conference on Artificial Intelligence*, pages 683–685, 1981.
- [10] J. Beng-Hee Ho, M. Behrmann, and D.C. Plaut. The interaction of spatial reference frames and hierarchical object representations: A computational investigation of drawing in hemispatial neglect. In *Proc. 17th Annual Conf. of the Cognitive Science Society*, pages 148–153, 1995.
- [11] E. Ladavas. Is hemispatial deficit produced by right parietal damage associated with retinal or gravitational coordinates. *Brain*, 110:167–180, 1987.
- [12] B. Mel. A neurally-inspired approach to 3-D visual object recognition. Presentation at Telluride Workshop on Neuromorphic Engineering, Telluride, Colorado, July 1994.
- [13] A. Pouget and T.J. Sejnowski. A model of spatial representations in parietal cortex explains hemineglect. In D. Touretzky, M. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*, pages 10–16. Cambridge, MA: MIT Press, 1996.
- [14] R.P.N. Rao and D.H. Ballard. Learning saccadic eye movements using multiscale spatial filters. In G. Tesauro, D.S. Touretzky, and T.K. Leen, editors, *Advances in Neural Information Processing Systems 7*, pages 893–900. Cambridge, MA: MIT Press, 1995.
- [15] R.P.N. Rao and D.H. Ballard. An active vision architecture based on iconic representations. *Artificial Intelligence (Special Issue on Vision)*, 78:461–505, 1995.
- [16] R.P.N. Rao and D.H. Ballard. A class of stochastic models for invariant recognition, motion, and stereo. Technical Report 96.1, National Resource Laboratory for the Study of Brain and Behavior, Department of Computer Science, University of Rochester, June 1996.
- [17] R.P.N. Rao and D.H. Ballard. Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation* (in press). Also, Technical Report 96.2, National Resource Laboratory for the Study of Brain and Behavior, Department of Computer Science, University of Rochester, 1996.
- [18] R.P.N. Rao and D.H. Ballard. Cortico-cortical dynamics and learning during visual recognition: A computational model. In J. Bower, editor, *Computation Neuroscience 1996*. New York, NY: Plenum Press, 1997.

- [19] R.P.N. Rao, G.J. Zelinsky, M.M. Hayhoe, and D.H. Ballard. Modeling saccadic targeting in visual search. In D. Touretzky, M. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems* 8, pages 830–836. Cambridge, MA: MIT Press, 1996.
- [20] P. Viola. Feature-based recognition of objects. In *AAAI Fall Symposium on Learning and Computer Vision*, 1993.