massachusetts institute of technology — artificial intelligence laboratory

Generalization Over Contrast and Mirror Reversal, But not Figure-ground Reversal, in an "Edge-based" Model of IT Neurons

Maximilian Riesenhuber

AI Memo 2001-034 December 2001
CBCL Memo 211

© 2001 massachusetts institute of technology, cambridge, ma 02139 usa — www.ai.mit.edu

Abstract

Baylis & Driver [1] have recently presented data on the response of neurons in macaque inferotemporal cortex (IT) to various stimulus transformations. They report that neurons can generalize over contrast and mirror reversal, but not over figure-ground reversal. This finding is taken to demonstrate that "the selectivity of IT neurons is not determined simply by the distinctive contours in a display, contrary to simple edge-based models of shape recognition", citing our recently presented model of object recognition in cortex [3]. In this memo, I show that the main effects of the experiment can be obtained by performing the appropriate simulations in our simple feedforward model. This suggests for IT cell tuning that the possible contributions of explicit edge assignment processes postulated in [1] might be smaller than expected.

Copyright © Massachusetts Institute of Technology, 2001

This report describes research done within the Center for Biological & Computational Learning in the Department of Brain & Cognitive Sciences and in the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology.

This research was sponsored by grants from: Office of Naval Research (DARPA) under contract No. N00014-00-1-0907, National Science Foundation (ITR) under contract No. IIS-0085836, National Science Foundation (KDI) under contract No. DMS-9872936, and National Science Foundation under contract No. IIS-9800032.

Additional support was provided by: AT&T, Central Research Institute of Electric Power Industry, Center for e-Business (MIT), Eastman Kodak Company, DaimlerChrysler AG, Compaq, Honda R&D Co., Ltd., ITRI, Komatsu Ltd., Merrill-Lynch, Mitsubishi Corporation, NEC Fund, Nippon Telegraph & Telephone, Oxygen, Siemens Corporate Research, Inc., Sumitomo Metal Industries, Toyota Motor Corporation, WatchVision Co., Ltd., and The Whitaker Foundation. M.R. is supported by a McDonnell-Pew Award in Cognitive Neuroscience.

Baylis and Driver [1] have recently presented data on how responses of IT neurons are affected by three different stimulus transformations. They report that neurons can generalize over contrast and mirror reversal, but not over figure-ground reversal. This finding is taken to demonstrate that "the selectivity of IT neurons is not determined simply by the distinctive contours in a display, contrary to simple edge-based models of shape recognition [2, 3]."

We have examined this claim by performing the appropriate simulations in our "edge-based" model of object recognition in cortex the authors refer to (which is actually no more "edge-based" than Hubel & Wiesel's simple cells are), using the same stimuli as in [1]. In particular, the model was as described in [3] with the addition that complex cell responses were made invariant to contrast reversal, to better capture the complex cell behavior reported in the literature [4] (see caption of Fig. 2).

The stimuli used, modified from [1], are shown in Fig. 1a. Fig. 1b shows one example shape (the upper left from Fig. 1) undergoing the three stimulus transformations (compare to Fig. 1 of [1]). We next defined four model IT units, each tuned to a different one of the four shapes of Fig. 1a. Figure 2 shows the response of one model unit (tuned to shape 3, in the lower left of Fig. 1a) to the 32 stimuli generated by performing the eight combinations of transformations for each of the four stimuli as shown in Fig. 1b. As in the case of the IT neuron shown in Fig. 3 of [1], comparing the appropriate histograms shows that the model unit's responses are similar across both contrast reversal and mirror-image transforms but not across figure-ground reversal, as can also be seen from the correlation plots in Fig. 3 (compare to Fig. 2 of [1]).

Average correlation values over all four model units were 1 for contrast reversal, 0.73 for mirror reversal (in agreement with our earlier report [3] that some model neurons show tolerance to mirror inversion), and 0.36 for figure-ground reversal. Thus, based on the responses of model units, the contrast-reversed version of a shape would be considered more similar than its mirror-reversed version. Most relevant, the figure-ground reversed version would be considered to be least similar to the original shape, all in agreement with the human similarity judgments and IT data reported in [1].

Interestingly, the low similarity of a shape and its figure-ground reversed transform was obtained without any special edge-assignment mechanisms, postulated in [1] to be necessary to distinguish a shape from its figure-ground-reversed transform. Rather, shape representations in the model are based on local features only, which are also affected by a figure-ground reversal (which leaves only the curved black/white center region constant) This is not to say that more elaborate edge-assignment computations [1] might not be used in cortex — after all, the IT neurons in [1] show a mean correlation coefficient for figure-ground reversal around zero, lower than the model neurons. Moreover, recent data on the neural correlates of border ownership in V2 [5] showing that visual information on stimulus borders far outside a neuron's traditional receptive field can modulate its response suggests a role for recurrent projections, at least for neurons in lower visual areas. It will be interesting to see if similar effects as found in [5] also hold for the much larger receptive fields of IT cells, which would provide more unequivocal support for the contribution of additional processes in the shaping of IT responses than the present experimental data. In any case, it is remarkable how a simple feed-forward model originally proposed to explain the invariances of IT neurons to scaling and translation captures the main effects caused by the very different manipulations of the Baylis & Driver study, suggesting that computational approaches may help to explain cognitive phenomena in terms of well-known neural mechanisms.

References

[1] Baylis, G. and Driver, J. (2001). Shape-coding in it cells generalizes over contrast and mirror reversal, but not figure-ground reversal. *Nat. Neurosci.* **4**, 937–942.

[2] Pinker, S. (1984). Visual cognition: an introduction. *Cognition* **18**, 1–64.

[3] Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**, 1019–1025.

[4] Spitzer, H. and Hochstein, S. (1985). Simple- and complex-cell response dependences on stimulation parameters. *J. Neurophys.* **53**, 1244–1265.

[5] Zhou, H., Friedman, H., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci.* **20**, 6594–6611.

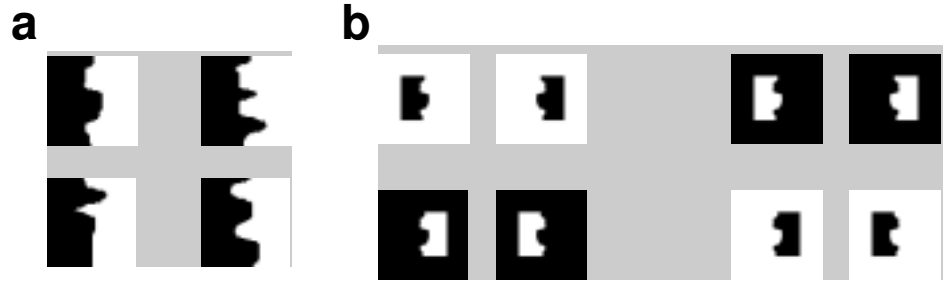

Figure 1: Shapes and transformations. **(a)** The four shapes used. They were scanned in from [1], thresholded, and rescaled to 64×64 pixels. **(b)** Shape transformations using shape 1 (upper left in (a)) as an example. The upper leftmost image shows the original 64×64 pixel stimulus embedded in a 128×128 frame. The right 2×2 block shows the contrast-reversed versions of the left block. In each 2×2 block, the right column is the mirror-reversed version of the left column. The bottom row is the figure-ground reversed version of the top row. Compare to Fig. 1 in [1].

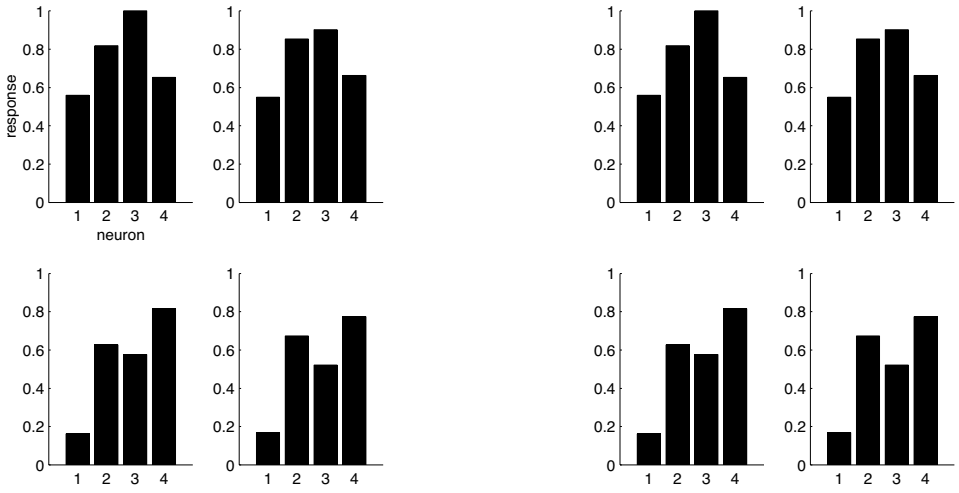

Figure 2: Responses of one model unit (unit 3, tuned to the stimulus in the lower left of Fig. 1a, presented as shown in the top left panel of Fig. 1b)) to the 4×8 stimuli. The layout is the same as in Fig. 1b, *i.e.*, each subplot shows the response (y-axis) of the neuron to the four shapes (x-axis) transformed like the corresponding panel in Fig. 1b. For instance, the lower left plot shows the unit's response to the figure-ground-reversed versions of the four shapes of Fig. 1a. The model was as described previously [3], except for that invariance of complex cells (and thus model IT units) to contrast-reversal was achieved by linearly rescaling the average intensity of image patches falling into a simple cell's receptive field to zero before calculating simple cell responses. Thus the response of a simple cell s_i to an image patch I in its receptive field was $s_i = |f_i \cdot I'|$ with $I' = (I - < I >)/||I - < I >||$, f_i the simple cell's optimal stimulus, $< \cdot >$ and $|| \cdot ||$ denoting the mean and Euclidean norm (to normalize simple cell responses to a range between 0 and 1), resp. Apart from an increased robustness to contrast-reversal, view-tuned units in this modified model show response properties similar to those reported previously [3] (not shown). View-tuned units, as described in [3], were each connected to the 50 C2 afferents most strongly activated by its preferred stimulus, with a tuning width of $\sigma = 0.4$. Results were robust with respect to the choice of σ.

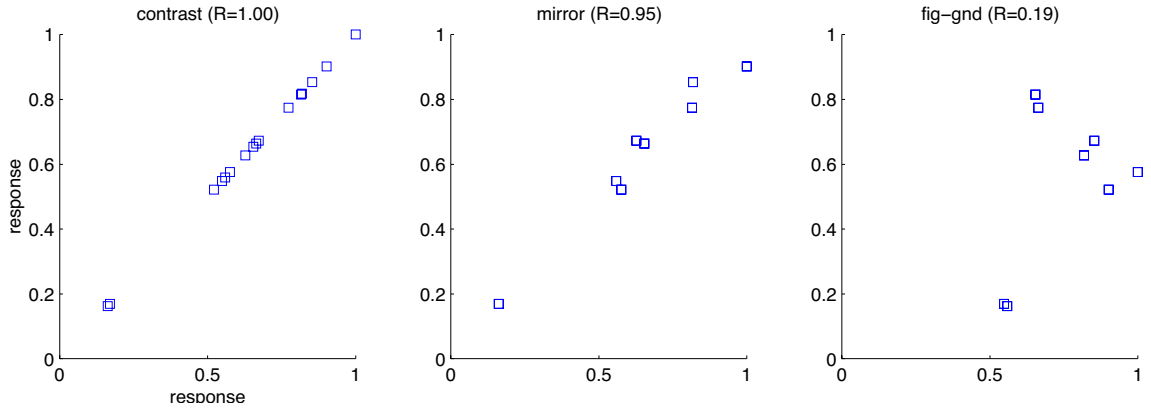

Figure 3: Correlation plot for the responses shown in Fig. 2. The plots show unit activation for particular stimuli on the x-axis and the transformed versions of the same stimuli along the y-axis. Left plot: Contrast reversal transform. Center plot: Mirror reversal. Right plot: Figure-ground reversal. Numbers above the plots show value of Spearman rank order correlation coefficient for the different transformations. While the correlation for both contrast and mirror reversal is significant, $p < 0.0001$, it is not significant for figure-ground reversal, $p = 0.24$.