Natural scene statistics of depth and motion pertinent to figure-ground segregation

Xin Huang[1], Chentao Wang[1], Bryce Arseneau[1], Thomas Yerxa[2], and Emily A. Cooper[3]

[1]Dept. of Neuroscience, School of Medicine & Public Health, University of Wisconsin - Madison, Madison, WI ; [2]Center for Neural Science, New York University, New York, New York ; [3]School of Optometry & Helen Wills Neuroscience Institute, University of California, Berkeley

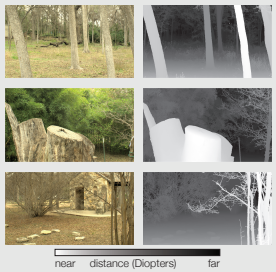
Introduction

- Segregating a figure object from the background, referred to as figure-ground segregation, is a fundamental function of vision.
- Perceptually, several important factors have been identified—notably motion & depth.
- We have been investigating how neurons in the middle-temporal (MT) cortex of macaque monkeys, an area important for motion and depth processing, represent multiple visual stimuli [1,2]. Our working hypothesis is that the properties of MT neurons reflect a strategy to exploit statistical regularities during natural vision to help achieve figure-ground segregation.
- To test this hypothesis, it is important to understand the properties of figure and ground during natural vision within spatial regions comparable to the receptive fields (RFs) of cortical neurons.
- In a natural scene statistics analysis, we asked whether figure and ground regions differ systematically in their depth (binocular disparity), and characterized the impact of observer self-motion and object motion on the relative retinal speeds of figure and ground.

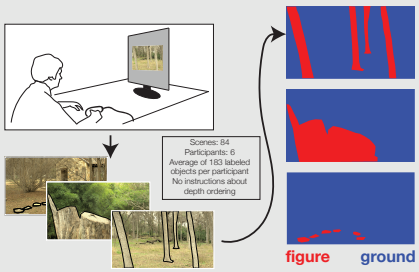
Methods

Data Generation

Database of natural images + laser range measurements [3]

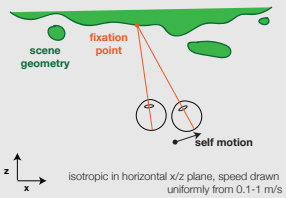


Humans view images and label "salient objects" to obtain figure/ground boundaries in each image

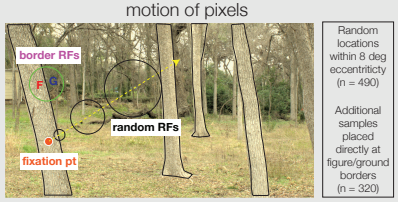


Viewer Simulation

Binocular observer fixating a scene point while in motion

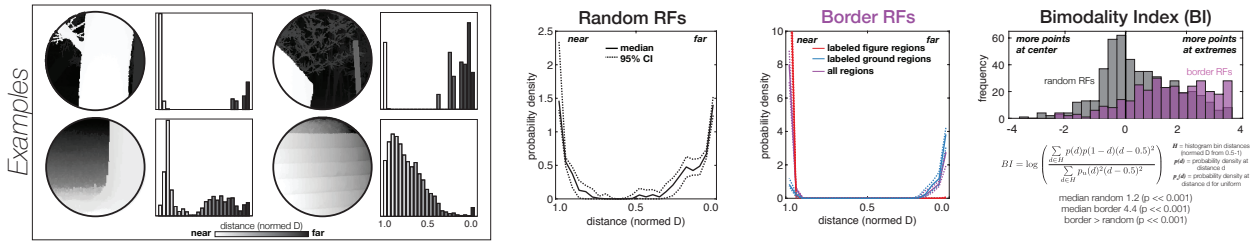


Place RFs at random retinotopic locations; calculate distance, binocular disparity, and retinal motion of pixels

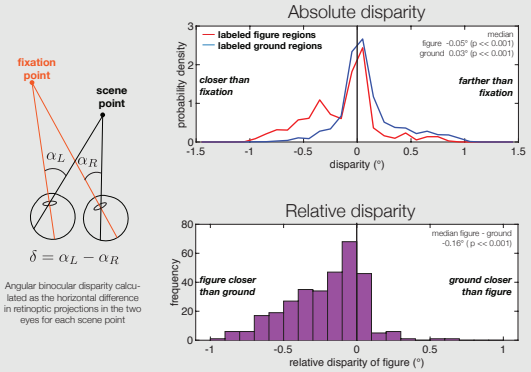


Results

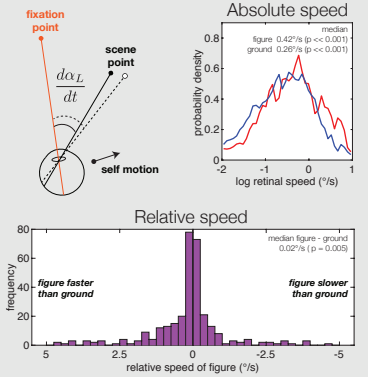
Average distance distribution in MT RFs is bimodal: multiple surfaces are often present in RFs during natural experience



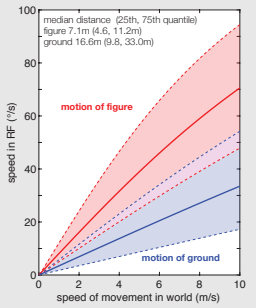
Surfaces labeled as *figure* tend to be closer, and therefore have more near absolute and relative disparity



A systematic, but small, speed difference between figure and ground caused by *self-motion*



Because figure regions tend to be nearer, *world motion* in these regions is associated with higher speed



Conclusions

Our results suggest that stereoscopic depth and motion in natural scenes have statistical regularities at the scale of MT receptive fields. Such regularities may be exploited by cortical neurons to help segregate figure from ground. In future work, we plan to:

- use different types of natural scenes and video clips to measure natural scene statistics of depth and object motion;
- simulate images in the RFs using real eye movements recorded from observers when they view natural scenes and perform a figure-ground segregation task;
- evaluate the statistics of retinal motion of figure and ground when combing self-motion with object motion in natural scenes.

Acknowledgment
This study is supported by NIH Grant R01EY022443 (Huang)

References

[1] Xiao & Huang, SFN 2017
[2] Wiesner, Baumgart & Huang, BioRxiv 2019
[3] Burge, McCann, & Geisler, J. Vision 2016