

Learning independent causes in natural images explains the spacevariant oblique effect

Constantin A. Rothkopf, Thomas H. Weisswange, and Jochen Triesch

Frankfurt Institute for Advanced Studies, Frankfurt, Germany.

Email: {rothkopf, weisswange, triesch}@fias.uni-frankfurt.de

Abstract—The efficient coding hypothesis posits that sensory processing increases independence between neural responses to natural stimuli by removing their statistical redundancy reflective of the structure present in the natural environment. While there is consensus on the role of the statistical structure of the physical environment in shaping the natural input to the sensory system, it is not well understood how the sensory apparatus itself and its active use during behavior determine the statistics of the input. To explore this issue, a virtual human agent is simulated navigating through a wooded environment under full control of its gaze allocation during walking. Independent causes for the images obtained during navigation are learned with algorithms that have been shown to extract computationally useful representations similar to those encountered in the primary visual cortex of the mammalian brain. The distributions of properties of the learned simple cell like units are in good agreement with a wealth of data on the visual system including the oblique effect, the meridional effect, properties of neurons in the macaque visual cortex, and functional Magnetic Resonance Imaging (fMRI) data on orientation selectivity in humans and monkeys. Finally, this analysis sheds new light on the discussion on orientation anisotropies based on carpented environments. Thus, when learning computational representations it is not sufficient to consider only the regularities of the environment but also the regularities imposed by the sensory apparatus and its use during behavior need to be taken into account.

I. INTRODUCTION

It has been proposed that a fundamental principle for the understanding of neuronal computations involving sensory data is that they have been shaped on evolutionary and developmental timescales by the regularities in the environment [1], [2]. Theories of the encoding of natural stimuli based on their statistical regularities have been successful at reproducing a number of properties of neurons in visual cortical areas in humans and animals [3], [4]. However, the dependence of these statistics on the sensory apparatus and its active use have been considered to a much lesser degree. This is surprising, given that the active character of vision is well established [5], [6], [7], [8]. Moreover, there is ample empirical data showing that this active selection process is much less involuntary and reactive as suggested by some experimental paradigms, if vision is studied in its ecologically valid context of extended, goal-directed, visuomotor behavior [9], [10], [11], [12].

Statistical models have been proposed in the past, which aim at reducing the statistical redundancy of the brain's representation of natural images by considering their higher order dependencies. Such models also can be viewed as generative

models, in which latent variables cause images according to probabilistic formulations of the image generation process. Applying such models to natural image ensembles has been shown to learn receptive field like units as hidden causes [13]. But such learning has usually assumed stationary statistics across the visual field and implicitly also independence from the active selection process due to eye movements as well as the imaging geometry. Such models therefore cannot explain that visual performance on a wide variety of tasks depends on the position within the visual field. As an example, perceptual studies in humans have d reduced discrimination ability for obliquely oriented patterns as compared to horizontal and vertical ones in the central area of the visual field [14]. This so called oblique effect varies with the position within the visual field in that with increasing eccentricity sensitivity to meridionally oriented stimuli increases [15].

A wealth of data on these psychophysical effects has been collected not only in humans, but across species [14]. The attenuation in performance for obliquely oriented stimuli compared to horizontal and vertical ones close to the fovea has its correspondence in a non-uniform distribution of preferred orientations in simple and complex cells in the primary visual cortices of several species including ferrets [16], cats [17], and monkeys [18]. These studies have also reported subtle differences in orientation selectivity across species and even across humans who have been raised in carpented versus non-carpented environments [19]. Furthermore, fMRI studies in monkeys and humans have shown a correlation between the blood-oxygen-level-dependent (BOLD) activation in V1 during orientation discrimination tasks and the performance on these tasks across different orientations [20].

While some aspects of the oblique effect can be understood in terms of the second order statistics of natural images, as quantified by their power spectrum [21], [22], the present study investigates the role of higher order dependencies and task behavior on properties of receptive fields across the visual field. The dependencies that go beyond second order are important, as removing the dependencies up to second order leaves the spatial structure such as edges intact and images with only correlational, i.e. second order structure such as that of natural images, do not contain features such as edges or contours. Indeed, receptive field like units that are spatially localized, oriented, and band-pass in different spatial frequency bands have been extracted from natural images by maximizing higher-order moments. Thus, receptive

field properties are strongly determined by the higher order dependencies present in the natural input to the visual system.

According to the generative model view, the visual cortex is involved in extracting statistical regularities present in its inputs. The extracted latent variables correspond to the causes of the natural images, encountered in the environment. Given a visual stimulus, the cortex inverts the model according to Bayes' theorem with the goal of inferring the most likely causes. Several algorithms have been proposed for learning basis sets of causes for natural images under the assumption of linear superposition. While the linear superposition assumption may not hold as a generative model for all processes leading to small image patches, it has been shown that by assuming a sparse distribution over causes, many qualities of cells present in the mammalian cortex can be extracted from natural image sets.

Here independent causes are learned separately at different positions across the visual field using the sparse coding algorithm [13]. To obtain natural image sequences while having full control over the gaze allocation during task execution, we decided to use image sequences obtained by simulating a virtual agent navigating through rendered wooded environments. The image statistics in such images agree well with those found in natural images. The advantage of using this methodology is that it allows for the complete control not only of the environment but also of the movements of the virtual agent during navigation and its active allocation of gaze.

The properties of the learned receptive fields are quantified through the parameters of best fitting Gabor functions. This analysis shows that the receptive fields have a preference for horizontal and vertical orientations at the center and increasingly meridional directions in the periphery. Furthermore, it is shown that the distribution of preferred orientations significantly depends on where gaze is directed within the visual field during walking. This analysis is therefore also relevant for the discussion on differences in orientation selectivity in humans raised in carpented versus non-carpented environments, because it is that when considering the regularities in the natural input to the visual system it is not sufficient to consider the statistical regularities in natural images without regard to the active use of the sensory apparatus.

II. METHODS

A. Image sequence generation

To generate the training dataset for the receptive field learning, the Unreal Tournament Engine 2.5 from Epic Games was used. For the forest scenery a structured surface with different densities of trees and bushes on a natural textured ground was build with the provided editor. A virtual human agent moved through this environment and image sequences from its perspective were collected. Each image of size 640x480 pixels corresponded to the momentary field of view of 120°.

To simulate two different gaze allocation strategies, sequences of images were collected in which the agent pointed its image sensor in the direction of navigation or tilted the head so as to point towards the ground plane at a distance of approximately 5m. Figure 1 shows typical images from both image sets demonstrating both gaze allocations.

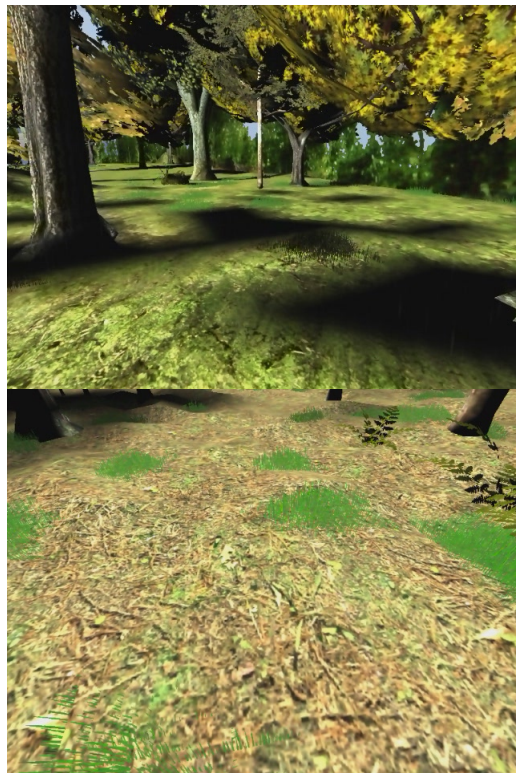


Fig. 1. Two typical views from the perspective of the simulated agent navigating through the wooded environment. Top: Gaze is aligned with the heading direction. Bottom: Gaze is directed to a point on the ground in front of the walking agent.

B. Learning algorithm

The generative model used in the present study is the classical sparse coding model first proposed by Olshausen and Field [13]. The latent variables x_i are linearly combined to result in a local image patch \mathbf{y} . The causes \mathbf{x} are modeled as being independent with a sparse prior distribution $P(x)$ and the image noise model is assumed to be Gaussian. For a Cauchy distribution as sparse prior one obtains:

$$P(\mathbf{x}) = \prod_{i=1}^N P(x_i) \quad (1)$$

$$P(x_i) = \text{Cauchy}(0, \gamma) \quad (2)$$

$$P(\mathbf{y}|\mathbf{x}) = \text{Normal}(G\mathbf{x}, \sigma^2\mathbf{I}) \quad (3)$$

While previous learning of such basis sets has assumed that natural images are stationary across the entire image, the present study learns separate basis sets i.e. the latent variables \mathbf{x} separately across the visual field for the two gaze allocations. Receptive fields are learned on these sets of image patches of size 8x8 pixels separately sampled from nine regions of size 64x64 pixels across the visual field. These regions are marked with colored squares in figure 2. In terms of the field of view used to render the natural scenes, the eccentricities considered here are at 0° and 55° horizontally and 0° and 40° vertically. Finally, we preprocessed the original image patches obtained from the rendered scenes with a whitening filter to equalize the variance in all directions of image space. Filtering was done in the frequency domain as described in [13].

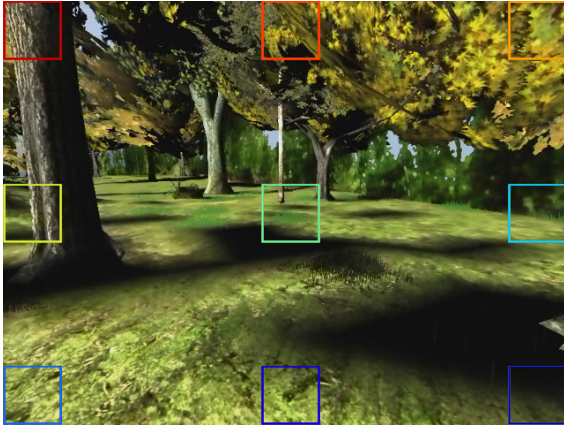


Fig. 2. Typical view of the simulated agent while navigating through the wooded environment. The nine regions from which images were sampled for the learning of independent causes are marked with different colors.

C. Characterization of learned receptive fields

To better quantify the properties of the learned units and to compare the results of the learning algorithm to experimental data obtained in primary visual cortex V1 of humans and animals, all generative fields were fit with Gabor functions. These functions are the product of a sinusoid with a Gaussian envelope. These localized basis filters are well suited to represent the spatial profile of many simple cells in V1 and are widely used in the literature.

Once the Gabor functions have been obtained, we calculate histograms of the properties of the learned basis functions such as their preferred orientation and spatial frequency distributions for the nine sets learned at different positions in the visual field. These histograms can be related directly to the distributions of such properties in populations of simple cells recorded in V1. The emphasis in the present study is on orientation, for which ample data has been collected neurophysiologically.

D. Relating receptive fields to performance measures

While it would be desirable to relate the distribution of properties of the learned generative fields directly to the psychophysical performance measures reported in the literature, there are several issues that make this difficult. First, the sparse coding algorithm learns the generative fields, while the neurophysiological data reports the responses of neurons in terms of receptive fields. For the overcomplete bases learned, inverting the generative model is highly nonlinear. It is therefore necessary to characterize the preferred stimuli of each unit using techniques such as reverse correlation.

Secondly, while it is possible to obtain orientation tuning curves for individual units from the parameters of the best fitting Gabor function analytically, this is not sufficient in determining psychophysical performance such as a discrimination threshold in an orientation discrimination task. The reason is, that some form of decision stage has to evaluate the responses and this introduces a large number of additional parameters, that are difficult to set in a principled way. It is

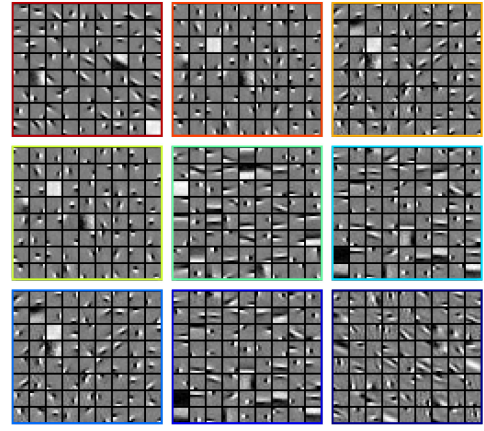


Fig. 3. Generative fields at different positions in the visual field as learned by sparse coding. The bases consisted of 72 units of size 8x8 each. The colors correspond to the regions marked in figure 2.

especially difficult to come up with well motivated internal noise models that could account for human task performance.

Furthermore, the learning of generative fields presented here does not take into account that acuity falls off with increasing eccentricity throughout the visual field. This is reflected by the fact that the number of generative fields is constant across the nine regions over which the learning was carried out. It is similarly difficult to change the number of generative fields in a principled manner to reflect the decrease in acuity.

An ideal observer analysis with the scope of estimating the lower bound on Fisher information through a linear discriminant is beyond the scope of this paper but it is nevertheless possible to get some well funded insights in the psychophysical performance by comparing the relative distributions of preferred orientations at different locations across the visual field. Fisher information J provides a measure of encoding accuracy because its inverse is the Cramer-Rao lower bound on the mean squared error of all possible unbiased estimators that can read out the variable in question from population activity without systematic error. Under the assumption of uncorrelated noise between neurons, Fisher information reduces to:

$$J(\theta) = \sum_i^N \frac{f'_i(\theta)^2}{\sigma_i^2} \quad (4)$$

where f_i is the tuning curve for an individual neuron i , θ is orientation, and σ_i is the variance of the noise in that unit. From this equation it can be seen, that the steeper the tuning curve of a unit, the more it contributes to the information gain on orientation, and the more units have steep portions of their tuning curve aligned with a particular orientation, the more the contribution to the total information.

III. RESULTS

Visual inspection of the learned basis sets depicted in Figure 3 already shows clear biases in the orientations of the units at different locations in the visual field. As an example, consider the two sets learned on the left side of the visual field, specifically the set at the bottom in the blue colored box and the one at the top in the red colored box. It can

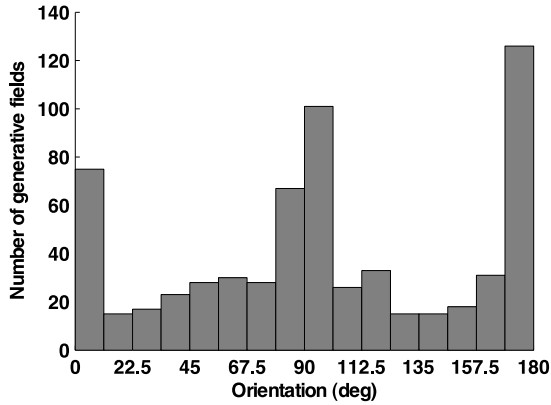


Fig. 4. Histogram of preferred orientations of the learned units across the entire visual field. The histogram was obtained by counting all preferred orientations as given by the best fitting Gabor function for all units and all positions.

be seen that both sets contain more units tuned to radial orientations compared to tangential orientations. Furthermore, when comparing these sets to the one obtained at the center of the field of view as marked by the light green box, one can discern the clear bias of horizontally and vertically tuned units at the center.

A. Distribution of preferred orientations

The distribution of preferred orientations obtained in the learning experiments presented here can directly be compared to the empirically measured distributions that were obtained in macaque monkeys. Figure 4 shows the histogram of preferred orientations as determined by the parameters of the best fitting Gabor function to all learned generative fields across the entire visual field, i.e. at all nine locations marked in Figure 2. This histogram shows that horizontal and vertical orientations are indeed more prevalent than oblique orientations.

This result agrees qualitatively with the aforementioned neurophysiological studies in macaque monkeys and cats that have shown that horizontally and vertically aligned simple cells are more common than simple cells at oblique orientations. It should nevertheless be noted, that the exact quantitative measures of these distributions are not only different across species, as is to be expected, but there is also considerable variability within species. Further investigations could clarify experimental biases in the selection of recorded units and also specify more details about the environment in which the animals were raised and are behaving.

Closer comparison of the histogram depicted in Figure 4 with the reported orientation selectivity distributions in neuronal populations suggests that the difference in the number of neurons coding for horizontal and vertical versus oblique orientations is stronger in the present simulations. Further work should clarify, which parameters in the present simulations are best at reproducing the empirical distributions most faithfully. Given the present results, this could include both differences in the visual environment as well as the gaze allocation strategy.

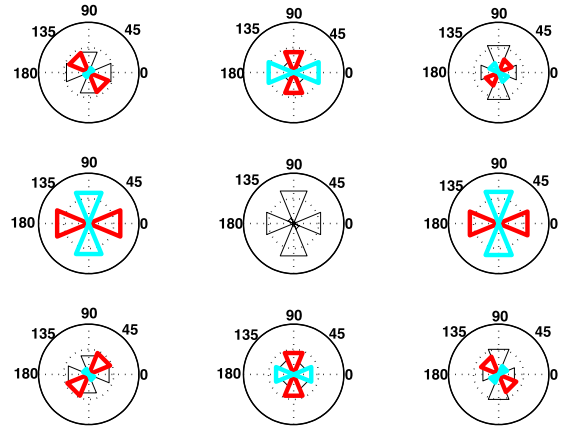


Fig. 5. Histogram of orientation selectivity for the nine separate positions in the visual field. The radial directions have been highlighted in red.

B. Explaining the oblique effect

The non-uniform distribution of preferred orientations at the fovea has its psychophysical correspondence in the better performance at horizontal and vertical orientations compared to oblique orientations in a large variety of tasks. These include orientation discrimination, line segment length estimation, and several acuity tasks. This attenuation of performance at oblique orientations at the fovea is called the oblique effect [23] and has been described in the human adult and child, and throughout the animal kingdom.

The non-uniform distribution of preferred orientations can be regarded as a good explanation of this effect. This connection has been drawn repeatedly in the literature. Indeed, recurring to equation 4 one can see that the consequence of an asymmetry in distribution of preferred orientations is accompanied by an asymmetry in Fisher information. Given that the majority of learned units are tightly tuned, they convey information for orientations very near to their preferred orientation. The bias for horizontal and vertical orientations is therefore the best candidate for explaining the oblique effect.

C. Explaining the meridional effect

While the oblique effect has been originally described through tasks primarily executed at the fovea, further studies have investigated how performance in tasks that require estimating or distinguishing orientations is distributed across the entire visual field. The psychophysical results reported by Rovamo et al. [15] showed a more complex variation of resolution limit for grating orientations at different meridians of the visual field.

The main characteristic of these results is that performance is still mostly better for horizontal and vertical orientations, but that additionally performance increases for orientations aligned with the meridian, i.e. performance for vertical gratings is better than for other orientations when presented along the vertical meridian, whereas performance is best for horizontal gratings when presented on the horizontal meridian. Furthermore, performance increases for oblique oriented gratings that are aligned along the radial orientation, while

performance decreases for tangentially oriented stimuli. This phenomenon has been called the meridional effect.

Figure 5 shows histograms of preferred orientations separately for the nine regions in the visual field at which learning was carried out. These histograms demonstrate the two separate phenomena, that are in good agreement with the psychophysical results described as oblique effect and meridional effect. First, the distributions of preferred orientations show a bias for horizontal and vertical orientations. This is in agreement with the previous result and has its correspondence in the oblique effect. The nine histograms furthermore show that the radial directions, which are colored in red in Figure 5, become more prevalent compared to tangential directions with increasing eccentricity. This is again in accordance with the meridional effect, when invoking arguments on Fisher information.

D. Influence of gaze allocation

When repeating the learning experiments with images obtained from the gaze allocation strategy that points the view vector on the ground plane in front of the walking agent, the distribution of preferred orientations differ significantly compared to the first strategy. Figure 6 shows the nine histograms corresponding to the nine locations across the visual field marked in Figure 2.

First, the histogram of the foveal distribution of preferred orientations at the center of Figure 6 shows a nearly uniform distribution. This is not what is observed neurophysiologically. Furthermore, while the results at the fovea for the agent pointing its gaze in the heading direction agree with the oblique effect, this distribution of preferred orientations can not be used to explain the oblique effect.

Secondly, while previously units were increasingly tuned to radial orientations compared to tangential ones with increasing eccentricity, this effect is lost in the lateral areas of the upper half of the visual field for the simulated agent pointing its gaze to the ground. Here more units are tuned to horizontal directions and tangential directions are slightly overrepresented compared to radial directions. Furthermore, the distributions obtained at the two lateral sides of the central regions, i.e. corresponding to the yellow and cyan boxes in Figure 2, neither agree with the oblique effect nor with the meridional effect.

IV. DISCUSSION

The efficient coding hypothesis suggests that sensory representations try to reduce the redundancies present in the sensory input. A successful way of formulating such redundancy reduction is through the use of generative models, which learn to extract latent variables that can be thought to cause the sensory input. The sparse coding algorithm has been successful in extracting representations that are similar to those encountered in the brain of a large number of species.

Here we learned independent causes using the sparse coding algorithm separately at different positions in the visual field for two different gaze allocation strategies. To have full control over the stimulus in terms of the natural layout of

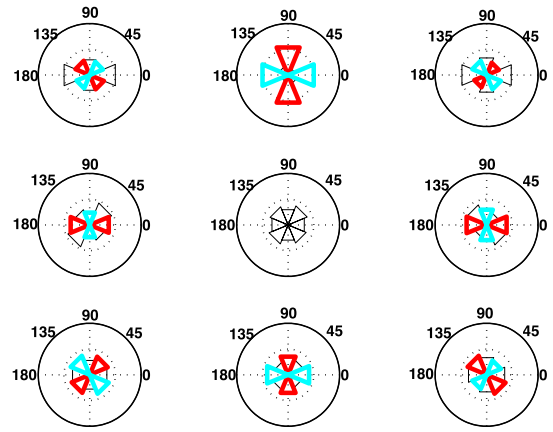


Fig. 6. Histogram of orientation selectivity for the nine separate positions in the visual field. The radial directions have been highlighted in red.

the environment and the gaze selection strategy, the images used for the learning were obtained by simulating a virtual agent in a computer rendered environment.

First, the results demonstrate that a large variety of studies reporting foveal anisotropies of orientation tuning in the visual cortex can be reproduced qualitatively by assuming an agent that aligns gaze with the heading direction. The results of learning independent causes for the natural images showed a strong bias for horizontal and vertical orientations across the entire visual field. This bias was strongest at the center of the visual field.

Furthermore, the space-variant anisotropies of orientation selectivity reported in the literature are also observed in the learned basis sets. While a bias for horizontally and vertically aligned orientations is maintained throughout the visual field, the radial directions tend to become more prevalent with increasing eccentricity, while the tangential orientations become more rare.

Both these observations can be brought in relation with the corresponding psychophysical phenomena, namely the oblique effect and the meridional effect. The observed biases in the distributions of preferred orientations agree qualitatively with the better discrimination performance in humans and monkeys for horizontal and vertical visual stimuli in the fovea. Similarly, the results also agree well with the observed increase in performance for radially oriented stimuli versus tangentially oriented ones in a wide variety of tasks.

These phenomena can be explained as a combined effect of the orientation statistics in the natural environment and its three dimensional layout, the statistics due to geometric perspective inherent in the optics of the visual system, and the statistics of gaze selection. Horizontal and vertical line segments are more often encountered in the wooded environment, and the geometric perspective results in an increase of radially oriented line segments with increasing eccentricity. Because of the difference between the upper and lower half of the visual field due to the ground plane, there may be slight differences in the distributions of radially oriented line segments. More importantly, when gaze is centered at the horizon, both the upper and lower field of view contain more radially oriented

line segments due to the perspective projection, whereas when gaze is directed towards the ground plane, the upper field of view is dominated by tangentially oriented line segments.

The present study also has interesting consequences for the carpentered environment hypothesis. It has been reported, that certain indigenous people have a less pronounced attenuation of discrimination for obliquely oriented stimuli than people from western cultures. It has been proposed to understand this phenomenon solely by the statistics of the physical world in which the subjects had been raised and in which they were immersed on a daily basis. The present study suggests that it is also important to consider what tasks the respective subjects do engage in, as the statistical properties of the input to the visual system cannot be considered without the active use of the visual system during natural tasks.

We conclude that the properties of model receptive fields not only depend on the statistics of visual scenes in the environment, but also on the statistics imposed on the stimulus by the imaging geometry and the statistics of the interaction with the environment during natural task execution. Taking all these determinants together, receptive fields can be learned that explain the space variant oblique effect.

ACKNOWLEDGMENT

The authors wish to thank Jörg Lücke for providing the code for matching Gabor functions to the learned receptive fields. This research was supported by EC MEXT-project PLICON and the German Federal Ministry of Education and Research within the "Bernstein Focus: Neurotechnology" research grant.

REFERENCES

- [1] F. Attneave, "Some informational aspects of visual perception," *Psychological Review*, vol. 61, pp. 183–193, 1954.
- [2] H. Barlow, *Sensory Communication*. Cambridge, MA, USA: MIT Press, 1961, ch. Possible principles underlying the transformation of sensory messages, pp. 217–234.
- [3] J. Daugman, "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields," *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 1, pp. 107–114, 1989.
- [4] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, pp. 1193–1216, 2001.
- [5] D. Ballard, "Animate vision," *Artificial Intelligence Journal*, vol. 48, pp. 57–86, 1991.
- [6] D. H. Ballard, M. M. Hayhoe, P. Pook, and R. Rao, "Deictic codes for the embodiment of cognition," *Behavioral and Brain Sciences*, vol. 20, pp. 723–767, 1997.
- [7] J. M. Findlay and I. D. Gilchrist, *Active Vision: The Psychology of Looking and Seeing*. Oxford, UK: Oxford University Press, 2003.
- [8] J. O'Regan and A. Noe, "A sensorimotor approach to vision and visual consciousness," *Behavioral and Brain Sciences*, vol. 24, pp. 939–973, 2001.
- [9] M. F. Land and D. N. Lee, "Where we look when we steer," *Nature*, vol. 369, pp. 742–744, 1994.
- [10] D. H. Ballard, M. M. Hayhoe, and J. Pelz, "Memory representations in natural tasks," *Journal of Cognitive Neuroscience*, vol. 7, no. 1, pp. 68–82, 1995.
- [11] M. Land and M. Hayhoe, "In what ways do eye movements contribute to everyday activities?" *Vision Research, special issue on Eye Movements and vision in the Natural World*, vol. 41, pp. 3559–3566, 2001.
- [12] M. Hayhoe, A. Shrivastava, R. Mruczek, and J. Pelz, "Visual memory and motor planning in a natural task," *Journal of Vision*, vol. 3, pp. 49–63, 2003.
- [13] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision Research*, vol. 37, pp. 3311–3325, 1997.
- [14] S. Appelle, "Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals," *Psychological Bulletin*, vol. 78, no. 4, pp. 266–278, 1972.
- [15] J. Rovamo, V. Virsu, P. Laurinen, and L. Hyvarinen, "Resolution of gratings oriented along and across meridians in peripheral vision," *Invest. Ophthalmol. Vis. Sci.*, vol. 23, no. 5, pp. 666–670, 1982.
- [16] D. M. Coppel, L. E. White, D. Fitzpatrick, and D. Purves, "Unequal representation of cardinal and oblique contours in ferret visual cortex," *Proceedings of the National Academy of Sciences of the U S A*, vol. 95, no. 5, pp. 2621–2623, 1998.
- [17] A. G. Leventhal and H. V. Hirsch, "Effects of early experience upon orientation sensitivity and binocularity of neurons in visual cortex of cats," *Proceedings of the National Academy of Sciences of the U S A*, vol. 74, no. 3, pp. 1272–1276, 1977.
- [18] J. A. J. Bauer, D. A. Owens, J. Thomas, and R. Held, "Monkeys show an oblique effect," *Perception*, vol. 8, no. 3, pp. 247 – 253, 1979.
- [19] R. C. Annis and B. Frost, "Human visual ecology and orientation anisotropies in acuity," *Science*, vol. 182, no. 4113, pp. 729–731, 1973.
- [20] Y. Sasaki, R. Rajimehr, B. W. Kim, L. B. Ekstrom, W. Vanduffel, and R. B. Tootell, "The radial bias: A different slant on visual orientation sensitivity in human and nonhuman primates," *Neuron*, vol. 51, pp. 661–670, 2006.
- [21] M. J. Wainwright, "Visual adaptation as optimal information transmission," *Vision Research*, vol. 39, no. 23, pp. 3960 – 3974, 1999.
- [22] N. D. Bruce and J. K. Tsotsos, "A statistical basis for visual field anisotropies," *Neurocomputing*, vol. 69, no. 10–12, pp. 1301 – 1304, 2006, computational Neuroscience: Trends in Research 2006.
- [23] M. M. Taylor, "Visual discrimination and orientation," *Journal of the Optical Society of America*, vol. 53, pp. 763–5, 1963.