Vision Research 120 (2016) 1–4

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Editorial

Understanding the statistics of the natural environment and their implications for vision

The job of the visual system is to extract useful information about the 3D world from the 2D image projections formed by our eyes. How do the statistics of visual scenes—encompassing all relevant aspects of surface geometry, material properties and illumination—affect the architecture and function of the visual system and its performance on specified tasks?

Natural scenes are complex and varied, and since von Helmholtz (1867) at least, it has been understood that vision requires inference. The problems are ill-posed—multiple different scenes might account for any given set of image observations—although only one of these accounts is correct. To perform well in natural environments, the visual system must therefore somehow embody an understanding of the likelihood of these various accounts.

While Helmholtz expressed these ideas qualitatively, with the development of information theory in the mid-twentieth century, a view of visual perception as a fundamentally statistical task began to emerge (Attneave, 1954; Barlow, 1972; Brunswik & Kamiya, 1953): optimal visual coding and inference demands knowledge of the statistics of real-world environments and how projection onto the image affects these statistics. The vast cortical and subcortical resources devoted to vision attests to its adaptive value and predicts that neural coding will have evolved and developed to embody these statistics, in order to perform well in the environments we commonly encounter.

The importance of these statistics to understanding visual perception is by now generally agreed. But the problem of identifying these statistics and understanding how exactly they have shaped the visual system is daunting. The challenge springs from two fundamental difficulties. The first is the Curse of Dimensionality: the number of degrees of freedom needed to describe the possible covariation of visual scenes and images over space and time is so enormous as to defeat standard approaches for statistical estimation and inference. Hope lies in the evidence that the effective dimensionality of real scenes and images is much lower than the full dimensionality of the embedding space (Chandler & Field, 2007), underlining the importance of efficient coding concepts and dimensionality reduction to understanding visual perception. But this leads to the second fundamental difficulty: this lower-dimensional manifold appears to be highly curved, hence defeating standard linear systems approaches which dominated much of vision science in the 20th century.

Given these massive challenges, the road to understanding the statistical foundations of visual perception is clearly long and winding, and the journey has just begun. Nevertheless, the ten papers in this special issue collectively represent an important step forward in understanding the statistics of natural scenes and how they have shaped visual coding, and in some cases at least, point to potential opportunities to overcome some of these challenges, a topic to which we will return at the end of this editorial.

1. Point statistics

Papers in the first half of this special issue address point statistics of the visual image, characterizing the local statistics of luminance and color contrasts. By marginalizing over (almost) all space and time dimensions, the analysis is restricted to one or two dimensions, and thus becomes tractable.

Galilei (1632) and von Helmholtz (1867) had both observed that spatial resolution is higher for dark stimuli than for light (Kremkow et al., 2014)—why is this? It has long been known that the distributions of luminance and contrast in natural scenes exhibits strong positive skewness (Laughlin, 1983; Richards, 1982). One consequence is that in natural images, negative contrasts are more numerous and carry more information than light contrasts. Given finite neural bandwidth, the principle of efficient coding (Barlow, 1961; Laughlin, 1981) predicts that this asymmetry would be matched by the brain, and recent evidence (Ratliff, Borghuis, Kao, Sterling, & Balasubramanian, 2010) suggests that it is: OFF retinal ganglion cells are more numerous and smaller than ON retinal ganglion cells.

Two papers in this issue show that these asymmetries are relevant to higher-level aspects of visual function as well. Sato et al. (2016) consider perceived blur: the tacit assumption has been that perceived blur is determined by the spatial frequency content of the image, treating darks and lights equally. However, the authors show that this is not the case, and that, along with the visual system's higher sensitivity and spatial resolution for darks, it gives darks higher weight in assessing blur.

While these results represent a concordance of visual tuning with natural scene statistics consistent with the efficient coding hypothesis, the results of Graham et al. (2016) do not. Graham et al. consider subjective preference for naturalistic images as a function of the skewness of the luminance distribution. Perhaps surprisingly, they find that observers prefer images when modified to have zero skewness (i.e., symmetric luminance distributions) rather than the positive skewness typically observed in natural scenes. What this seems to suggest is that aesthetic preference does not necessarily follow the principles of efficient coding.

Three papers address the point statistics of color. Provenzi et al. (2016) identify a simplifying feature of second-order statistics in the spatio-chromatic domain: they show that spatial and

http://dx.doi.org/10.1016/j.visres.2016.01.003

chromatic covariance matrices of natural images commute with each other. This formalizes a sense in which spatial and chromatic statistics are separable, enabling a simplified description of the second-order spatio-chromatic statistics of natural images.

The other papers concerning color attempt to understand ecological aspects of the spatiochromatic variation of natural images. Color features in the image are co-determined by surface reflectance and illumination. It is generally agreed that the former is of greater interest to the observer, and so the latter appears largely as a nuisance variable. How then can the observer factor out the effects of varying illumination to obtain invariant estimates of surface color?

One possible assumption is that the illuminant can be approximated as roughly uniform within a scene, so that color variations in the image can mainly be attributed to variations in surface color. Examining the hyperspectral statistics of natural scenes over space and time, Nascimento et al. (2016) find that in fact the instantaneous variation in illumination color within a scene typically varies roughly as much as illumination varies over time or latitude. These findings argue against the use of overly simplistic models of illumination in natural scenes.

Foster et al. (2016) explore the consequences of these illumination variations on color measurements made in the image. A potential way for the visual system to attain a degree of invariance with respect to these illumination variations is to rely not upon absolute but upon relative color measurements. The latter are computed as ratios of cone responses measured at different points in the image, which are invariant to purely temporal variations of the global illumination. The authors show that in fact, due to the complex geometry of natural scenes, the local illuminant and hence these ratios vary considerably. However, placing either spatial or temporal limits on these ratios can reduce variation below detectability, suggesting that cone ratios may serve as useful pseudo-invariants for more localized spatiotemporal neighborhoods within a scene.

2. Oriented filter statistics

These first five studies focused on first- and second-order point statistics of the luminance and color distributions of natural images. However, important features in the scene, such as the boundaries of objects, result in locally oriented structure in the image, and since Hubel & Wiesel's discovery that most neurons in primate area V1 are tuned for orientation (Hubel & Wiesel, 1959, 1968), the local processing of oriented information in the image has been seen as a primary goal of early visual cortex.

Hubel & Wiesel sometimes referred to cells in V1 as edge- and line-detectors, which suggests that their receptive fields should not only be oriented, but should have specific selectivity for the detection of edges and lines versus other forms of oriented structure. In tension with this proposal was the spatial frequency account of early visual coding (Campbell & Robson, 1968), which envisioned these neurons as uniformly tiling a two-dimensional Fourier representation of the image. In such a representation, tuning for edges and lines should manifest as a bias toward particular bandwidths and phases (odd and even). While early physiological studies in cat (e.g., Field & Tolhurst, 1986) did not confirm this prediction, later studies in primate (e.g., Ringach, 2002) did.

MaBouDi et al. (2016) address whether the statistics of natural images may account for these physiological findings. In particular, using a standard quadrature filter model of an oriented V1 complex cell, they analyze the phase distribution of local image patches as a function of the frequency tuning, bandwidth and aspect ratio parameters of these filters. While for most parameter values they find a relatively uniform distribution of phases, for a specific range of filters they find that phase response is bimodal, consistent with tuning for specific local features such as edges and lines. They speculate that the visual system makes use of these inhomogeneities to code visual inputs efficiently: phase-independent complex cells for configurations that correspond to uniform local phase distributions, and phase-tuned complex cells whose configurations correspond to bimodal phase distributions.

The nonlinearity involved in the quadrature filter model of complex cells is just one example of the many nonlinear properties of early visual cortical neurons that complicate statistical analysis. Golden et al. (2016) seek to relate these nonlinearities to principles of sparse coding (Bell & Sejnowski, 1997; Olshausen & Field, 1996). Employing a representation of neural tuning in curved coordinate frames introduced by Zetzsche and colleagues (Zetzsche, Krieger, and Wegmann (1999), Zetzsche and Nuding (2005) and Zetzsche and Rohrbein (2001)), they show that the sparse coding network of Olshausen and Field produces a form of curvature that accounts for nonlinearities that give rise to key selectivity and invariance properties of oriented V1 neurons. Since the objective function used to learn these sparse networks does not explicitly pursue this curvature, these findings provide insight into the connections between sparsity, response nonlinearity and curved representational manifolds for neural coding.

The prominence of orientation tuning in early visual cortex is often attributed to the importance of identifying the boundaries of objects in the scene. However these boundaries are signaled not just by gradients in luminance and color but also in motion and binocular disparity. How important is each of these modalities to the detection of an object boundary? Mély et al. (2016) address this question by training classifiers that use individual modalities and ensembles. While motion and binocular disparity are often cited as most diagnostic for the discrimination of object boundaries, the authors find that in fact color and luminance play dominant roles. However, they also find that more accurate boundary extraction can be achieved by combining all cues, and, importantly, motion and stereoscopic disparity play a much larger role when combined with the other cues than would be expected from their performance in isolation. This suggests a tighter coupling of these modalities than might be suggested by a strict modular view of visual cortex (Livingstone & Hubel, 1988).

3. Disparity statistics

The detection of object boundaries is only one expression of the general role stereoscopic disparity plays in revealing the 3D structure of visible surfaces in the scene. Hibbard et al. (2016) consider the statistical cues in the natural visual environment that are available to compute disparity. While early "first-order" models relied upon luminance correlations between corresponding image patches in the two eyes, psychophysical (Hess & Wilcox, 1994) and physiological (Tanaka & Ohzawa, 2006) work has revealed mechanisms that allow stereoscopic disparity to be computed from stimuli with no first-order inter-ocular correlations. Such mechanisms must rely upon nonlinear processing (e.g., rectification or squaring) to reveal "second-order" disparity cues. Hibbard et al. show that such cues are indeed present in natural images and, while correlated with first-order cues, can improve the accuracy of disparity estimation. These findings suggest that stereoscopic mechanisms have evolved to take advantage of the diverse cues to stereoscopic matching available in complex natural scenes.

Despite this richness of cues, recovering an accurate global surface estimate from stereo is one of the classic ill-posed vision problems; there are many ways to interpolate a surface through noisy local disparity measurements, particularly when the measurements are sparse. To perform well, the visual system must therefore incorporate a statistical prior over these possible surfaces. The question of how these scene priors can be learned

neurally is therefore of great importance. Recent work (Samonds, Potetz, & Lee, 2009; Samonds, Potetz, Tyler, & Lee, 2013) shows the potential of Markov random fields (MRF) as models of neural disparity processing. Using a Boltzmann machine model (Hinton & Sejnowski, 1986), Zhang et al. (2016) show that such MRF models can be tuned to natural scene disparity statistics in unsupervised fashion, leading to network properties that mimic certain aspects of the biology, including positive connectivity between neurons tuned to similar disparities at nearby locations.

4. Outlook

Despite the progress demonstrated in these papers and other recent studies, one cannot escape the sense that our understanding of how the visual system is shaped by the statistics of our environment is still in its early stages, and much remains to be done. This is not unexpected, given the challenges presented by the Curse of Dimensionality and the nonlinear structure of natural scenes and inference. What could be the way forward?

Should we look to the computer vision literature for insight? Here we find that generic, deep, feed-forward nonlinear network models that date from the Neocognitron of the early 1980s (Fukushima, 1980) have finally come of age (LeCun, Bengio, & Hinton, 2015), defining the state of the art for a broad range of computer vision datasets. These deep feed-forward models work well because we now have the powerful computers and large datasets required to train their many parameters using variations on classical back-propagation methods. Through this training process, these networks implicitly learn the statistics of (selected) natural image sets, as well as some of the invariance and selectivity properties relevant to the specific tasks they are assigned. Unfortunately, the mathematical and physical principles underlying their performance are left scattered over thousands of parameters, leaving scientists little insight into these principles.

These deep hierarchical networks learn by feedback but inference is strictly feedforward. Although visual cortex is sometimes described as hierarchical, feedback connections appear to outweigh feedforward connections (Van Essen, Anderson, & Felleman, 1992). For example, the "input" neurons in the primary visual cortex of the primate receive most of their inputs from cortical, rather than thalamic sources (Peters, Payne, & Budd, 1994). These feedback connections are not just involved in learning but play an important role in visual inference (Lamme & Roelfsema, 2000).

There are also profound differences between the job of the human visual system and tasks that these computer vision systems are assigned. Biological systems are inherently general-purpose in nature, called on to solve a diversity of problems (scene layout, object recognition, navigation. . .), often concurrently. This encourages the development of efficient, generative, general-purpose models that serve multiple functions. Modern computer vision algorithms, on the other hand, target narrowly defined problems associated with specialized datasets, leading to specific models that can limit representations to capture only the most relevant features. It is thus not clear how relevant these specialized deep networks are to biological vision systems, although some similarities have been demonstrated (DiCarlo et al., 2014).

If big data and deep artificial neural networks do not provide a pathway to complete understanding, where shall we look? Some clues are offered by the papers in this special issue. Simple theoretical principles: the geometry of objects, the physics of image formation, efficient and sparse coding, these are all ideas used by papers in this issue that can be further developed and understood to yield greater insights while keeping the dimensionality of our models low. We believe these principles will increasingly be coupled to the principles of perceptual organization first outlined by Gestalt psychologists (Koffka, 1935). Despite the passing of time, these principles have not only persisted, but have had profound influence on modern computational theory and statistical accounts of perception (e.g., Elder & Goldberg, 2002; Geisler, Perry, Super, & Gallogly, 2001; Martin, Fowlkes, & Malik, 2004; Parent & Zucker, 1989). We believe a continuing confluence of these simple but powerful ideas will ultimately lead to a more complete understanding of the statistics of the natural environment and their implications for vision.

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*(3), 183–193.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*. Cambridge, MA: MIT Press.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception, 1*(4), 371–394.

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research, 37*(23), 3327–3338.

Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of 'proximity' and of other Gestalt factors. *American Journal of Psychology, LXVI*, 20–32.

Campbell, F. W., & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551–566.

Chandler, D. M., & Field, D. J. (2007). Estimates of the information content and dimensionality of natural scenes from proximity distributions. *Journal of the Optical Society of America A, 24*(4), 922–941.

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision, 2*(4), 324–353.

Field, D. J., & Tolhurst, D. J. (1986). The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 228*, 379–400.

Foster, D. H., Amano, K., & Nascimento, S. M. C. (2016). Time-lapse ratios of cone excitations in natural scenes. *Vision Research, 120*, 45–60.

Fukushima, K. (1980). Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36*(4), 193–202.

Galilei, G. (1632). *Dialogo Sopra i Due Massimi Sistemi del Mondo (Battista Landini, Florence, Italy)*. .

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurence in natural images predicts contour grouping performance. *Vision Research, 41*(6), 711–724.

Golden, J. R., Vilankar, K. P., Wu, M. C. K., & Field, D. J. (2016). Conjectures regarding the nonlinear geometry of visual neurons. *Vision Research, 120*, 74–92.

Graham, D., Schwarz, B., Chatterjee, A., & Leder, H. (2016). Preference for luminance histogram regularities in natural scenes. *Vision Research, 120*, 11–21.

Hess, R. F., & Wilcox, L. M. (1994). Linear and non-linear filtering in stereopsis. *Vision Research, 34*, 2431–2438.

Hibbard, P. B., Goutcher, R., & Hunter, D. W. (2016). Encoding and estimation of first- and second-order binocular disparity in natural images. *Vision Research, 120*, 108–120.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cogniton. Volume 1: Foundations* (Vol. 1). Cambridge, MA: MIT Press.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology, 148*, 574–591.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*(1), 215–243.

Koffka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt, Brace & World.

Kremkow, J., Jin, J., Komban, S. J., Wang, Y., Lashgari, R., Li, X., et al (2014). Neuronal nonlinearity explains greater visual spatial resolution for darks than lights. *Proceedings of the National Academy of Sciences of the United States of America, 111*(8), 3170–3175. http://dx.doi.org/10.1073/pnas.1310442111.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571–579.

Laughlin, S. B. (1981). A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung, 36*, 910–912.

Laughlin, S. B. (1983). Matching coding to scenes to enhance efficiency. In O. J. Braddick & A. C. Sleigh (Eds.), *Physical and biological processing of images*. Berlin: Springer.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. http://dx.doi.org/10.1038/nature14539.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240*(4853), 740–749.

MaBouDi, H., Shimazaki, H., Amari, S.-i., & Soltanian-Zadeh, H. (2016). Representation of higher-order statistical structures in natural scenes via spatial phase distributions. *Vision Research, 120*, 61–73.

Martin, D., Fowlkes, C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26*(5), 530–549.

Mély, D. A., Kim, J., McGill, M., Guo, Y., & Serre, T. (2016). A systematic comparison between visual cues for boundary detection. *Vision Research, 120*, 93–107.
Nascimento, S. M. C., Amano, K., & Foster, D. H. (2016). Spatial distributions of local illumination color in natural scenes. *Vision Research, 120*, 39–44.
Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*(6583), 607–609.
Parent, P., & Zucker, S. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11* (1989), 823–839.
Peters, A., Payne, B. R., & Budd, J. (1994). A numerical analysis of the geniculocortical input to striate cortex in the monkey. *Cerebral Cortex, 4*(3), 215–229.
Provenzi, E., Delon, J., Gousseau, Y., & Mazin, B. (2016). On the second order spatiochromatic structure of natural images. *Vision Research, 120*, 22–38.
Ratliff, C. P., Borghuis, B. G., Kao, Y. H., Sterling, P., & Balasubramanian, V. (2010). Retina is structured to process an excess of darkness in natural scenes. *Proceedings of the National Academy of Sciences of the United States of America, 107*(40), 17368–17373. http://dx.doi.org/10.1073/pnas.1005846107.
Richards, W. A. (1982). Lightness scale from image intensity distributions. *Applied Optics, 21*(14), 2569–2582.
Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology, 88*(1), 455–463.
Samonds, J. M., Potetz, B. R., & Lee, T. S. (2009). Cooperative and competitive interactions facilitate stereo computations in macaque primary visual cortex. *Journal of Neuroscience, 29*(50), 15780–15795. http://dx.doi.org/10.1523/JNEUROSCI.2305-09.2009.
Samonds, J. M., Potetz, B. R., Tyler, C. W., & Lee, T. S. (2013). Recurrent connectivity can account for the dynamics of disparity processing in V1. *Journal of Neuroscience, 33*(7), 2934–2946. http://dx.doi.org/10.1523/JNEUROSCI.2952-12.2013.
Sato, H., Motoyoshi, I., & Sato, T. (2016). On–Off asymmetry in the perception of blur. *Vision Research, 120*, 5–10.
Tanaka, H., & Ohzawa, I. (2006). Neural basis for stereopsis from second-order contrast cues. *Journal of Neuroscience, 26*, 4370–4382.
Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science, 255* (5043), 419–423.
von Helmholtz, H. (1867). Handbuch der physiologischen optik. In K. G. Leipzig (Ed.). *Allgemeine Encyklopadie der Physik* (Vol. 2). Germany: Voss.
Zetzsche, C., Krieger, G., & Wegmann, B. (1999). The atoms of vision: Cartesian or polar? *Journal of the Optical Society of America A, 16*, 1554–1565.
Zetzsche, C., & Nuding, U. (2005). Nonlinear and higher-order approaches to the encoding of natural scenes. *Network, 16*(2–3), 191–221.
Zetzsche, C., & Rohrbein, F. (2001). Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network, 12*(3), 331–350.
Zhang, Y., Li, X., Samonds, J. M., & Lee, T. S. (2016). Relating functional connectivity in V1 neural circuits and 3D natural scenes using Boltzmann machines. *Vision Research, 120*, 121–131.

James H. Elder
Department of Electrical Engineering & Computer Science, Department of Psychology, Centre for Vision Research, York University, 4700 Keele Street Toronto, Ontario M3J 1P3, Canada
E-mail address: jelder@yorku.ca

Jonathan Victor
Feil Family Brain and Mind Research Institute, Weill Cornell Medical College, 1300 York Avenue, New York, NY 10065, USA
E-mail address: jdvicto@med.cornell.edu

Steven W. Zucker
Depts. of Computer Science and Biomedical Engineering, Yale University, 51 Prospect St., New Haven, CT 06520-8285, USA
E-mail address: steven.zucker@yale.edu

Available online 8 February 2016