JOSA COMMUNICATIONS

Communications are short papers. Appropriate material for this section includes reports of incidental research results, comments on papers previously published, and short descriptions of theoretical and experimental techniques. Communications are handled much the same as regular papers. Proofs are provided.

Images, statistics, and textures: implications of triple correlation uniqueness for texture statistics and the Julesz conjecture: comment

Jonathan D. Victor

Department of Neurology and Neuroscience, Cornell University Medical College, 1300 York Avenue, New York, New York 10021

Received June 4, 1993; revised manuscript received December 8, 1993; accepted December 8, 1993

The mathematical framework for studies of texture perception is discussed. Textures correspond to statistical ensembles, whereas images are spatially finite samples of a texture. The main ideas underlying the use of visual textures in experimental and theoretical analyses of preattentive vision are summarized, with an emphasis on the distinction between texture ensembles and images. The Julesz conjecture [Perception **2**, 391 (1973)] is that preattentive discrimination of textures is possible only for textures that have different second-order correlation statistics. Recently Yellot [J. Opt. Soc. Am. A **10**, 777 (1993)] claimed that the triple correlation uniqueness (TCU) theorem, a mathematical result that every monochromatic image of finite size is uniquely determined (up to translation) by its third-order statistics, makes higher-order variants of the Julesz conjecture trivial. However, the TCU theorem applies to individual images, and not to texture ensembles, and thus is of limited relevance to the study of texture perception.

Key words: isodipole textures, texture perception.

In a recent paper Yellott[1] presents some mathematical results concerning correlation statistics and discusses the relevance of these results for the Julesz[2] conjecture and related matters. The Julesz conjecture states that preattentive discrimination of textures is possible only for textures that have different second-order correlation statistics. Many counterexamples to this conjecture have subsequently been discovered (principally by Julesz and co-workers), but the Julesz conjecture retains its value as a starting point for theories of texture perception.

Yellott makes two claims in his paper: (1) that the triple correlation uniqueness (TCU) theorem, a mathematical result that every monochromatic image of finite size is uniquely determined (up to translation) by its third-order statistics, makes higher-order variants of the Julesz conjecture trivial and (2) that previous counterexamples to the Julesz conjecture are flawed by statistical irregularities that are overcome by a new construction, presented by Yellott. This Communication disputes both claims.

1. IMAGE STATISTICS, IMAGES, AND TEXTURES

The TCU result discussed by Yellott applies to finite images. Clearly, if it were not possible to make distinct textures with identical third-order statistics, then higher-order variants of the Julesz conjecture would indeed be trivial: discrimination of textures with identical third- (or higher-) order textures could not be tested, because distinct textures could not be produced. But this conclusion is justified only if we consider visual textures and visual images to be equivalent mathematical objects.

Julesz never provided a formal mathematical definition of a visual texture. However, it is implicit in his work, as well as in that of subsequent researchers in the field, that a visual texture is not a specific image but rather an ensemble of images that share certain statistics. For example, textures may be constructed by placement of elements at random locations or in random orientations[2–5] or by independent and random assignment of gray levels from a particular histogram to each pixel[6] or by Markov-like processes.[7,8] Thus an adequate theoretical treatment of visual textures must begin by recognition of the distinction between a texture ensemble (a set of textures, each infinite in spatial extent, with an assigned probability structure) and images (individual finite samples of examples drawn from the ensemble).

It is straightforward to construct distinct texture ensembles that have identical statistics up to order n, but not at order $n + 1$, for arbitrary n.[8] At first glance this would seem to be inconsistent with the TCU theorem for $n > 2$, which states that third-order statistics uniquely specify images. However, these constructions are fully consistent with the TCU theorem, because the TCU theorem applies to individual images, not to texture ensembles.

In an attempt to relate the TCU theorem, which ap-

0740-3232/94/051680-05$06.00

plies to images, to the Julesz conjecture, which applies to textures, Yellott argues that mathematical results concerning images reduce questions concerning texture ensembles to matters that can be settled in a trivial way by pure thought. Here it is argued that the concept of a texture ensemble, as distinct from that of an image, makes good sense, both mathematically and biologically. The TCU theorem does not apply to texture ensembles and therefore does not trivialize the Julesz conjecture. Rather, texture ensembles (as distinct from images) constitute an important tool for the experimental analysis of biological vision.

Studies of texture perception aim to determine how the visual system discriminates, segregates, and classifies textures. The underlying view is that visual information is encoded not in a pixel-by-pixel fashion but rather by neural computations that extract certain pieces of information from the image and discard others. For example, the overall amount of contrast is a visually salient feature, but the precise position of each texture element does not appear to be encoded. How can the nature of these neural computations be determined? It does not suffice to observe the response to a single image, because distinct computational mechanisms may produce identical results for certain images. Thus a hypothesis for a neural computation is best tested by examination of responses to multiple images, that is, an ensemble.

In this Communication the term image is used to mean a specific, usually finite, pattern of light and dark. The term texture ensemble is used to mean a collection of spatially infinite patterns of light and dark, together with an assignment of probabilities to each element of the collection. The term texture is used to mean the rule that determines the texture ensemble. As discussed in Section 3 below, texture ensembles that are useful for the study of vision have a key property: any average over the entire spatially infinite ensemble may be replaced either by an infinite spatial average over a single typical member of the ensemble or by a finite spatial average over all members of the ensemble. In view of this property, one may think of textures in several ways: as an infinite collection of finite images, as a single infinite image, or as an infinite collection of infinite images.

Use of the term texture has led to confusion in the past, as it is often applied to a particular image drawn from a texture ensemble. The present formalization is intended to clear up this confusion: use of the term texture in relation to a single image has always implied membership of that image in a collection of similar images with which it shares certain statistical properties. Here, the term texture sample is used to denote an image drawn from a texture ensemble.

2. AN ILL-POSED PROBLEM?

Although the point of view presented here provides a theoretical argument for the role of ensembles in the analysis of texture perception, it might be argued (as Yellott does in his discussions of hypotheses H2–H6) that it is not possible to reduce this idea to practice in a biologically meaningful way. The main problem is that since a particular texture example may belong to many different ensembles, it would appear that a judgment that requires an assignment of a texture example to an ensemble is an ill-defined or impossible task. However, whether the task is indeed ill-defined is an empirical issue, not a theoretical one. Consider a task in which a subject is asked to segregate or sort multiple images and is not given any information on what criteria are to be used. One might find that different subjects segregated the samples in different ways and that there was no consistency in the subjects' performances. This would indeed be consistent with the task's being ill defined. Alternatively, one might find strong consistency between subjects. This would indicate that, for the particular samples used, some built-in neural computation provided a natural partition. One could hypothesize what this neural computation was and then create another set of examples to test this hypothesis.

There are two fundamental ideas here: (1) that the computational machinery that a subject uses to perform a psychophysical task reflects biology and not just formal mathematical considerations and (2) that the experimenter can test computational hypotheses only by examination of responses to multiple images. Neither of these ideas is specific to texture segregation; rather, both are deeply rooted in psychophysics. Consider, for example, the space–time image of a moving sine grating. As is well known, this image is consistent with a range of velocity vectors, provided that the components of the velocity vectors perpendicular to the grating's bars are identical. This phenomenon is known as the aperture problem.[9] Despite the mathematical ambiguity, subjects do not see a full-field grating stimulus as an ambiguous one. Rather, they report a consistent, unique velocity percept: the perpendicular velocity. In order to be sure that the subject is judging velocity, not spatial phase, distance of motion, temporal frequency, or other confounding factors, one must repeat such judgments with stimuli that differ in initial phase, duration, spatial frequency, etc. Thus, reliable judgments are made despite the fact that the stimulus is formally consistent with a continuum of velocities, and examination of responses to a set (ensemble) of stimuli rules out the possibility that some other computation on the stimulus is being performed. More generally,[10] biological vision consists of a cluster of ill-posed problems, such as shape from shading, structure from motion, figure from ground, etc. The visual system necessarily makes inferences from partial information, and the discovery of how these inferences are made, for example, by Bayesian inference, is what the study of biological vision is all about.

3. ROLE OF ERGODICITY

It is easy to construct a stimulus set for which texture segregation is performed in a rapid and consistent manner across subjects: for example, a stimulus set consisting of several examples drawn from the even texture paired with examples drawn from the random texture, or, of samples drawn from many pairs of textures that differ in their second-order statistics. As discussed above, since reproducible texture segregation can be achieved for some texture ensembles, the task is not in principle ill defined or impossible. In his discussion of hypotheses H2–H6, Yellott proposes several examples of the texture-segregation task in which performance is guaranteed to be variable across subjects. His examples do not negate the usefulness of a texture-segregation paradigm as applied to ensembles; they merely show that it is possible to take an informative paradigm and find a stimulus set for which the results will be uninterpretable.

This is where the proviso of ergodicity is relevant. In formal terms, ergodicity means that averages performed over the ensemble of textures can be replaced by spatial averages over a single (spatially infinite) example drawn from the ensemble. Informally, ergodicity means that each texture sample, if considered over its entire (infinite) extent, will typify the entire ensemble. Two-dimensional Markov textures (including the even and odd textures[8]), independent identically distributed (IID) textures,[6] and textures consisting of randomly placed, randomly oriented tokens are ergodic.

Without ergodicity, there is no basis for estimation of ensemble statistics from a single image, even if the image is large. For nonergodic textures, estimates of ensemble statistics from image statistics are guaranteed to be misleading. Consequently, subjects will fail to perform reliably on texture-segregation tasks based on nonergodic ensembles (such as those considered by Yellott[1] in hypotheses H2 and H3 or the heterogeneous ensemble proposed above), but this failure will be uninformative. Conversely, ergodicity gives the subject the chance to succeed, so that success or failure provides information about the capabilities of the visual system.

In his discussion of hypotheses H4–H6, Yellott[1] trivializes the mathematical concept of ergodicity to the notion that the statistics of a single, spatially finite, image must be equal to those of the ensemble. Were this the case, then indeed there would be no distinction between the statistics of finite images and ergodic ensembles. But this is not what ergodicity means; ergodicity specifies replacement of an ensemble average by an average over a single, *spatially infinite*, example.

Since a subject can see only a finite number of spatially restricted views of texture examples, the subject cannot determine whether an ensemble is ergodic. But this does not render the concept useless or vacuous. Mathematical objects and concepts are almost never rigorously realized in the laboratory. The lines and points of real visual images have nonzero spread and are therefore not the lines and points of Euclid; the sinusoids and Gaussian envelopes of visual images have finite extent and therefore do not rigorously have the properties of mathematical sinusoids or Gaussians; white noises used in systems-analysis methods[11] are not spectrally flat. Similarly, the notion of ergodicity must be considered in an analogous fashion: it motivates the design of stimuli, but it cannot be rigorously achieved in the laboratory.

4. ISSUE OF NONUNIQUENESS

Yellott[1] argues (in his discussion of hypothesis H2) that one difficulty with viewing texture-segregation tasks as applying to ensembles is that a single texture example may belong to many ensembles. For example, at least occasionally a finite image drawn from the random ensemble would also be an even texture, just by chance alone. This would lead to subjects classifying as even a texture sample that the experimenter obtained from the random ensemble. However, the chance of this occurring is very slim: for a 32×32 sample of the random texture, the chance that it will be consistent with having been selected from the even ensemble is 1 in 2^{961}, or approximately 10^{-289}. Not surprisingly, subjects spontaneously reject this interpretation. Somewhat more likely is the possibility that an atypical sample such as an all-white image or a checkerboard happened to be included in the even stimulus set. This might cause subjects' performance to be more variable and difficult to interpret. For a 32×32 sample of the textures, the chance of this occurring is 1 in 2^{63}, or approximately 10^{-19}. One is not likely to be led astray by such occurrences, and if one were concerned that performance depended on the choice of specific texture examples, it would be a simple matter to make another stimulus set from the same ensembles. Indeed, experimental studies concerning the even, odd, and random textures explicitly included exploration of the texture ensembles.[12-15]

5. VALIDITY OF THE EVEN, ODD, AND RANDOM TEXTURES AS COUNTEREXAMPLES

In Section 4 of his paper, Yellott[1] calculates second- and third-order statistics for some images that are examples of the even, odd, and random (coin-toss) textures. The motivation for these calculations is to illustrate the fact that although these statistics are matched across the ensemble, they are not matched for specific examples. This point had already been made by Gagalowitz[16] and by Victor.[12] Yellott claims that these imbalances invalidate the even, odd, and random textures as counterexamples to the Julesz conjecture. But one can accept this claim only if one neglects the crucial distinction between images and ensembles. (As we will see below, this distinction is necessary even for the purportedly cleaner counterexamples proposed by Yellott[1]).

Variances and covariances of sample estimates of second-order statistics rigorously correspond to ensemble fourth-order statistics, and variances and covariances of sample estimates of third-order statistics rigorously correspond to ensemble sixth-order statistics. [This may be seen as follows: a variance V_n of a sample estimate of an nth-order statistic E_n will be an average of the form $V_n = \langle (E_n - \langle E_n \rangle)^2 \rangle$. Since E_n is nth order, V_n is a statistic of order $2n$. The argument for covariances is similar.] Consequently, ensembles that differ in their fourth-order (and higher even-order) statistics will necessarily differ in the variances or covariances in estimates of second- and third-order statistics from individual images. Thus the findings that for the even texture, estimates of second- and third-order statistics converge more slowly to their sample mean and that estimates of distinct second- and third-order statistics are highly correlated is a necessary consequence of the higher even-order correlations. In his Fig. 8, Yellott[1] constructs matched examples of even and odd textures in which three fourths of the pixels are of the same luminance and the remaining one fourth are of opposite luminance. These texture samples have more nearly equal third-order statistics than unmatched texture samples but remain readily discriminable—an

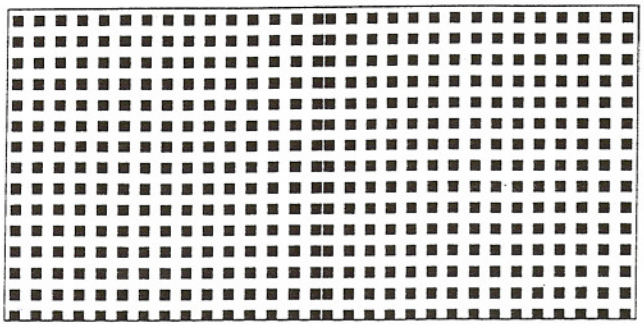

Fig. 1. Demonstration of segregation of two regions with identical statistics.

approach that was taken early in the study of these textures (see Fig. 2 of Ref. 12). Discriminability of such matched texture samples provides empirical evidence that the deviations of sample estimates of third-order statistics from their ensemble means are not crucial for texture discrimination.

The usefulness of the calculations of estimates of second- and third-order statistics is thus not in demonstrating the flaws in these textures as counterexamples, but rather in their potential relevance to understanding the neural computations used in texture discrimination. Indeed, this approach has already been taken. Victor and Conte[14] manipulated long-range fourth-order correlations of the even textures (and consequently the rate of convergence of estimates of second-order statistics) by introducing various kinds of decorrelation into the even texture. Analysis of responses to these textures defined the spatial extent of correlations relevant to texture discrimination.

6. COUNTEREXAMPLES WITHOUT ENSEMBLES?

Yellott[1] argues that the counterexamples that he presents to the second-order (original) Julesz conjecture are superior to those previously presented, because they obviate the need to consider textures as ensembles. There is no doubt that his construction indeed yields discriminable images and textures with identical second-order statistics and thus is new evidence that statistical outliers are not responsible for segregation of isodipole textures. Nevertheless, there are two problems.

The first problem is that, in principle, bipartite stimuli may be segregated by differences in local second-order statistics at their boundary. For example, Fig. 1 shows a bipartite field whose halves have identical statistics of all orders but in which a spatial-phase anomaly at the border produces a rapid segregation.

Intuitively, it is highly unlikely that such edge effects are responsible for segregation in the constructions introduced by Yellott.[1] Nevertheless, if one is to maintain strict rigor, the possible contribution of such edge effects must be ruled out empirically. One way to minimize these edge effects is to replace the segregation task by a classification task in which texture samples are shown one at a time, a strategy that my co-workers and I have adopted.[14,15] Another way to rule out the contribution of these edge effects is to show that segregation does not depend on the spatial phases of the texture samples. That is, discrimination has to be examined for multiple instances of the textures, which differ in initial spatial phase (for the discrimination task) or in relative spatial phases across the border (for the segregation task). However, any such strategy returns us to the notion that the counterexample is fundamentally a statement about ensembles, not about examples. That is, the concept, as well as the use, of a texture ensemble is a natural consequence of the need to control for the artifacts that inevitably occur in the use of a real stimulus as an example of a mathematical ideal. It is interesting to note that the even/odd counterexample of Julesz *et al.*[7] provides a striking example of texture segregation without introducing second- or third-order anomalies at the texture boundary. In this case the border is not anomalous, but the texture samples must be considered to be elements of ensembles.

The second problem with reliance on an exact match of image statistics (rather than an ensemble match) is more fundamental: this approach will never allow analysis of higher-order interactions. The well-documented presence of nonlinearities at both retinal and cortical levels implies that visual mechanisms are not restricted to interactions of formal order 3 or less. Yet the TCU theorem[1] shows that such mechanisms cannot be probed by individual images considered in isolation. On the other hand, textures considered as ensembles lead to a detailed analysis of higher-order mechanisms.[14,15]

7. TEXTONS AND GLOBAL STATISTICS

It is proper to put this debate, which originates with a long-disproved conjecture concerning texture discrimination, in the context of more-recent thinking on the subject. Julesz[17] sums up his texture studies as follows: "In essence, we found that texture segmentation is not governed by global (statistical) rules, but rather depends on local, nonlinear features (textons), such as color, orientation, flicker, motion,.... Particularly important is the realization that—contrary to common belief—texture segmentation cannot be explained by differences in power spectra" (p. 757). I concur with this view.[15]

Even though the neural computations are local, a library of textures with carefully defined global statistical properties is a useful tool for their analysis. This is because many windows on neural activity (e.g., psychophysics, electrical and magnetic evoked potentials, functional imaging) superimpose a spatial average on the inherently local computational processes. By discovering what texture statistics survive this averaging process, one can draw inferences about the underlying local processes.[6,7,14,15]

8. SUMMARY

Visual textures are statistical structures, and one must consider them as such to interpret the Julesz conjecture and related notions properly. Yellott[1] presents a result concerning the statistics of individual images, but this result is of limited relevance to the study of texture perception because of the distinction between spatially finite samples of a texture (an image) and the ensemble from which the sample is drawn. This distinction makes sense both logically and empirically. The assumptions and approximations that are required for linking the mathematical notion of a texture ensemble to laboratory practice are no more severe or unnatural than those required in the application of other useful mathematical structures to biology. Interpretation of texture perception in terms of images rather than of ensembles leads to an impoverished experimental and theoretical analysis.

REFERENCES

1. J. Yellott, "Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture," J. Opt. Soc. Am. A **10,** 777–793 (1993).
2. B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch, "Inability of humans to discriminate between visual textures that agree in second-order statistics—revisited," Perception **2,** 391–405 (1973).
3. T. Caelli and B. Julesz, "On perceptual analyzers underlying visual texture discrimination. Part I," Biol. Cybern. **28,** 167–175 (1978).
4. T. Caelli, B. Julesz, and E. N. Gilbert, "On perceptual analyzers underlying visual texture discrimination. Part II," Biol. Cybern. **29,** 201–214 (1978).
5. J. D. Victor and S. Brodie, "Discriminable textures with identical Buffon needle statistics," Biol. Cybern. **31,** 231–234 (1978).
6. C. Chubb and M. S. Landy, "Orthogonal distribution analysis: a new approach to the study of texture perception," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, eds. (MIT Press, Cambridge, Mass., 1991), pp. 201–230.
7. B. Julesz, E. N. Gilbert, and J. D. Victor, "Visual discrimination of textures with identical third-order statistics," Biol. Cybern. **31,** 137–149 (1978).
8. E. N. Gilbert, "Random colorings of a lattice on squares in the plane," SIAM J. Alg. Discr. Meth. **1,** 152–159 (1980).
9. J. A. Movshon, E. H. Adelson, M. S. Gizzi, and W. T. Newsome, "The analysis of moving visual patterns," in *Pattern Recognition Mechanisms*, Exp. Brain Res. Suppl. **11,** C. Chagas, R. Gattass, and C. Gross, eds. (Springer-Verlag, Berlin, 1985), pp. 117–151.
10. T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," Nature **317,** 314–319 (1985).
11. P. Z. Marmarelis and V. Z. Marmarelis, *Analysis of Physiological Systems: The White-Noise Approach* (Plenum, New York, 1978).
12. J. D. Victor, "Complex visual textures as a tool for studying the VEP," Vision Res. **25,** 1811–1827 (1985).
13. J. D. Victor and V. Zemon, "The human visual evoked potential: analysis of components due to elementary and complex aspects of form," Vision Res. **25,** 1829–1844 (1985).
14. J. D. Victor and M. M. Conte, "Cortical interactions in texture processing: scale and dynamics," Vis. Neurosci. **2,** 297–313 (1989).
15. J. D. Victor and M. M. Conte, "Spatial organization of nonlinear interactions in form perception," Vision Res. **31,** 1457–1488 (1991).
16. A. Gagalowitz, "A new method for texture fields synthesis: some applications to the study of human vision," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-3,** 520–523 (1981).
17. B. Julesz, "Early vision and focal attention," Rev. Mod. Phys. **63,** 735–772 (1991).