Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Vision Research 46 (2006) 299–309

Vision Research

www.elsevier.com/locate/visres

Brief communication

Texture synthesis and perception: Using computational models to study texture representations in the human visual system

Benjamin J. Balas *

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, E25-225, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

Received 27 January 2005; received in revised form 28 April 2005

Abstract

Traditionally, texture perception has been studied using artificial textures made of random dots or repeated shapes. At the same time, computer algorithms for natural texture synthesis have improved dramatically. We seek to unify these two fields through a psychophysical assessment of a particular computational model, providing insight into which statistics are most vital for natural texture perception. We employ Portilla and Simoncelli's texture synthesis algorithm, a parametric model that mimics computations carried out in human vision. We find an intriguing interaction between texture type (periodic, structured, or 3-D textures) and image statistics (autocorrelation function and filter magnitude correlations), suggesting different representations may be employed for these texture families under pre-attentive viewing.

Keywords: Texture; Perception; Computational vision; Natural images

1. Introduction

The visual perception of textures has been an area of interest spanning a wide variety of disciplines from art to computer science. The fields of computer vision, perception, and graphics have each made significant contributions to our overall understanding of texture perception and representation, albeit in quite different ways.

1.1. Psychophysical studies of texture perception

Psychophysicists are of course most interested in what representations and rules the human visual system uses to process textures. In this endeavor, Bela Julesz stands out as one the earliest and arguably most important contributors to the field. The "Julesz conjecture" (Julesz, 1962) represents one of the first hypotheses concerning what image statistics were represented in the human visual system. The original hypothesis was that textures differing only in third-order or higher pixel statistics would be indiscriminable by human observers. This early version of the conjecture was proved false by Julesz himself years later (Julesz, 1975) and the hypothesized "bar" for human discriminability of textures has been pushed past third-order statistics (Julesz, Gilbert, & Victor, 1978) to a possible resting place at fourth-order statistics (Klein & Tyler, 1986). However, recent working analyzing the formalism of creating extreme-order textures (Tyler, 2004a) suggests that the global statistics should not be the sole focus of texture research. Local processes that human observers use to compare different texture samples may be of more importance (Tyler, 2004b). Indeed, most recent models of texture perception rely on linear filter banks rather than higher-order pixel statistics (Malik & Perona, 1990).

* Tel.: +1 617 252 1815; fax: +1 617 253 2964.
E-mail address: bjbalas@mit.edu.

0042-6989/$ - see front matter
doi:10.1016/j.visres.2005.04.013

Many studies concerned with the psychophysics of human texture perception make use of random-dot textures or structured textures composed of repeated symbols like oriented bars, or T, L and X-shaped elements. When using artificial textures such as these, pixel-level texture analysis and simple filter-based strategies are relevant tools. Though useful as a model world for examining texture processing strategies, these artificial textures are not representative of the set of natural textures we encounter in everyday experience. Indeed, these textures violate key features of natural images, specifically the redundancy of natural images (Attneave, 1954; Barlow, 1961). Human observers have implicit knowledge of this redundancy (Kersten, 1987), suggesting that it may be better to study natural textures that match statistical properties of the real world. Natural images have been used to study what higher-level image qualities are used to group textures along salient dimensions (Rao & Lohse, 1996), but little effort has been made to examine low-level representations of photographic textures using psychophysical methods.

1.2. Analysis and synthesis of photographic textures

Machine vision research regarding texture analysis and synthesis is a useful body of work to consider as a means of resolving this difficulty. All of these algorithms share the goal of using small samples of some original texture as a starting point for the reconstruction of arbitrarily large amounts of the same texture. The end result should ideally be indistinguishable from the true texture, although no algorithm can truly remove all artifacts of the synthesis process. Rather than random-dot textures, these algorithms are most often applied to natural textures and have been very successful at creating convincing images for graphics applications. Given that these algorithms operate on natural textures, we will consider them as a useful vehicle for studying the perception of such images by human observers.

The quality of the final reconstruction produced by any of these algorithms informs us as to the utility of both the representation used for the original texture and the process by which that representation is used to generate novel images. However, for us to truly feel confident in relating the computational procedure used for texture synthesis to human perceptual processes it is helpful if the algorithm uses representations employed by the human visual system. For this reason, several texture synthesis strategies that produce strikingly good reproductions of target textures will not be considered here. For example, "image quilting" strategies (Efros & Freeman, 2001) have no true "representation" of a texture, in that patches of the original image are reassembled to make the synthetic version. In a sense, the original image is the only representation of the texture used. Likewise, pixel-growing strategies (Efros & Leung, 1999) are equally problematic in that they represent texture in terms of the distribution of individual pixels in the original image. Synthesis requires a time-consuming search process through the sample provided for analysis. While both of these procedures are extremely useful for graphics applications, we do not believe that they easily relate to human vision.

To achieve a deeper insight as to what statistics are important for the visual processing of natural textures, we turn instead to parametric models of texture analysis and synthesis. These models utilize the idea that filters resembling those found in early visual cortex provide information useful for texture segmentation and classification (Bergen & Adelson, 1988; Bergen & Adelson, 1986). Texture analysis by such filters has proven quite successful at modeling pre-attentive segmentation performance (Malik & Perona, 1990). Filter-based analysis has also contributed to a formal definition of Julesz' "textons" (Julesz, 1981) in terms of clustered filter outputs (Malik, Belongie, Leung, & Shi, 2001).

In terms of texture synthesis, Heeger and Bergen's model (Heeger & Bergen, 1995) demonstrated the utility of "steerable filters" (Simoncelli & Freeman, 1995) for the synthesis of stochastic textures that lacked global structure or distinct textural sub-regions. Distributions of filter coefficients at multiple scales and orientations are extracted from a target image, and a synthetic image can be created by forcing a white-noise field to have matching distributions. The resulting images are quite convincing for some kinds of textures, but fail to capture long-range relationships or inhomogeneous textures. Despite these limitations, this model fulfills two important criteria to be useful as a tool for studying human texture perception. It can be used to synthesize natural textures, and the representation it relies upon (oriented derivative-of-gaussian filters) is motivated by receptive fields found in early stages of visual processing. For the current study, we shall employ a model which is similar to Heeger and Bergen's, but which is able to produce high-quality syntheses across a range of different kinds of texture.

1.3. The model of Portilla and Simoncelli

Since its initial presentation, the basic Heeger–Bergen model has been improved in many ways. In particular, to overcome the inability of the original model to reproduce extended contours and other large-scale structures in the target texture, additional constraints across scales and orientations were introduced by Portilla and Simoncelli (Portilla & Simoncelli, 1999; Portilla & Simoncelli, 2000; Simoncelli & Portilla, 1998; Simoncelli, 1997). We opt in the current study to use their model as a basis for exploring the necessary and sufficient statistics for the successful synthesis of various kinds of photographic texture. There are several reasons for this choice. First, Portilla and Simoncelli's model produces very high-quality images. Second, synthesis can be achieved relatively quickly, meaning a library of synthesized textures can be created in a reasonable time frame. This is in contrast to the FRAME model of texture synthesis (Zhu, Wu, & Mumford, 1996, 1997), which is very powerful, but slower. Finally, the implementation of the algorithm allows for "lesioning" of the code to remove certain parameters from the synthesis process. This last aspect of the model makes it particularly attractive for our purposes, as it allows us to synthesize textures lacking certain statistical constraints. We may then assess how well the final image approximates the target texture.

The Portilla–Simoncelli model utilizes four large sets of parameters to generate novel texture images from a specified target. In all cases, a random-noise image is altered such that its distributions of these parameters match those obtained from the target image. The first of these parameter sets is a series of first-order constraints (marginals) on the pixel intensity distribution derived from the target texture. The mean luminance, variance, kurtosis and skew of the target are imposed on the new image, as well as the range of the pixel values. The skew and kurtosis of a low-resolution version of the image is also included in this set. Second, the local autocorrelation of the target image's low-pass counterparts in the pyramid decomposition is measured (coeff. corr), and matched in the new image. Third, the measured correlation between neighboring filter magnitudes is measured (mag. corr). This set of statistics includes neighbors in space, orientation, and scale. Finally, cross-scale phase statistics are matched between the old and new images (phase). This is a measure of the dominant local relative phase between coefficients within a sub-band, and their neighbors in the immediately larger scale. Portilla and Simoncelli report on the utility of each of these parameter subsets in their description of the model, but offer no clear perceptual evidence beyond the visual inspection of a few example images. The current study aims to carry out a true psychophysical assessment, in the hopes that doing so will more clearly demonstrate which statistics are perceptually important for representing natural textures.

We present the results of two experiments, designed to test the aforementioned parameter subsets value in producing textures that are indiscriminable from the target texture under pre-attentive conditions. We note that this is markedly different than analyzing the resulting images under full scrutiny. This is because the kinds of artifacts and errors that may seem glaring given an attentive analysis of an image may be invisible under pre-attentive conditions. Our strategy is to first produce synthetic textures that are not matched to the target texture for one or more of the parameter families previously mentioned. We then determine how discriminable synthetic textures are from original textures under brief presentation. In so doing, we explicitly assume a local windowing model of texture processing similar to a recently proposal of Tyler's (Tyler, 2004b). We compare discriminability of "lesioned" textures to the discriminability of synthetic textures created using the full set of statistical parameters in the model. This allows us to determine how much each parameter subset contributes to the final synthesis. Further, we break down our target textures into three families ("periodic", "structured," and "3-D asymmetric" textures) to see whether or not different statistics are needed to convincingly synthesize specific categories of images.

2. Methods

2.1. Subjects

A total of 16 subjects participated in the two experiments described here, eight in each of our two experiments. Subject age ranged from 19 to 27 years, and all subjects had normal or corrected-to-normal vision.

2.2. Stimuli

Original textures—18 256×256 texture samples were chosen from a set of textures available via the NYU Laboratory for Computational Vision (http://www.cns.nyu.edu/~eero/software.html). Several textures are Brodatz images (Brodatz, 1996) while the remainder are original photographs collected by the NYU laboratory. The images were selected to conform to three preconceived visual categories, pseudoperiodic, structured, and 3-D textures with asymmetric luminance gradients.

The first two categories were selected because both the presence of periodicity and the presence of structured elements have been suggested as useful criteria for classifying textures in the computer vision literature (Haralick, 1979; Portilla, Navarro, Nestares, & Tabernero, 1996). While there are four classes of texture that can be obtained by crossing the presence or absence of periodicity with the presence or absence of structured elements, we have opted to include only two of those classes here (periodic and non-structured as well as non-periodic and structured textures). Of the four possibilities available to us, we believe that the two we have selected are most likely to require different statistics for successful synthesis. For our purposes, we will consider pseudoperiodic textures to be images based on a spatially regular repeated pattern, which may vary slightly across the image. Structured textures are defined as those textures composed of discrete elements that are not repeated in a predictable way across the image.

The third category of textures we shall examine, 3-D textures with asymmetric luminance gradients, is included specifically to examine how important cross-scale phase information is to pre-attentive texture perception. "3-D asymmetric" textures are images that contain strong lighting effects which suggest depth. In particular, these images contain luminance gradients of the same sign and similar orientation across the image surface. These image conditions make cross-scale phase statistics very relevant for viewing synthetic images with scrutiny, and we wish to explore if this dependency holds for pre-attentive viewing. Previous studies have indicated that local phase differences are not pre-attentively discriminable, but primarily use very simple stimuli to evaluate this claim (Malik & Perona, 1990; Rentschler, Hubner, & Caelli, 1988; Sun & Perona, 1996). It may be the case that in natural images local phase relationships provide useful information under pre-attentive viewing. We note that though this last set of textures looks very heterogeneous, this does not necessarily mean that the human visual system does not use a common statistical mechanism to represent them.

All of our target textures, grouped into the three families described here, are displayed in Fig. 1.

"Lesioned" textures—Five synthetic versions of each original texture image were created using Portilla and Simoncelli's algorithm. The first four images were created by choosing to ignore one family of statistical measurements taken from the original image while performing the synthesis procedure. In order, marginal statistics, raw autocorrelation statistics, filter magnitudes, and cross-scale phase measurements were removed from consideration one at a time for each condition. The fifth category of synthesized textures was created by synthesizing each texture using the full set of statistical constraints. Each synthesized image was 256×256 pixels in size, using parameters extracted from a 192×256 pixel patch taken from the original texture. These slightly smaller patches were used to remove the text credits that appeared in the lower left corner of some images. Examples of the synthesized textures created from a particular target texture are displayed in the top row of Fig. 2.

"Pair-wise impoverished" textures—For Experiment 2, we create four new categories of texture images by synthesizing texture patterns using the marginal statistics alone, and also the marginal statistics plus each of the three remaining parameter subsets added in one at a time. While the images in Experiment 1 allow us to discuss the necessity of each subset of parameters for texture synthesis, these images are designed to give us insight as to the sufficiency of these subsets for successful texture reconstruction. The reason for using "pair-wise" images rather than synthesizing textures using each parameter subset in isolation is that in inspecting the top-left image of Fig. 2, it is obvious that those images lacking the same first-order statistics as their parent textures are strikingly different from the other lesioned images. This is the case because encapsulated in those first-order measurements are highly salient global image properties like overall contrast and mean luminance of the image. From this, we expect that first-order properties will certainly prove to be necessary for good synthesis in Experiment 1. This means that testing sets of images that lack these prop-

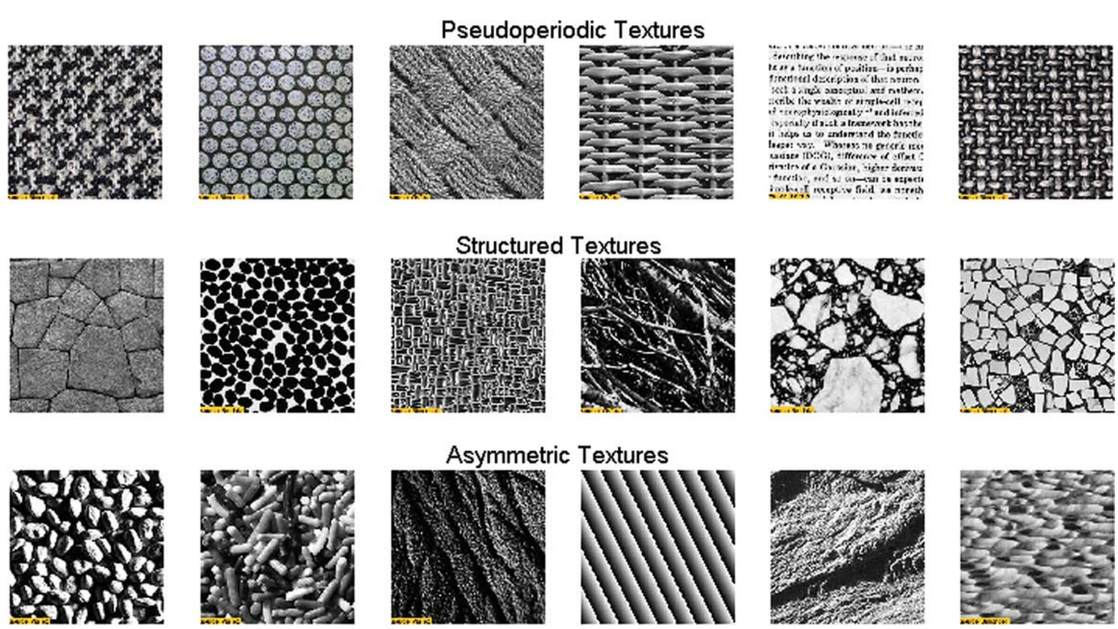

Fig. 1. The collection of textures used to create synthetic images for Experiments 1 and 2. The top row contains textures that have strong periodicity. (We consider text pseudoperiodic because of the even spacing of rows.) The middle row contains textures that are composed of repeated structural elements but lack strong periodicity or global structure. The bottom row contains textures with asymmetric lighting effects suggesting depth.

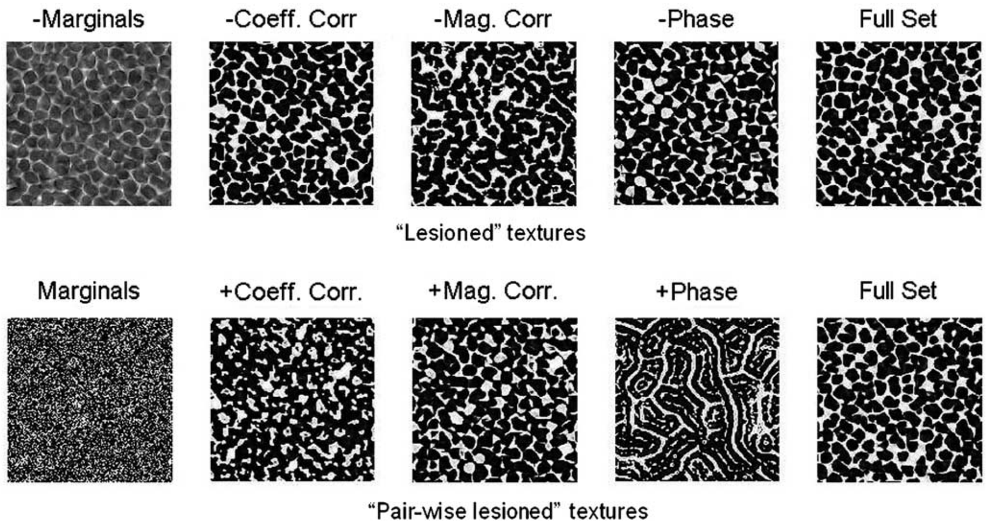

Fig. 2. (Top row) "Lesioned" texture images created using the Portilla and Simoncelli algorithm. Synthesized textures from our original images were created using either the full set of statistical parameters (far right) or using all but one subset of those parameters. From left to right, the images in this figure were constructed without explicit matching of first-order constraints (mean, range, variance, kurtosis and skew), sub-band coefficient correlation, sub-band magnitude correlation, and cross-scale local phase information. (Bottom row) "Pair-wise" lesioned images created by including the first-order statistics in all synthetic textures with the addition of: (from left to right) nothing additional, sub-band coefficient correlation, sub-band magnitude correlation, cross-scale phase information, and all parameters in the Portilla and Simoncelli algorithm.

erties will be relatively useless. Instead, we include these parameters in all cases, allowing us to test the first-order properties themselves for sufficiency as well as the remaining parameter subsets (with the caveat that pixel distributions are always matched). Examples of these synthesized textures are displayed in the bottom row of Fig. 2.

2.3. Procedure

Subjects were seated approximately 100 cm from a 17″ Dell Ultrasharp monitor. All stimulus display and response recording functions were controlled via the Matlab Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

In both experiments, subjects were to perform a 3AFC "oddball" task, in which three unique texture patches were presented on each trial. The "oddball" image was drawn from a random location within either the original texture or the synthesized version of that texture. On each trial, two non-overlapping distractor patches were then drawn from either the synthetic or original image, respectively. By randomly sampling our patches from the larger images at each trial, we have access to a very large set of possible stimuli, making the memorization of individual patches impossible. Using two non-overlapping distractors on each trial also ensures that common features within the two distractor images cannot contribute to task performance. Finally, given that the oddball image on each trial can be either real or synthetic, subjects must compare all three images to each other to perform well on each trial. In Experiment 1, we will be looking for cases in which the removal of a statistical constraint *improves* detection of the "oddball" image. This will indicate that the "lesioned" constraint carried information that is necessary for a good synthetic image. Conversely, in Experiment 2, we will be looking for cases where the imposition of a statistical constraint results in poor detection of the oddball. This will indicate that the included constraints carried sufficient information for a good synthetic image.

Each image patch was windowed with a circular mask to remove any orientation-specific interactions between the contours of the image frame and contours within the texture itself. Subjects were not familiarized with the textures previously, and all three texture patches in a given trial were distinct images. These measures were taken to ensure that neither high-level information nor pictorial matching strategies could contribute to subjects' performance.

On each trial, the three images were displayed at the vertices of an equilateral triangle such that the distance between each image and central fixation was approximately 3.5° of visual angle. Each stimulus was approximately 2° of visual angle in diameter (approximately 64 pixels/degree), and the entire stimulus triad was onscreen for 250 ms and then removed. Responses were collected after the stimulus triad disappeared. Subjects indicated the location of the oddball texture patch via the "1", "2", and "3" keys to indicate left, top, and right respectively. Response time was not recorded, and no feedback was provided to the subjects. Presentation order was randomized for each subject.

Subjects completed 144 trials per "lesion" condition for each of our three texture families for a total of 2160 trials. Breaks were scheduled every 720 trials.

3. Results

Experiment 1– In our first experiment, we are looking for evidence that subsets of statistical constraints collected by the Portilla and Simoncelli algorithm are differentially important for the successful synthesis of our three texture families. In particular, this experiment assesses the degree to which each subset of parameters is necessary for the synthesis of each type of texture by removing one set of constraints at a time.

A 2-way ANOVA (with repeated measures) was run on the number of accurate responses. The data revealed a highly significant of lesioning condition ($p < 10^{-4}$) as well as a highly significant interaction between texture category and lesion ($p < 10^{-5}$). There was no main effect of texture category ($p > 0.4$).

In Fig. 3, we see that as we predicted the first-order statistics of our texture distributions are clearly necessary for successful synthesis. Subjects are at ceiling at detecting the "oddball" texture when these constraints are removed. Further, the interaction between lesion and texture category appears to be driven by the differential importance of raw coefficient correlation and magnitude correlation for our three families of textures. To be more specific, pseudoperiodic textures seem to rely relatively equally (and weakly) on both of these sets of parameters, given that the removal of each does not cause a large increase in the number of correct detections. In contrast, the magnitude correlation statistics are clearly quite necessary for successful synthesis of structured textures, while the coefficient correlations seem to contribute almost nothing to the full synthesis. This pattern of results is also observed with the 3-D asymmetric textures, although the effect of removing the magnitude correlations is less pronounced. We note that the constraints on cross-scale phase do not appear to be necessary for any of our three texture categories, indicating that under pre-attentive conditions these constraints matter very little.

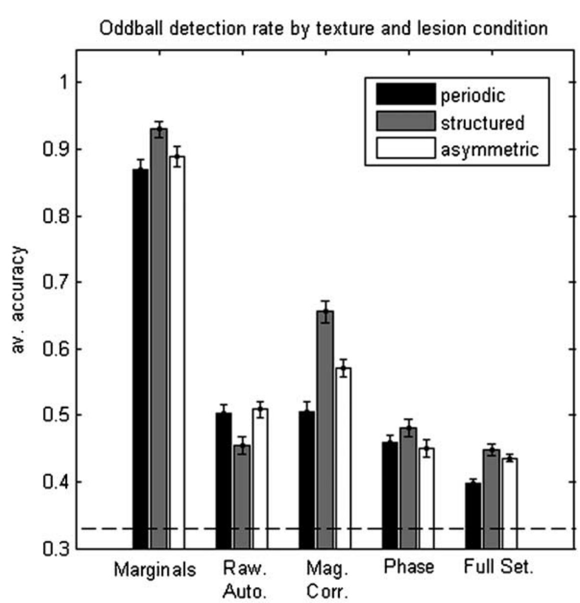

Fig. 3. Plot of the average performance on the oddball detection task as a function of both texture category and texture lesion (mean values ±1 standard error across subjects). Greater accuracy at oddball detection indicates greater necessity of the lesioned statistical constraints. Note both the clear importance of first-order statistics at left, as well as the interaction between the necessity of coefficient and magnitude correlation for periodic, structured, and asymmetric textures.

To confirm this assessment of the results, we conducted post-hoc Tukey–Kramer tests within each texture category between each of the 4 "lesion" conditions and the "full set" condition. We find that for pseudoperiodic textures, only the removal of the first-order statistics produces a rate of oddball detection significantly greater than the "full set" images ($p < 0.05$). However, for the structured textures and the 3-D asymmetric textures, we find that both the removal of the first-order statistics and the removal of the magnitude correlation statistics produce rates of oddball detection significantly greater than that of the "full set" textures ($p < 0.01$, and $p < 0.05$ respectively).

Experiment 2—In this second experiment, we are testing the sufficiency of both first-order statistics in isolation and pair-wise combinations of first-order information and the remaining three parameter subsets for producing successful synthetic texture images. In these results we will be looking for cases where the inclusion of parameter subsets gives rise to low rates of oddball detection. This will indicate that the subsets included may be sufficient for producing synthetic textures viewed under pre-attentive conditions.

As in Experiment 1, we ran a 2-way ANOVA with repeated measures on subjects' accuracy, with lesion condition and texture category as factors. As before, we find no effect of texture category ($p > 0.3$) but a significant effect of lesion condition ($p < 10^{-5}$) and a significant interaction between these two factors ($p < 0.01$).

We note in Fig. 4 that the inclusion of first-order statistical constraints alone results in a rate of oddball detection that is at ceiling. This indicates that though these parameters are necessary for synthesis, they are certainly not sufficient. Of more interest however, is the relationship between the other three parameter subsets. Specifically, we notice that for all three of our texture families magnitude correlation proves to be quite useful for synthesis, producing rates of oddball detection comparable to the "full set" images. Raw coefficient correlation, by contrast, appears to only be

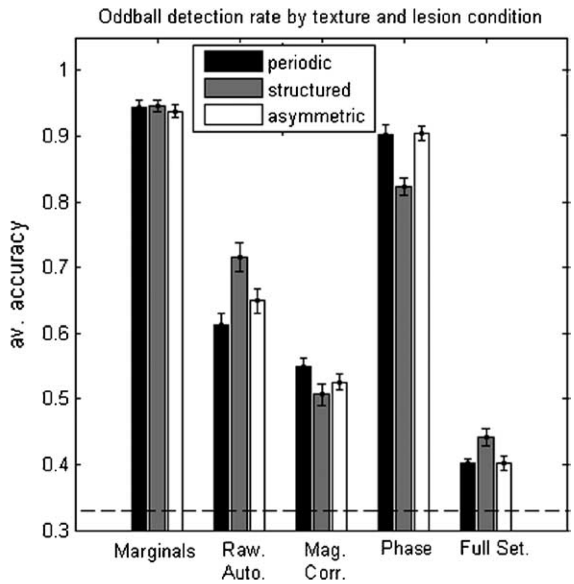

Fig. 4. Rates of oddball detection for all three texture families as a function of statistics included in the synthesis process (mean values ±1 standard error across subjects). Poorer accuracy at oddball detection indicates greater sufficiency of the included statistical constraint. We note that both marginal statistics alone, and the pair-wise inclusion of marginal and cross-scale phase constraint provide poor syntheses. In contrast, magnitude correlations and marginal statistics together provide for relatively good synthesis of all three textures. The raw coefficient correlation is only weakly sufficient, and appears to contribute most effectively to pseudoperiodic and asymmetric textures.

weakly sufficient, and more useful for pseudoperiodic and asymmetric textures than structured images. Finally, we also note that oddball detection rates are very high when cross-scale phase statistics are included with the first-order measurements. For structured textures, this rate is somewhat lower than ceiling (perhaps an indication that cross-scale phase provides some small amount of useful information), but overall demonstrative of the insufficiency of phase information for texture synthesis.

As before, post-hoc Tukey–Kramer tests were run to confirm our intuitions regarding the interaction of inclusion condition and texture family. In comparing the "full set" responses to the other conditions within texture families, we find that for structured and 3-D asymmetric textures three lesion conditions differ significantly ($p < 0.05$) from the "full set" rate of oddball detection, with the sole exception being magnitude correlation. For pseudoperiodic textures, we find that all conditions differ significantly ($p < 0.05$) from the "full set". From inspection of the graph, raw coefficient correlation could be considered weakly sufficient in spite of this analysis. However, it is our belief that the efficacy of first-order statistics and magnitude correlation for all of our textures is most clearly indicated by this experiment. For all three texture families, images that incorporate these statistics generate the lowest rates of oddball detection. Moreover, these rates do not significantly differ from "full set" detection rates in two cases.

4. Discussion

We have found in our pre-attentive discrimination task that the necessity of various statistical parameters for high-quality synthesis is different for pseudoperiodic, structured, and 3-D asymmetric texture images. We find that first-order pixel statistics such as the mean, variance, and range of luminance values are vitally important for creating perceptually matched textures from any target image. This is hardly surprising given how easily human observers can discriminate between different brightness and contrast levels. Of more interest is the reliance of each texture family on autocorrelation and filter magnitude correlation statistics. Periodic textures demonstrate no strong need for either of these measures, but rather rely weakly and almost equally on both. Structured textures, by comparison, appeared to rely quite heavily on the magnitude correlation statistics, while demonstrating no need for preservation of the local autocorrelation statistics. 3-D asymmetric textures inhabit a middle ground between these two extremes, relying significantly on magnitude correlation (though less so than structured textures) and showing little need for preservation of the raw autocorrelation.

None of our texture classes appeared to rely heavily on cross-scale phase statistics for synthesis. This suggests that these measurements may only be important for texture images that undergo scrutiny, or classes of texture not represented here. It is our belief that this latter possibility is more likely. We selected the 3-D asymmetric textures specifically with the hope of finding a reliance on phase statistics, but this does not mean that another texture class not examined here does not make use of these measurements. Also, task demands may unfairly limit the extent to which both these sets of statistics can be extracted by the visual system. Slightly longer viewing times might make the errors brought on by these lesions more apparent, while still remaining in the realm of pre-attentive texture perception.

In terms of the sufficiency of our parameter subsets, we find that preserving only first-order measurements of the pixel distribution is clearly not enough to create a convincing synthetic image. Again, this is not surprising given that the human visual system is known to have strong representations of higher-order features (like edges) that will not be preserved through balancing only pixel-based statistics. Also, as we might expect from the results of Experiment 1, cross-scale phase statistics combined with proper first-order measurements result in extremely poor syntheses. Again, it is the imposition of the autocorrelation and coefficient magnitude constraints that prove most useful in this task. Autocorrelation statistics prove weakly sufficient for pseudoperiodic and asymmetric textures, but comparatively less effective for producing structured textures. When we include magnitude correlation statistics instead, two of our texture categories (structured and 3-D asymmetric) give rise to detection rates that are not statistically different from those obtained with the full set of constraints. We note that pseudoperiodic textures are also well-specified when only magnitude correlations and first-order properties are preserved, though detection rates are still above "full set" rates. This is in good agreement with the data from Experiment 1, in that both structured and 3-D asymmetric textures seem to be well-represented by magnitude correlations, while making less use of the raw autocorrelation. Also as before, it appears that pseudoperiodic textures make use of these two statistics relatively equally, and to a lesser degree than either of the other two texture categories.

The necessity and insufficiency of first-order image properties is not a new or surprising contribution. Of more importance is the perceptual role of cross-scale phase statistics and the magnitude correlations introduced by Portilla and Simoncelli. In the first case, we point out that neither the inclusion or absence of cross-scale phase information affected the synthesis process in any way that indicated this information was of perceptual use under pre-attentive viewing. This supports results obtained with mirror-image gabor-like stimuli (Malik & Perona, 1990; Rentschler et al., 1988), suggesting that both for schematic and natural stimuli local phase statistics do not contribute to pre-attentive processing. In the second case, we note that the magnitude correlations appear to be extremely important for the perceptual integrity of structured and 3-D asymmetric textures under pre-attentive conditions. It is quite interesting that these statistics are so important, as they suggest predominantly local measures (in space, orientation, and scale) support the perception of quite complex textures. Indeed, for these two texture families, matching only these parameters and first-order properties one can create synthetic images that are not of significantly lower quality than those made with the entire set of constraints.

Also noteworthy is the fact that none of the parameter subsets examined here proved "necessary" for the synthesis of pseudoperiodic textures, save for first-order pixel statistics. This suggests that these textures might be represented by statistics that are more evenly distributed across the subsets considered here. This is somewhat at odds with previous results concerning periodic textures, specifically with regard to the importance of the autocorrelation function (Fujii, Sugi, & Ando, 2003). However, pre-attentive viewing of natural textures may prove quite different from viewing the same images with scrutiny. Importantly, one should not conclude from our results that pseudoperiodic textures are not well-represented by the statistics utilized in the Portilla–Simoncelli model. Rather, the information for synthesizing such textures is not distributed in a heavily asymmetric way with regard to the four parameter subsets described here.

There are several additional caveats that must be raised, as well. The first of these concerns the discriminability of the "full set" images from the target textures. In our 3AFC task, chance performance was 33%, a rate of oddball detection lower than that displayed by all but a few of our subjects. Overall, this indicates that even in the most difficult condition our synthetic textures were still reliably discriminable from their respective targets. In all cases, we are only able to consider the necessity and sufficiency of the parameters included relative to this baseline. We do not see this as especially problematic, but it does indicate that there is still a fair amount of work to be done as far as creating more powerful parametric texture synthesis algorithms. We are limited to testing the statistical constraints imposed by this particular model, and though they seem both reasonable and useful we must remember that there remains an infinite number of image statistics that may prove useful in the future.

Second, we must mention that the sets of statistics we have considered in these experiments are not completely independent. There are many examples of redundancy between some of these sets, most notably between the raw autocorrelation and the magnitude correlation (Portilla & Simoncelli, 2000). This does not render our necessity and sufficiency tests invalid, but it does alter how we should interpret the data. For example, neither the raw autocorrelation nor the magnitude correlation statistics proved necessary for the synthesis of pseudoperiodic textures. Naively, one might think that this implies that both of these parameter subsets could be removed without perceptual consequences. When we do so, however, we find that the resulting synthetic image is quite poor. We suggest therefore that these results be interpreted as indicating only the relative contributions of each parameter subset, not an absolute record of which statistics one should feel free to leave out when creating synthetic textures.

Related to the redundancies between the parameter subsets used here is a more serious concern regarding our method for assessing necessity and sufficiency through the use of "lesioned" textures. We have generated our stimuli by asking the Portilla–Simoncelli algorithm to preserve some sets of statistics, while not making any effort to preserve other sets. We have no guarantee however, that the algorithm will not match the "lesioned" statistics accidentally despite making no explicit attempt to do so. If it is the case that some statistics are accidentally matched by the algorithm when they are lesioned (while others are not), our interpretation of these results is potentially erroneous. Parameters determined to be "unimportant" in our necessity experiment may instead simply be the victims of accidental matching, rendering them as indiscriminable as images synthesized using all of the parameters.

To determine if this has indeed occurred for the stimuli presented here, we present a brief analysis of the extent to which each "lesioned" statistic is accidentally reconstructed. For three of our single lesion conditions (raw autocorrelation, magnitude correlation, and phase) we use a signal-to-noise utility provided with the Portilla–Simoncelli algorithm to measure parameter fidelity in a synthetic texture relative to the original image. This utility measures the SNR of each statistic set in a synthetic image relative to the original image. We apply this measure to our "lesioned" images and the "full set" images. By taking the difference in SNR between "full set" and "lesioned" images, we can roughly assess how well each statistic is numerically matched to the original parameters when it is lesioned as compared to when it is retained. When this difference is large, we can be relatively certain that those statistics are not being matched very well in the lesioned image. Conversely, when this difference is very small, we must assume that the lesioned statistics are being well-matched.

The advantage of using a difference score between lesioned and non-lesioned synthetic images is that it allows us to put the different parameter subsets on more equal footing. The SNR measure only reflects the numerical difference between the parameters obtained from the original image and those obtained from the synthetic image. That said, if we are unaware of how well these values are matched in the best circumstances ("full set" synthesis) than this value has little meaning. For example, imagine that the SNR of one set of lesioned parameters is 3 dB and that of another is 10 dB. This appears to indicate that the second set of parameters is being well-matched accidentally and the first is not. However, it may be the case that this second set of parameters is simply easier to match during the synthesis process. This would result in a high SNR relative to other parameters in all the lesioned images, even if there is information being lost when we fail to constrain these statistics. By using a difference score, we are able to account for a baseline shift such as this that might otherwise corrupt our understanding of how much "damage" each lesion does. That said, since the SNR is only a measure of the numerical difference between two sets of numbers there is a subtler issue concerning the relevant scale for each parameter subset. We cannot say for sure whether a difference score of 2 dB in SNR for one lesioned statistic is exactly equivalent to the same score for another lesioned statistic. However, this measure at least gives us an ability to talk about the extent of numerical matching that is occurring for lesioned statistics.

We only analyze the SNR differences for these three lesion conditions because we believe that first-order statistics are being reliably lesioned by inspection of our images. Further, the high rates of oddball detections for these images support the successful lesioning of these parameters. Since most of our interesting data comes from the other three lesion conditions, we shall only inspect their SNR differences. In particular, we wish to know if either the autocorrelation parameters or the phase parameters are being accidentally reconstructed. This would suggest that the reported inefficacy of these parameters is due to an incomplete lesioning process rather than perceptual processing. We also wish to see if there are any interactions between SNR differences and texture categories that would predict the interactions we see in the oddball task. If so, this would suggest that our effects are being driven by the weaknesses of the lesioning procedure. If we see no interactions that mimic the perceptual data, we can be more confident in our behavioral data. The results of this analysis are displayed in Fig. 5.

We first note that in general the algorithm matches non-lesioned parameters equally well in the full set condition and each lesion condition, as evidenced by the near-zero differences in SNR. Likewise, the statistics that are "ignored" are never matched as well in the lesioned images as in the full syntheses, as evidenced by the positive SNR differences. However, it is also immediately apparent that when raw autocorrelation statistics are lesioned, the relevant differences in SNR are numerically smaller than in the other lesion conditions. This may indicate that these statistics are well-matched even when we ask the Portilla–Simoncelli algorithm to ignore them, explaining in part the lack of strong necessity we have observed for the raw autocorrelation statistics. Even when lesioned, these statistics may be well-preserved, leading to synthetic images that are almost as good as the "full set" images.

For the moment, we must allow the possibility that the overall lack of any strong need for the autocorrelation statistics in our experiments may reflect either varying efficacy of the lesioning procedure, or a real perceptual phenomenon. This means that we must leave open the possibility that these parameters may be more vital to all of our texture categories than we have implied here.

We note however, that the overall lack of a dependency on phase is likely not caused by accidental preservation of those statistics in the phase-lesion condition. Phase SNR differences in this condition are numerically large, suggesting that these measurements are not well preserved during lesioning.

Further, there are no significant main effects of texture category or interactions between texture category and the SNR differences within each set of lesioned images (three 2-way ANOVAs, $p > 0.3$ in all tests). This

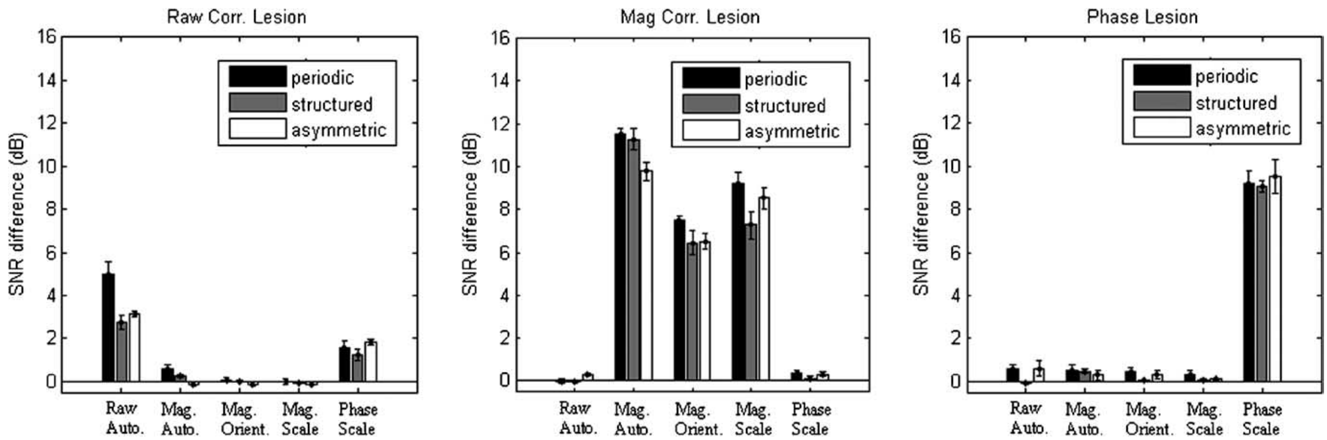

Fig. 5. Differences in SNR between "full set" images and "lesioned" images for raw autocorrelation, magnitude correlation, and phase lesions. We include a measure of SNR difference for the raw autocorrelation parameters, all three subsets of the magnitude correlation parameters (spatial neighbors, orientation neighbors, and scale neighbors), and the phase parameters. Raw autocorrelation-lesioned images (left), magnitude correlation-lesioned images (middle), and phase-lesioned SNR differences (right) are all averaged within each texture family, and the mean SNR difference ±1 standard error is displayed. We note the lack of any main effects of texture category or interactions between texture family and statistics subset in any of our three graphs. However, the raw autocorrelation graph indicates numerically smaller differences between the full set SNR and the lesioned SNR relative to the other lesion conditions.

suggests to us that the interactions we report in our data are not predicted by "accidental matching". Future efforts to put all lesioned images on some sort of perceptually equal footing would be extremely valuable to this enterprise, however.

As a final note, we must reflect on whether we have limited our scope too much by studying one particular model of texture synthesis when there are so many possibilities to choose from. Do these experiments tell us anything about texture processing beyond the limitations of one particular computational model for creating synthetic textures? We suggest that they do. Wavelet-like representations of images are employed in numerous computer vision applications, and the Portilla–Simoncelli model makes use of them in principled ways. In particular, these experiments tell us what contributions early vision might make to representing natural textures. The different sets of "lesioned" statistics are no more than different processing steps applied to the basic measurement of multi-scale oriented contrast, and therefore the Portilla–Simoncelli model helps us understand how we might use the information in V1 to do useful work for texture recognition. Of course, as models for synthesis develop further, so too should psychophysical assessment of those models continue in tandem.

5. Conclusions

We have used a parametric model of texture synthesis as a tool for examining the necessity and sufficiency of different statistical measures for the perceptual similarity of texture images. We have found that different requirements apply for periodic textures as opposed to structured textures, notably in the need for autocorrelation measurements and conditional histograms of edge-like filter magnitudes. Cross-scale phase statistics were found to be of little use under pre-attentive conditions, while first-order pixel properties were demonstrated to be vital for capturing global image similarity. These results demonstrate the value of using computational models for texture synthesis to address perceptual questions regarding texture processing. It is hoped that this may help to bridge the gap between the communities of graphics, machine vision, and psychophysical texture research. Moreover, the 3AFC task presented here represents a modest contribution towards the formulation of texture discrimination tasks that make explicit the importance of local texture analysis in the human visual system.

Acknowledgments

The author would like to thank Bernd Heisele, Yuri Ivanov, and Pawan Sinha for their support of this work. Also, Erin Conwell and David Cox read earlier drafts of this manuscript and made many helpful comments. Two anonymous reviewers also provided many helpful suggestions and insights. B.J.B. is supported by a National Defense Science and Engineering Graduate fellowship.

References

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

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. Rosenblith (Ed.), *Sensory communication*. Cambridge, MA: MIT Press.

Bergen, J., & Adelson, E. (1988). Early vision and texture perception. *Nature, 333*, 363–364.

Bergen, J. R., & Adelson, E. H. (1986). Visual texture segmentation based on energy measures. *Journal of the Optical Society of America A, 3*, 99.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Brodatz, P. (1996). *Textures: A photographic album for artists and designers*. New York: Dover.

Efros, A. A., & Freeman, W. T. (2001). Image quilting for texture synthesis and transfer. *Paper presented at the SIGGRAPH '01, Los Angeles, CA*.

Efros, A. A., & Leung, T. K. (1999). Texture synthesis by non-parametric sampling. *Paper presented at the Proceedings of the International Conference on Computer Vision, Corfu*.

Fujii, K., Sugi, S., & Ando, Y. (2003). Textural properties corresponding to visual perception based on the correlation mechanism in the visual system. *Psychological Research, 67*, 197–208.

Haralick, R. M. (1979). Statistical and structural approach to texture. *Proceedings of the IEEE, 67*, 786–804.

Heeger, D., & Bergen, J. (1995). Pyramid-based texture analysis/synthesis. *Paper presented at the ACM SIGGRAPH*.

Julesz, B. (1962). Visual pattern discrimination. *IRE Transactions of Information Theory, IT, 8*, 84–92.

Julesz, B. (1981). Textons, the elements of texture perception and their interactions. *Nature, 290*, 91–97.

Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American, 232*, 34–43.

Julesz, B., Gilbert, E. N., & Victor, J. D. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics, 31*, 137–140.

Kersten, D. (1987). Predictability and redundancy of natural images. *Journal of the Optical Society of America A, 4*, 2395–2400.

Klein, S. A., & Tyler, C. W. (1986). Phase discrimination of compound gratings: generalized autocorrelation analysis. *Journal of the Optical Society of America A, 3*, 868–879.

Malik, J., Belongie, S., Leung, T., & Shi, J. (2001). Contour and texture analysis for image segmentation. *International Journal of Computer Vision, 43*(1), 7–27.

Malik, J., & Perona, J. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A, 7*, 923–932.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Portilla, J., Navarro, R., Nestares, O., & Tabernero, A. (1996). Texture synthesis-by-analysis based on a multiscale early-vision model. *Optical Engineering, 35*(8), 2403–2417.

Portilla, J., & Simoncelli, E. (1999). Texture modeling and synthesis using joint statistics of complex wavelet coefficients. *Paper presented at the IEEE workshop on statistical and computational theories of vision, Fort Collins, CO*.

Portilla, J., & Simoncelli, E. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40*(1), 49–71.

Rao, A. R., & Lohse, G. L. (1996). Towards a texture naming system: identifying relevant dimensions of texture. *Vision Research, 36*, 1649–1669.

Rentschler, I., Hubner, M., & Caelli, T. (1988). On the discrimination of compound Gabor signals and textures. *Vision Research, 28*, 279–291.

Simoncelli, E., & Portilla, J. (1998). Texture characterization via joint statistics of wavelet coefficient magnitudes. *Paper presented at the fifth IEEE international conference on image processing, Chicago*.

Simoncelli, E. P. (1997). Statistical models for images: compression, restoration and synthesis. *Paper presented at the 31st Asilomar conference on signals, systems and computers, Pacific Grove, CA*.

Simoncelli, E. P., & Freeman, W. T. (1995). The steerable pyramid: a flexible architecture for multi-scale derivative computation. *Paper presented at the second international conference on image processing, Washington, DC*.

Sun, J., & Perona, P. (1996). Early computation of shape and reflectance in the visual system. *Nature, 11*, 165–168.

Tyler, C. W. (2004a). Beyond fourth-order texture discrimination: generation of extreme-order and statistically-balanced textures. *Vision Research, 44*, 2187–2199.

Tyler, C. W. (2004b). Theory of texture discrimination based on higher-order perturbations in individual texture samples. *Vision Research, 44*, 2179–2186.

Zhu, S. C., Wu, Y. N., & Mumford, D. (1996). Filters, random fields and maximum entropy (FRAME)—towards the unified theory for texture modeling. *Paper presented at the IEEE conference on computer vision and pattern recognition*.

Zhu, S. C., Wu, Y. N., & Mumford, D. (1997). Minimax entropy principle and its application to texture modeling. *Neural Computation, 9*, 1627–1660.