Published in: Proceedings of Fifth International Conference on Image Processing,
Vol I, Chicago, IL, 4-7 October 1998.

Texture Characterization via Joint Statistics of Wavelet Coefficient Magnitudes

Eero P. Simoncelli
Center for Neural Science, and
Courant Institute of Mathematical Sciences
New York University
New York, NY 10003

Javier Portilla
Instituto de Optica
Consejo Superior de Investigaciones Cientificas
Serrano 121
28006 Madrid, SPAIN

We present a parametric statistical characterization of texture images in the context of an overcomplete complex wavelet frame. The characterization consists of the local autocorrelation of the coefficients in each subband, the local autocorrelation of the cofficent magnitudes, and the cross-correlation of coefficient magnitudes at all orientations and adjacent spatial scales. We develop an efficient algorithm for sampling from an implicit probability density conforming to these statistics, and demonstrate its effectiveness in synthesizing artificial and natural texture images.

Many applications in image processing, computer graphics, and computer vision can benefit from a statistical model for visual images. But the dimensionality of the space of images is overwhelmingly large, and thus density inference is very difficult unless one makes several restrictive assumptions. The most common assumptions are locality (the characterization is specified on local spatial neighborhoods), stationarity (the statistics depend only on relative spatial position), and restricted forms of density (e.g., Gaussian). The subclass of images that we commonly call "visual texture" seems most directly amenable to local stationary (i.e. Markov random field) density models, since most textures can be described as a set of repeating structural elements subject to some randomness in location, size, color, orientation, etc.

Julesz pioneered the statistical characterization of textures by proposing that the Nth-order statistics (for some unspecified N) of texture pixels, when considered as samples of a stationary source, might suffice to partition textures into classes that are indistinguishable by a human observer [15]. First and second order statistics (of pixels, or of coefficients in a fixed linear basis) have been used extensively for describing and synthesizing textures [e.g., 8, 12, 10, 5, 11]. In addition, some authors have used adaptive linear representations, in which the basis set is adjusted according to the image statistics, either by tuning the filters to the dominant frequencies of the image [2, 7, 24], or using adaptive filter bandwidths [19]. Second-order statistical approaches, however, are clearly unable to capture significant structures that occur in many textures.

A number of authors have noted that image wavelet subbands have non-Gaussian densities with long tails, and sharp peaks at zero [9, 16, 21, 17, 29]. This is presumably due to the fact that images consist of smooth areas interspersed with occasional edges or other "features". Several recent advances in texture representation are based on the marginal statistics of the responses of a set of filters [8, 25, 13, 29]. In particular, Heeger and Bergen [13] used a fixed overcomplete linear basis to synthesize textures by iteratively alternating between matching the subband histograms, and matching the pixel histogram. This method can reproduce the random features of many natural textures, as well as the dominant scales and orientations, but fails to reproduce extended structural elements (e.g., straight or curved contours) or highly regular patterns. Zhu et. al. [29] used Gibbs sampling to draw from the maximal-entropy density with marginals matching those estimated from subbands of an example image. The subband filters are chosen adaptively to minimize entropy. This technique may be used with both linear and non-linear operators and gives excellent results for a variety of textures, but its computational cost is excessive.

Although the set of marginals along all possible axes are sufficient to uniquely constrain a high-dimensional probability density[1], the marginals of a fixed finite linear basis are often insufficient. In particular, long-range structures (such as straight or curved contours), pseudo-periodic patterns, and second-order textures are not well represented in typical bases. Figure 1 shows an example of two images whose marginal statistics in a wavelet decomposition are identical. Nevertheless, the synthesized image on the right does not capture the texture shown in the left image.

In recent work, we have studied the joint statistics of wavelet coefficients [3, 20]. In particular, we have examined the joint histograms of pairs of coefficient *amplitudes*

[1] This statement is a variant of the Fourier projection-slice theorem used for tomographic reconstruction.

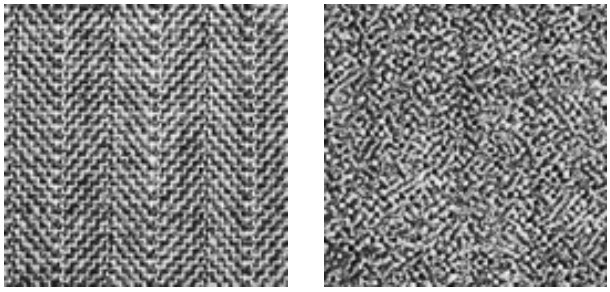

Figure 1. Textures with matching marginal statistics.

at adjacent spatial locations, orientations, and scales of an orthonormal wavelet basis. We find that these are highly correlated, even when the raw coefficients are uncorrelated. There is an intuitive explanation for this: the "features" of real images give rise to large coefficients in local spatial neighborhoods, as well as at adjacent scales and orientations. These relationships (in addition to the marginal statistics) may thus be important for characterizing structural patterns in texture.

The use of joint statistics of rectified subband coefficients for texture analysis appears often in the human vision literature in the form of "second-order" texture analyzers and models [e.g., 1, 27, 4]. In addition to motivation from human visual experiments, there are many textures for which this sort of processing is sensible. The classic example is a set of locally oriented patterns arranged spatially, such as the herringbone fabric shown in figure 1. Recent nonlinear joint models have given impressive synthesis results. Popat and Picard [18] have developed a probability model for densities of local coefficient clusters (including those at different scales), which they have applied to compression, classification, restoration, and synthesis. DeBonet and Viola [6] recently developed a fast heuristic synthesis technique which captures such joint relationships across scale. Their method is very successful at capturing repeating structures.

In this paper, we describe a synthesis technique that is capable of capturing both structural and random aspects of textures. In particular, we characterize textures in terms of a set of statistical measurements on a complex analytic wavelet representation: (1) the local spatial correlation of coefficients within each subband, (2) the local spatial correlation of coefficient magnitudes, (3) the cross-correlation between coefficient magnitudes at adjacent scales and all orientations, and (4) The first few moments of the pixel histogram. Note that we have assumed both spatial locality and stationary, but not Gaussianity. We develop an efficient algorithm for synthesizing images subject to these constraints via iterative projection onto solution sets. We show striking examples of texture synthesis and constrained texture synthesis, demonstrating the power and flexibility of the model.

Wavelet Decomposition

Our texture characterization is based on a fixed overcomplete linear decomposition whose basis functions are spatially localized, oriented, and roughly one octave in bandwidth. We chose to use a "steerable pyramid" [23, 22], since this transform has nice reconstruction properties (specifically, it is a tight frame), in addition to properties of translation- and rotation-invariance. Similar representations have been used by Unser for texture segmentation [26]. Although our algorithm may be implemented using a real-valued pyramid, we have found that the results are improved when we utilize complex analytic (i.e., quadrature pair) filters. That is, the real and imaginary parts of the filters form a Hilbert transform pair.

The transform is implemented using a set of oriented complex analytic filters that are polar-separable when expressed in the Fourier domain:

$$F_{n,k}(r,\theta) = B_n(r)G_k(\theta), \quad n \in [0,N], k \in [0,K-1],$$

where

$$B_n(r) = \begin{cases} \cos\left(\frac{\pi}{2}\log_2\left(\frac{2^n r}{\pi}\right)\right), & \frac{2^n r}{\pi} \in [\frac{1}{2}, 2] \\ 0, & \text{otherwise} \end{cases}$$

$$G_k(\theta) = \begin{cases} \left[\cos(\theta - \frac{\pi k}{K})\right]^{K-1}, & |\theta - \frac{\pi k}{K}| < \frac{\pi}{2} \\ 0, & \text{otherwise,} \end{cases}$$

where r, θ are polar frequency coordinates. Subbands are subsampled by a factor of 2^n along both axes. In addition, one must retain (non-oriented) highpass and lowpass residual bands, which are computed using the following filters:

$$H(r) = \begin{cases} \cos\left(\frac{\pi}{2}\log_2(\frac{r}{\pi})\right), & r \in [\frac{\pi}{2}, \pi] \\ 0 & \text{otherwise} \end{cases}$$

$$L(r) = \begin{cases} \cos\left(\frac{\pi}{2}\log_2(\frac{2^{(N+1)}r}{\pi})\right), & r \in [\frac{\pi}{2^{N+1}}, \frac{\pi}{2^N}] \\ 1 & r < \frac{\pi}{2^{N+1}} \\ 0 & r > \frac{\pi}{2^N}. \end{cases}$$

We used $K = 4$ orientation bands, and $N = 4$ pyramid levels for our examples. The transformation may be inverted by convolving each complex subband with its associated complex-conjugated filter and adding the results. Alternatively, one may reconstruct from either the real or imaginary portions alone.

Texture Parameterization

Our texture description is based on a set of statistical measurements on the image pixels, the raw wavelet coefficients, and the wavelet coefficient magnitudes:

- Image pixel statistics: Mean, variance, skewness, kurtosis, minimum and maximum values (6 parameters).
- Raw coefficient statistics: Central samples of the autocorrelation of each subband ($N \cdot K \cdot \frac{M^2+1}{2}$ parameters), and

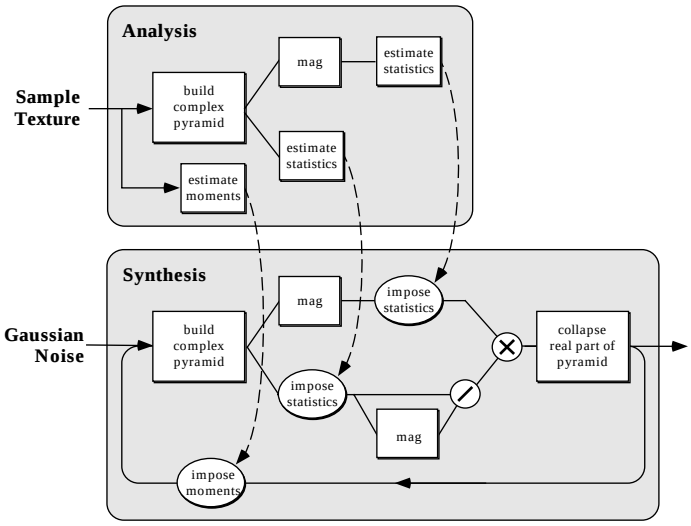

Figure 2. Diagram of texture synthesis algorithm.

mean, variance, minimum and maximum of lowpass and highpass residual bands (8 parameters). These characterize the regularity (linear predictability) of the texture.

• Coefficient magnitude statistics: Central samples of the autocorrelation of each subband (magnitude) ($N \cdot K \cdot \frac{M^2+1}{2}$ parameters), cross-correlation of each subband with other orientations at the same scale ($N \cdot \frac{K(K-1)}{2}$ parameters), and cross-correlation of each subband with other orientations at a coarser scale ($4K^2(N-1)$ parameters). These represent significant structures in images, such as edges, bars, repeated patterns and "second order" structure.

For our texture examples, we have made choices of N=4, K=4 and M=7, resulting in a total of 1034 parameters. This set represents the union of the parameters we found to be necessary for each of our example textures. Comparable results may be achieved for many of the individual textures using substantially fewer parameters.

Synthesis Technique

We now address the problem of sampling from a density that is constrained by the set of statistical measurements taken from an example image. Ideally, we would like to sample from a density that has maximal entropy subject to these constraints, but the computational cost is prohibitive. Instead, we start with an image of Gaussian white noise, and force this image to satisfy the measurements listed above. Our technique uses repeated projections onto the statistical constraint surfaces, and bears a close resemblance to the projection onto convex sets (POCS) approaches that have been used in image restoration [28, 14]. The order of constraint enforcement is illustrated in figure 2. For the adjustment of magnitude statistics, we first adjust the autocorrelation, and then the joint correlation with other orientation bands and the next coarser scale.

Consider first the problem of imposing joint correlations between a set of coefficient magnitudes at different orientations. We use a linear transform to impose correlations, and we choose this transform so as to minimize the expected change in the coefficients. That is, we seek an orthogonal projection onto the constraint surface. Let $\vec{x}$ be a random vector corresponding to a set of coefficient magnitudes at the set of K orientations, but at the same image position and scale. We wish to modify $\vec{x}$ such that the new vector has correlation matrix C_x. We seek matrix M such that $\mathcal{E}\left(M\vec{x}\vec{x}^T M^T\right) = C_x$, where $\mathcal{E}(\cdot)$ is the expectation operator. Let $B_x = \mathcal{E}\left(\vec{x}\vec{x}^T\right)$, and let $\{E_C, E_B, D_C, D_B\}$ be appropriate eigenvector/eigenvalue matrices such that:

$$B_x = E_B D_B D_B^T E_B^T, \qquad C_x = E_C D_C D_C^T E_C^T. \quad (1)$$

Then the complete set of solutions may be specified as:

$$M = E_C D_C O D_B^{-1} E_B^T, \quad (2)$$

where O is any orthonormal matrix. We would like to choose O to minimize the expected change in the vector $\vec{x}$. A reasonable (but not optimal) choice is: $O = E_B^T E_C$. An analogous solution exists for the problem of adjusting $\vec{x}$ to have an appropriate correlation matrix, as well as an appropriate joint correlation with another random vector $\vec{y}$, corresponding to the set of coefficient magnitudes at the next coarser scale. We applied this solution to the orientation bands at each scale of the pyramid, proceeding from coarse to fine.

We also impose local $M \times M$ autocorrelation within each subband (both on the raw coefficients, and their magnitudes). We solve for a $M \times M$ convolution kernel, $h_{\alpha,\beta}$, that most nearly satisfies

$$D_{n,m} = \sum_{\alpha,\beta} h_{\alpha,\beta} C_{n-\alpha,m-\beta},$$

in a least-squares sense. Here, $C_{n,m}$ is the initial autocorrelation image, and $D_{n,m}$ is the desired autocorrelation. We then perform the convolution in the Fourier domain by multiplication with the (positive) square root of the kernel Fourier transform.

Finally, the pixel mean and variance are adjusted by subtracting and dividing the image by the appropriate constants. The skew and kurtosis are imposed (sequentially) by moving in the direction of the gradient (of the moment) until the desired moment is achieved.

Although we make no formal claim of convergence, we have not observed any failures. Our examples typically settled in 50-100 iterations, requiring a few minutes of computation on a typical computer workstation.

Examples

We show some example textures in figure 3. Note that the algorithm successfully captures randomness and repeating structures. It even captures the "visual flavor" of an image of a crowd of people, although this would not typically

be considered a texture. Figure 4 shows two examples of synthesis failures. In the first example, an image of soap bubbles, the synthesized image fails to capture the three-dimensional appearance of the original texture. In the second, a synthetic image of polygonal gray patches, the synthesized image contains bright and dark lines in addition to intensity edges. Figure 5 shows examples of constrained synthesis, in which we extend an image beyond its boundaries. We accomplish this by including an extra projection step in our iterative loop. After collapsing the pyramid, we replace the central pixels of the synthesized image by those of the correct image.

We have presented a characterization of visual texture, and developed an algorithm for synthesizing textures from this characterization. The resulting synthesized images are quite striking, especially when compared against typical second order synthesis results in the literature. The technique is flexible, and may prove useful in the context of other applications such as compression or "super-resolution" enhancement. Nevertheless, the parameterization needs refinement: The failure images (especially the polygonal patches) indicate that some textural aspects are missing. In addition, one would like a more concrete test of the quality of the results. The only true test of texture synthesis system is whether the results appear (to a human observer) to be "the same" as the original example. Thus, the performance of the algorithm should be experimentally verified through subjective measurements.

References

[1] J. R. Bergen and E. H. Adelson. Early vision and texture perception. *Nature*, 333:363–364, 1988.

[2] A. C. Bovik, M. Clark, and W. S. Geisler. Localized measurements of emergent image frequencies by Gabor wavelets. *IEEE Pat. Anal. Mach. Intell.*, 38:691–712, 1992.

[3] R. W. Buccigrossi and E. P. Simoncelli. Progressive wavelet image coding based on a conditional probability model. In *ICASSP*, volume IV, pages 2957–2960, Munich, Germany, April 1997.

[4] C. Chubb and G. Sperling. Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *J. Opt. Soc. Am. A*, 5:1986–2007, 1988.

[5] G. Cross and A. Jain. Markov random field texture models. *IEEE Trans PAMI*, 5:25–39, 1983.

[6] J. De Bonet and P. Viola. A non-parametric multi-scale statistical model for natural images. In *Adv in Neural Info Processing*, volume 9. MIT Press, December 1997.

[7] D. Dunn, W. E. Higgins, and J. Wakeley. Texture segmentation using 2-D Gabor elementary functions. *IEEE Pat. Anal. Mach. Intell.*, 16(2), 1994.

[8] O. D. Faugeras and W. K. Pratt. Decorrelation methods of texture feature extraction. *IEEE Pat. Anal. Mach. Intell.*, 2(4), 1980.

[9] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987.

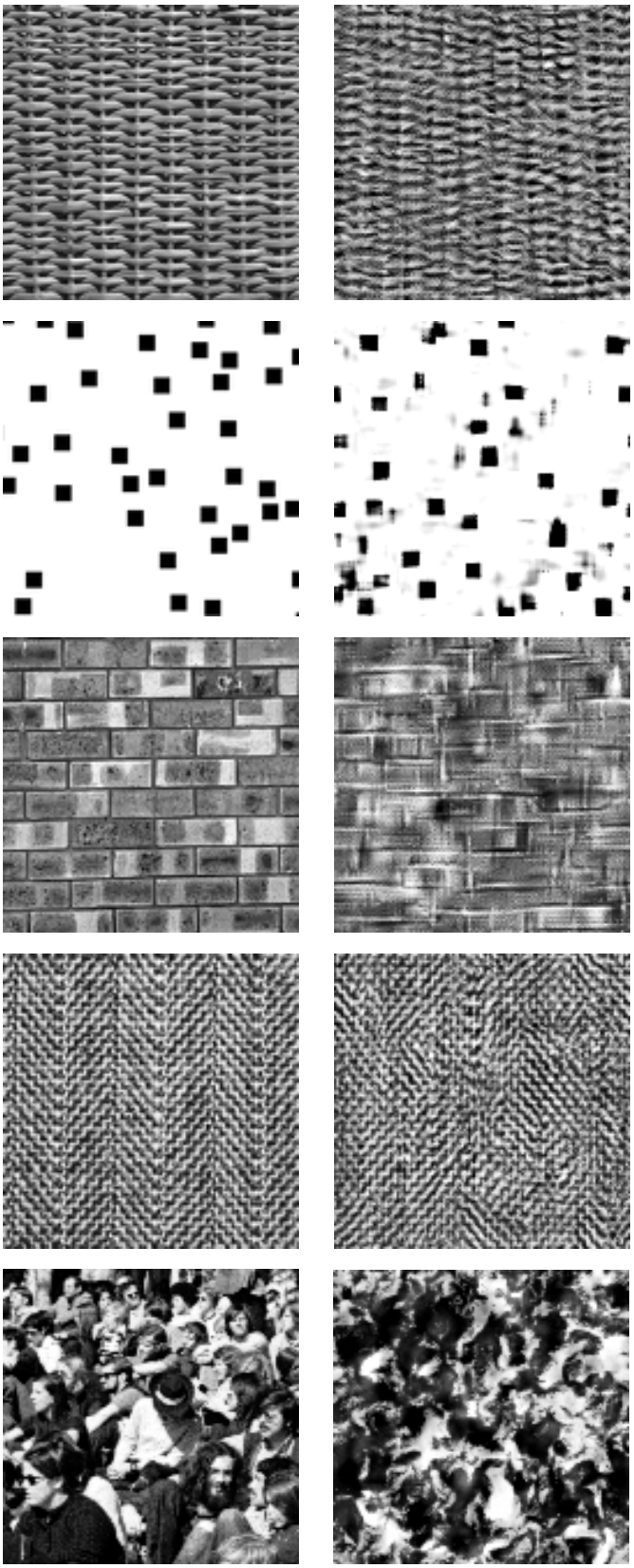

Figure 3. Examples. Left: Original texture. Right: Synthesized texture.

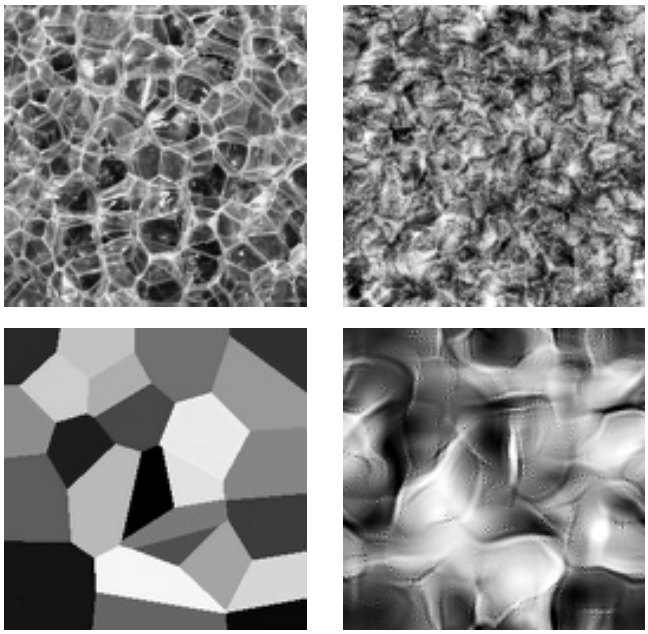

Figure 4. Synthesis failures.

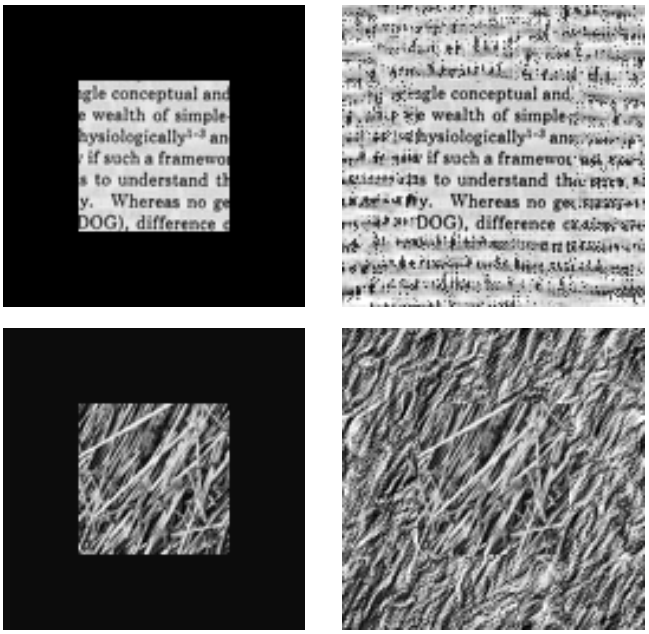

Figure 5. Constrained texture synthesis (extension of an image beyond its boundaries). Left: Central portion of original image. Right: Same image, with black border replaced by synthesized texture.

[10] A. Gagalowicz. A new method for texture fields synthesis: Some applications to the study of human vision. *IEEE Pat. Anal. Mach. Intell.*, 3(5), 1981.

[11] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Pat. Anal. Mach. Intell.*, 6:721–741, 1984.

[12] M. Hassner and J. Sklansky. The use of Markov random fields as models of texture. *Comp. Graphics Image Proc.*, 12:357–370, 1980.

[13] D. Heeger and J. Bergen. Pyramid-based texture analysis/synthesis. In *Proc. ACM SIGGRAPH*, August 1995.

[14] A. N. Hirani and T. Totsuka. Combining frequency and spatial domain information for fast interactive image noise removal. In *ACM SIGGRAPH*, pages 269–276, 1996.

[15] B. Julesz. Visual pattern discrimination. *IRE Trans Info Theory*, IT-8, 1962.

[16] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Pat. Anal. Mach. Intell.*, 11:674–693, July 1989.

[17] B. A. Olshausen and D. J. Field. Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7:333–339, 1996.

[18] K. Popat and R. W. Picard. Cluster-based probability model and its application to image and texture processing. *IEEE Trans Im Proc*, 6(2):268–284, 1997.

[19] J. Portilla, R. Navarro, O. Nestares, and A. Tabernero. Texture synthesis-by-analysis based on a multiscale early-vision model. *Optical Engineering*, 35(8), 1996.

[20] E. P. Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conf on Signals, Systems and Computers*, pages 673–678, Pacific Grove, CA, November 1997.

[21] E. P. Simoncelli and E. H. Adelson. Noise removal via Bayesian wavelet coring. In *Third Int'l Conf on Image Proc*, volume I, pages 379–382, Lausanne, September 1996.

[22] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Second Int'l Conf on Image Proc*, volume III, pages 444–447, Washington, DC, October 1995.

[23] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multi-scale transforms. *IEEE Trans Information Theory*, 38(2):587–607, March 1992. Special Issue on Wavelets.

[24] A. Teuner, O. Pichler, and B. J. Hosticka. Unsupervised texture segmentation of images using tuned matched Gabor filters. *IEEE Trans. Image Proc.*, 4(6), 1995.

[25] J. K. Tugnait. Estimation of linear parameteric models of non-Gaussian discrete random fields with applications to texture synthesis. *IEEE Trans. Image Proc.*, 3(2), 1994.

[26] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Trans. Image Proc.*, 4(11), 1995.

[27] H. Voorhees and T. Poggio. Computing texture boundaries from images. *Nature*, 333:364–367, 1988.

[28] D. C. Youla. Generalized image restoration by the method of alternating orthogonal projections. *IEEE Trans. Circuits and Systems*, 25:694–702, 1978.

[29] S. Zhu and D. Mumford. Prior learning and Gibbs reaction-diffusion. *IEEE Pat. Anal. Mach. Intell.*, 19(11), 1997.