Texture Recognition Using a Non-parametric Multi-Scale Statistical Model

Proceedings IEEE Conf. on Computer Vision and Pattern Recognition, 1998

Jeremy S. De Bonet & Paul Viola
Artificial Intelligence Laboratory
Learning & Vision Group
545 Technology Square Massachusetts Institute of Technology
Cambridge, MA 02139

EMAIL: jsd@ai.mit.edu & viola@ai.mit.edu
HOMEPAGE: http://www.ai.mit.edu/projects/lv

Abstract

We describe a technique for using the joint occurrence of local features at multiple resolutions to measure the similarity between texture images. Though superficially similar to a number of "Gabor" style techniques, which recognize textures through the extraction of multi-scale feature vectors, our approach is derived from an accurate generative model of texture, which is explicitly multi-scale and non-parametric. The resulting recognition procedure is similarly non-parametric, and can model complex non-homogeneous textures. We report results on publicly available texture databases. In addition, experiments indicate that this approach may have sufficient discrimination power to perform target detection in synthetic aperture radar images (SAR).

1 Introduction

The notion of texture is difficult to capture formally. Textures usually can be described informally as the output of some physical process wherein local structure is repeated seemingly at random. Two texture patches are considered to be the same if they appear to have been generated by the same process. Previous approaches to texture classification make strong limiting assumptions about the structure of the physical processes. These assumptions allow one to efficiently generalize from a limited set of example images at the cost of failing to model textures for which the assumptions are invalid. For example, many approaches fail to accurately model textures with long-scale structure that is neither local in space nor local in frequency (e.g. long edges or sharp discontinuities). Other approaches, because they recognize textures using simple local measures, implicitly assume that textures are self-similar at every scale. This excludes even very common textures like bricks, or polka dots which are not self-similar at short scales – the space between the bricks looks very different from the bricks themselves.

Here we present a non-parametric multi-scale statistical texture model which addresses both of these limitations. It models the joint distribution of pixels in an image as a coarse to fine probabilistic process, where decisions made at high frequencies are conditioned on those made at lower frequencies. Multi-resolution conditional distributions can model long-scale structure as conjunctions of low and high frequency features. The conditional distributions are estimated from example images in a flexible non-parametric fashion which can capture complex and multi-modal distributions. In this way the model can capture both the bricks and the space between them in a single distribution.

1.1 Probabilistic Modeling

The probabilistic modeling of images is an endeavor which reaches back to the 60's and 70's (Duda and Hart, 1973). The allure of statistical approaches is that they provide a unified view of model estimation, classification and generation. The perfect statistical model would provide $P(I|C)$, a function of images which returns the likelihood of an image I given that it is an example of class C. With such a distribution, Bayes' Law can be used to estimate

the likelihood of a particular class given some image. Furthermore, $P(I|C)$ provides a principled mechanism for generating or sampling "typical" examples of the class. Finally, most distributions have a set of parameters W, called sufficient statistics, for which $P(I|C, W_C) = P(I|W_C)$. Theses parameters may be estimated from a small set of example images and often generalize to the entire class. Though the advantages of a statistical approach are compelling, to date, a generic, efficient, and unified statistical model for natural images has yet to appear. Nevertheless, many approaches have shown good performance with limited classes of textures.

Perhaps the most influential statistical model for generic images is the Markov random field (MRF) (Geman and Geman, 1984). MRFs can be used to describe a very wide variety of image distributions using information about local regularities between pixels. The parameters of an MRF determine the conditional distribution of a pixel given its neighbors: $P(I(x,y)|N(x,y))$ where x, y range over the pixel indices, $I(x,y)$ is the pixel value, and $N(x,y)$ is the subset of I which neighbors x, y. In their most general form, reasoning with MRFs can require vast amounts computation, often causing convergence to be infeasibly slow.

The complexity of generic MRFs is largely eliminated if the conditional distributions are assumed to be Gaussian. Such a process is called a Gaussian Markov random field (GMRF) (Chellappa and Chatterjee, 1985). The conditional distributions which define a GMRF can be estimated from the correlations between pixels. Since the joint distribution of the pixels in an image under a GMRF model is also Gaussian, the computation of the likelihood of an image given a class is a simple operation. The assumption of Gaussian statistics does however limit the scope of the modeled textures. For example, every marginal distribution or linear function of a GMRF distribution must be unimodal and Gaussian. Therefore GMRFs cannot model image classes which contain discrete structures, like a checkerboard pattern. Moreover, GMRFs cannot distinguish between two images with the same second-order statistics.

Besides MRF's, another influential class of approaches use Gabor-like filters to analyze textures (Santini and Jain, 1996; Fogel and Sagi, 1989). These approaches analyze texture images by convolving with a set of "texture features", frequently derivatives of Gaussians. Using a set of feature detectors $f_1 ... f_n$, n response images can be computed $F_j = f_j \otimes I$, where $\otimes$ denotes convolution. At every image point x, y a feature vector is defined

$$V(x,y) = [F_1(x,y), F_2(x,y), ..., F_n(x,y)] .$$

For some natural textures only few 'typical' feature vectors occur, and these texture classes are often modeled with one, or perhaps a few of these typical feature vectors. Textures are then recognized when these typical features are detected in new images.

The Gabor and GRMF techniques are not as dissimilar as they may at first appear. The first step for Gabor techniques is to convolve the texture image with a set of Gabor filters; the first step in GMRF techniques is to project the image onto the eigenvectors of the covariance matrix. Both these operations tend to *whiten* or decorrelate the pixels of the resulting image (Daugman, 1985; Olshausen and Field, 1996; Duda and Hart, 1973). The main difference between GMRF and Gabor approaches is that GMRFs assume that the texture has a Gaussian distribution once whitened, while the Gabor approaches, which are essentially non-parametric, assume that the distribution of images is better approximated with a few sample points. Not surprisingly, Gabor techniques are more successful than GMRF's precisely when the texture class has a non-Gaussian or multi-modal distribution.

These same principles underlie recent approaches using wavelets to model realistic natural images. These approaches first use a wavelet transform – which is often closely approximated by a set of Gabor filters – to whiten the image. The wavelet coefficients are then modeled with histograms which allows for distributions which are distinctly non-Gaussian. Under these models, images are considered members of a class if the observed histograms of features matches the nominal histogram for the class: $h_j = H(F_j)$, where $H(\cdot)$ is the pixel histogram of a feature image. From the histogram it is possible to compute $p_j^*(v) = \frac{h_j(v)}{\sum_k h_k(w)}$, an estimate for the statistical distribution of pixels in the feature image F_j.

By making the assumption that each of the feature responses is statistically independent – a much stronger, and more limiting assumption – a texture class can be modeled with the distribution:

$$\prod_{j,x,y} p_j^*(F_j(x,y))$$

These insights have been used for noise reduction (Donoho and Johnstone, 1993; Simoncelli and Adelson, 1996), and example driven texture synthesis (Heeger and Bergen, 1995). Essentially, these approaches assume that natural images are caused by

the independent cooccurence of wavelet features. The assumption of independence makes these approaches tractable, but it is also their main weakness. We conjecture that there in fact strong cross-scale dependence between the wavelet coefficients of an image (which is consistent with observations in (Buccigrossi and Simoncelli, 1997))[1].

In many types of natural images there are long-scale features which span both space and frequency (e.g. extended edges). The appearance of such features requires the cooccurence of many localized band-pass features – a cooccurence that is extremely unlikely if we assume that the features are statistically independent. So while the coefficients of the wavelet transform are often uncorrelated, they are not independent. A model which recognizes these statistical relationships and exploits them will be able to more effectively generalize over the image class. We approximate the joint distribution of coefficients as a chain, in which coefficients that occur higher in the wavelet pyramid condition the distribution of coefficients at lower levels (i.e. low frequencies condition the generation of higher frequencies). This allows for the modeling of long-scale relationships within images.

We have introduced the generic model elsewhere (De Bonet, 1997; De Bonet and Viola, 1997). In this paper we will show that it can be used to perform texture recognition on both homogeneous textures and more complex textures, which contain multiple types of non-local visual structures. We will also demonstrate that this technique can be used to solve difficult real problems in the analysis of synthetic aperture radar (SAR) images. Because of the physics of the imaging process, these images challenge classical vision approaches because there is no region of the image which is not textured.

2 The Non-parametric Multi-scale Statistical Model

The multiresolution wavelet transform is most efficiently computed as an iterative convolution using a bank of filters. First a "pyramid" of low frequency downsampled images is created: $G_0 = I$, $G_1 = 2\downarrow(g \otimes G_0)$, and $G_{i+1} = 2\downarrow(g \otimes G_i)$, where $2\downarrow$ downsamples an image by a factor of 2 in each dimension and g is a low pass filter. At each level a series of filter functions are applied: $F_j^i = f_i \otimes G_j$, where the f_i's are compact band-pass filters. Computation of the F_j^i's is a linear transformation that can thought of as a single matrix W. With careful selection of g and f_i this matrix can be constructed so that $W^{-1} = W^\top$ (Mallat, 1989; Simoncelli et al., 1992) [2].

For every pixel in an image define the *parent vector* of that pixel:

$$\vec{V}(x,y) = \Big[F_0^0(x,y), F_0^1(x,y), \ldots, F_0^N(x,y), F_1^0(\lfloor\frac{x}{2}\rfloor,\lfloor\frac{y}{2}\rfloor), F_1^1(\lfloor\frac{x}{2}\rfloor,\lfloor\frac{y}{2}\rfloor), \ldots, F_1^N(\lfloor\frac{x}{2}\rfloor,\lfloor\frac{y}{2}\rfloor), \ldots F_M^0(\lfloor\frac{x}{2^M}\rfloor,\lfloor\frac{y}{2^M}\rfloor), F_M^1(\lfloor\frac{x}{2^M}\rfloor,\lfloor\frac{y}{2^M}\rfloor), \ldots F_M^N(\lfloor\frac{x}{2^M}\rfloor,\lfloor\frac{y}{2^M}\rfloor)\Big] \quad (1)$$

where M is the top level of the pyramid and N is the number of features. Rather than generating each of these coefficients independently, we define a chain across scale. In this chain the generation of the lower levels depend on the higher levels:

$$p(\vec{V}(x,y)) = p(\vec{V}_M(x,y)) \times p(\vec{V}_{M-1}(x,y)|\vec{V}_M(x,y)) \times p(\vec{V}_{M-2}(x,y)|\vec{V}_{M-1}(x,y),\vec{V}_M(x,y)) \times \ldots \times p(\vec{V}_0(x,y)|\vec{V}_1(x,y),\ldots,\vec{V}_{M-1}(x,y),\vec{V}_M(x,y)) \quad (2)$$

where $\vec{V}_l(x,y)$ is the a subset of the elements of $\vec{V}(x,y)$ computed from G_l. Usually we will assume that textures are ergodic processes, i.e. that $p(\vec{V}(x,y))$ is independent of x and y. This assumption is critical if we are to observe enough data to estimate this distribution.

It is important to note that this probabilistic model is not made up of a collection of independent chains, one for each $\vec{V}(x,y)$. Parent vectors for neighboring pixels have substantial overlap as coefficients in the higher pyramid levels (which are lower resolution) are shared by neighboring pixels at lower pyramid levels. Thus, the generation of nearby pixels will be strongly dependent. In a related approach a similar arrangement of generative chains has been termed a Markov tree (Basseville et al., 1992).

In previous work we have demonstrated that textures generated by sampling from the above distribution are of very high quality. The samples are perceptually indistinguishable from the original image, while simultaneously exhibiting many variations which are characteristic of the source texture class.

[1] This assumption of independence has also been challenged in the recent work of (Zhu, Wu and Mumford, 1996; Buccigrossi and Simoncelli, 1997).

[2] Computation of the inverse wavelet transform is algorithmically similar to the computation of the forward wavelet transform.

2.1 Estimating the Conditional Distributions

The additional descriptive power of our model does not come without cost. The conditional distributions that appear in equation (2) must be estimated from observations. We do this directly from the data in a non-parametric fashion. Given a sample of parent vectors $\left\{\vec{S}(x,y)\right\}$ from an example image we estimate the conditional distribution as a ratio of Parzen window density estimators:

$$p(\vec{V}_l(x,y)|\vec{V}^M_{l+1}(x,y)) = \frac{p(\vec{V}^M_l(x,y))}{p(\vec{V}^M_{l+1}(x,y))} \approx \frac{\sum_{x',y'} R(\vec{V}^M_l(x,y), \vec{S}^M_l(x',y'))}{\sum_{x',y'} R(\vec{V}^M_{l+1}(x,y), \vec{S}^M_{l+1}(x',y'))} \quad (3)$$

where $\vec{V}^k_l(x,y)$ is a subset of the parent vector $\vec{V}(x,y)$ that contains information from level l to level k, and $R(\cdot)$ is a function of two vectors that returns maximal values when the vectors are similar and smaller values when the vectors are dissimilar. We have explored various $R(\cdot)$ functions. In the results presented the $R(\cdot)$ function returns a fixed constant $1/z$ if all of the coefficients of the vectors are within some threshold θ and zero otherwise.

2.1.1 Incorporating multiple model images

With little modification information from multiple examples of the target texture can be incorporated into the density estimator. When multiple images are available the Parzen conditional distributions (equation (3)) can be constructed using the additional parent vectors observed from these images.

3 Texture Comparison Using Cross Entropy

Texture recognition requires that we evaluate which of several texture classes is most likely to have generated a test image. One natural approach is to construct one probabilistic model for each texture class. New images are then classified, using equation (2) and Bayes' Law to find the model which makes the image most likely. Experiments have shown that this approach works well, but it is not significantly better than many published techniques.

If we were to take the view that each texture as a single sample from a probability distribution, the simple Bayesian approach would be the optimal strategy. However, this contradicts our above assumption that textures are in fact ergodic – an assumption that is necessary if we hope to learn the cross-scale conditional distributions from a single texture. In doing so, we have assumed that a texture is not a single sample from a distribution, but is actually many samples from a *spatially ergodic* distribution. When this is true the direct Bayesian approach is no longer valid, and a more robust method is needed.

Here is simple example which illustrates why this is so: consider a process which creates random binary images by flipping a single biased coin such that a pixel is colored white with an independent probability of 0.75. Suppose you ware asked to decide which of two images is more likely to have been generated by this process: the first has 75 white pixels and 25 black ones; the second image has 100 white pixels (and 0 black). Intuitively, it seems more likely that the first image was generated by this process. But if we treat each image as single sample from a joint distribution the probabilities indicate otherwise. The probability of generating the first image is much lower than that of the all white image (roughly 3×10^{-25} compared to roughly 3×10^{-13}). Why does this approach fail to pick out the correct image? It does not take into account that the overwhelming majority of samples which are generated by this process will have about 75 white pixels. While it is true that the first image is more likely than the second, it is much less *typical*. Formally, typical images are those whose entropy is approximately the same as the entropy of the class to which they belong. The fact that most images are typical is known as the Asymptotic Equipartition Property (Cover and Thomas, 1991).

A better way to decide which of these two images was generated by the above process is to measure which is more typical. This is done using the Kullback-Liebler (KL) divergence or *cross-entropy*[3]. The cross-entropy is a measure of the difference between two distributions:

$$D(p||q) = \int p(x) log \frac{p(x)}{q(x)} \quad (4)$$

$$= \int p(x) \log p(x) - \int p(x) \log q(x) \quad (5)$$

$$= -H(p) - \int p(x) \log q(x). \quad (6)$$

It can be viewed as the difference between two expected log likelihoods: the log likelihood of samples of $p(x)$ under the distribution $p(x)$, and the log likelihood of samples of $p(x)$ under $q(x)$. For the first image we estimate the probability of white pixels to

[3]The minimum value of cross-entropy is zero, so it is not symmetric, and is therefore, not a metric.

be $p = 0.75$. For the second image set we estimate $p = 1.0$. The true probability of a white pixel is $q = 0.75$. The cross-entropy between $p = 1.0$ and $q = 0.75$ is 0.28, while the cross-entropy between the $p = 0.75$ and $q = 0.75$ is 0.0, a perfect fit.

We use cross-entropy to measure the difference between a texture distribution and the observed distribution in a test image. Given an example image I_E we can estimate a model probability distribution $P_E()$ using the Parzen density estimator described above. From a test image I_T we estimate $P_T()$. By replacing the integration above with a monte-carlo sampling the following approximation results:

$$D\left(I_E||I_T\right) \approx \sum_{x,y} \left\{ \begin{array}{l} log\left[\sum_{x',y'} R(\vec{V}_E(x,y), \vec{V}_E(x',y'))\right] \\ -log\left[\sum_{x',y'} R(\vec{V}_E(x,y), \vec{V}_T(x',y'))\right] \end{array} \right\} \quad (7)$$

where $V_i(x,y)$ is the parent vector from image I_i at location (x,y). In equation (7) we have used the Parzen estimator derived from I_E to both approximate the entropy of I_E and the cross-likelihood of I_T.

The cross-entropy compares the ability of the model distribution to explain the image from which is was derived (I_E) to its ability to explain the test image (I_T).

In some of the experiments below we classify textures into one of two classes. By selecting the model with the lowest cross-entropy we minimize the probability of incorrect classification. But in real applications we may not wish to simply minimize total misclassification. Instead, we may wish to lower either the number of false-positives or false-negatives. In this case a threshold, η, is compared to the differences between the cross-entropies. By manipulating η, false-positive errors are traded off with false-negative errors. The curve generated by varying η, known as the receiver operating characteristics (ROC) curve, will be shown for several of the experiments described below.

4 Experiments

4.1 Standardized tests

On "easy" data sets, such as the MeasTex `Brodatz` texture test suite, performance is slightly higher than other techniques, our approach achieved 100% correct classification compared to 97% achieved by a GMRF approach (Brodatz, 1966; Smith and Lovell, 1995; Chellappa and Chatterjee, 1985). However, it is unlikely that this result is statistically significant. On the MeasTex `lattice` test suite, which contains textures which are highly non-Gaussian our approach achieved 96% correct classification while a Gabor Convolution Energy method achieved 89% and GMRF's achieve only 79% (Fogel and Sagi, 1989). However, because performance is so close to perfect on these relatively easy tasks it is difficult to get an accurate estimate of the relative performance of the current system.

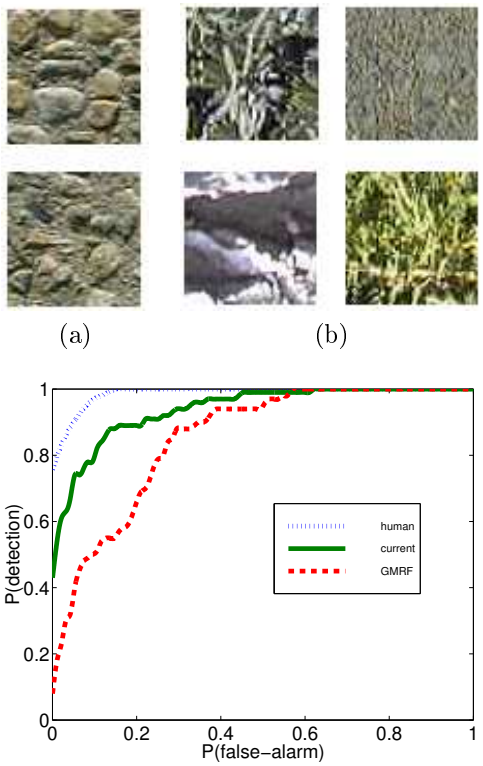


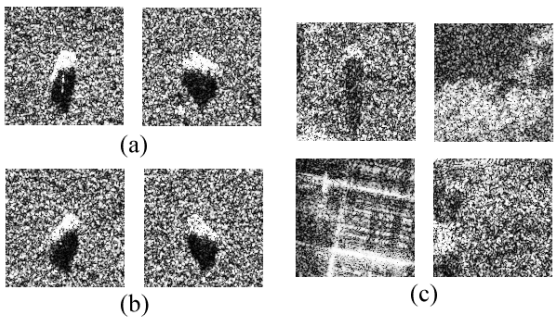
Figure 1: (a) Two images which contain the same texture. (b) Images which contain different textures. (c) ROC classification curves for human observers (top) and for the current model (bottom).

4.2 Natural textures

To measure performance on far more difficult textures, we collected a set of 20 types of natural texture and compared the classification power of this model to that of human observers (humans discriminate textures extremely accurately.) Examples of each texture were acquired by extraction of image patches from a single image of a roughly texturally homogeneous natural scene.

An example of two images of the same texture are shown in Figure 1 (a), and of several textures are shown in (b). In this dataset, some images in the same class look very different, and some in different classes look similar.

Three ROC curves are shown in Figure 1. For each



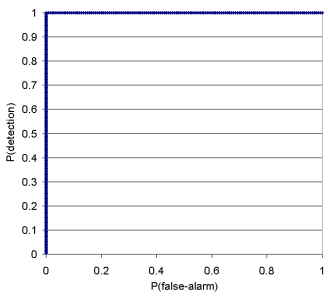


Figure 2: LEFT: 128×128 SAR images of (a) T72 tanks, (b) BMP2 personnel carriers, and (c) clutter images which contain no vehicles. RIGHT: ROC curve for discriminating full resolution (128×128) T72, or BMP2 vehicles from clutter images.

curve, probability of correctly identifying each target texture, $P(detection)$, is plotted against probability of misclassification $P(false-alarm)$. A curve which passes through $(0,1)$ indicates perfect performance, while a diagonal line from $(0,0)$ to $(1,1)$ indicates chance.

The top curve, which achieved the best performance, was generated by the averaged results for 7 human observers. The middle curve was generated by the current model, which does not perform as well as human observers, but is substantially better than GMRFs which generate the bottom curve. The maximum likelihood decision classification yields 93% accuracy for human observers, 87% for current model, and 75% for GMRFs (chance = 5%.)

4.3 Vehicle detection in SAR data

In this section we present the preliminary results of using this model for vehicle detection in synthetic aperture radar images. Using a model constructed from four SAR images of a particular vehicle type, we classify a data set consisting of three types of images: images of the target vehicle, images of a second vehicle, and clutter images which contain no vehicles. The data was acquired from the Model Based Vision Lab MSTAR data set (Model Based Vision Lab,). In each class there were 140 images; two examples of each are shown in Figure 2. The target vehicle was a T72 tank (a), distractor images consisted of images of a BMP2 personnel carrier (b), and clutter images (c).

Using this data set we ask two questions: how well can we discriminate between images containing clutter and those containing a vehicle; and how well can we discriminate between vehicle types, given only models of one, or both of them.

The ROC plot on the right of Figure 2 shows the performance of the current technique at discriminating between clutter images and images containing *either* vehicle type, given two model images of *only* the T72. The point 100% detected versus 0% false-alarm is reached indicating that at some threshold, perfect performance, over this limited data set, is achieved. However, this measure is only preliminary; in real applications the the maximum acceptable P(false-alarm), (Neyman-Pearson criterion) is below the $\pm \sim 1\%$ precision of these experiments.

4.4 SAR vehicle classification system

We now turn to the question of discriminating between the T72 and BMP2 vehicles. Using just examples of one vehicle type, only 67% of the vehicles are correctly classified. By combining the information provided by both pairs of model images performance can be improved.

To determine whether a test image is more likely to have been generated by one model than by the other we take the difference of the cross-entropy measures we can obtain a new measure: $L_{classify} = D(I_{E_1}||I_T) - D(I_{E_2}||I_T)$ To remove biases we compare $L_{classify}$ to a threshold $\eta_{classify}$ using a binary decision rule:

$$L_{classify} \underset{\text{class 2}}{\overset{\text{class 1}}{\gtrless}} \eta_{classify} \tag{8}$$

By varying the value chosen for $\eta_{classify}$ we can maximize the accuracy obtained by the system.

The left plot of Figure 3 is the accuracy achieved by the system as a function of the threshold. In the right, the output of the classification system is shown for the two types of images. The top curve are the sorted responses to the T72, and the bottom to the BMP2[4] From the peak on the left plot, we see that the maximum accuracy obtained is 78%. At this point, we

[4]When performing classification, of course, the system does not have access to these labels, as finding them is the objective!

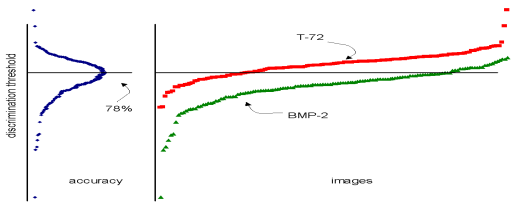


Figure 3: By searching over $\eta_{classify}$ the maximum accuracy achievable by the system can be found.

find the maximum-accuracy threshold value $\eta_{classify}$, which produces a linear discriminator shown by the horizontal line; on the right plot, all images which fall below the line are classified as BMP2, and those above as T72.

5 Discussion

In this paper we have described a new texture model which is based upon conditional coarse to fine scale conditional distributions that are estimated non-parametrically from example images. By capturing the joint cross-resolution characteristics of the wavelet distribution it can model long-scale visual characteristics, such as lines, edges, and complex structures. Further, by building the distribution in a non-parametric fashion, it can be complex and multi-modal, allowing it to simultaneously capture different types of local structure.

In earlier work we showed how textures could be synthesized from such a distribution (De Bonet, 1997). Here we show how to use these distributions to measure the cross-entropy between texture images.

Results on discrimination using simple texture sets, performance is slightly better than standard techniques such as GMRFs. Experiments with more difficult textures indicate performance which, though less than that achievable by human observers, is significantly better than GMRFs. We have tested this model on a real-world problem of the classification of synthetic aperture radar imagery. Though SAR classification rates shown here are only preliminary, more experiments are underway and these results are promising.

References

Basseville, M., Benveniste, A., Chou, K. C., Golden, S. A., Nikoukhah, R., and Willsky, A. S. (1992). Modeling and estimation of multiresolution stochastic processes. *IEEE Transactions on Information Theory*, 38(2):766–784.

Brodatz, P. (1966). *Textures: a Photographic Album for Artists and Designers*. Dover Publications, New York.

Buccigrossi, R. W. and Simoncelli, E. P. (1997). Image compression via joint statistical characterization in the wavelet domain. Technical Report 414, GRASP Laboratory, University of Pennsylvania.

Chellappa, R. and Chatterjee, S. (1985). Classification of textures using gaussian markov random fields. In *Proceedings of the International Joint Conference on Acoustics, Speech and Signal Processing*, volume 33, pages 959–963.

Cover, T. and Thomas, J. (1991). *Elements of Information Theory*. John Wiley & Sons.

Daugman, J. G. (1985). Uncertainty relation for resolution in space spatial frequency and orientation optimized by two–dimensional visual cortical filters. *Journal of the Optical Society of America*, A 2:1160–1169.

De Bonet, J. S. (1997). Multiresolution sampling procedure for analysis and synthesis of texture images. In *Computer Graphics*. ACM SIGGRAPH.

De Bonet, J. S. and Viola, P. (1997). Rosetta: An image database retrieval system. In *Proceedings from Image Understanding Workshop*, San Mateo, CA. Morgan Kaufmann.

Donoho, D. L. and Johnstone, I. M. (1993). Adaptation to unknown smoothness via wavelet shrinkage. Technical report, Stanford University, Department of Statistics. Also Tech. Report 425.

Duda, R. and Hart, P. (1973). *Pattern Classification and Scene Analysis*. John Wiley and Sons.

Fogel, I. and Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, 61:103–113.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Heeger, D. J. and Bergen, J. R. (1995). Pyramid based texture analysis/synthesis. In *Computer Graphics*, pages 229–238. ACM SIGGRAPH.

Mallat, S. G. (1989). Multiresolution approximation and wavelet orthonormal bases of L2(R). *Trans. Amer. Math. Soc.*, 315:69–87.

Model Based Vision Lab. Mbvlab. URL: http://www.mbvlab.wpafb.af.mil.htm.

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

Santini, S. and Jain, R. (1996). Gabor space and the development of preattentive similarity. In *Proceedings of ICPR 96*. International Conference on Pattern Recognition, Vienna.

Simoncelli, E. P. and Adelson, E. H. (1996). Noise removal via bayesian wavelet coring. In *IEEE Third Int'l Conf on Image Processing*, Laussanne Switzerland. IEEE.

Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J. (1992). Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2):587–607.

Smith, G. M. and Lovell, B. C. (1995). Metrics for texture classification algorithms. In *Proceedings DICTA-95 Digital Image Computing: Techniques and Applications*, pages 223–227.

Zhu, S. C., Wu, Y., and Mumford, D. (1996). Filters random fields and maximum entropy(frame): To a unified theory for texture modeling. *To appear in Int'l Journal of Computer Vision*.