REACTION-DIFFUSION EQUATIONS AND LEARNING[1]

Jayant Shah
Mathematics Department
Northeastern University, Boston, MA
email: shah@neu.edu
Tel: 617-373-5660 FAX: 617-373-5658

[1]This work was supported by NIH Grant I-R01-NS34189-04 and NSF Grant DMS-9531293

Abstract

A system of coupled differential equations is formulated which learns priors for modelling "preattentive" textures. Learning is driven by the feature residuals computed from the observed values and the values calculated by the system from a synthesized image which is generated by means of a reaction-diffusion equation.

I. Introduction

An effective approach for modelling many problems in Computer Vision is provided by variational calculus. In this approach, an energy functional is formulated containing a linear combination of terms (also called potentials in the probabilistic context), each of which is a nonlinear transformation of the output of a linear filter such as the gradient or the laplacian of the smoothed image intensity (see §2). Diffusion equations are derived by gradient descent to find solutions minimizing the energy functional. In this paper, this approach is employed for modelling textures. The main problem is that of estimating the potentials and their coefficients in the linear combination and is solved by applying the maximum likelihood principle.

Reaction-diffusion equations were studied by Turing for modelling pattern formation and applied recently by Sherstinsky and Picard [1] to image processing. However, it is not clear how to design these equations. Recently, Zhu and Mumford [2] have introduced a new reaction-diffusion equation for synthesizing textures. It is derived as the gradient flow of an energy functional in which all the nonlinear transformations are obtained from a single transformation involving just three parameters, its center, scale and its rate of growth. Surprisingly, the basic form of this potential is qualitatively the same as some of the ad-hoc potentials already in use such as the Blake-Zisserman [3] and the Perona-Malik potentials [4] and the potentials used by Geman and Graffigne for stochastic modelling of textures [5]. It is also similar to the "edge-strength" function encountered in the segmentation problem [6] and the sigmoid function used in neural nets (see §3). A fundamental requirement is that the potential must exhibit saturation for large values of the filter output, a phenomenon also observed in animal vision. The question is how to estimate the center, the scale and the rate of growth of each potential. The approach of Zhu and Mumford is to use the method of entropy minimax and employs the Gibbs sampler of Geman and Geman [7] in the process to synthesize images. In the present paper, a reaction-diffusion equation is formulated to replace the Gibbs sampler.

A brief description of the entropy minimax method of Zhu, Wu and Mumford [8] is given in §2. The entropy minimax principle is a powerful principle first formulated by Christensen [9],[10],[11] in the context of pattern recognition and statistical inference. The problem is to model the probability distribution on the feature space or the space of images. Since entropy is inversely related to the amount of information in the model, its maximization ensures that the model contains no more information than what is present in the observed sample. Minimization of entropy is used to find the model that captures the maximum amount of information from the sample.

To apply the entropy minimax principle, both Christensen and Zhu et al partition the feature space and approximate the probability distribution by a piecewise constant function. The problem is thus reduced to a problem in parametric statistics, albeit with a greatly increased number of parameters. The principle of maximum entropy is used to estimate the probability distribution corresponding to a given partition. Christensen uses the principle of minimum entropy to find the optimum partitioning of each feature space which consumes most of the computational effort. Zhu, Wu and Mumford propose the principle of minimum entropy to implement feature pursuit so that features are introduced in the order of their importance and usually the first few features suffice to represent the sample adequately. The minimum entropy principle works here even with fairly crude estimates of entropy because it is used for feature pursuit rather than feature selection. If the features are chosen in a wrong order because the entropy estimate is not accurate enough, the only penalty is an increase in the computational burden, possibly by a large amount. (The first experiment described in §5 illustrates this behavior.) Consequently, the main computation in their formulation is in estimating the parameters by the principle of maximum entropy which amounts to maximizing the log-likehood function by gradient ascent. The computationally intensive part is concerned with the synthesis of a sample image by the Gibbs sampler from the current estimates of the parameters during each update.

In this paper the unknown potential of each feature is represented as a linear combination of a large number of fixed potentials obtained by shifting and scaling a smooth "mother" potential (§4). Again, the total number of parameters to be estimated is greatly increased. But since the potentials are now analytic functions, it is possible to use gradient descent instead of the Gibbs sampler for synthesizing images. The result is a system of coupled differential equations, one for updating each of the parameters and one for updating the synthesized image. However, the use of reaction-diffusion equation for synthesizing images produces most likely images rather than typical images. A justification for the use of the maximum likelihood principle in this context is given in §4. Illustrative examples given by Zhu and Mumford in [2] indicate that such equations are capable of synthesizing blobs and stripes. The objective of this paper is to investigate whether it is possible for a such a system to learn and to generate more complex textures. Experiments described in §5 lend support to the feasibility of this approach. A possible explanation for the success of these experiments is that each individual filter in the bank of filters produces a pattern of blobs or stripes and the system combines this collection of blobs and stripes for generating more interesting textures.

A preliminary version of this paper appeared in [12].

II. Entropy Minimax

What follows is a brief summary of the entropy minimax formulation of Zhu, Wu and Mumford. Details may be found in [8]. In practice, the method amounts to an application of the maximum likelihood principle and feature pursuit based on

residuals. Start with the Gibbs form of probability distribution:

$$p(I) = \frac{1}{Z} e^{-U(I)} \tag{1}$$
$$U(I) = \sum_{\alpha} \int_{D} \phi^{(\alpha)} \left(I^{(\alpha)} \right)$$

where I is an image, $I^{(\alpha)}$ is a linear transform of I, $\phi^{(\alpha)}(\xi)$ is a nonlinear function, D is the image domain and Z is the partition function. Of course, $U(I)$ is the corresponding energy functional. The problem is to estimate the potential functions $\phi^{(\alpha)}(\xi)$. Consider one of the $\phi^{(\alpha)}(\xi)$'s and to simplify the notation, denote it by ϕ, omitting the superscript. Divide the domain of ϕ into M bins. Let χ_i denote the characteristic function of the i^{th} bin:

$$\chi_i = \begin{cases} 1 & \text{if } \xi \in i^{th} \text{ bin} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Let χ_i denote the coordinate of the center of the i^{th} bin. Let $|D|$ denote the area of the image domain. Then,

$$\phi(\xi) \approx \sum_{i=1}^{M} \phi(\xi_i) \chi_i(\xi) \tag{3}$$
$$\int_{D} \phi(\xi(\mathbf{x})) d\mathbf{x} \approx \sum_{i=1}^{M} \lambda_i f_i(\xi)$$

where $\lambda_i = |D| \phi(\xi_i)$ and f_i is the normalized frequency:

$$f_i(\xi) = \frac{1}{|D|} \int_{D} \chi_i(\xi(\mathbf{x})) d\mathbf{x} \tag{4}$$
$$\approx \frac{1}{|D|} \# \left\{ \mathbf{x} \in \mathbf{D} : \xi(\mathbf{x}) \in i^{th} \text{ bin} \right\}$$

The approximate probability distribution function

$$p(\Lambda, I) \approx \frac{1}{Z} e^{-\sum_{\alpha} \sum_{i} \lambda_i^{(\alpha)} f_i^{(\alpha)}(I)} \tag{5}$$

now depends only on the parameters $\{\lambda_i^{(\alpha)}\}$. The normalized frequency $f_i^{(\alpha)}(I)$ depends only on $I^{(\alpha)}$ and not on ϕ.

The same form for the probability distribution function is derived in [8] by applying the principle of maximum entropy as follows. The problem is cast as that of maximizing the entropy

$$\int p(I) \log \frac{1}{p(I)} dI \tag{6}$$

with respect to p, subject to the constraints

$$E_p[f_i^{(\alpha)}] = f_{i,obs}^{(\alpha)} \tag{7}$$

where $E_p[\cdot]$ denotes the expected value of the feature with repsect to p and $f_{i,obs}^{(\alpha)}$ is the observed mean value of the feature. Then application of the method of Lagrange multipliers shows that the probability distribution function must have form (1).

The entropy is maximized indirectly with respect to the parameters λ_i^α by applying the maximum likelihood principle as follows. By applying gradient ascent to the log-likelihood function, we get

$$\frac{d\lambda^{(\alpha)}}{dt} = E_{p(\Lambda)}[f_i^{(\alpha)}] - f_{i,obs}^{(\alpha)} \tag{8}$$

where Λ denotes the set of current values of the parameters λ_i^α. Since the log-likelihood function here is convex, there is a unique maximizing Λ and hence a unique Λ satisfying Eq. (7). Therefore the Λ maximizing the likelihood function also solves the constrained entropy maximization problem.

It is not feasible to compute the expected values in Eq. (7). However, they may be estimated using the ergodicity theorem of Geman and Geman [7]: Synthesize a typical image $I_{syn,\Lambda}$ from the distribution $p(\Lambda)$ and use $f_i^{(\alpha)}(I_{syn,\Lambda})$ as an estimate for $E_{p(\Lambda)}[f_i^{(\alpha)}]$. The main computation is now that of $I_{syn,\Lambda}$ and the transforms $I_{syn,\Lambda}^{(\alpha)}$. Computation of the Gibbs sampler is made managable by keeping the number of allowed pixel values and hence the number of local characteristics that must be computed low.

In order to reduce the number of features used and thereby the number of image transforms to be calculated, feature pursuit is used. At each stage of the feature pursuit, optimum value of Λ is found for the current set of chosen features. The next feature chosen should be the one which maximizes the reduction in entropy which is equivalent to minimizing the Kullback-Leibler distance (that is, the relative entropy) from the true probability distribution. However, it is hard to estimate entropies. Since the purpose of the gradient ascent Eq. (8) is to drive down the residual on the right hand side to zero, a heuristic strategy is to select at each step the feature with the largest residual vector. In essence, we use the residuals to indicate the distance of $p(\Lambda)$ from the true probability distribution. Let S denote the set of features already selected. Let Λ_S denote the set of values obtained by setting λ_i^α equal to its maximum likelihood estimate if the feature α belongs to S and zero otherwise. Initially, S is empty and $I_{syn,\emptyset}$ consists of uniform noise. Define

$$d(\beta) = \sum_{i=1}^{M} |f_i^{(\beta)}(I_{syn,\Lambda_S}) - f_{i,obs}^{(\alpha)}| \tag{9}$$

Choose β such that $d(\beta)$ is maximum over the complement of the set S. The use of the L^1–norm in Eq. (9) instead of the L^2–norm or higher norms is recommended by Zhu et al as it gave the best results in their analysis of natural scenes.

III. The Basic Potential

Zhu and Mumford derive their reaction-diffusion equation using potentials of the form:

$$\phi^{(\alpha)}(\xi) = a_\alpha \frac{(|\xi - c_\alpha|/b_\alpha)^{q_\alpha}}{1 + (|\xi - c_\alpha|/b_\alpha)^{q_\alpha}} \tag{10}$$

They arrive at this form by fitting curves to the piecewise constant potentials they found empirically by analyzing a large number of natural scenes. The potential is symmetric about c_α and asymptotically reaches a constant value a_α monotonically. Introduction of the shift parameter c_α is new and necessary because there is no reason why a particular feature should behave symmetrically with respect to the origin. To understand the behavior of such potentials, consider the segmentation functional:

$$E_{MS}(I, B) = \sigma^q \int_{D-B} \|\nabla I\|^q + \gamma^q |B| + \int_D |I - I_{obs}|^q \tag{11}$$

where B is the segmenting curve, $|B|$ is its length and $1 < q < \infty$. First look at the GNC algorithm of Blake and Zisserman ($q = 2$). They replace the first two terms in functional (11) by a function of $\|\nabla I\|$ which is a smoothed version of a truncated parabola and thus has essentially the same shape as given by Eq. (10) (see [13]). The Blake-Zisserman potential is symmetric about the origin. For small values of $\|\nabla I\|$, it behaves like $\sigma^2 \|\nabla I\|^2$ and the behavior is governed by diffusion. Parameter σ plays the role of $\sqrt{a_\alpha}/b_\alpha$. As $\|\nabla I\| \to \infty$, the potential saturates, monotonically approaching the constant value γ^2. The ratio γ/σ is the scale parameter corresponding to b_α in (10). It marks the transition region between diffusion and saturation. The diffusion equation of Perona and Malik may also be derived from a similar potential [13]. The trouble of course is that these approximate functionals have zero infimum and the corresponding gradient descent equations are unstable. Recently, Braides and Dal Maso have regularized the Blake and Zisserman functional by replacing $\|\nabla I\|^2$ in the functional by its average value over a neighborhood and show that the regularized functional, suitably normalized, converges to the segmentation functional (11) as the averaging radius tends to zero [14].

The form (10) is closely related also to the edge-strength function implicit in the approximation of functional (11) formulated by Ambrosio and Tortorelli [15]:

$$E_{AT}(I, v) = \sigma^q \int_D \left[(1 - v)^2 \|\nabla I\|^q + \gamma^q \left(\rho \|\nabla v\|^2 + \frac{v^2}{\rho} \right) + |I - I_{obs}\}^q \right] \tag{12}$$

which is valid also for the case $q = 1$. The edge-strength function v which minimizes $E_{AT}(I, v)$ is a smoothing of

$$\frac{2\rho \|\sigma \nabla I/\gamma\|^q}{1 + 2\rho \|\sigma \nabla I/\gamma\|^q} \tag{13}$$

which is identical in form to the potential (10) with zero shift.

The exponent q determines the type of diffusion which occurs when the gradient is small. The gradient flow for minimizing the q–norm of the gradient is governed by $(q-1)I_{nn} + I_{ss}$ where I_{nn} is the second derivative in the direction of ∇I and I_{ss} in the direction orthogonal to it. When $q = 2$, the diffusion is isotropic, governed by the laplacian. In the limiting case when $q = 1$, smoothing occurs by curvature–dependent evolution of the level curves of I and the gradient flow of functional (12) develops shocks [6]. As $q \to \infty$, the flow in the limit is purely in the direction of the gradient and has been analyzed by Jensen [16]. Potential (10) assumes (double) sigmoidal shape used in neural nets as $q_\alpha \to \infty$ and becomes a purely thresholding function in the limit. The scaling parameter b_α may now be thought of as a threshold for the feature in the sense of neural nets.

Potentials of the same kind in the form of edge-strength functions are also employed in the newly developed faster methods for segmenting images, notably, the method of curve evolution, which is intimately related to the segmentation functionals (11) and (12), (see [6]), and a more recent graph-theoretic method proposed by Shi and Malik [17]. The two methods are in fact closely related; the latter may be interpreted as an approximation of the former [18]. In both approaches, the increased speed of computation is achieved basically by delinking determination of the edge-strength function from boundary detection. The edge-strength function is calculated in advance of boundary detection by means of ad-hoc potentials similar to the basic potential (10). The object boundaries are now determined one closed curve at a time.

A very important consequence of the saturation of potentials is that the functional

$$U(I) = \int_D \sum_\alpha \phi^{(\alpha)}\left(I^{(\alpha)}\right) \tag{14}$$

has not only an infimum, but also a supremum. Hence, the weights a_α may be allowed to be negative, a possibility very effectively exploited by Zhu and Mumford in [2] for synthesizing textures and removing clutter. For an illustration, consider the case of a single potential. Denote the corresponding filter by F so that the integrand in Eq. (14) is $\phi(F * I)$. The gradient flow is given by the equation

$$\frac{\partial I}{\partial t} = F_- * \phi'(F * I) \tag{15}$$

where $F_-(\mathbf{x}) = -\mathbf{F}(-\mathbf{x})$ When $a > 0$, I accentuates the feature values near the center c and the flow is a diffusion flow. If $a < 0$, instead of diffusion we get sharpening of features, a reactive behavior; in the steady state, feature values are near saturation rendering the steady state insensitive to the value of q. Consider for example the case where F smooths the image by convolving it with a Gaussian with standard deviation equal to $3/\sqrt{2}$ and then computes the laplacian of the smoothed image. The initial image consists of uniform noise. Fig.1 shows two examples. The frame on the left shows the result with $a = 1, b = 10, c = -6$ and $q = 2$. Smearing

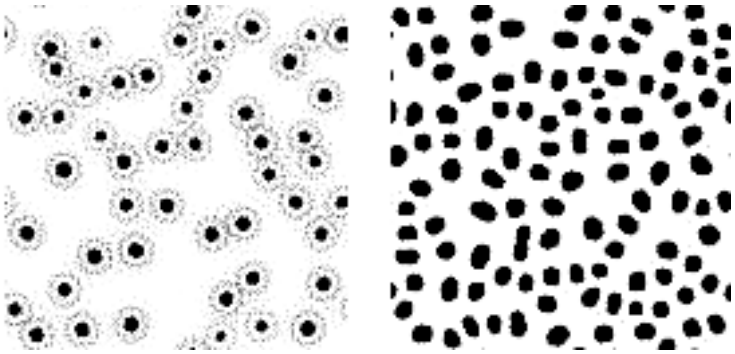

Figure 1: Synthesized textures using Eq. (15). Left frame: Diffusive behavior, Right Frame: Reactive behavior.

due to diffusion is clearly seen. In contrast, the result depicted in the frame on the right corresponds to the case with $a = -1$, $b = 10, c = 6$ and $q = 2$. The absolute values of the laplacian are driven towards satuaration, with positive values dominating since the center is positive, producing a pattern of black blobs. Since the saturation acts like thresholding, the image is nearly piecewise constant and the boundaries of the blobs are sharp.

IV. Learning by Diffusion

As already discussed in the introduction, a way to estimate directly the parameters in the reaction–diffusion equation is to represent each unknown potential as a linear combination of fixed smooth potentials. We create such a set of potentials by shifting and scaling the "mother" potential

$$\psi(\xi) = \frac{|\xi|^q}{1 + |\xi|^q} \tag{16}$$

Function ψ is "universal" in the following sense. Let Ψ denote the set of functions of the form

$$a_o + \sum_{finite} a_i \psi\left(\frac{\xi - c_i}{b_i}\right) \tag{17}$$

Then, for each $-\infty < h < k < \infty$, the restriction of Ψ to the interval $[h, k]$ is dense in the set $C^0[h, k]$ of continuous functions on $[h, k]$ with respect to the uniform norm. This follows from a result of Leshno, Lin, Pinkus and Schoken [19] which asserts that if ψ is continuous almost everwhere and bounded on every compact set, then it is universal if and only if it is not a polynomial. Thus, for the purpose of approximating the potentials $\phi^{(\alpha)}$ by elements in Ψ, any universal function will suffice. The choice (16) is suggested by the results of Zhu and Mumford and by the form of the potentials encountered in segmentation functionals.

Let $R = \max\{\xi_{\max} - \xi_{mean}, \xi_{mean} - \xi_{\min}\}$. Define

$$\psi_{m,k}(\xi) = \psi\left(m\frac{\xi}{R} - k\right) \tag{18}$$

where m, k are integers, $m > 0$ and $|k| \leq m$. For a fixed m, range of ξ is covered by $2m+1$ potentials and the potentials get narrower and narrower as m increases. The situation is analogous to the multiscale representation of a function by wavelets except that our basis potentials are not orthogonal.

The unknown potential $\phi(\xi)$ may be approximated as

$$\phi(\xi) \approx \frac{1}{|D|}\sum_{m,k}\theta_{m,k}\psi_{m,k}(\xi) \tag{19}$$

In view of the universality of ψ, it is expected that with a sufficiently large set of fixed potentials, a single value of q will suffice for approximating the unknown potentials. Hence, q was set equal to 2 in all the experiments. Note that the main effect of q is on the type of diffusion.

The probability distribution is now given by

$$p(I, \Theta) = \frac{1}{Z}e^{-U(I,\Theta)} \tag{20}$$

where

$$U(I,\Theta) = \int_D \sum_\alpha \sum_{m,k} \frac{1}{|D|}\theta_{m,k}^{(\alpha)}\psi_{m,k}^{(\alpha)}\left(I^{(\alpha)}\right) = \sum_\alpha \sum_{m,k}\theta_{m,k}^{(\alpha)}v_{m,k}^{(\alpha)}(I) \tag{21}$$

$$\text{where} \quad v_{m,k}^{(\alpha)}(I) = \frac{1}{|D|}\int_D \psi_{m,k}^{(\alpha)}\left(I^{(\alpha)}\right)$$

Quantities $v_{m,k}^{(\alpha)}$ may be thought of as a set of nonlinear features derived from the original linear filters. The equations for gradient ascent to estimate $\theta_{m,k}^{(\alpha)}$ may be derived as before, but the synthesized image is computed by applying gradient descent to the energy functional $U(I,\Theta)$. We get a system of coupled differential equations[1]:

$$\frac{d\theta_{m,k}^{(\alpha)}}{dt} = v_{m,k}^{(\alpha)}(I_{syn}) - v_{m,k}^{(\alpha)}(I_{obs}) \tag{22}$$

$$\frac{\partial I_{syn}}{\partial t} = \sum_\alpha F_-^{(\alpha)} * \sum_{m,k}\frac{\theta_{m,k}^{(\alpha)}}{|D|}\psi_{m,k}'^{(\alpha)}\left(F^{(\alpha)} * I_{syn}\right)$$

where $F^{(\alpha)} * I_{syn}$ is the linear transform.

[1]It has been pointed out by a reviewer that similar equations have been used by Anderson and Langer [20] using the absolute value function for all α in place of $\psi_{m,k}$ and disregarding the shift and scaling parameters.

These equations are degenerate in the direction of the current vector Θ in the sense that the infimum of $U(I,\Theta)$ with respect to I is invariant with respect to scaling of the Θ and hence Θ must be normalized. The scaling of Θ corresponds to setting the temperature in simulated annealing. As the temperature tends to zero, the annealing method converges to the most likely image. The probability distribution in the limit is uniform over the set of the minima of $U(I,\Theta)$ and zero elsewhere [7]. Thus, as the temperature tends to zero, the probability mass becomes more and more concentrated around the global minima and the most likely images become also typical images. In this sense, the use of the maximum likelihood principle here is consistent.

Normalization of Θ may be done by restricting Θ to the Euclidean length of one. The first equation in (22) may then be replaced by

$$\frac{d\theta_{m,k}^{(\alpha)}}{dt} = \left[v_{m,k}^{(\alpha)}\left(I_{syn}\right) - v_{m,k}^{(\alpha)}\left(I_{obs}\right)\right]^{\perp} \tag{23}$$

where the right hand side is the component of the residuals orthogonal to Θ.

Again, feature pursuit may be used as described in §2.

V. Experiments

The following choices were made for the three experiments described below.

The filter bank $\{F^{(\alpha)}\}$ is a subset of the filter bank used by Zhu, Wu and Mumford, consisting of 73 linear filters:

$$LG(T) = \frac{4}{\pi T^4}\left[\left(\frac{x}{T}\right)^2 + \left(\frac{y}{T}\right)^2 - 1\right] e^{-\left[\left(\frac{x}{T}\right)^2+\left(\frac{y}{T}\right)^2\right]} \tag{24}$$

where $T = \sqrt{2}/2, 1, 2, 3, 4, 5, 6$ and Gabor filters

$$\begin{aligned} G\cos\left(T,\nu\right) &= \frac{1}{\pi T^2} e^{-\frac{1}{2}\left[\left(\frac{2x'}{T}\right)^2+\left(\frac{y'}{T}\right)^2\right]} \cos\frac{2\pi x'}{T} \\ G\sin\left(T,\nu\right) &= \frac{1}{\pi T^2} e^{-\frac{1}{2}\left[\left(\frac{2x'}{T}\right)^2+\left(\frac{y'}{T}\right)^2\right]} \sin\frac{2\pi x'}{T} \end{aligned} \tag{25}$$

where $x' = x\cos\nu + y\sin\nu$, $y' = -x\sin\nu + y\cos\nu$, $T = 2, 4, 6, 8, 10, 12$, and $\nu = 0°, 30°, 60°, 90°, 120°, 150°$. The filters $G\sin(2,\cdot)$ were omitted because $G\sin(2,0)$ is identically zero at the lattice points.

The set of potentials $\psi_{m,k}$ consisted of 35 potentials with $m = 1, 2, 4, 8, 16$ and $|k| \le m/2$. The full range of centers k was not used since the centers of the unknown potentials should be near the mean values of the corresponding features.

The observed images were normalized so that the pixel values ranged from 0 to 255. The uniform noise was sampled from this range of values. The input images and the synthesized images are 128×128 pixels in size except the last input image is 79×142 pixels in size. The range of values of each feature was computed by

combining the range obtained from the observed image with the range obtained from the image consisting of uniform noise.

The size of the time step Δt was empirically chosen as follows. After the first feature was chosen, Θ was set equal to the residuals with respect to that feature, with the length adjusted to one. The time step was then chosen so that the first update vector $\delta\Theta$ had length equal to 0.1 and this value of the time step was maintained during all the subsequent updates. After each new feature was chosen, Θ was updated 10 times. Each time Θ was updated, I_{syn} was computed using the second update equation in (22). The time step Δt for updating I_{syn} was set such that during the first update of the uniform noise, the maximum change in the pixel values was equal to 2. The image was updated 60 times before introducing the next feature. Note that it is not crucial to drive down the residuals to zero before a new feature is introduced. It is sufficient to make the residuals small enough compared to the residuals of the new feature. In order to avoid boundary effects, toroidal topology was assumed.

In the first experiment, the system was given the synthetic image shown in Fig. 1b as the image to be learned. The system selected six filters in the following order: $LG(4)$, $LG(0)$, $Gcos(4, 90)$, $LG(6)$, $Gcos(4, 0)$, $Gcos(8, 90)$. Fig. 2a shows the uniform noise with which the system begins. The synthesized images after 1, 4 and 6 filters were selected are shown in Figs. 2b, 2c and 2d respectively. Interestingly, although the input image was synthesized with a single filter $LG(3)$ using Eq. (15), the system (22) chose $LG(4)$ instead as its first filter. The values of $d(\beta)$ of the two filters are very close with the latter having a slightly higher value.

The second test image shown in Figure 3a depicts animal fur. Figure 3b shows the result after the system had selected 8 filters in the following order: $Gcos(2, 60)$, $Gcos(2, 0)$, $Gcos(6, 150)$, $LG(\sqrt{2}/2)$, $Gsin(12, 0)$, $Gcos(12, 120)$, $Gcos(2, 90)$ and $Gsin(6, 60)$.

The last experiment is shown in Figure 4. Figure 4a is the input image showing cheetah skin. Figure 4b shows the image synthesized by the system after it chose 13 filters in the following order: $LG(1)$, $Gcos(12, 150)$, $Gcos(12, 120)$, $Gsin(12, 60)$ $Gcos(10, 90)$, $Gcos(12, 0)$, $Gsin(12, 30)$, $G(6, 120)$, $LG(4)$, $Gsin(6, 30)$, $Gsin(6, 0)$, $Gsin(6, 60)$ and $Gsin(6, 150)$.

It is not possible to give meaningful estimates of the running times for the system. Since the objective of the work described in this paper was to test the feasibility of this approach, the experiments were carried out without any kind of numerical optimization on an old workstation . Each experiment was allowed to run for many hours to test the numerical stability of the system even though the images looked acceptable earlier on.

References

[1] A. Sherstinsky and R.W. Picard, M-latice: from morphogenesis to image processing", *IEEE Trans. on Image Processing* **5(7)**, July, 1996

[2] S.C. Zhu and D. Mumford, Prior learning and Gibbs reaction-diffusion, *IEEE Trans. PAMI* **19(11)**, Nov. 1997.

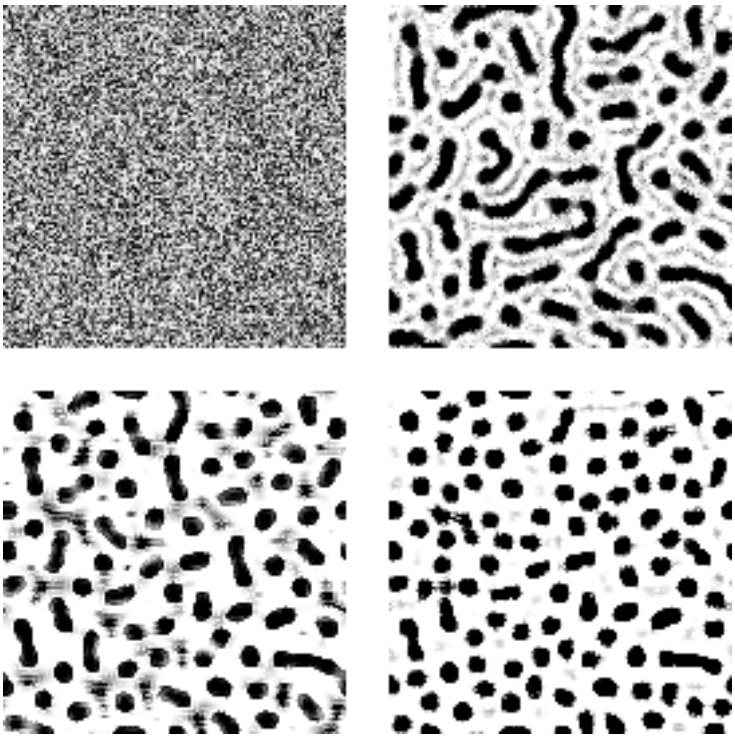

Figure 2: Learned texture. (a) Top left: uniform noise generated by the system initially, (b) Top right: Texture generated after the first filter was chosen, (c) Bottom left: synthesis using 4 filters, (d) Bottom right: synthesis using 6 filters. (Compare with the right frame in Fig. 1.)

[3] A. Blake and A. Zisserman: Visual Reconstruction, MIT Press, 1987.

[4] P. Perona and J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. PAMI* **12(7)**, July, 1990.

[5] S. Geman and C. Graffigne, Markov random field image models and their applications to computer vision, *International Congress of Mathematicians*, 1986, 1496-1517.

[6] J. Shah, A common framework for curve evolution, segmentation and anisotropic diffusion", *IEEE Conf. on Computer Vision and Pattern Recognition,* June, 1996.

[7] S. Geman and D. Geman, Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images", *IEEE Trans. PAMI* **6**, 1984, 721-741.

[8] S.C. Zhu, Y. Wu and D. Mumford, Filters, random fields and maximum entropy (FRAME), *Int'l J. of Computer Vision* **27(2)** , March-April 1998, 1-20.

[9] R. Christensen, A pattern discovery program for analyzing qualitative and quantitative data", *Behavioral Science* **13(5)**, , September, 1968, 423-424.

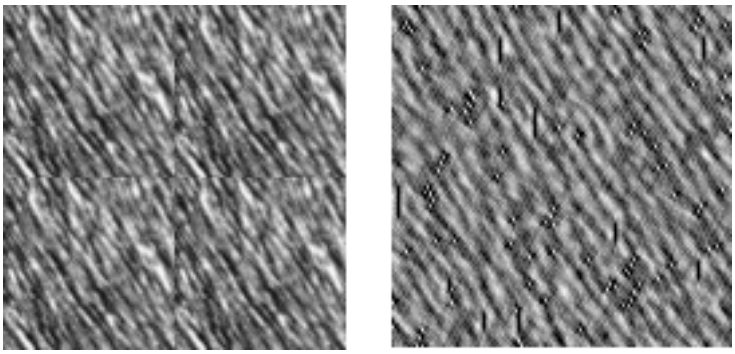

Figure 3: (a) Left: Texture to be learned, (b) Texture synthesized by the system.

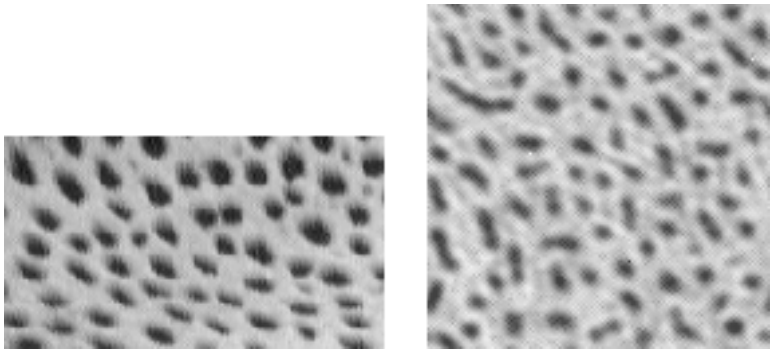

Figure 4: (a) Left: Texture to be learned, (b) Texture synthesized by the system.

[10] R. Christensen: Entropy minimax, a non-bayesian approach to probability estimation from empirical data", *Proc. of the 1973 International Conference on Cybernetics and Society, IEEE Systems, Man and Cybernetics Society*, November 1973, 321-325.

[11] R. Christensen, Entropy minimax sourcebook, general description, **1,** Entropy Limited, Lincoln MA, 1981.

[12] J. Shah, Minimax entropy and learning by diffusion, *IEEE Conf. on Computer Vision and Pattern Recognition,* June, 1998.

[13] J. Shah, Parameter estimation, multiscale representation and algorithms for energy-minimizing segmentations, *IEEE Conf. on Computer Vision and Pattern Recognition*, June, 1990.

[14] A. Braides and G. Dal Maso, Nonlocal approximation of the Mumford-Shah functional, *Calc. Var.* **5**, 1997, 293-322.

[15] L. Ambrosio and V.M. Tortorelli, On the approximation of functionals depending on jumps by quadratic, elliptic functionals, *Boll. Un. Mat. Ital* 1992.

[16] R. Jensen, Uniqueness of Lipschitz extensions: minimizing the sup norm of the gradient, *Arch. Rational Mechanics* **123**, 1993, 51-74.

[17] J. Shi and J. Malik, Normalized cuts and image segmentation, *IEEE Conf. on Computer Vision and Pattern Recognition,* June, 1997.

[18] J. Shah, Riemannian drums, curve evolution and segmentation, in *Scale-Space Theories in Computer Vision, Lecture Notes in Computer Science* **1682**, Ed: M. Nielsen, P. Johansen, O.F. Olsen and J. Weickert, Springer, 1999.

[19] M. Leshno, V.Ya. Lin, A. Pinkus and S. Schoken, Multilayer feed-forward networks with a non-polynomial activation function can approximate any function, *Neural Networks*, 1993.

[20] C.H. Anderson and W.D. Langer, Statistical Models of Image Texture, *Technical Report*, Washington University Medical School, 1997.

VI. Biography

Jayant Shah is a professor of Mathematics at Northeastern University, Boston, MA.