IEEE Conference on Computer Vision and Pattern Recogntion, June, 1998.

Minimax Entropy and Learning by Diffusion

Jayant Shah

Mathematics Department, Northeastern University, Boston, MA 02115 (shah@neu.edu)

Abstract[1]

A system of coupled differential equations is formulated which learns priors for modelling "preattentive" textures. It is derived from an energy functional consisting of a linear combination of a large number of terms corresponding to the features that the system is capable of learning. The system learns the parameters associated with each feature by applying gradient ascent to the log-likelihood function. Updates of each parameter are thus governed by the residual with respect to the corresponding feature. A feature residual is computed from its observed value and the value generated by the system. The latter is calculated from a synthesized sample image which is generated by means of a reaction-diffusion equation obtained by applying gradient descent to the energy functional.

1. Introduction

A very 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 or potentials, 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. Diffusion equations are derived by gradient descent to find solutions minimizing the energy functional. Estimation of the coefficients in the linear combination, that is the parameters of the system, is computationally expensive and so most of the time, these parameters are chosen empirically. A more fundamental problem is that the potentials are usually chosen in an ad hoc manner. The objective of this paper is to formulate a system of differential equations to address both of these problems.

The formulation derived in this paper is illustrated by application to texture modelling by reaction-diffusion equations. This type of equation was first studied by Turing and applied recently by Sherstinsky and Picard [14] to image processing. However, it is not clear how to design these equations in general. Recently, Zhu and Mumford [16] 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. Remarkably, 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 [2] and the Perona-Malik potentials [10] and the potentials used in stochastic modelling of textures [8]. It is also similar to the "edge-strength" function encountered in the segmentation problem [12] 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 such that parameters in the equation may be estimated by the maximum likelihood principle.

A brief description of the entropy minimax method of Zhu, Wu and Mumford [17] is given in §2. The entropy minimax principle is a powerful principle first formulated by Christensen [4,5] 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 use the principle of minimum entropy to implement feature pursuit so that features are introduced in the order of their importance and usu-

[1]This work was partially supported by NIH Grant I-R01–NS34189–01 and NSF Grant DMS-9531293.

ally 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 an enormous amount. 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 from the current estimates of the parameters during each update. However, since the potentials are approximated by non-smooth functions, gradient descent cannot be used to synthesize images and the Gibbs sampler must be used.

The solution proposed in this paper (§4) to the problem of parameter estimation is to represent the unknown potential of each feature as a linear combination of a large number of fixed potentials obtained by shifting and scaling a smooth "mother" potential. The method is analogous to the one proposed by Christensen [6] in which the unknown probability distribution is appoximated by a linear combination of Gaussian potentials. 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 the sample image from a given set of values of the parameters. The result is a system of coupled differential equations, one for updating each of the parameters and one for updating the synthesized image.

2. Entropy Minimax

What follows is a brief summary of the entropy minimax formulation of Zhu, Wu and Mumford. In their setting, the maximum entropy principle is equivalent to the maximum likelihood principle applied to the Gibbs form of probability distribution and since feature pursuit based on feature residuals works well, it is not necessary to invoke the principle of minimum entropy either. Hence, the description given below is in the framework of the maximum likelihood principle and does not use entropy minimax explicitly.

Start with the Gibbs form of probability distribution:

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

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 $\phi^{(\alpha)}$. (For the sake of notational clarity, the weight associated with each feature $I^{(\alpha)}$ is absorbed in $\phi^{(\alpha)}(\xi)$.) Consider one of the $\phi^{(\alpha)}$'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:

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

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

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

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

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

The probability distribution function now depends only on the parameters $\{\lambda_i^{(\alpha)}\}$ which may be estimated by the maximum likelihood principle. By applying gradient ascent to the log-likelihood function, we get

$$\frac{d\lambda_i^{(\alpha)}}{dt} = E_{p(I,\Lambda)}\left[f_i^{(\alpha)}(I)\right] - f_i^{(\alpha)}(I_{obs}) \tag{5}$$

where Λ denotes the set of current values of the parameters $\lambda_i^{(\alpha)}$, $E_{p(I,\Lambda)}[f_i^{(\alpha)}(I)]$ denotes the expected value of $f_i^{(\alpha)}$ and the last term is the observed value of $f_i^{(\alpha)}$. In this paper, we assume that the observed image is sufficiently large so as to provide a good estimate for the observed frequencies.

It is not feasible to compute the expected values in Equation (5). Instead, Geman and Geman suggest the following estimator based on their ergodicity theorem [7]: Synthesize a sample image $I_{syn,\Lambda}$ from the distribution $p(I,\Lambda)$ and use $f_i^{(\alpha)}(I_{syn,\Lambda})$ as an estimate for $E_{p(I,\Lambda)}[f_i^{(\alpha)}(I)]$. The main computation is now that of $I_{syn,\Lambda}$ and the transforms $I^{(\alpha)}_{syn,\Lambda}$. 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. Features are introduced in the order of their importance. Since the purpose of gradient ascent (5) is to drive down the residuals on the right hand side of the equation to zero, a heuristic strategy is to select at each step the feature with the largest residual vector. Let S denote the set of features already selected. Let Λ_S denote the set of values obtained by setting $\lambda_i^{(\alpha)}$ 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

$$d(\beta) = \sum_{i=1}^{M} \left| f_i^{(\beta)}(I_{syn,\Lambda_S}) - f_i^{(\beta)}(I_{obs}) \right| \tag{6}$$

Choose β such that $d(\beta)$ is maximum over the complement of the set S. The use of the L^1–norm in Equation (6) 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.

3. The Basic Potential

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

$$\phi^{(\alpha)}(\xi) \approx a_\alpha \frac{(|\xi - c_\alpha|/b_\alpha)^{p_\alpha}}{1 + (|\xi - c_\alpha|/b_\alpha)^{p_\alpha}} \tag{7}$$

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 the 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:

$$E_{MS}(I, B) = \int_{D-B} ||\nabla I||^p + \gamma^p |B| + \frac{1}{\sigma^p} \int_D |I - I_{obs}|^p \tag{8}$$

where B is the segmenting curve, $|B|$ is its length and $1 < p < \infty$. To see the relation of this functional to the Zhu–Mumford potential, first look at the GNC algorithm of Blake and Zisserman. They replace the last two terms in the functional (8) by a function of $||\nabla I||$ which has essentially the same shape as given by equation (7) with $p = 2$, (see [11]). The diffusion equation of Perona and Malik may also be derived from a similar potential [11]. The trouble of course is that the new 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 $||\nabla I||$ in the functional by $||\nabla I_{ave}||$ where I_{ave} is the image intensity averaged over a neighborhood and show that the regularized functional, suitably normalized, converges to the segmentation functional (8) as the averaging radius tends to zero [3]. Another approximation of functional (8) is due to Ambrosio and Tortorelli [1]:

$$\begin{aligned} E_{AT}(I, v) = \int_D & [(1 - v)^2 ||\nabla I||^p \\ & + \gamma^p \left(\rho ||\nabla v||^2 + \frac{v^2}{\rho} \right) + \frac{1}{\sigma^p} |I - I_{obs}|^p] \end{aligned} \tag{9}$$

The minimizing edge-strength function v is a smoothing of

$$\frac{2\rho ||\nabla I/\gamma||^p}{1 + 2\rho ||\nabla I/\gamma||^p} \tag{10}$$

which is identical in form to the potential (7) with zero shift.

The role of the exponent p_α is also interesting. The gradient flow for minimizing the p–norm of the gradient is governed by $I_{ss} + (p-1)I_{nn}$ where I_{ss} is the second derivative along the level curves of I and I_{nn} is along the gradient direction. In the limiting case when $p = 1$, we get smoothing by curvature–dependent evolution of the level curves of I and the gradient flow of functional (9) develops shocks [12]. As $p \to \infty$, the flow in the limit is purely in the direction of the gradient and has been analyzed by Jensen [9]. Potential (7) assumes sigmoidal shape used in neural nets as $p_\alpha \to \infty$ 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. Each feature space is partitioned into three segments.

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 (8) and (9), (see [12]), and a more recent graph-theoretic method proposed by Shi and Malik [15]. The two methods are in fact closely related; the latter may be interpreted as an approximation of the former [13]. 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 (7). The object boundaries are now determined one closed curve at a time.

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

$$U(I) = \int_D \sum_\alpha \phi^{(\alpha)}\left(I^{(\alpha)}\right) \tag{11}$$

has not only an infimum, but also a supremum. Hence, the weights a_α may be allowed to be negative, a possibility discussed by Geman in [8] and very effectively exploited by Zhu and Mumford in [16] to construct Gibbs reaction-diffusion equations for synthesizing textures and removing clutter. As a simple illustration, consider the case of a single potential. Denote the corresponding filter by F so that the integrand in Equation (11) is $\phi(F * I)$. The gradient flow is given by the equation

$$\frac{\partial I}{\partial t} = F_- * \phi'(F * I) \tag{12}$$

where $F_-(\mathbf{x}) = -F(-\mathbf{x})$. When $a > 0$, I accentuates the feature values near the center c and the flow is a diffusion flow. The kind of diffusion we get depends on the value of p. If $a < 0$, instead of diffusion we get sharpening of features, a reactive behavior; in the steady state, feature values near saturation rendering the steady state insensitive to the value of p. 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. Figure 1 shows two examples. Figure 1a shows the result with $a = 1, b = 10, c = -6$ and $p = 2$. Smearing due to diffusion is clearly seen. In contrast, the result depicted in Figure 1b corresponds to the case with $a = -1, b = 10, c = 6$ and $p = 2$. 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.

4. Learning by Diffusion

The challenge is now to find a way to estimate directly the parameters in the reaction–diffusion equation. As already discussed in the introduction, a way to achieve this 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

$$\psi(\xi) = \frac{|\xi|^p}{1 + |\xi|^p} \tag{13}$$

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

$$\psi_{m,k}(\xi) = \psi\left(m\frac{\xi}{R} - k\right) \tag{14}$$

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 multiscale representation of a function by wavelets except that our basis potentials are not orthogonal.

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

$$\phi(\xi) \approx \frac{1}{|D|}\sum_{m,k}\theta_{m,k}\psi_{m,k}(\xi) \tag{15}$$

It is expected that with a sufficiently large set of fixed potentials, a single value of p will suffice for approximating the unknown potentials. Hence, p was set equal to 2 in all the experiments.

The probability distribution is now given by

$$p(I, \Theta) = \frac{1}{Z}\exp\left(-U(I, \Theta)\right) \tag{16}$$

where

$$\begin{aligned} U(I,\Theta) &= \int_D \sum_\alpha \sum_{m,k} \frac{1}{|D|}\theta_{m,k}^{(\alpha)}\psi_{m,k}^{(\alpha)}\left(I^{(\alpha)}\right) \\ &= \sum_\alpha \sum_{m,k} \theta_{m,k}^{(\alpha)} v_{m,k}^{(\alpha)}(I) \\ \text{where } v_{m,k}^{(\alpha)}(I) &= \frac{1}{|D|}\int_D \psi_{m,k}^{(\alpha)}\left(I^{(\alpha)}\right) \end{aligned} \tag{17}$$

Quantities $v_{m,k}^{(\alpha)}$ may be thought of as a set of nonlinear features derived from the original linear filters. 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:

$$\begin{aligned} \frac{d\theta_{m,k}^{(\alpha)}}{dt} &= v_{m,k}^{(\alpha)}(I_{syn}) - v_{m,k}^{(\alpha)}(I_{obs}) \\ \frac{\partial I_{syn}}{\partial t} &= \sum_\alpha F_-^{(\alpha)} * \sum_{m,k} \frac{\theta_{m,k}^{(\alpha)}}{|D|}\psi_{m,k}'^{(\alpha)}\left(F^{(\alpha)} * I_{syn}\right) \end{aligned} \tag{18}$$

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

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 independent of the size of Θ and hence Θ must be normalized. This may be done by restricting Θ to Euclidean length of one. The first equation in (18) may then be replaced by

$$\frac{d\theta_{m,k}^{(\alpha)}}{dt} = \left[v_{m,k}^{(\alpha)}(I_{syn}) - v_{m,k}^{(\alpha)}(I_{obs})\right]^\perp \tag{19}$$

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

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

5. 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:

(20)
$$LG(T) = \frac{4}{\pi T^4}\left[\left(\frac{x}{T}\right)^2 + \left(\frac{y}{T}\right)^2 - 1\right] e^{-\left[\left(\frac{x}{T}\right)^2 + \left(\frac{y}{T}\right)^2\right]}$$

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

(21)
$$G\cos(T,\nu) = \frac{1}{\pi T^2} e^{-\frac{1}{2T^2}\left(4x'^2 + y'^2\right)} \cos\frac{2\pi x'}{T}$$
$$G\sin(T,\nu) = \frac{1}{\pi T^2} e^{-\frac{1}{2T^2}\left(4x'^2 + y'^2\right)} \sin\frac{2\pi x'}{T}$$

where $x' = x\cos\nu + y\sin\nu$, $y' = -x\sin\nu + y\cos\nu$ and $T = 2, 4, 6, 8, 10, 12$, $\nu = 0^o, 30^o, 60^o, 90^o, 120^o, 150^o$. The filters $G\sin(2,.)$ 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 $|k| \leq m/2$. 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 stability properties of the system (18) are unknown. 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 (18). 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.

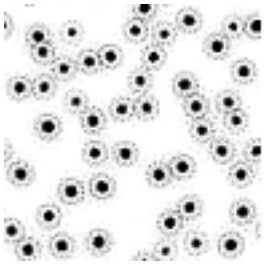

FIGURE 1a

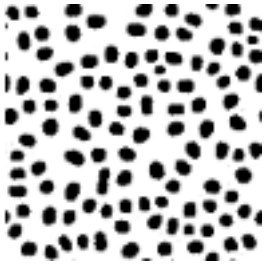

FIGURE 1b

FIGURE 2a

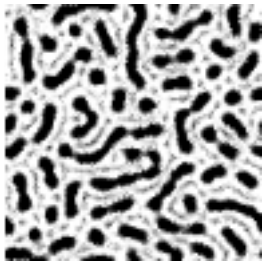

FIGURE 2b

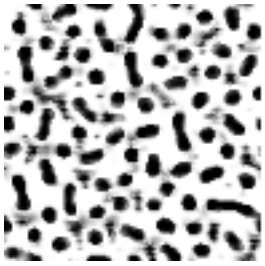

FIGURE 2c

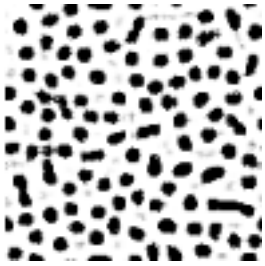

FIGURE 2d

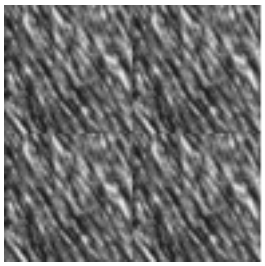

FIGURE 3a

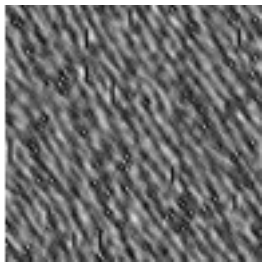

FIGURE 3b

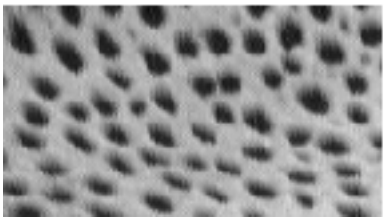

FIGURE 4a

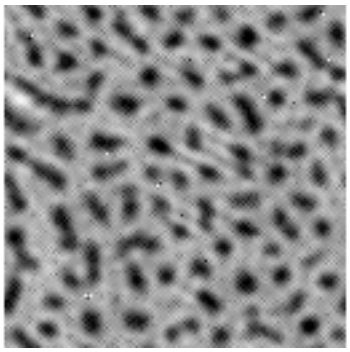

FIGURE 4b

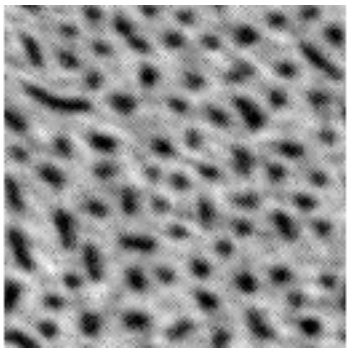

FIGURE 4c

In the first experiment, the system was given the synthetic image shown in Figure 1b as the input image. The system selected six filters in the following order: LG(4), LG(0), Gcos(4,90), LG(6), Gcos(4,0), Gcos(8,90). Figure 2a shows the uniform noise with which the system begins. The synthesized images after 1, 4 and 6 filters were selected are shown in Figures 2b, 2c and 2d respectively. Interestingly, although the input image was synthesized with a single filter $LG(3)$ using reaction-diffusion equation (12), the system (18) chose $LG(4)$ instead as its first filter. 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). (Figures fail to reproduce finer details due to size reduction and 300 dpi printer.)

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 9 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). Figure 4c shows the result after the system chose 4 additional filters, Gsin(6,30), Gsin(6,0), Gsin(6,60) and Gsin(6,150).

6. References

1. L. Ambrosio and V.M. Tortorelli: "On the Approximation of Functionals depending on Jumps by Quadratic, Elliptic Functionals", Boll. Un. Mat. Ital. (1992).
2. A. Blake and A. Zisserman: *Visual Reconstruction*, MIT Press, (1987).
3. A. Braides and G. Dal Maso: "Nonlocal Approximation of the Mumford-Shah Functional", Calc. Var. 5, pp.293-322, (1997).
4. R. Christensen: "A Pattern Discovery Program for Analyzing Qualitative and Quantitative Data", Behavioral Science, v.13, n.5, pp.423-424, (September, 1968).
5. 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, pp.321-325, (November 1973).
6. R. Christensen: *Entropy Minimax Sourcebook, Volume 1: General Description*, Entropy Limited, Lincoln MA, (1981)
7. S. Geman and D. Geman: "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images", IEEE Trans. PAMI 6, pp.721-741, (1984).
8. S. Geman and C. Graffigne: "Markov Random Field Image Models and their Applications to Computer Vision", International Congress of Mathematicians, pp.1496-1517, (1986).
9. R. Jensen: "Uniqueness of Lipschitz Extensions: Minimizing the Sup Norm of the Gradient", Arch. Rational Mechanics, v.123, pp.51-74, (1993).
10. P. Perona and J. Malik: "Scale-Space and Edge Detection using Anisotropic Diffusion", IEEE trans. PAMI, v.12, n.7, (July, 1990).
11. J. Shah: "Parameter Estimation, Multiscale Representation and Algorithms for Energy-Minimizing Segmentations", IEEE Conf. on Computer Vision and Pattern Recognition, (June, 1990).
12. J. Shah: "A Common Framework for Curve Evolution, Segmentation and Anisotropic Diffusion", IEEE Conf. on Computer Vision and Pattern Recognition, (June, 1996).
13. J. Shah: "Riemannian Drums, Curve Evolution and Segmentation", to appear.
14. A. Sherstinsky and R.W. Picard: "M-latice: from morphogenesis to image processing", IEEE Trans. on Image Processing, v.5, n.7, (July, 1996).
15. J. Shi and J. Malik: "Normalized Cuts and Image Segmentation", IEEE Conf. on Computer Vision and Pattern Recognition, (June, 1997).
16. S.C. Zhu and D. Mumford: "GRADE: Gibbs Reaction Diffusion And Diffusion Equation", Sixth International Conference on Computer Vision, (January, 1998).
17. S.C. Zhu, Y. Wu and D. Mumford: "Filters, Random fields And Maximum Entropy (FRAME)", IEEE Conf. on Computer Vision and Pattern Recognition, (June, 1996).