Generalized Perona-Malik Equation Based on Entropy Maximization

Hesan Z. Rafi[1], Hamid Soltanian-Zadeh[1,2]
[1] Control and Intelligent Processing Center of Excellence
Electrical and Computer Engineering Department
University of Tehran, Tehran 14395-515,Iran
Email:hesanrafi@yahoo.com
[2] Medical Image Analysis Lab.
Radiology Department
Henry Ford Health System
Detroit, MI 48202,USA
Email: hamids@rad.hfh.edu

***Abstract*— This paper presents a new multiscale filter based on a relatively new entropy manipulation method. The method inspires using a generalized neighborhood operation, which updates all the pixels inside the neighborhood window in such a way that the spatial entropy inside the neighborhood window is increased.**

The interesting similarity between the new proposed filter and Perona-Malik's filter encourages considering it as a generalization to the Perona-Malik equation.
The new proposed filter is compared with the Perona-Malik filter and the results show that the new method outperforms Perona-Malik method.

I. Introduction

Multiscale description of images has gained a lot of attention in image processing community; therefore a prolific research activity has been conducted in this area during the two past decades. Despite the complicated concept of many published ideas, the basic idea behind multiscale analysis is quite simple; an image is simplified via iterative filtering. At each scale (the number of iterations) a coarser version of input image is generated. These filters are usually formulated as the evolution of the initial image under a suitable linear or nonlinear diffusion process.

These filters should satisfy a very important property "any feature at a coarse level of resolution is required to possess a (not necessarily a unique) cause at a finer level of resolution, although, the reverse needs not to be true."

er:1 This property is called causality and can also be stated from an information theoretic point of view "at each scale we only have loss of information and no new information should be added to the images during evolution" i.e. At each scale the entropy of the image is increased.[7]

The entropy of an image is obtained by assuming an image as a realization of spatially distributed light quanta; as a result the intensity of each point is the unnormalized probability of light.[7]

$$p\ (x) = \frac{I(x)}{\sum_{x \varepsilon X} I(x)} \tag{1}$$

Then using Shannon-Wiener entropy measure we can estimate the entropy of the image.

$$H\ (x) = -\sum_{x \varepsilon X} p\ (x) \log p\ (x) \tag{2}$$

At coarse resolution, the spatial smoothing is high and the spatial probability distribution is close to being uniform, as a result the spatial entropy (uncertainty) is high. Although the mathematical infrastructure of information theory is believed to be the best possible approach to deal with manipulation of data[3], so far entropy maximization has only been used as a constraint to restore images. It has been shown that restoration under this constraint is equivalent to Maximum A Posterior (MAP) estimation under the assumption that site intensities are i.i.d random variables.[2] It worth noting that this method does not produce scale-space images and is a parametric method in the sense that the statistic's of the noise should be modelled properly.

This entropy measure has other applications in multiscale image analysis. For example Sporring [7] applied the Shannon-Wiener entropy in linear multiscale filters to perform scale selection in textures, Wiekert [4] proved monotony of the Shannon-Wiener entropy in linear and nonlinear multiscale filters and used this measure for a uniform sampling of the scale axis with respect to information content. Although this probability perspective corresponds very nicely with the theory of scale-space [7], it does not provide any idea to restore an image directly using this measure.

Fortunately in recent years Xu et al have proposed a new entropy estimator based on Renyi's alternative entropy which not only estimates the entropy directly through data but also provides an easy method to manipulate data to increase or decrease the entropy of the data set.

In this paper we propose a new multiscale filter that generates scale-space images by increasing the entropy of the image at each scale using this entropy estimator. We believe that this insight leads to development of a new paradigm in multiscale image analysis. For example the new multispectral filter which filters multispectral images mutually is a generalization of this idea. This filter generates multispectral scale space images by increasing the entropies of images individually while it increases the mutual information between multispectral images. The organization of this paper is as follows; the entropy estimator will be briefly reviewed in section II. In section III the principles of this new filter are described. The interesting similarity between the new proposed filter and Perona-Malik equation encourages considering it as a generalization to the Perona-Malik equation. This section also illustrates a mechanical model for this filter compares it with the mechanical model of the Perona-Malik equation. Section IV contains experimental results; the results show that the new method outperforms Perona-Malik method.

II. Quadratic Entropy Estimator

In this section we briefly review a relatively new type of entropy estimator recently proposed by Xu et al which not only enables us to estimate the entropy of a realization set of a RV directly, but also suggests how to manipulate these data to increase or decrease the entropy of that set. This entropy estimator is based on the integration of Parzen window pdf estimator, coupled with Renyi's quadratic entropy.

In 1960, Alfered Renyi, famous Hungarian mathematician, proposed a generalized entropy measure, called Renyi's entropy. Renyi's entropy with order a is obtained by the following equation.

$$H_{R_\alpha} = \frac{1}{1-\alpha} \log(\sum_{k=1}^{N} p_k^\alpha) \tag{3}$$

$$\alpha \geq 0, \alpha \neq 1$$

Renyi's entropy measure constitutes a family of entropy measures which are monotonic decreasing functions of parameter α. The limit of this measure as α approaches unity is Shannon's entropy[5]

$$\lim_{\alpha \to 1} H_{R_\alpha} = H_S$$

Thus Shannon's entropy can be regarded as one member of Renyi's entropy family. Further, as Kapur pointed out, this measure is equivalent to Shannon's measure with regards to entropy minimization and maximization.[5]

Similar to the Boltzman-Shannon differential entropy $\int_{-\infty}^{\infty} f_Y(y) log f_Y(y) \mathrm{d}y$ continuous form of the (3) for continuous random variable Y with pdf $f_Y(y)$, is obtained by the following equation

$$H_{R_\alpha} = \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} f_Y(y)^\alpha \mathrm{d}y \tag{4}$$

When $\alpha = 2$, H_{R_2} is called quadratic entropy, because of the dependence of the entropy quantity on quadratic form of probability distribution. This quadratic form

$$H_{R_\alpha} = -\log \int_{-\infty}^{\infty} f_Y(y)^2 \mathrm{d}y \tag{5}$$

gives us more convenience as we will see later.

Now suppose that we have a set of observations; $A = \{a_i | 1 \leq i \leq N\}$ whose elements are the realization of a RV and we want to estimate its entropy. Regardless of the method used to estimate the entropy measure, we have to estimate the pdf of that RV first. Therefore, any pdf estimator (parametric or non-parametric) can be used to achieve this goal, but since the problem should be solved generally it's more reasonable to choose a nonparametric method which doesn't assume any model for the underlying pdf. Among nonparametric methods Parzen window with Gaussian kernel is one of the most popular methods because:
1-Parzen window estimator is consistent for estimating a density from a wide class of densities.
2-The asymptotic rate of convergence for Parzen estimator is optimal for smooth densities.
According to this method one may obtain the pdf through the following formula

$$f_X(x) = \frac{1}{N} \sum_{i=1}^{N} G(x - a_i, \sigma^2) \tag{6}$$

Where G is a Gaussian kernel and δ is the standard deviation. (6) means that each point is occupied by a kernel function, and the whole density is the average of all kernel functions.
Now if we notice the following identity where a and b are constants

$$\int_{-\infty}^{\infty} G(x - a, \sigma_1^2) G(x - b, \sigma_2^2) \mathrm{d}x = G(a - b, \sigma_1^2 + \sigma_2^2) \tag{7}$$

we can simply calculate the integral of the Renyi's quadratic entropy (5)

$$\begin{aligned} V &= \int_{-\infty}^{\infty} f_X(x)^2 \mathrm{d}x \qquad (8) \\ &= \int_{-\infty}^{\infty} [\frac{1}{N} \sum_{i=1}^{N} G(x - a_i, \sigma^2)][\frac{1}{N} \sum_{j=1}^{N} G(x - a_j, \sigma^2)] \mathrm{d}x \\ &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} G(a_i - a_j, 2\ \sigma^2) \end{aligned}$$

Thus Renyi's entropy can be easily calculated by $H_{R_2}(X|A) = -\ logV$

These equations interestingly inspire a physical interpretation.[1] Let us assume that we place physical particles in the locations prescribed by a_i and a_j. The integration of the product of two Gaussian kernels representing some kind of mass density can be regarded as the interaction between particles and a_i, a_j which results in the potential energy $G(a_i - a_j, 2\sigma^2)$. Notice that it is always positive and is inversely proportional to the square of the distance between the particles. We can consider that a potential field exists for each particle in the space where the field strength is defined by the Gaussian kernel; i.e., an exponential decay with the square of the distance. In the real world, physical particles interact with the potential energy proportional to the inverse of the distance between them, but here the potential energy abides by a different law which in fact is determined by the kernel in pdf estimation. V in (8) is the overall potential energy including each pair of data particles. As pointed out previously, these potential energies are related to ”information” and thus are called ”information potentials” (IP). Accordingly, data samples will be called ”information particles” (IPT).

Just like in mechanics, the derivative of the potential energy is a force, in this case an information driven force that moves the data samples in the space of the interactions to change the distribution of the data and thus the entropy of the data. Therefore,

$$\frac{\partial}{\partial a_i} G(a_i - a_j, 2\ \sigma^2) = -\frac{(a_i - a_j)}{2\ \sigma^2} G(a_i - a_j, 2\ \sigma^2) \quad (9)$$

can be regarded as the force that a particle in the position of sample a_j impinges upon a_i and will be called an information force[1]. If all the data samples are free to move in a certain region of the space, then the information forces between each pair of samples will drive all the samples to a state with minimum information potential. If we add all the contributions of the information forces from the ensemble of samples on ai we have the overall effect of the information potential on sample a_i; i.e.

$$\frac{\partial V}{\partial a_i} = \frac{-1}{\sigma^2 N^2} G(a_i - a_j, 2\sigma^2)(a_i - a_j) \quad (10)$$

The Information force is the realization of the interaction among ”information particles.” The entropy will change towards the direction (for each information particle) of the information force. Accordingly, Entropy maximization or minimization could be implemented in a simple and effective way.

III. Generalized Perona-Malik Filter

Neighborhood operations are the central tools for low level image processing. Proper combination of neighboring pixels can perform quite different image processing tasks such as detection of simple local structures (i.e. edges, corners, and lines), motion determination, reconstruction of images taken with indirect imaging techniques (tomography), and restoration. A neighborhood operation takes the values of pixels in the neighborhood of a point, performs some operations with them, and writes the results back on to the point.[6] A 3×3 neighborhood window as shown in Fig.1 is selected to implement the new filter. The entropy of the gray level intensities inside the neighborhood window, according to (8) can be calculated through the following formula

$$H_{R_2}(I) = \frac{1}{9^2} \sum_{i=1}^{9} \sum_{j=1}^{9} G(a_i - a_j, 2\sigma^2) \quad (11)$$

It should be noted that when all the pixels inside the neighborhood window have the same gray level intensities, the entropy of gray level intensity is minimized (equal to zero), but the entropy of the image as we defined in (2) is maximized. According to the causality principle,in scale-space images the entropy of image should be increased at each iteration; thus the entropy of the gray level intensities inside the neighborhood window should be decreased.

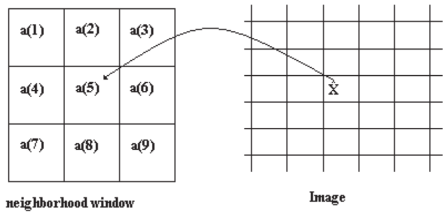

Fig. 1. A common 3×3 neighborhood window, which is selected to implement new generalized filter.

Since in neighborhood operations the central pixel receives the result of operation, the $a(5)$ should be modified in such a way that (11) is decreased. The modification of the central pixel could be obtained using the gradient descent method which is appropriate for this purpose

$$a^{t+1}(5) = a^t(5) + \gamma \frac{\sum_{j=1}^{9} G(a_i - a_j, 2\sigma^2)(a_i - a_j)}{\sum_{i=1}^{9} \sum_{j=1}^{9} G(a_i - a_j, 2\sigma^2)} \quad (12)$$

Where t is the index of scale and γ is a constant coefficient known as learning coefficient and plays an important role in stability of the filter. This equation resembles Perona-Malik equation for an 8 neighborhood window which is

$$a^{t+1}(5) = a^t(5) + \gamma \frac{\sum_{j=1}^{9} G(a_i - a_j, 2\sigma^2)(a_i - a_j)}{\sum_{j=1, j \neq 5}^{9} G(a_i - a_j, 2\sigma^2)} \quad (13)$$

As one can see, only the denominator of (13) and (12) differ slightly. In fact, by proper choosing of learning coefficient and the same variance parameter, these two filters can restore images with approximately same PSNR improvement.

An important question arises here. Why should only the central pixel in the neighborhood window be modified? There is a classical answer for this question ”Neighborhood operations are designed to estimate the true value of a pixel from its surrounding pixels. In fact neighborhood operations make use of pixel dependencies in a small region of image (neighborhood window), and since the pixel dependencies are supposed to be isotropic, with regard to symmetry the central pixel of an odd sized neighborhood window receives the result of operation.[6] But as we want to decrease the entropy of the gray level intensity in a neighborhood window there is no logical reason to restrict modification only to the central pixel. In fact this restriction causes two major drawbacks that could be understood better using the mechanical model described in the pervious section.

If we consider the gray level intensities in a neighborhood window as information particles, there is an interaction force (like gravity) among these particles, which according to the sign of force, moves these particles to one of the two stable states which are the minimum and maximum entropy state. This is just like the Spring-Mass model proposed for Perona-Malik equation [**?**],where the movement of the particles are made non-conservative by stopping them after a small period of time $\triangle t$ and re-starting movement with zero velocity in the next iteration. Now if we only permit the IPT in the central pixel to move, the information forces from the IPTs in the neighborhood window move that IPT in the direction of decreasing entropy until the net force on it becomes zero. There are two major drawbacks with this. First when the central pixel reaches to the equilibrium point further decrease in entropy of the neighborhood window is not possible, clearly this point is not the global minimum of the entropy because the net forces on the other neighborhood pixels are not zero and the system is in equilibrium because these particles are fixed in their positions. The second draw back is even more serious, suppose that we want to decrease the entropy of the neighborhood window with a determined value, clearly the position of the central pixel when all other pixels are fixed is not equal with its position when all pixels are free to move. Which means that the estimated value for the central pixel is not correct. . These problems can also be investigated using diffusion process concept. It is well known that Perona-Malik's filter simulates a discretized diffusion process in a neighborhood window, when we permit only the central pixel to vary; it means that in a neighborhood window the intensity diffused from the central pixel to the neighboring pixels leaves the central pixel, but does not reach to the neighboring pixels or when some intensity is diffused from neighboring pixels the intensity in central pixel increases but the neighboring pixels do not lose that amount of intensity.Of course since the neighborhood window rasters the whole image these filters are mean preserving filters, but this is why Perona-Malik filters with adaptive threshold value are not mean preserving filters. To overcome these drawbacks,we propose to make all the pixels inside the neighborhood window free to move. Considering the neighborhood window in Fig1, the new value for each pixel is obtained through the following equation

$$a^{t+1}(k) = a^{t}(k) + \gamma \frac{\sum_{j=1}^{9} G(a_i - a_j, 2\sigma^2)(a_i - a_j)}{\sum_{i=1}^{9}\sum_{j=1}^{9} G(a_i - a_j, 2\sigma^2)} \quad (14)$$

$$1 \leq k \leq 9$$

There is another advantage over the conventional neighborhood operations. The window needs not to be an odd sized one any more. It worth noting that like before, the image pixels are rastered one by one.

IV. Experimental Results

For experimental evaluation of the new proposed filter with Perona-Malik's filter we have used a 256×256 Lena image Fig.2 This image is degraded by a white Gaussian noise to model the noise of the imaging system Fig.3; as a result the Peak Signal to Noise Ration (PSNR) of the image is reduced to 25db . The PSNR which is used to measure the quality of image, is calculated as

$$PSNR = \frac{127 \times 127}{\frac{1}{256\times256}\sum_i \sum_j (I(i,j) - U(i,j)^{2})} \quad (15)$$

where I(i,j) and U(i,j) are the original image and noisy image samples, respectively. We have used a threshold value equal to 25 for both filters. Fig.4 shows the restored Lena image produced by using the Perona-Malik's filter after 6 iterations and Fig.5 shows the restored Lena image using the new proposed filter after the same number of iterations. It can be seen that the edges and details of images are preserved better in the Fig.5. PSNR of Figs 4 and 5 are 30.52 and 31.85, respectively. Fig.6 shows the curves of PSNR improvement versus iteration number. It is evident from this figure that the conventional method cannot reach to the improvement of the new proposed filter even if we increase the number of iterations. Fig.7 shows the restored Lena image using an unusual 2×2 neighborhood window.The PSNR of this image is 30.1 after 6 iterations.

Fig. 2. A 256×256 original Lena image.

Fig. 3. Original Lena image degraded by white gaussian additive noise.

Fig. 4. Degraded image restored by Perona-Malik method.PSNR=30.52

Fig. 5. Degraded image restored by new proposed filter .PSNR=31.85

V. CONCLUSION

We have proposed a new multiscale filter based on an alternative entropy measure. This filter uses a generalized neighborhood operation,which updates all pixels inside the neighborhood window. Since the new proposed filter is very similar to Perona-Malik filter, it can be considered as a generalization to this filter. One of the advantages of the generalized neighborhood operations is that,there is no necessity any more to use only odd sized windows.

REFERENCES

[1] D. Xu ,*"Energy, Entropy and Information Potential in NeuroComputing"* Ph.D. Dissertation, U. of Florida, 1998.

[2] C. Chang and S. Chatterjee,"Image Restoration Based on Generalized Maximum Entropy Criterion"

[3] C.E. Shannon,*"A mathematical theory of communication"*Ḃell Sys. Tech. J. 27, 1948,pp379-423, 623-653

[4] I. Pollak,*Segmentation and Restoration vai Multiscale Filtering*,IEEE Signal Processing Mag.Sep2002.pp26-36.

[5] J.N. Kapur,*"Measures of Information and Their Applications"*J̇ohn Wiley&Sons. 1994

[6] B. Jahne,*"Digital Image Processing"* 5^{th}revision.Springer 2002

[7] J. Sporring and J. Weickert *"Information Measures In Scale-Spaces"* IEEE Tran. Informatin Theory,Vol 45,No 3,Apr1999,pp1051-1058.

[8] P. Perona and J. Malik *" Scale-Space and Edge Detection Using Anisotropic Diffusion"* IEEE Trans. PAMI Vol12,No7,July 1990,pp 629-639

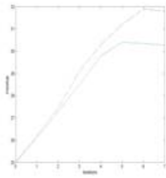

Fig. 6. The PSNR of scale-space images produced by Perona-Malik Equation(line) and new proposed information theoretic based filter versus scale(dashed line).

Fig. 7. Lena degraded image restored by new proposed filter using an unusual 2×2 neighborhood window.PSNR=30.1