Mutual Information Restoration of Multispectral Images Using A Generalized Neighborhood Operation

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- Information theory based techniques for signal and image processing are now considered as a viable alternative to other popular techniques. This paper presents a new multispectral filter based on mutual information maximization to mutually restore multispectral images. For the sake of simplicity we consider only two multispectral images, but the idea can be generalized to more images. Since multispectral images contain analogous information about a scene, as a rule their mutual information is assumed to be maximal; but noise and other independent artifacts decrease their mutual information. A generalized neighborhood operation based on an alternative mutual information measure is used to increase the mutual information between the two neighborhood windows, sliding simultaneously on both images. The main feature of this generalized neighborhood operation is that it updates all pixels inside the neighborhood window. This filter does not assume any specific relation among the gray level intensities of images, and uses both inter-frame and intra-frame information to suppress noise. Application of the proposed method to simulated images shows the outperformance of this method compared with Perona-Malik method which has received much attention in recent years because of its capability in both noise reduction and edge enhancement.

I. INTRODUCTION

Information fusion of multispectral images is a very important issue in remote sensing and medical image analysis. Fusion process submits the multispectral images to some preprocessing steps. Registration (i.e. spatial realignment) of images is the most important phase of preprocessing. As another preprocessing step, images are passed through filters to reduce noise and increase performance of fusion process.

Since filtration is of great importance in image processing, a huge number of filtration methods have been proposed over years. These methods can usually be considered as neighborhood operations on a *single* image [6]. That is to say they are not devised to mutually restore multispectral images. On the other hand, information theory based techniques for signal and image processing are now considered as a viable alternative to other popular techniques.

In this paper, we propose a new multispectral filter based on Mutual Information (MI) maximization. By definition MI is the amount of information that one variable conveys about the other[4]; therefore, since registered multispectral images are informative of one and the same scene, they should have maximum possible MI, but noise and other independent artifacts decrease MI between them. This filter is a generalized neighborhood operation which increases the MI between two sliding windows of the same coordinates. In contrast to conventional neighborhood operations, which update only the central pixel of an odd sized neighborhood window [6], the new generalized neighborhood operation updates all pixels inside the neighborhood window. This filter does not assume any specific relation among the gray level intensities of images, and uses both inter-frame and intra-frame information to suppress noise.

In section II, we briefly review a relatively new type of MI measure which has been proposed by Xu *et al* [1] and enables us to estimate the MI of two small data sets using a closed mathematical formula directly through data. In section III, we describe the new generalized neighborhood operation which filters multispectral images. Finally, in section IV, we present experimental results and compare the new proposed filter with Perona-Malik filter which is widely used to preprocess multispectral images before multispectral segmentation [7]. This filter has received much attention in recent years because of its capability in both noise reduction and edge enhancement [6].

II. ALTERNATIVE MUTUAL INFORMATION (MI) MEASURES

The MI maximization has been successfully used by Viola [2] to register multispectral images. If multispectral images are not registered, their MI decreases. This registration process finds a transform that maximizes the MI between images. As Viola has pointed out, estimating Shannon's MI by pdfs is an inordinately difficult task. So he estimates Shannon's MI using sample mean method which requires a large amount of data. This estimation method is not suitable to estimate the MI between small data sets like two neighborhood windows where we have a small amount of data and we need to know the influence of each sample on the overall MI.

In this section we briefly review a relatively new alternative MI measure proposed by Xu *et al* [1] which enables us to estimate the MI between two small data sets directly through data samples using a closed mathematical formula.

MI measures the relationship between two variables; in other words, MI is the measure of uncertainty removed from one variable when the other is given. Following Shannon [3],[4] the MI between two RV's X_1 and X_2 is defined as

$$I_S(X_1,X_2)= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\log\frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)\times f_{X_2}(x_2)}dx_1dx_2 \quad (1)$$

This measure could also be regarded as the Kullback-Lieber divergence between the joint pdf $f_{X_1,X_2}(x_1, x_2)$ and the factorized marginal pdf's $f_{X_1}(x_1)$, $f_{X_2}(x_2)$.The Kullback-Lieber divergence between two pdfs $f(x)$ and $g(x)$ is defined as

$$D_{K\_L}(f,g)=\int_{-\infty}^{\infty} f(x)\log\frac{f(x)}{g(x)}dx \quad (2)$$

As pointed out by Kapur [6] there is no reason to restrict MI only to this distance measure. Another possible distance measure is based on Cauchy-Schwartz inequality

$$D_{CS}(f,g)=\log\frac{(\int_{-\infty}^{\infty}(f(x)^2dx)(\int_{-\infty}^{\infty}g(x)^2dx)}{(\int_{-\infty}^{\infty}f(x)g(x)dx)^2} \quad (3)$$

Obviously, $D_{CS}(f,g) \geq 0$ with equality iff f(x)=g(x) almost everywhere. Thus with D_{CS} as a measure of distance, we may define Cauchy-Schwartz Quadratic Mutual Information (CS-QMI) between two variables X_1 and X_2 as

$$I_{CS}(X_1,X_2)=D_{CS}(f_{X1,X2}(x_1,x_2),f_{X1}(x_1)f_{X2}(x_2)) \quad (4)$$

Therefore, for the given data set $\{a(i)=(a_1(i),a_2(i))^T \mid 1\leq i\leq N\}$ of a random variable $X = (x_1, x_2)$ to estimate the CS-QMI of x_1 and x_2 we must estimate the joint and the marginal pdfs of x_1 and x_2. Parzen window method with Gaussian kernel is used to estimate these pdfs.

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2)&=\frac{1}{N}\sum_{i=1}^{N}G(x_1-a_1(i),\delta^2)\times G(x_2-a_2(i),\delta^2)\\ f_{X_1}(x_1)&=\frac{1}{N}\sum_{i=1}^{N}G(x_1-a_1(i),\delta^2)\\ f_{X_2}(x_2)&=\frac{1}{N}\sum_{i=1}^{N}G(x_2-a_2(i),\delta^2)\end{aligned} \quad (5)$$

Where $G(x,\delta^2)$ is a Gaussian kernel. Using the following identity where a and b are considered to be constants

$$\int_{-\infty}^{\infty}G(x-a,\delta^2)\times G(x-b,\delta^2)dx = G(a-b,2\delta^2) \quad (6)$$

we have

$$I_{CS}(x_1,x_2)=\log\frac{V_JV_M}{V_C^2} \quad (7)$$

where

$$\begin{aligned} V_J&=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}G(a_1(i)-a_1(j),2\delta^2)G(a_2(i)-a_2(j),2\delta^2)\\ V_M&=\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}G(a_1(i)-a_1(j),2\delta^2)\right)\times\\ &\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}G(a_2(i)-a_2(j),2\delta^2)\right)\\ V_C&=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N}G(a_1(i)-a_1(j),2\delta^2)\right)\times\\ &\left(\frac{1}{N}\sum_{j=1}^{N}G(a_2(i)-a_2(j),2\delta^2)\right)\end{aligned} \quad (8)$$

So we can easily estimate the I_{CS} of two variables using these formulae. For further study about this MI measure and some other generalized information measures and their applications, interested reader is referred to [2].

III. MUTUAL INFORMATION MAXIMIZATION USING GENERALIZED NEIGHBORHOOD OPERATIONS

Neighborhood operations are the central tools for low level image processing. These operations are used to extract certain features from an image. That is why the image resulting from a neighborhood operation is also called a feature image. 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 [6].

The most important characteristic of a neighborhood operation is the size of the neighborhood window. Although neighborhood operations are usually defined on a single image [6], we can generalize neighborhood operations to N images.

For the sake of simplicity, we consider the case $N = 2$ and select a 3×3 neighborhood window. We can consider these two multispectral images as a single image whose pixels are 2×1 vectors. The vectors contain the gray level of multispectral pixels with the same coordinates.

As shown in 'Figure 1', each neighborhood window is actually composed of two neighborhood windows sliding simultaneously over both images. Since they ideally contain the same information, their MI should be high; but noise causes a decrease in MI. The same notations as the pervious section are used to denote the gray level values of the pixels inside the windows; therefore the MI between these two windows can be calculated using (7).

In conventional neighborhood operations, we try 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. But as we want to increase the MI of the gray level in a neighborhood window there is no logical reason to restrict modification only to the central pixel.

Therefore we intend to modify the pixels gray level in order to increase the MI. The gradient descent method can be used to increase MI; therefore, the derivative of MI with respect to each pixel is calculated as follows

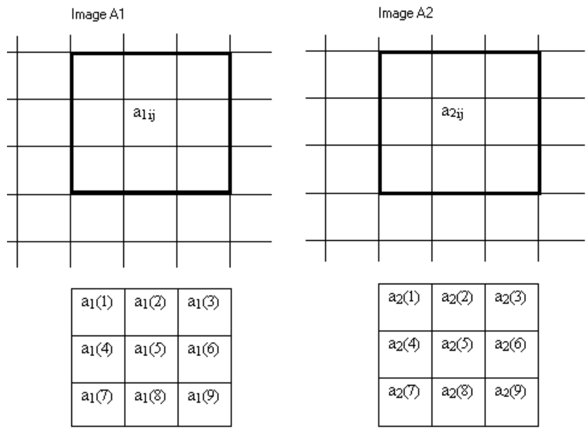


Figure 1. Two neighborhood windows with the same coordinates slide over multispectral images *A*1 and *A*2, and the neighborhood operation increases MI between these neighborhood windows.

$$\frac{\partial I_{CS}}{\partial a_k(p)} = \frac{1}{V_J}\frac{\partial V_J}{\partial a_k(p)} + \frac{1}{V_M}\frac{\partial V_M}{\partial a_k(p)} - \frac{2}{V_C}\frac{\partial V_C}{\partial a_k(p)} \qquad (9)$$
$$k = 1,2$$
$$1 \le p \le 9$$

The new values for the pixels are calculated by adding the initial values and (9) multiplied by a coefficient known as learning coefficient which plays an important role in maximization process. Choosing values greater then 0.5 usually causes the MI oscillate around its initial value.

$$a_{k,new}(p) = a_{k,old}(p) + \gamma \frac{\partial I_{CS}}{\partial a_k(p)} \qquad (10)$$

By scanning the whole image using this neighborhood operation the MI between the feature images increases and, as a result, the two images become mutually restored. Since we have used the gradient descent method to maximize the MI between windows we should iteratively subject the resulting feature images to this filter; in other words this new filter is an iterative filter.

IV. EXPERIMENTAL RESULTS

In this section, we compare the experimental results of the new proposed filter with that of Perona-Malik filter which has received much attention in recent years, and is widely used as a preprocessing step in many multispectral segmentation methods [7],[8].

This filter achieves both noise reduction and edge enhancement through the use of an anisotropic diffusion equation which in essence acts as an unstable inverse diffusion near edges and as a stable linear-heat-equation –like-diffusion in homogeneous regions.

This filter has been implemented using neighborhood operations. Considering the first image A_1 and its corresponding neighborhood window in ‘Figure 1’, the central pixel is updated according to the following formula [8]

$$a_{k,new} = a_{k,old} + \gamma \frac{\sum_{i=1}^{9}(a_k(i)-a_k(5))e^{-\frac{(a_k(i)-a_k(5))^2}{4\delta^2}}}{\sum_{i=1,i\neq 5}^{9} e^{-\frac{(a_k(i)-a_k(5))^2}{4\delta^2}}} \qquad (11)$$

In order to compare these two methods, we have generated two simple simulated multispectral images which represent the same scenes as shown in ‘Figure 2.a’ and ‘Figure 2.b’. The gray level intensities of these two images resemble the gray level intensities of the gray matter, the white matter and the cerebro spinal fluid in T1 and T2 multispectral MR images. To model the noise generated by the imaging system, we have added a white Gaussian noise to the simulated images resulting in the images ‘Figures 3.a and 3.b’. ‘Figures 4.a and 4.b’ represent the subtraction of the simulated images from the noisy images showing the added actual white Gaussian noise.

‘Figures 5.a and 5.b’ show the multispectral noisy images restored by the proposed method and ‘Figures 6.a and 6.b’ represent the images restored by Perona-Malik method. To evaluate the quality of restored images, the noisy images have been subtracted from the restored images; then the cross correlation between the estimated noise and the actual noise is calculated.

‘Figures 7.a and 7.b’ show the estimated noise using the proposed method. The cross correlations between these images and images represented in ‘Figures 4.a and 4.b’ are respectively 0.86 and 0.91. ‘Figures 8.a and 8.b’ show the estimated noise using Perona-Malik filter.
The cross correlations between these images and the images shown in ‘Figures 4.a and 4.b’ are respectively 0.65 and 0.82.

As the results show, the noise estimation using the proposed method is more similar to the actual noise. It is worth noting that the low contrast edges in Perona-Malik’s method are diminished because every image is restored individually but in the proposed method the edges are preserved better because both images are mutually restored. As it can be observed there is a high correlation between the estimated noise (‘Figure 8.a’) and the original image (‘Figure 4.a’) which shows that the image in ‘Figure 8.a’ is not a good estimation.

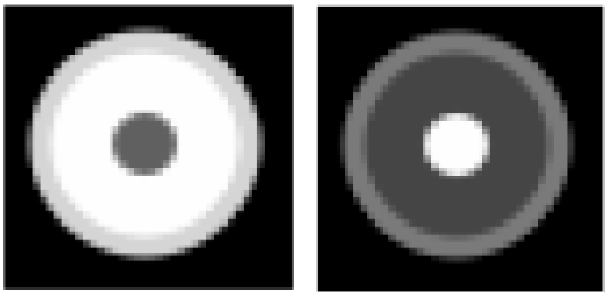


Figure 2.a **Figure 2.b**

Two multispectral images with gray level intensities similar to the gray level intensities of the gray matter, the white matter and the cerebro spinal fluid in multispectral T1 and T2 MRI images.

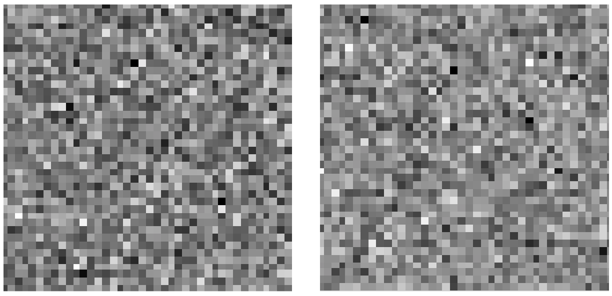


Figure 4.a **Figure 4.b**

Subtraction of noisy images from ideal images which show the noise on ‘Figures 3.a and 3.b’

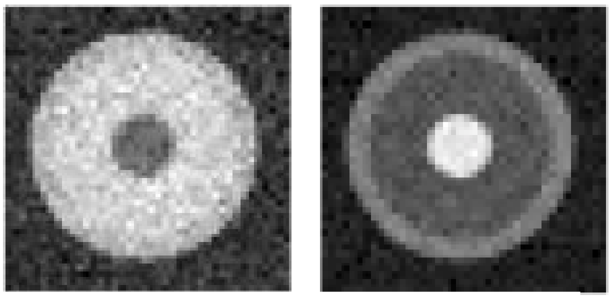


Figure 3.a **Figure 3.b**

Simulated multispectral images generated by imaging system.

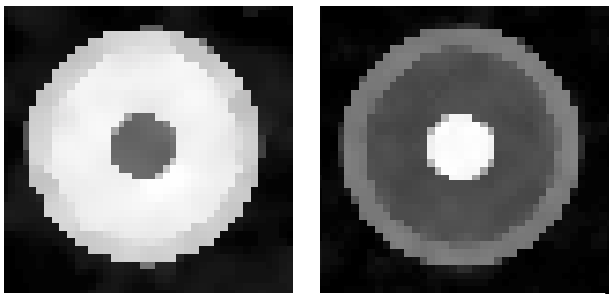


Figure 5.a **Figure 5.b**

Multispectral images restored by proposed method

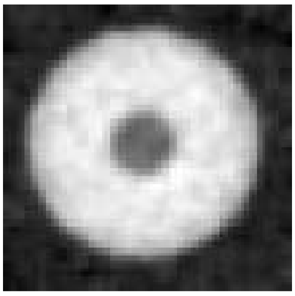
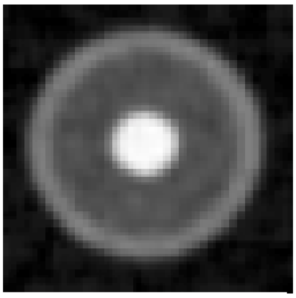
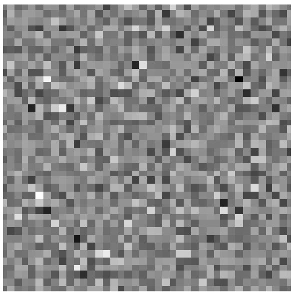
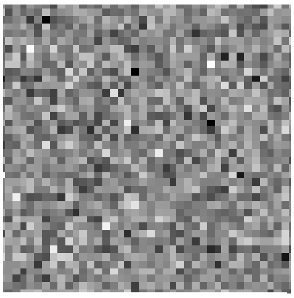



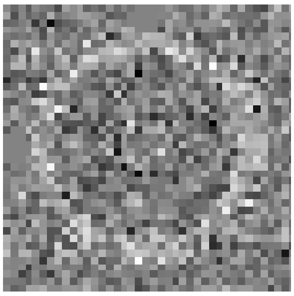
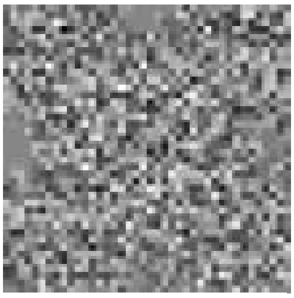
Figure 6.a **Figure 6.b**

Multispectral images restored by Perona-Malik method

Figurer 7.a **Figure 7.b**

Subtraction of noisy images from restored images using the proposed method. The cross correlation between these images and ‘Figures 4.a and 4.b’ are respectively 0.86 and 0.91.

Figurer 8.a **Figure 8.b**

Subtraction of noisy images from restored images using the proposed method. The cross correlation between these images and ‘Figures 4.a and 4.b’ are respectively 0.65 and 0.82.

V. CONCLUSION

We have proposed a new multispectral filter based on an alternative MI measure using a generalized neighborhood operation. One of the main advantages of this filter is using both inter and intra frame information to suppress noise. Two simulated multispectral images have been used to compare the proposed method with Perona-Malik’s filter. The experimental results show that Perona-Malik filter fails to preserve low contrast edges but the proposed method may preserve them since it uses both multispectral images to restore them simultaneously.

REFERENCES

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

[2] P. Viola, N. Schraudolph, T. Sejnowski , *”Empirical entropy manipulation for real-world problems”* Proc. Neural Info. Proc. Sys. (NIPS 8) Conf., 851-857, 1995.

[3] C. E. Shannon, ”*A mathematical theory of communication”* Bell Sys. Tech.J. 27, pp 379-423, 1948.

[4] C. Shannon and W. Weaver ,*”The mathematical theory of communication”*, University of Illinois Press, 1949.

[5] J. N. Kapur, *”Measures of Information and Their Applications”* John Wiley & Sons, San Diego, 1994.

[6] B. Jahne,*”Digital Image Processing”* 5*th* revision. Springer, Berlin, 2002.

[7] I. N. Bankman *”Handbook of Medical Images Processing and analysis”* Academic Press , New York 2000.

[8] P. Perona and J. Malik, *”Scale-Space and Edge Detection Using Anisotropic Diffusion”* IEEE Trans. PAMI vol 12, no 7, pp 629-639, July 1990