Mutual Information Restoration of Multispectral Images

Hesan Z.Rafi
Electrical and Computer
Engineering Department
University of Tehran
Tehran 14399,Iran
Email: hesanrafi@yahoo.com

Hamid Soltanian-Zadeh
Radiology Department
Henry Ford Hospital
Detroit,MI,48202,USA
Email: hamidis@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 could be generalized to more images.**
Since multispectral images contain analogous information about a scene, as a rule they should have maximum mutual information; but noise and other independent artifacts decrease their mutual information measure. A neighborhood operation based on a relatively new generalized mutual information measure is used to increase the mutual information between the two neighborhood windows, sliding simultaneously on both images.
From an image processing point of view, 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.
Applying this new proposed method to phantom 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 defined as 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 neighborhood operation which increases the MI between two sliding windows of the same coordinate.
From an image processing point of view, 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 III we present experimental results and compare the new proposed filter with Perona-Malik filter which has received much attention in recent years because of its capability in both noise reduction and edge enhancement.

II. Generalized Mutual Information (MI) Measures

So far MI maximization has been successfully used by Viola[2] to register multispectral images. If the 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 MI measure proposed by Xu *et al* 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 RVs X_1 and X_2 is defined as

$$I_s(X_1,X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\log\frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)f_{X_2}(x_2)}\mathrm{d}x_1\mathrm{d}x_2 \quad (1)$$

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

$$D_{\text{K-L}}(f,g) = \int_{-\infty}^{\infty} f(x)\log\frac{f(x)}{g(x)}\mathrm{d}x \quad (2)$$

As pointed out by Kapur[5] there is no reason to restrict MI only to this distance measure, another possible distance measure is based on Cauchy-Schwartz inequality

$$D_{\text{C-S}}(f,g) = \log\frac{(\int_{-\infty}^{\infty} f(x)\mathrm{d}x)^2(\int_{-\infty}^{\infty} g(x)\mathrm{d}x)^2}{(\int_{-\infty}^{\infty} f(x)g(x)\mathrm{d}x)^2} \quad (3)$$

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

$$I_{C-S}(X_1,X_2) = D_{C-S}(f_{X_1,X_2}(x_1,x_2), f_{X_1}(x_1)f_{X_2}(x_2)) \quad (4)$$

Therefore for the given data set $\{a(i) = (a_1(i),a_2(i))^T|1 \leq i \leq N\}$ 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 pdf's of x_1 and x_2. Parzen window method with Gaussian kernel is used to estimate these pdfs.

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

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

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

we have

$$I_{C-S}(X_1,X_2) = \log\frac{V_J\ V_M}{V_C^2} \quad (7)$$

where

$$V_J = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{G}(a_1(i)-a_1(j),2\delta^2)\boldsymbol{G}(a_2(i)-a_2(j),2\delta^2)$$

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

$$V_C = \frac{1}{N}\sum_{i=1}^{N}[\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{G}(a_1(i)-a_1(j),2\delta^2)]$$

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 [1].

III. Mutual Information Maximization using 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, lines), motion determination, reconstruction of images taken with indirect imaging techniques (tomography), and restoration [6].

The first characteristic of a neighborhood operation is the size of the neighborhood window. Its second characteristic is the position of the pixel that receives the result of operation. If both of these characteristics are known we will have a neighborhood system. The neighborhood operations are usually defined on a single image[6], but we can generalize neighborhood operations to N images.

For the sake of simplicity, we consider the case N=2. 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. We select the most common neighborhood system which is a 3×3 neighborhood window whose central pixel receives the result of operation.

As shown in Fig.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 could be calculated using (7).

We intend to modify the central pixels $(a_1(5),a_2(5))$in order to increase the MI. The gradient descent method could

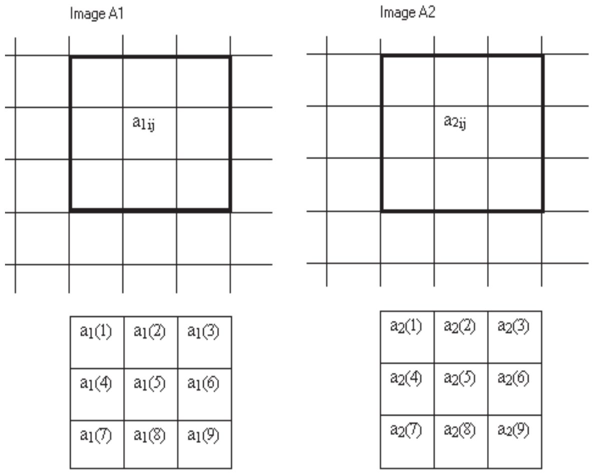

Fig. 1. Two neighborhood windows with the same coordinates slide over multispectral images A_1 and A_2, and the neighborhood operation increases MI between these neighborhood windows

be used to increase MI; therefore, the derivative of MI with respect to central pixels should be calculated as follows

$$\frac{\partial I_{CS}}{\partial a_k(5)} = \frac{\partial V_J}{\partial a_k(5)}\frac{1}{V_J} + \frac{\partial V_M}{\partial a_k(5)}\frac{1}{V_M} - 2\frac{\partial V_C}{\partial a_k(5)}\frac{1}{V_C} \qquad (9)$$

$$k = 1, 2$$

The new values for the central 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 .5 usually causes the MI oscillate around its initial value.

$$a_{k,new}(5) = a_{k,old}(5) + \gamma\frac{\partial I_{CS}}{\partial a_k(5)}$$

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 in many multispectral segmentation methods[7]. This filter achieves both noise reduction and edge enhancement through 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 Fig.1 the central pixel is updated according to following formula [8]

$$\triangle a_1(5) = \gamma\frac{\sum_{i=1}^{9}(a_1(j) - a_1(i))e^{\frac{-(a_1(j)-a_1(i))^2}{k^2}}}{\sum_{i=1}^{9} e^{\frac{-(a_1(j)-a_1(i))^2}{k^2}}} \qquad (10)$$

Although Perona-Malik have used a 4-neighborhood system and have announced that an 8-neighborhood window does not significantly change results,we implemented their filter with an an 8-neighborhood window. In order to compare these two methods, we have generated two simple multispectral phantom images which represent the same scene as shown in Fig2-a and Fig2-b. To model the noise generated by the imaging system, we have added a white Gaussian noise to figures Fig2-a and Fig2-b resulting in figures Fig3-a and Fig3-b.The SNRs of these degraded images are respectively 3.56 and 3.54. Fig4-a and Fig4-b show the restored images using Perona-Malik method after 50 iterations. The SNRs of these restored images are respectively 10.45 and 10.51.As one can tell by images images, in each image the low contrast edge has been severely diminished. Figures Fig5-a and Fig5-b show the results of the implementation of the new proposed filter after the same number of iterations. The SNRs of Fig5-a and Fig5-b are respectively 12.03 and 12.27.

It could be seen from the images that our method has preserved the edges better than Perona-Malik method. Of

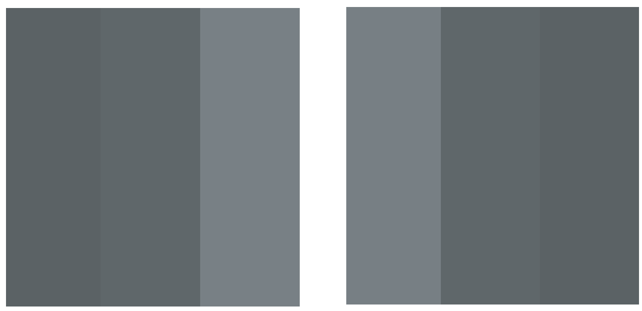

Fig.2-a Fig.2-b

Fig. 2. Tow multispectral phantom images, representing the same scene.

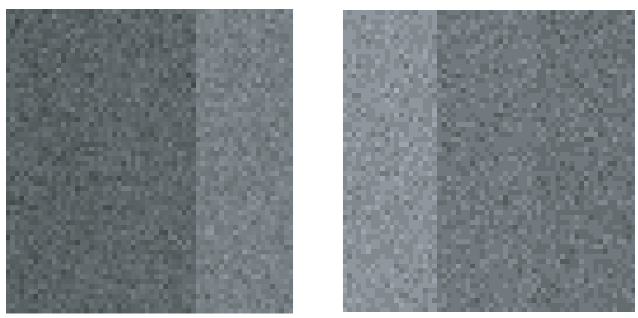

Fig3-a Fig3-b

Fig. 3. Multispectral images degraded by noise

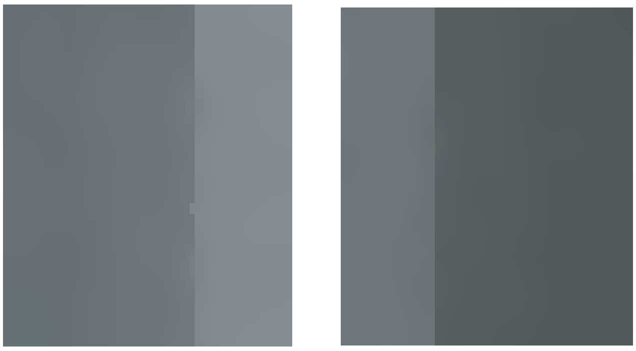

Fig4-a Fig4-b

Fig. 4. Mutispectral degraded images, restored by Perona-Malik method.

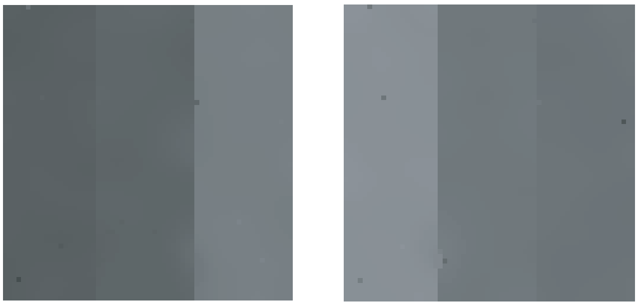

Fig5-a Fig5-b

Fig. 5. Multispectral degraded images, restored by the new proposed method.

course it should be noted that the phantoms are generated in such a way that the edge with low contrast in one image has a high contrast in the other image one and vice versa. Since Perona-Malik method restores images independently it fails to preserve these low contrast edges but the proposed method uses both mutispectral images to restore them simultaneously.

V. Conclusion

We have proposed a new multispectral filter based on a generalized MI measure. One of the main advantages of this filter is using both inter and intra frame information to suppress noise. The experimental results also show that the SNR Improvement of the restored images using the proposed method is higher than that of Perona-Malik method.

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, 1948,pp379-423, 623-653

[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. 1994

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

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

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