http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 6 (8): 1224-1230, 2007
ISSN 1812-5638

A Comparative Analysis of Feature Based Image Fusion Methods

[1]M. Sasikala and [2]N. Kumaravel
[1]Department of Instrumentation Engineering,
Madras Institute of Technology, Anna University, Chennai, India
[2]Department of Electronics and Communication Engineering, Anna University, Chennai, India

Abstract: The objective of image fusion is to combine the source images of the same scene to form one composite image that contains a more accurate description of the scene than any one of the individual source images. A comparison of various feature based fusion schemes is presented in this study. Feature extraction plays a major a role in the implementation of feature-level fusion approaches. Prior to the merging of images, salient features, present in all source images, are extracted using an appropriate feature extraction procedure. Then, fusion is performed using these extracted features. The performance of image fusion is evaluated by normalized least square error, entropy, overall cross entropy, standard deviation and mutual information. The experimental results show that the images fused with salience match measure, gradient and gradient match measure gives better performance.

Key words: Image fusion, multiresolution decomposition, multiresolution reconstruction, pixel based fusion, feature based fusion

INTRODUCTION

In medical imaging, various modalities provide different features of the internal body as they use different physical principles for imaging. These different physical principles are sensitive to different parameters. Hence, various modalities of medical imaging provide complementary information. The problem associated with these imaging systems is that the spatial and spectral resolution of imaging systems limits the information content of a single image. To use relevant information from different modalities image fusion is required. Image fusion can be performed at three different processing levels according to the stage where the fusion takes place: pixel level, feature-level and decision level as suggested by Varshney (1997) and Pohl and Van Genderen (1998). In this study, feature based fusion schemes are developed. Some generic requirements can be imposed on the fusion result: (a) the fused image should preserve, as closely as possible, all relevant information contained in the input images, (b) the fusion process should not introduce any artifacts or inconsistencies, which can distract or mislead the human observer, or any subsequent image processing steps and (c) in the fused image, irrelevant features and noise should be suppressed to a maximum extent. The application areas of image fusion include remote sensing proposed by Daniel and Willsky (1997), medical imaging proposed by QU *et al.* (2001), automated machine vision proposed by Slamani *et al.* (1997) and aviation. The source images are registered prior to fusion.

In recent years, many image fusion methods have been exploited as in Yang and Blum (2006) and Min *et al.* (2006). The simplest image fusion on pixel level is to sum and average the original images pixel by pixel. However when this method is applied, several undesired effects including reduction in contrast of feature would appear. If the source images are RGB color images, the methods of pixel level fusion also include Intensity-Hue-Saturation (IHS) transform and Principal Components Substitution (PCS) and so on. Zhang and Blum (1999) recognized that multiscale transforms are very useful for analyzing the information content of images for the purpose of fusion and some sophisticated approaches based on multiscale transforms, such as High-Pass Filtering (HPF) method, Laplacian pyramid, gradient pyramid, morphological pyramid and wavelet transform, have been proposed. For example, the HPF method has proven itself to be more efficient than IHS and PCS in preserving spectral features of enhanced bands. Image fusion based on Laplacian pyramid has been used in recent years. Li *et al.* (1995) pointed out that the Laplacian pyramid based image fusion techniques have certain drawbacks in the regions where the multi-modal data are significantly different.

Wavelets are a mathematical tool for hierarchical decomposing functions. After many successful applications in signal processing, wavelets have also

Corresponding Author: M. Sasikala, Department of Instrumentation Engineering, Madras Institute of Technology, Anna University, Chromepet Chennai, 600 044, Tamilnadu, India
Tel: 044-26564099 Fax: 044-22232403

been accepted as a powerful image processing technique among image fusion society. Wavelet transform can provide efficient localization in both space and frequency domains. Comparing with other multiscale transforms, Mallat (1989) suggested that wavelet transform is more compact and able to provide directional information in the low-low, low-high, high-low and high-high bands and contains unique information at different resolutions. Image fusion based on wavelet transform can provide better performance than those based on other multiscale methods. The wavelet representation provides directional information whereas the Laplacian pyramid does not supply spatial orientation in the decomposition. Since the wavelet basis function can be chosen orthogonal, the information at each layer of decomposition is unique. In this study, many feature based fusion schemes are compared. In terms of objective evaluation criteria, experimental results show that the images fused with salience match measure, gradient and gradient match measure improves the quality of the fused image, compared with other approaches.

IMAGE FUSION USING MULTIRESOLUTION DECOMPOSITION

Discrete wavelet transform: Image wavelet multi-resolution analysis applies Discrete Wavelet Transform (DWT) and it's a powerful tool for image analysis. The original space V_0 (scale space) can be decomposed into a lower resolution scale subspace V_1, the difference between V_0 and V_1 can be represented by the complementary wavelet subspace W_1. Similarly, we can continue to decompose V_1 into scale subspace V_2 and wavelet subspace W_2. For an N-level decomposition, we will obtain N+1 subspaces with one coarsest resolution subspace V_0 and N different subspace W_i, i is from 1 to N. Each digital signal in the space V_0 can be decomposed into some components in each subspace. In many cases, it is much easier to analyze these components rather than analyze the original signal itself. DWT can only decompose scale space, but it cannot decompose wavelet subspace anymore. Namely:

$$V_0 = V_1 \oplus W_1 = V_2 \oplus W_2 \oplus W_1 = V_3 \oplus W_3 \oplus W_3\ W_2 \oplus W_1 = .. \quad (1)$$

Fusion method: Different coefficient fusion methods yield different performances. After DWT decomposition, the low frequency coefficients reflect the gross approximations of the source images. The most frequently used method to choose the low frequency coefficients complies with the maximum absolute value principle. Firstly, compare the absolute values of the low frequency coefficients from the different source images and then choose the one with the greater absolute value. In order to compare the performance of different high frequency coefficient selecting methods, all the low frequency coefficients are decided by this approach in this study.

To simplify the description of the different fusion rules, we make an assumption that A and B are just two source images and the fused image is F. The symbol $C_j(A,p)$ and $C_j(B,p)$ is used to denote the wavelet transform coefficients of the source image A and B respectively and the symbol $C_j(F,p)$ is used to denote the wavelet transform coefficient of the fused image F, where j is the decomposition level and p is the location of the current coefficient. All the approaches mentioned in this paper can also be used in the case of more than two source images. There are some methods to choose the high frequency coefficients as follows:

GENERIC IMAGE FUSION SCHEMES

Pixel-based image fusion:

- In the pixel-based fusion scheme, the subband signal $C_j(F,p)$ of the fused image is simply acquired by picking the high frequency coefficient with greater absolute value.

$$C_j(F,p) = \begin{Bmatrix} C_j(A,p), & |C_j(A,p)| \geq |C_j(B,p)| \\ C_j(B,p), & |C_j(B,p)| > |C_j(A,p)| \end{Bmatrix} \quad (2)$$

Since the useful features in the image usually are larger than one pixel, the pixel-by-pixel maximum fusion rule may not be the most appropriate method.

- In the lowest spatial resolution, the subband signal $C_j(F,p)$ is acquired by averaging $C_j(A,p)$ and $C_j(B,p)$ of A and B.

$$C_j(F,p) = 0.5 * C_j(A,p) + 0.5 * C_j(B,p) \quad (3)$$

Feature based fusion

Fusion scheme based on salience measure: This fusion scheme is the weighted average scheme suggested by Burt and Kolezynski (1993). The salient features are first identified in each source image. The salience of a feature is computed as a local energy in the neighborhood of a coefficient.

$$E(A,p) = \sum_{\varphi = Q} w(q) C_j^2(A,q) \quad (4)$$

Where, w(q) is a weight and $\sum_{\varphi = Q} w(q) = 1$. In practice the neighborhood Q is small (typically 5×5 or 3×3) window centered at the current coefficient position. The closer the point q is near the point p, the greater w(q) is. E(B,p) can also be obtained by this rule. The selection mode is implemented as:

$$C_j(F,p) = \begin{Bmatrix} C_j(A,p), \ E(A,p) \ \geq \ E(B,p) \\ C_j(B,p), \ E(B,p) \ > E(A,p) \end{Bmatrix} \quad (5)$$

This selection scheme helps to ensure that most of the dominant features are incorporated into the fused image.

Fusion scheme based on salience match measure: In this fusion method, the salience measure of each source image is computed using Eq. 4. At a given resolution level j, this fusion scheme uses two distinct modes of combination: selection and averaging. In order to determine whether the selection or averaging will be used, the match measure M(p) is calculated as:

$$M(p) = \frac{2\sum_{\varphi = Q} w(q)C_j(A,q)C_j(B,q)}{E(A,p)+E(B,p)} \quad (6)$$

If M(p) is smaller than a threshold T, then the coefficient with the largest local energy is placed in the composite transform while the coefficient with less local energy is discarded. The selection mode is implemented as:

$$C_j(F,p) = \begin{Bmatrix} C_j(A,p), \ E(A,p) \geq E(B,p) \\ C_j(B,p), \ E(B,p) > E(A,p) \end{Bmatrix} \quad (7)$$

Else if M(p)$\geq$ T, then in the averaging mode, the combined transform coefficient is implemented as:

$$C_j(F,p) = \begin{Bmatrix} W_{max}C_j(A,p) + \ W_{min}C_j(B,p), \\ E(A,p) \geq E(B,p) \\ W_{max}C_j(B,p) + \ W_{min}C_j(A,p), \\ E(B,p) > E(A,p) \end{Bmatrix} \quad (8)$$

Where:

$$W_{min} = 0.5 - 0.5\left(\frac{1-M(p)}{1-T}\right) \text{and } W_{max} = 1 - W_{min} \quad (9)$$

In this study, the fusion methods are implemented using the parameters such as a window size of 5×5 and a T-value of 0.95. This value of T gave better performance.

Fusion scheme based on local deviation: This method was suggested by Qiang *et al*. (2003).The high frequency coefficient $C_j(F,p)$ is obtained by choosing the corresponding coefficient with the greater local deviation (Min *et al*., 2006).

$$C_j(F,p) = \begin{Bmatrix} C_j(A,p), \ \sigma_j(A,p) \geq \sigma_j(B,p) \\ C_j(B,p), \ \sigma_j(A,p) < \sigma_j(B,p) \end{Bmatrix} \quad (10)$$

Fusion scheme based on convolution: This fusion scheme was proposed by Wang HAI-HUI (2004). In this fusion scheme, the subbands are convolved with a feature extracting operator F (a matrix) and a pixel with larger output is selected to be the corresponding coefficient of composite subbands. Because DWT provides directional information in low-high, high-low and high-high subbands blocks, the operator F_1, F_2 and F_3 convolve horizontal, vertical and diagonal directions subbands block, respectively. F_1, F_2 and F_3 extract the edge information of the subimages based on direction:

Firstly, a local feature S is defined as:

$$S(A,p) = F_1 * C_j(A,p), F_2 * C_j(A,p), F_3 * C_j(A,p) \quad (11)$$

Where, the asterisk means convolution and

$$F_1 = \{\{-1,-1,-1\},\{2,2,2\},\{-1,-1,-1\}\} \quad (12)$$

$$F_2 = \{\{-1,2,-1\},\{-1,2,-1\},\{-1,2,-1\}\} \quad (13)$$

$$F_3 = \{\{-1, 0,-1\}, \{0, 4, 0\}, \{-1, 0,-1\}\} \quad (14)$$

S(B,p) can also be acquired by this rule. Finally, choose the high frequency coefficient of the fused image:

$$C_j(F,p) = \begin{Bmatrix} C_j(A,p), \ S(A,p) \geq S(B,p) \\ C_j(B,p), \ S(B,p) > S(A,p) \end{Bmatrix} \quad (15)$$

Fusion scheme based on local gradient: For a function f (x,y) it is common practice to approximate the magnitude of the gradient by using absolute values instead of squares and square roots:

$$\nabla f = |G_X| + |G_Y| = |f(x,y) - f(x+1,y)| + |f(x,y) - f(x,y+1)| \quad (16)$$

This equation is simpler to compute and it still preserves relative changes in gray levels. In image processing, the difference between pixel and its neighbors reflect the edges of the image. Firstly compute the differences between the low frequency coefficient at the point p and its eight neighbors, respectively. The value E is acquired by summing squares of all the differences. At last, choose the low frequency coefficient with the greater value E as the corresponding coefficient of the fused image. This method can maintain the information of edges. So it can improve the quality of the fused image. The algorithm is as follows:

$$E(A,p) = \sum_{\varphi = Q} \left|C_j(A,q) - C_j(A,p)\right|^2 \tag{17}$$

$$E(B,p) = \sum_{\varphi = Q} \left|C_j(B,q) - C_j(B,p)\right|^2 \tag{18}$$

Finally, select the corresponding high frequency coefficient of the fused image:

$$C_j(F,p) = \begin{Bmatrix} C_j(A,p),\ E(A,p) \geq E(B,p) \\ C_j(B,p),\ E(B,p) > E(A,p) \end{Bmatrix} \tag{19}$$

Fusion scheme based on gradient match measure: This method is based on salience match measure and gradient method. A local energy is defined as:

$$E(A,p) = \sum_{\varphi = Q} \omega(q)\left|C_j(A,q) - C_j(A,p)\right|^2 \tag{20}$$

Where, w(q) is a weight. The closer the point q is near the point p, the greater w(q) is. E(B,p) can also be acquired by this way.

The match measure M(p) is defined as:

$$M(p) = \frac{2\sum_{\varphi = Q} w(q)\left|C_j(A,q) - C_j(A,p)\right|\left|C_j(B,q) - C_j(B,p)\right|}{E(A,p) + E(B,p)} \tag{21}$$

The selection mode is implemented as in Eq. 7, 8 and 9.

Fusion scheme based on edges: This fusion scheme was proposed by Petrovic (2003). In this fusion framework, feature-level information in the form of edges and image segment boundaries are used to guide the signal level fusion process. The input signals are processed prior to fusion in order to identify the meaningful structures they contain. These feature sets are then combined using feature-level fusion methods in order to obtain a clear, unified picture of what entities should be preserved in the fused images. This information is then fed back into the signal level fusion process, which produces the fused image according to its specifications. The feature extraction part of the process uses a single input image (A) to produce an edge map output E_A that indicates, in a binary way, a presence or a lack of, an edge at each pixel in the output image. For a two input image fusion, edge map binary images are defined for both inputs A and B (E_A and E_B).

The edge maps are not fully exclusive as they represent the same scene; it is likely that many of the boundaries will be present in both maps. As the primary goal of the fusion process is to ensure all the exclusive region boundaries from the input images are present in the fused image (common boundaries will be automatically included through signal level fusion) a different method of XOR fusion can be applied. In this case a boundary point in E_A or E_B is kept only if it is exclusive to its image. Otherwise, it is removed which leaves the pyramid fusion mechanism to resolve which of the two input boundaries is more significant. This is expressed for input A through Eq. 22 where XOR and AND represent the logical exclusive OR and AND operations on the input edge maps. The fused exclusive edge map for input image B, X_B is also calculated.

$$X_A = (E_A \text{ XOR } E_B) \text{ AND } E_A \tag{22}$$

The values of High pass sub-band signals represent local saliency and direct selection maps are produced as in the case of direct signal-level fusion by selecting the coefficient with the larger absolute value for the fused image:

$$S_p = \begin{Bmatrix} 1, \text{ if } \ \left|C_j(A,p)\right| > \left|C_j(B,p)\right| \\ 0, \text{ otherwise} \end{Bmatrix} \tag{23}$$

Where, S_p is the selection for pixel at location p. (S = 1 indicates that the fused pixel should come from A and 0 indicates that the fused pixel should come from B). Selection maps obtained in this way are fused with the fused feature-level maps for each input image through a simple logical OR operation. This produces two separate selection maps, one for each input sub-band, that indicate which of its pixels are to be passed on into the fused signal. Since image boundaries cause significant values in whole neighbourhoods of sub-band coefficients around their original position, fused boundary maps are morphologically dilated by a simple square element in order to include all the necessary elements to reconstruct the boundary in the fused sub-band. The fused selection maps thus become:

$$S^A{}_P = S_P \mid \text{dilate } (X_A, k\times k)_p \tag{24}$$

$$S^B_P = (1\text{-}S_P) \mid \text{dilae}(X_B, k\times k)_P \quad (25)$$

Where, | signifies a logical OR operation on binary maps and dilate (I, k×k) a morphological dilate operation over signal I with a square block k×k (a value of k = 3 is used in our experiments). Practically, this means that every sub-band pixel on or around the desired input image boundary is included in the fused sub-band even if it would not otherwise be selected. The dilation operation followed by the OR operation also means that at some locations, both input sub-bands are to be considered as sources for the fused pyramid. In order to achieve this, the fused sub-band signal is finally produced from the input signals using a simple binary weighted sum, Eq. 26. This process is repeated for every pair of input sub-band signals. Once sub-band (pyramid) fusion is completed, the respective image reconstruction processes produce the fused images.

$$C_j(F,p) = S_P{}^A C_j(A,p) + S_P{}^B C_j(B,p) \quad (26)$$

RESULTS AND DISCUSSION

Evaluation criteria: The normalized least square error (NLSE) between the reference image R and the fused image F:

$$\text{NLSE} = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}[R(m,n) - F(m,n)]^2}{\sum_{m=1}^{M}\sum_{n=1}^{N}[R(m,n))]^2} \quad (27)$$

Entropy (H):

$$H = -\sum_{i=0}^{L} h(i)\log_2 h(i) \quad (28)$$

Where:

h = The normalized histogram of the fused image to be evaluated

L = The maximum value for a pixel in the image

L = 255

The entropy is used to measure the overall information in the fused image. The larger the value is, the better fusion results we get.

Overall across entropy (OCE):

$$\text{OCE}(X,Y;Z) = \frac{\text{CE}(X;Z) + \text{CE}(Y;Z)}{2} \quad (29)$$

Where, X,Y are the source images, Z is the fused image, CE(X;Z) CE(Y:Z) is the cross entropy of the source image X (Y) and the fused image Z:

$$\text{CE}(X;Z) = \sum_{i=0}^{L} h_x(i)\log_2\left|\frac{h_x(i)}{h_z(i)}\right| \quad (30)$$

The overall cross entropy is used to measure the difference between the source images and the fused image. Qu *et al*. (2002) suggested that the less the value of OCE is, the better fusion result we get.

Standard Deviation (SD): The standard deviation which is the square root of the variance reflect the spread in the data. So a high contrast image will have a high variance and low contrast image will have a low variance.

Mutual Information (MI): The Mutual Information fusion criterion presented here states that the MI of the image intensity values of respective pixel pairs is maximal if the images are well fused. The image fusion performance metric is:

$$M_F{}^{AB} = I_{FA}(f,a) + I_{FB}(f,b) \quad (31)$$

Where:

$$I_{FA}(f,a) = \sum_{f,a} P_{FA}(f,a)\log\frac{P_{FA}(f,a)}{P_F(f)P_A(a)} \quad (32)$$

$$I_{FB}(f,b) = \sum_{f,b} P_{FB}(f,b)\log\frac{P_{FB}(f,b)}{P_F(f)P_B(b)} \quad (33)$$

Where, $p_A(a)$, $p_B(b)$ and $p_F(f)$ are the probability density functions in the individual images. $p_{FA}(f,a)$ and $p_{FB}(f,b)$ are the joint probability density function.

The various image fusion schemes are tested on several pairs of CT and MRI images. Biorthogonal 5.5 wavelet filter is used. The pixel level and the feature-level fusion algorithms are compared. Figure 1a and b show the source images. Figure 1c-k are the fused images based on pixel level fusion (Algorithm 1 and 2), salience, salience match measure, local deviation, convolution, gradient, gradient match measure and edges, respectively.

The performances of the pixel based and the feature based image fusion methods based on DWT is presented in Table 1. The images fused with salience match measure, gradient and gradient match measure have higher entropy,

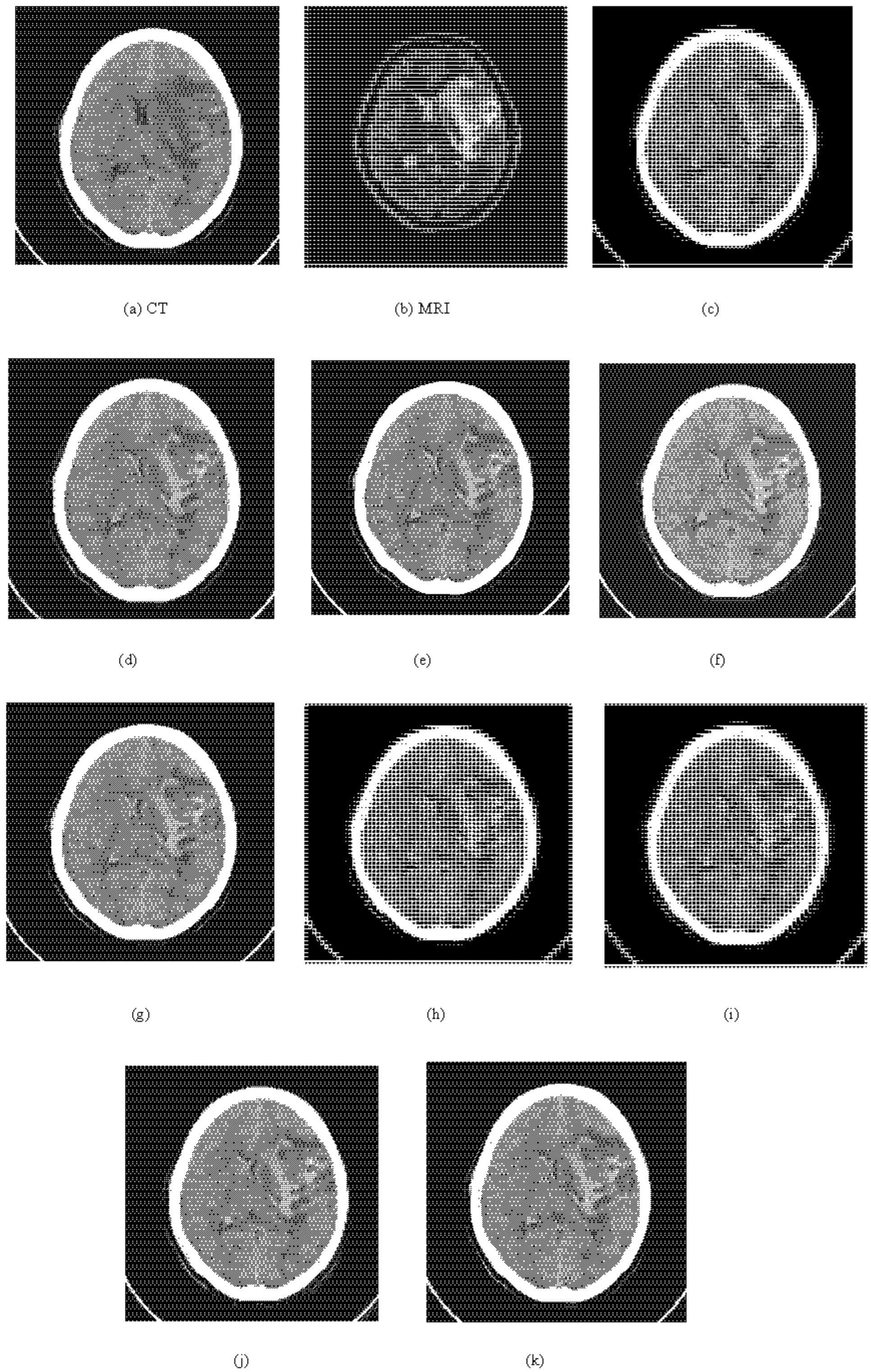

Fig. 1: Source images and fused images based on DWT

Table 1: Performance evaluation of various fusion algorithms

Methods	NLSE	H	OCE	SD	MI
Pixel level (Algorithm 1)	0.1151	6.5183	0.71980	64.3124	7.3349
Pixel level (Algorithm 2)	0.1160	6.5682	0.70850	65.6902	7.3637
Salience	0.1151	6.4744	0.72570	63.6629	7.3384
Salience match measure	0.1333	7.9374	0.63337	68.7832	7.5367
Deviation	0.1150	6.4740	0.71620	63.6553	7.3387
Convolution	0.1156	6.5794	0.76870	64.8294	7.3405
Gradient	0.1214	7.6450	0.69000	68.5233	7.5689
Gradient match measure	0.1272	7.8261	0.64320	68.0698	7.4858
Edge	0.1134	6.5491	0.72390	64.7177	7.3272

lower overall cross entropy, higher standard deviation and higher mutual information compared with the pixel level and other feature-level algorithms. Images fused with salience match measure, gradient and gradient match measure gives better performance among the feature based and pixel based methods.

CONCLUSION

The feature based fusion methods are compared in this paper. Fusion is performed based on DWT. For the medical image fusion, it has been shown that an image fusion technique based on multi-resolution wavelet decomposition is a wonderful trade-off between spectral and spatial information. The images fused with features such as salience match measure, gradient and gradient match measure are better than the other methods according to the objective evaluation criteria. Experimental results show that these methods give better results for image fusion as image contrast, average information content and detail information of fused images are increased. The feature based image fusion schemes provides significant improvement in reliability of the fusion process (increases robustness), reduces the loss of information and eliminates the presence of artifacts.

REFERENCES

Burt, P.J. and R.J. Kolezynski, 1993. Enhanced image capture through fusion. In: Proceedings of 4th International Conference on Computer Vision, Berlin, Germany, pp: 173-182.

Daniel, M.M. and A.S. Willsky, 1997. A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing. Proc. IEEE., 85: 164-180.

Li, H., B.S. Manjunath and S.K. Mitra, 1995. Multisensor image fusion using the wavelet transform. Graphic. Models Image Process., 57: 235-245.

Mallat, S.G., 1989. Multifrequency channel decomposition of images and wavelet models. IEEE. Trans. on Acoustics, Speech and Signal Processing, 37: 2091-2110.

Min, L., C. Wei and T. Zheng, 2006. A region-based multi-sensor image fusion scheme using pulse-coupled neural network. J. Pattern Recognit. Lett., 27: 1948-1956.

Petrovic, V., 2003. Multi-level image fusion. In: Proceedings of SPIE., 5099: 87-96.

Pohl, C. and J.L. Van Genderen, 1998. Multisensor image fusion in remote sensing: Concepts, methods and applications. Int. J. Remote Sens., 19: 823-854.

Qiang, Zan-Xia, Peng Jia-Xiong and Wang Hong-qun, 2003. Remote sensing image fusion based on local deviation of wavelet transform. J. Huazhong Univ. Sci. Technol. (Nature Science Edn.), 31: 89-91.

Qu, G., D. Zhang and Y. Pingfan, 2001. Medical image fusion using two dimensional discrete wavelet transform. Proc. SPIE, 4556: 86-95.

Qu, G., D. Zhang and P. Yan, 2002. Information measure for performance of image fusion. Electron. Lett., 38: 313-315.

Slamani, M.A., L. Ramao, M. Uner, P. Varshney, D.D. Weiner, M. Alford, D. Ferris and V. Vannicola, 1997. Enhancement and fusion of data for concealed weapon detection. Proc. SPIE., 3068: 8-19.

Varshney, P.K., 1997. Multisensor data fusion. Electron. Commun. Eng. J., 9: 245-253.

Wang, HAI-HUI., 2004. Multisensor image fusion by using discrete multiwavelet transform. In: Proceedings of the 3rd International Conference on Machine Learning and Cybernetics. Shanghai, pp: 26-29.

Yang, J. and R.S. Blum, 2006. A region-based image fusion method using the Expectation-Maximization Algorithm. In: 40th Annual Conference on Information Sciences and Systems, pp: 468-473.

Zhang, Z. and R.S. Blum, 1999. A categorization of multiscale-decomposition based image fusion schemes with a performance study for a digital camera application. In: Proceedings of the IEEE., 87: 1315-1326.