Discrete Cosine Transform-based Image Fusion

V.P.S. Naidu

National Aerospace Laboratories, Bangalore
E-mail: vpsnaidu@gmail.com

ABSTRACT

Image fusion by multi-resolution discrete cosine transform (MDCT) algorithm has been implemented and evaluated. The performance of this algorithm is compared with that of well known image fusion technique using wavelets. It is observed that image fusion by MDCT perform almost similar to that of wavelets. It is computationally very simple and it could be well suited for real time applications.

Keywards: Multi-sensor image fusion, multi-resolution DCT, image fusion performance evaluation metrics, discrete cosine transform

1. INTRODUCTION

Multi-sensor image fusion (MIF) is a technique that combines two or more registered images to increase the spatial resolution of acquired low detail multi-sensor images and preserving their spectral information. Of late MIF has emerged as an innovative and promising research area in image processing. The benefiting fields from MIF are viz. military, remote sensing, machine vision, robotic, surveillance, enhanced vision system, and medical imaging, etc. The problem that MIF tries to solve is to merge the information content from several images (or acquired from different imaging sensor modalities) taken from the same scene in order to accomplish a fused image that contains the finest information coming from the different original source images[1]. Hence, the fused image would provide enhanced superiority image than any of the original source images. Depending on the merging stage, MIF could be performed at three different levels viz. pixel level, feature level, and decision level[2,3]. In this paper, pixel-level based MIF is presented that represents a fusion process generating a single combined image containing an additional truthful description than individual source image.

The simplest MIF is to take the average of the grey level source images pixel by pixel. This technique would produce several undesired effects and reduced feature contrast in the fused image. To overcome these problems, multi-scale transforms, such as wavelets[1,4-12], image pyramids[3,13-16], spatial frequency[17], statistical signal processing[18-21] and fuzzy set theory[22] have been proposed. Multi-resolution wavelet transforms could provide good localisation in both spatial and frequency domains. Discrete wavelet transform would provide directional information in decomposition levels and contain unique information at different resolutions[4,5]. In this paper, the multi-resolution discrete cosine transform (MDCT) is applied to fuse the source images.

One of the important prerequisites to be able to apply fusion techniques to source images is the image registration i.e., the information in the source images needed to be adequately aligned and registered prior to fusion of the images. In this paper, it is assumed that the images to be fused are already registered.

2. DISCRETE COSINE TRANSFORM

Discrete cosine transform (DCT) is an important transform in image processing. Large DCT coefficients are concentrated in the low frequency region; hence, it is known to have excellent energy compaction properties.

The 1D discrete cosine transform X(k) of a sequence x(n) of length N is defined as[23-28]:

$$X(k) = \alpha(k) \sum_{n=0}^{N-1} x(n) \cos\left(\frac{\pi(2n+1)k}{2N}\right), \quad 0 \le k \le N-1 \quad (1)$$

where

$$\alpha(k) = \begin{cases} \sqrt{\dfrac{1}{N}} & k = 0 \\[2ex] \sqrt{\dfrac{2}{N}} & k \ne 0 \end{cases} \quad (2)$$

One can observe that for $k=0$, the Eqn (1) becomes $X(0) = \sqrt{\dfrac{1}{N}} \sum_{n=0}^{N-1} x(n)$. The first transform coefficient is the average of all samples in the sequence and is known as DC coefficient, and other transform coefficients are known as AC coefficients.

The inverse discrete cosine transform is defined as:

$$x(n) = \sum_{k=0}^{N-1} \alpha(k) X(k) \cos\left(\frac{\pi(2n+1)k}{2N}\right), \quad 0 \le n \le N-1 \quad (3)$$

Eqn (1) is generally called as analysis formula or forward transform and Eqn (3) is called as synthesis formula or

inverse transform. The basis sequence $\cos\left(\frac{\pi(2n+1)k}{2N}\right)$ is real and discrete time sinusoids.

The 2D DCT is a direct extension of 1D DCT. The 2D discrete cosine transform $X(k_1,k_2)$ of an image or 2D signal $x(n_1,n_2)$ of size $N_1 \times N_2$ is defined as:

$$X(k_1,k_2) = \alpha(k_1)\alpha(k_2)\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} x(n_1,n_2)\cos\left(\frac{\pi(2n_1+1)k_1}{2N_1}\right)$$

$$\cos\left(\frac{\pi(2n_2+1)k_2}{2N_2}\right), \quad \begin{array}{l} 0 \le k_1 \le N_1-1 \\ 0 \le k_2 \le N_2-1 \end{array} \quad (4)$$

where $\alpha(k_1)$ and $\alpha(k_2)$ are similar to Eqn (2).

Similarly, the 2D inverse discrete cosine transform is defined as:

$$x(n_1,n_2) = \sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1} \alpha(k_1)\alpha(k_2)X(k_1,k_2)\cos\left(\frac{\pi(2n_1+1)k_1}{2N_1}\right)$$

$$\cos\left(\frac{\pi(2n_2+1)k_2}{2N_2}\right), \quad \begin{array}{l} 0 \le n_1 \le N_1-1 \\ 0 \le n_2 \le N_2-1 \end{array} \quad (5)$$

Both DCT and IDCT are separable transformation and the advantage of this property is that 2D DCT or 2D IDCT can be computed in two steps by successive 1D DCT or 1D IDCT operations on columns and then on rows of an image $x(n_1,n_2)$ as shown in Fig. 1.

3. MULTI-RESOLUTION DCT

Multi-resolution DCT (MDCT) is very similar to wavelets transform, where signal is filtered separately by low-pass and high-pass finite impulse response (FIR) filters and the output of each filter is decimated by a factor of two to achieve first level of decomposition. The decimated low pass filtered output is filtered separately by low-pass and high-pass filter followed by decimation by a factor of two provides second level of decomposition. The successive levels of decomposition can be achieved by repeating this procedure. The idea behind the MDCT is to replace the FIR filters with DCT[21].

The information flow diagram of MDCT (one level of decomposition) is shown in Fig. 2. The image to be decomposed is transformed into frequency domain by applying DCT in column-wise. Take the IDCT on first 50 % of points (0 to 0.5π) to get the low passed image L. Similarly, take the IDCT on second 50 % of points (0.5π to π) to get the high passed image H. The low passed image L is transformed into frequency domain by applying DCT in row wise. Take the IDCT on first 50 % of points (in row wise) to get low passed image LL and similarly take IDCT on the remaining 50% to get the high passed image LH. The high passed image H is transformed into frequency domain by applying DCT in row wise. Take the IDCT on first 50 % of points (in row wise) to get low passed image HL and similarly take IDCT on the remaining 50% to get the high passed image HH. The LL contains the average image information corresponding to low frequency band of multi scale decomposition. It could be considered as smoothed and sub sampled version of the source image. It represents the approximation of source image. LH, HL and HH are detailed sub images which contain directional (horizontal, vertical and diagonal) information of the source image due to spatial orientation. Multi resolution could be achieved by recursively applying the same algorithm to low pass coefficients (LL) from the previous decomposition.
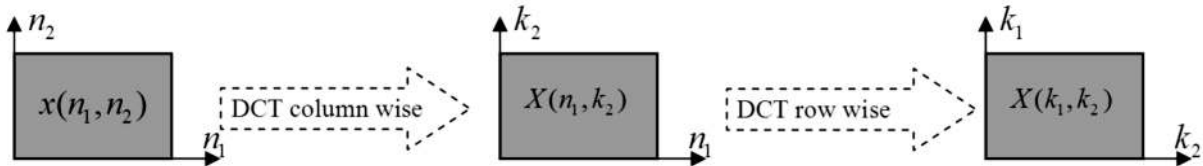

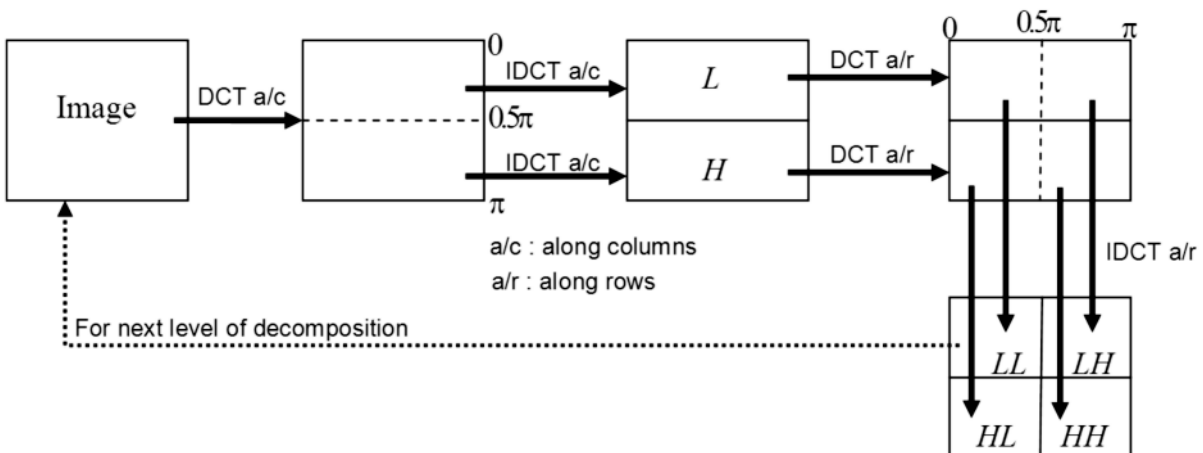**Figure 1. Computation of 2-D DCT using separability property.**

Figure 2. Multi-resolution image decomposition structure using DCT.

(a)

(b)

(i) First level of decomposition

(ii) Second level of decomposition

(c)

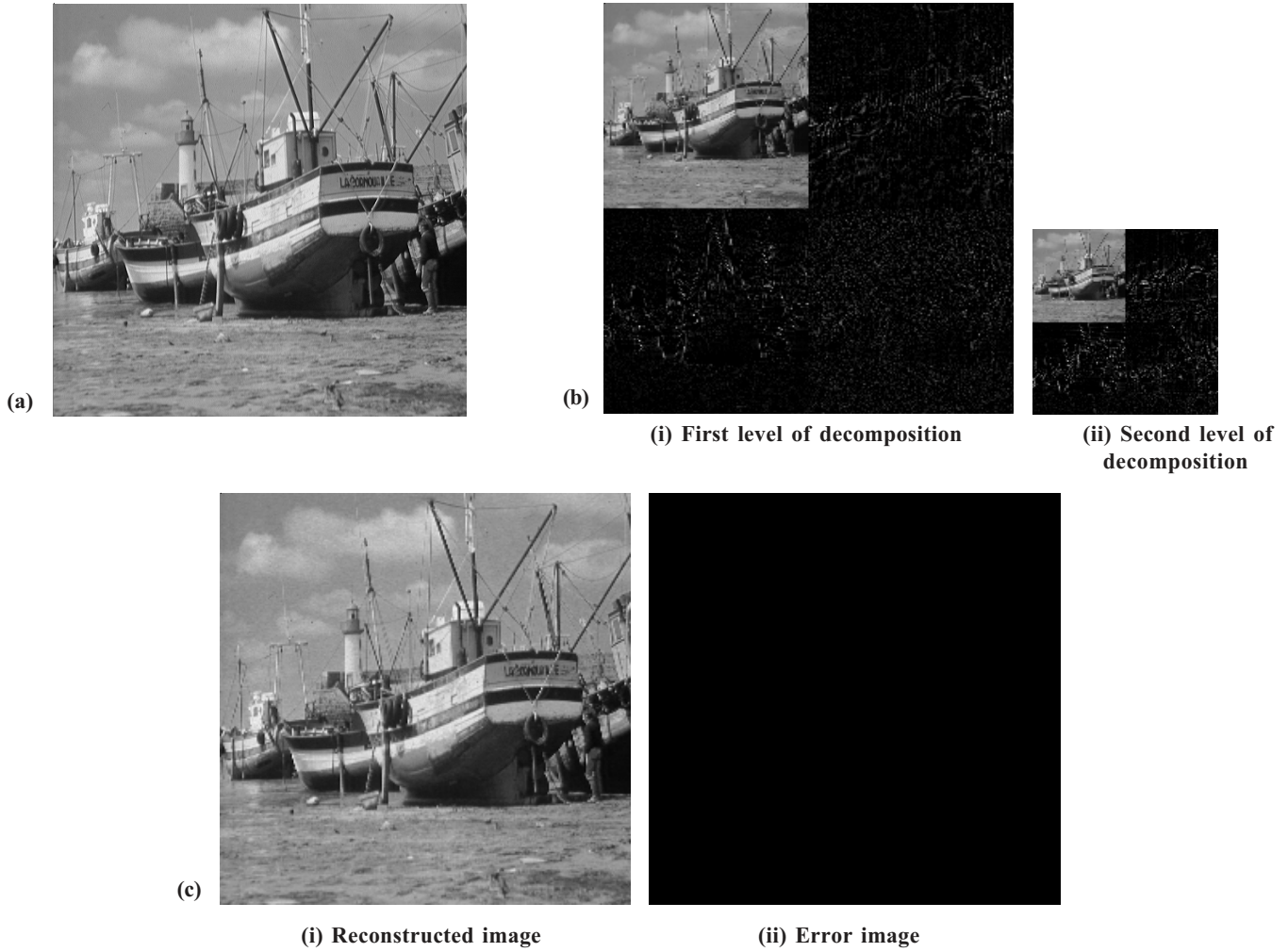**(i) Reconstructed image**

(ii) Error image

Figure 3. (a) Ground truth image; (b) Multi-resolution image decomposition; (c) Reconstructed image from 2nd level of decomposition and the error image.

3. FUSION

The schematic diagram for the MDCT based pixel level image fusion scheme is shown in Fig. 4. One can observe that the modification of the present scheme is the use MDCT instead of wavelets or pyramids. The images to be fused I_1 and I_2 are decomposed into D $(d = 1, 2, ..., D)$ levels using MDCT. The resultant decomposed images from I_1 are $I_1 \rightarrow \left\{ {}^1 LL_D, \left\{ {}^1 LH_d, {}^1 HH_d, {}^1 HL_d \right\}_{d=1,2,...,D} \right\}$ and from I_2 are $I_2 \rightarrow \left\{ {}^2 LL_D, \left\{ {}^2 LH_d, {}^2 HH_d, {}^2 HL_d \right\}_{d=1,2,...,D} \right\}$. At each decomposition level $(d = 1, 2, ..., D)$, the fusion rule will select the larger absolute value of the two MDCT detailed coefficients, since the detailed coefficients are corresponds to sharper brightness changes in the images such as edges and object boundaries etc. These coefficients are fluctuating around zero. At the coarest level $(d = D)$, the fusion rule take average of the MDCT approximation coeficients since the approximation coeficents at coarser level are the smoothed and subsampled verion of the original image. The fused image I_f can be obtained using:

$$I_f \leftarrow \left\{ {}^f LL_D, \left\{ {}^f LH_d, {}^f HH_d, {}^f HL_d \right\}_{d=1,2,...,D} \right\} \quad (6)$$

4. PERFORMANCE EVALUATION

4.1 With Reference Image

When the reference image is available, the performance of image fusion algorithms can be evaluated using the following metrics:

1. Percentage fit error[29] (PFE)

$$PFE = \frac{norm(I_r - I_f)}{norm(I_r)} * 100 \quad (7)$$

where, *norm* is the operator to compute the largest singular value.

It is computed as the norm of the difference between the corresponding pixels of reference and fused image to the norm of the reference image. This will be zero when both reference and fused images are exactly alike and it will be increased when the fused image is deviated from the reference image.
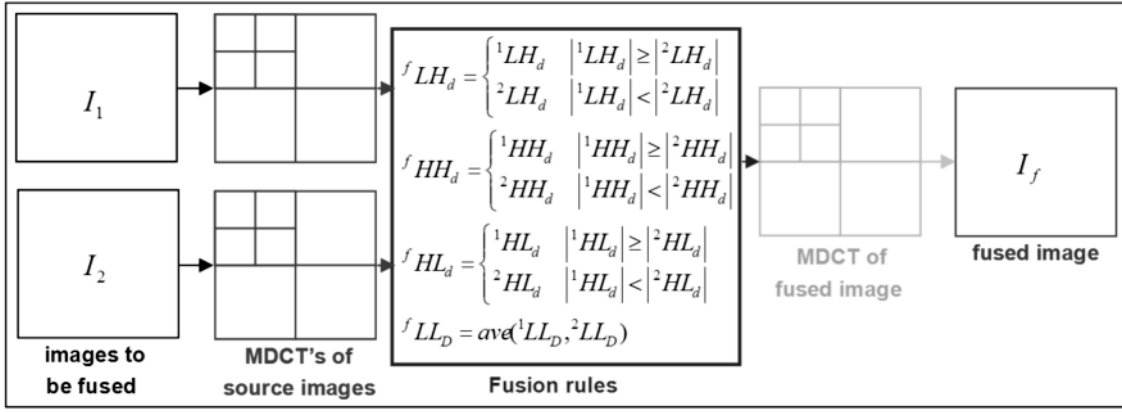
Figure 4. Schematic diagram for the MDCT based pixel level image fusion scheme.

2. Peak signal to noise ratio[30] (PSNR)

$$PSNR = 20\log_{10}\left(\frac{L^2}{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_r(i,j)-I_f(i,j)\right)^2}\right) \quad (8)$$

where, L in the number of gray levels in the image.

Its value will be high when the fused and reference images are similar. Higher value implies better fusion.

3. Measure of structural similarity[31, 32] (SSIM)

$$SSIM = \frac{\left(2\mu_{I_r}\mu_{I_f}+C_1\right)\left(2\sigma_{I_rI_f}+C_2\right)}{\left(\mu_{I_r}^2+\mu_{I_f}^2+C_1\right)\left(\sigma_{I_r}^2+\sigma_{I_f}^2+C_2\right)} \quad (9)$$

where, C_1 is a constant that is included to avoid the instability when $\mu_{I_r}^2+\mu_{I_f}^2$ is close to zero and C_2 is a constant that is included to avoid the instability when $\sigma_{I_r}^2+\sigma_{I_f}^2$ is close to zero

where, $\mu_{I_r} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}I_r(i,j)$

$$\mu_{I_f} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}I_f(i,j)$$

$$\sigma_{I_r}^2 = \frac{1}{MN-1}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_r(i,j)-\mu_{I_r}\right)^2$$

$$\sigma_{I_f}^2 = \frac{1}{MN-1}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_f(i,j)-\mu_{I_f}\right)^2$$

$$\sigma_{I_rI_f} = \frac{1}{MN-1}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_r(i,j)-\mu_{I_r}\right)\left(I_f(i,j)-\mu_{I_f}\right)$$

Natural image signals would be highly structured and their pixels reveal strong dependencies. These dependencies would carry vital information about the structure of the object. It compares local patterns of pixel intensities that have been normalized for luminance and contrast.

4.2 Without Reference Image

When the reference image is not available, the following metrics could be used to test the performance of the fused algorithms.

1. Standard deviation[33] (SD)

$$\sigma = \sqrt{\sum_{i=0}^{L}(i-\bar{i})^2 h_{I_f}(i)} \;, \quad \bar{i}=\sum_{i=0}^{L}ih_{I_f} \quad (10)$$

where, $h_{I_f}(i)$ is the normalized histogram of the fused image $I_f(x,y)$ and L number of frequency bins in the histogram.

It is known that standard deviation is composed of the signal and noise parts. This metric would be more efficient in the absence of noise. It measures the contrast in the fused image. An image with high contrast would have a high standard deviation.

2. Cross entropy[34] (CE)

Overall cross entropy of the source images I_1, I_2 ,and the fused image I_f is:

$$CE(I_1,I_2;I_f) = \frac{CE(I_1;I_f)+CE(I_2;I_f)}{2} \quad (11)$$

where, $CE(I_1;I_f) = \sum_{i=0}^{L} h_{I_1}(i)\log\left(\frac{h_{I_1}(i)}{h_{I_f}(i)}\right)$

$$CE(I_2;I_f) = \sum_{i=0}^{L} h_{I_2}(i)\log\left(\frac{h_{I_2}(i)}{h_{I_f}(i)}\right)$$

3. Spatial frequency[35-36] (SF)

Spatial frequency criterion is:

$$SF = \sqrt{RF^2+CF^2} \quad (12)$$

where, the row frequency of the image is:

$$RF = \sqrt{\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=2}^{N}[I_f(x,y)-I_f(x,y-1)]^2}$$

and column frequency of the image is:

$$CF = \sqrt{\frac{1}{MN}\sum_{y=1}^{N}\sum_{x=2}^{M}[I_f(x,y)-I_f(x-1,y)]^2}$$

This frequency in spatial domain indicates the overall activity level in the fused image. (x, y) is the pixel index. The fused image with high SF would be considered.

5. RESULTS AND DISCUSSION

The National Aerospace Laboratories' indigenous aircraft SARAS, shown in (Fig. 5(a)), is considered as a reference image I_r to evaluate the performance of the proposed fusion algorithm. The complimentary pair input images I_1 and I_2 are taken to evaluate the fusion algorithm

(a) Reference image I_r (b) First source image I_r (c) Second source image

Figure 5. Reference and source images.

Figure 6. Fused and error image with one level *D*=1 of decomposition using MDCT.

Figure 7. Fused and error image with one level *D*=1 of decomposition using wavelets.

Figure 8. Fused and error image with two levels *D*=2 of decomposition using MDCT.

Figure 9. Fused and error image with two levels *D*=2 of decomposition using wavelets.

Table 1. Performance evaluation metrics

Levels of decomposition	Algorithm	With reference image			Without reference image		
		PFE	PSNR	SSIM	CE	SD	SF
$D = 1$	MDCT	3.8772	38.6038	0.9673	8.8438	46.1094	12.4658
$D = 1$	Wavelets	3.7899	38.7026	0.9736	10.6944	46.2320	13.0840
$D = 2$	MDCT	3.3175	39.2808	0.9635	10.3484	46.8292	15.4992
$D = 2$	Wavelets	3.2027*	39.4338*	0.9764*	11.3181*	46.9953*	15.6984*

and these images are shown in Figs. 5(b)-5(c). The complementary pair has been created by blurring the reference image of size 512 x 512 with a Gaussian mask using diameter of 12 pixels. The images are complementary in the sense that the blurring occurs at the top half and the bottom half respectively. The first column in Figs. 6 - 9. shows fused images and the second column shows the error images. The error (difference) image is computed by taking the corresponding pixel difference of reference image and fused image, i.e., $I_e(x,y) = I_r(x,y) - I_f(x,y)$. The fused and error images by one level of decomposition using MDCT and wavelet fusion algorithms are shown in Fig. 6 and Fig. 7 respectively. Similarly the fused and error images by two levels of decomposition using MDCT and wavelet are shown in Fig. 8 and Fig. 9 respectively. It is observed that the fused images of both MDCT and wavelet are almost similar for these images. The reason

could be because of taking the complementary pairs. The performance metrics for evaluating the image fusion algorithms are shown in Table 1. The metrics shown in table with asterisk (*) mark, are better among others. The performance of MDCT is almost similar to that of wavelets. Higher level of decomposition performs superior fusion.

6. CONCLUSION

Pixel level image fusion by MDCT algorithm has been implemented and evaluated. The performance of this algorithm is compared with well known image fusion technique by wavelets. It is concluded that image fusion by MDCT is almost similar to that of wavelets. It is computationally very simple and it could be well suited for real-time applications. Image fusion by higher level of decomposition provides better fusion results.

REFERENCES

1. Pajares, Gonzalo & Cruz, Jesus Manuel de la. A wavelet-based image fusion tutorial. *Pattern Recognition,* 2007, **37**, 1855-872.

2. Varsheny, P.K. Multisensor data fusion. *Elec. Comm. Eng. J.,* 1997, **9**(12), 245-53.

3. Burt, P.J. & Lolczynski, R.J. Enhanced image capture through fusion. *In* Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany, 1993. pp. 173-82,

4. Mallet, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.,* 1989, **11**(7), 674-93,

5. Wang, H.; Peng J. & Wu, W. Fusion algorithm for multisensor image based on discrete multiwavelet transform. *IEE Pro. Vis. Image Signal Process.,* 2002, **149**(5).

6. Li, H.; Manjunath, B.S. & Mitra. Multisensor image fusion using wavelet transform. *Graph. Models Image Process.*, 1995, **57**(3), 235-45.

7. Pu, T. & Ni, G. Contrast-based image fusion using discrete wavelet transform. *Optical Engineering,* 2000, **39**(8), 2075-082.

8. Yocky, D.A. Image merging and data fusion by means of the discrete two-dimensional wavelet transform. *J. Opt. Soc. Amer.,* 1995, **12**(9), 1834-841.

9. Nunez, J.; Otazu, X.; Fors, O.; Prades, A.; Pala, V. & Arbiol, R. Image fusion with additive multiresolution wavelet decomposition: applications to spot1 landsat images. *J. Opt. Soc. Amer.,* 1999, **16**, 467-74.

10. Rockinger, O. Image sequence fusion using a shift invariant wavelet transform. *In* Proceedings of the IEEE International Conference on Image Processing, 1997, **13**, pp. 288-91.

11. Qu, G.H.; Zang, D.L. & Yan, P.F. Medical image fusion by wavelet transform modulus maxima. *J. Opt. Soc. Amer.,* 2001, **9**, 184-90.

12. Chipman, L.J.; Orr, T.M. & Graham, L.N. Wavelets and image fusion. *Proceedings SPIE,* 1995, **2529,** 208-19,

13. Jahard, F.; Fish, D.A.; Rio, A.A. & Thompson, C.P. Far/near infrared adapted pyramid-based fusion for automotive night vision. *In* IEEE Proceedings of the 6th International Conference on Image Processing and its Applications (IPA97), 1997. pp. 886-90.

14. Ajazzi, B.; Alparone, L.; Baronti, S. & Carla, R. Assessment pyramid-based multisensor image data fusion. *Proceedings SPIE,* 1998, **3500,** 237-248.

15. Akerman, A. Pyramid techniques for multisensory fusion. *Proceedings SPIE,* 1992, **2828**, 124-31.

16. Toet, A.; Ruyven, L.J. Van & Valeton, J.M. Merging thermal and visual images by a contrast pyramid. *Optical Engineering,* 1989, **28**(7), 789-92.

17. Li, Shutao; Kwok, James T. & Wang, Yaonan. Combination of images with diverse focuses using the spatial frequency. *Information fusion,* 2001, **2**(3), 167-76.

18. Blum, Rick S. Robust image fusion using a statistical signal processing approach. *Information fusion,* 2005, **6**, 119-28.

19. Yang, J. & Blum, Rick S. A statistical signal processing approach to image fusion for concealed weapon detection. *In* IEEE International Conference on Image Processing, 2002, Rochester, NY. pp. 513-16,

20. Naidu, V.P.S. & Raol, J.R. Fusion of out of focus images using principal component analysis and spatial frequency. *J. Aerospace Sci. Technol.,* Aug. 2008, **60**(3), 216-25.

21. Naidu, V.P.S. & Raol, J.R. Pixel-level image fusion using wavelets and principal component analysis – a comparative analysis. *Def. Sci. J.,* May 2008, **58**(3), 338-52.

22. Nejatali, Abdilhossein & Ciric, L.R. Novel image fusion methodology using fuzzy set theory. *Optical Engineering,* 1998, **37**(2), 485-91.

23. Ahmed, N.; Natarajan, T. & Rao, K.R. Discrete Cosine Transform. *IEEE Trans. Comp.*, **32**, 1974, 90-93.

24. Discrete cosine transform. http:/documents.wolfram.com/ applications/digitalimage/ UsersGuide/ImageTransform/ ImageProcessing8.2.html.

25. Dimitrov, Vassil & Khan, Wahi. Multiplierless DCT algorithm for image compression applications. *Int. J. Info. Theo. Appli.,* 2004, **11**, 162-69,

26. Gonzalez, R.C. & Wintz, P. Digital image processing. Addison-Wesley, MA, 1987.

27. Khayam, Syed Ali. Information theory and coding. March 2003. Lecture Notes: ECE802-602

28. Strang, G. The discrete cosine transform. *SIAM Review,* 1999, **41**, 135-47.

29. Naidu, V.P.S.; Girija G. & Raol, J.R. Evaluation of data association and fusion algorithms for tracking in the presence of measurement loss. *In* AIAA Conference on Navigation, Guidance and Control, 11-14 August 2003, Austin, USA.

30. Arce, Gonzalo R. Nonlinear signal processing–A statistical approach. Wiley-Interscience Inc., Publication, USA, 2005.

31. Wang, Z. & Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.,* 2002, **9**(3), 81-84.

32. Wang, Z.; Bovik, A.C.; Sheikh, H.R. & Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.,* 2004, **13**(4), 600-12.

33. Blum, Rick, S. & Liu, Zheng. Multi-sensor image fusion and its applications. Taylor & Francis Group, CRC Press, NW, 2006.

34. Leung, Lau Wai; King, Bruce & Vohora, Vijay. Comparison of image data fusion techniques using entropy and INI. *In* 22nd Asian Conference on Remote Sensing, 5-9 Nov. 2001, Singapore.

35. Eskicioglu, A.M. & Fisher, P.S. Image quality measures and their performance. *IEEE Trans. Commu.,* 1995, **43**(12), 2959-965.

36. Li, Shutao; Kwok, James T. & Wang, Yaonan. Combination of images with diverse focuses using the spatial frequency. *Information Fusion,* 2001, **2,** 169-76.