A variational approach for multi-focus image fusion and visualization enhancement

Yuan-Xiang Li[1], Ze-Ming Zhou[2], Jin-Ze Song[1], Ning Ma[2], Chun He[2], Peng Zhang[2]
1 School of Aeronautics & Astronautics, Shanghai Jiao Tong University, Shanghai, China
2 Institute of Meteorology, PLA University of Science and Technology, Nanjing, China
Email: {yuanxli@sjt.edu.cn, zeming_zhou@yahoo.com.cn}

Abstract - A variational model is presented to perform multi-focus image fusion and contrast enhancement for visualization. The fusion process can be formulized as minimizing a weighted energy functional which is constructed with the geometry and the coherence of the input multi-focus images. The total variation constraint is applied in the functional for the smoothness of the solution and noise reduction. The variational problem is solved with the gradient descent flow and the corresponding Euler-Lagrange equation is modified by introducing the terms for histogram equalization and edge enhancement. The model is compared with the Laplacian-based, wavelet transform-based and contrast-based fusion approaches visually and quantitatively. Experimental results show that the model can recover an everywhere-in-focus image while enhancing its contrast and improving its perceptual effects efficiently. Our model provides a good alternative for multi-focus image fusion.

Keywords: Multi-focus image fusion, variational model, visualization enhancement.

1 Introduction

According to the optical imaging formula, when an object in the scene is in focus, another one with a different depth may be out of focus. So, for the same scene, when two or more pictures are taken from an optical lens with different focuses, the regions of the scene are in focus or blurring, which means those images contain complementary and redundant information. An everywhere-in-focus image about the scene is very useful for advanced vision tasks such as object recognition and analysis.

In recent years, many image fusion approaches have been applied in remote sensing application and multi-focus image processing. Multi-resolution analysis [1] is one of the efficient tools which decomposes the input images in different scale space and then designs the suitable rules for fusion. Discrete wavelet transform (DTW) has been widely used for image fusion but the artifacts near the edges usually appear in the result. This impairment is effectively addressed by a trous wavelet transform [2], curvelet transform [3], discrete wavelet frame transform and support vector machine [4]. Recently, image fusion approaches based on partial differential equation (PDE) have been proposed and show some prospective results. Socolinsky [5] presented a contrast-based fusion model, in which the contrast of a multiband image is defined and the variational approach is used to find the fused image. This work was extended [6] by combining the contrast of the input images with perceptual enhancement for better visualization. Kumar [7] proposed a total variation model for pixel-level image fusion. Wang [8] presented a variational model for fusion and denoising of multi-focus images.

In this paper, we propose a variational and PDE model for multi-focus image fusion and visualization enhancement. The energy functional is designed to remove the noise and preserve the geometry and correlation of the input images. The terms of histogram equalization and edge enhancement are employed in the model. In section 2, we present the fusion model and give the discrete scheme. In section 3, we show some fusion results on multi-focus and multi-modalities images and compare with the conventional approaches in terms of fusion quality metrics. Conclusion is given in section 4.

2 Fusion and enhancement model

In this section, we briefly introduce the contrast-based fusion model and then propose a variational fusion approach for multi-focus images according to the fusion requirements such as the geometry and coherence information preserving. To improve the visualization effects of the everywhere-in-focus image, Euler-Lagrange equation deduced from the functional is modified for image fusion and contrast enhancement simultaneously.

2.1 Energy functional

The geometry of the multi-focus images can be expressed with the structure tensor and contrast vector field [5] [6]. Let $I_i, i = 1,...,n$ be the input images, the structure tensor at P is defined as

$$\chi^2(P) = \begin{bmatrix} \sum_{i=1}^{n} (\frac{\partial I_i}{\partial x})^2 & \sum_{i=1}^{n} \frac{\partial I_i}{\partial x}\frac{\partial I_i}{\partial y} \\ \sum_{i=1}^{n} \frac{\partial I_i}{\partial x}\frac{\partial I_i}{\partial y} & \sum_{i=1}^{n} (\frac{\partial I_i}{\partial y})^2 \end{bmatrix}$$

where P is the pixel in image I_i .Clearly, the matrix is symmetric with real entries and thus its eigenvalues are both real and non-negative. Let θ_p be the eigenvector of the largest eigenvalue λ_p and the contrast vector field $\mathbf{V}$ is constructed with all V_p :

$$V_p = \lambda_p \theta_p sign(\theta_p \cdot \sum_{i=1}^{n} c(i) \nabla I_i) \tag{1}$$

$$c(i) = \frac{|\nabla I_i|}{(\sum_{i=1}^{n} |\nabla I_i|^2)^{1/2}} \tag{2}$$

$\mathbf{V}$ encodes the first order spectral information and intensity information of the multi-focus images [5].

In the contrast-based model, the energy functional is defined as

$$E(f) = \int_{\Omega} |\nabla f - \mathbf{V}|^2 dx \tag{3}$$

whose minima is the fused image f , and it means the gradient of f is approximate to $\mathbf{V}$.

We modify the above model and define the contrast energy term as

$$E_{ct}(f) = \int_{\Omega} |\nabla f - \alpha \mathbf{V}|^2 dx \tag{4}$$

where α is a constant to weight the effects of the contrast on the fused image. Clearly, if $\alpha > 1$, the gradient level of f is magnified, which means the enhancement of edges, but it would probably lead to the coherence distortion.

As for the fusion model in (4), the contrast is the only information to recover the fused image. To preserve the image coherence better between the fused image and the input images, we define the image coherence energy term as

$$E_{co}(f) = \int_{\Omega} (f - \sum_{i}^{n} c(i) I_i)^2 dx \tag{5}$$

where $I_i, i = 1,2,...,n,$ are the multi-focus images. The minima of the (5) indicate that the pixel intensity of the fused image is equal to the weighted average value of the input images approximately.

The perceptual contrast model is sensitive to the noise, since its discrete scheme uses the second order differential operator of f and $\mathbf{V}$ is calculated with the local pixel intensity variation. Thus, we define the regularity term for the smoothness of the solution and image noise removal, similar to the total variation model:

$$E_{nr}(f) = \int_{\Omega} |\nabla f| dx \tag{6}$$

Considering the above factors for image fusion, we define the fusion model as follows:

$$E(f) = \lambda_1 \int_{\Omega} |\nabla f - \alpha \mathbf{V}|^2 dx + \lambda_2 \int_{\Omega} (f - \sum_{i}^{n} c(i) I_i)^2 dx + \lambda_3 \int_{\Omega} |\nabla f| dx \tag{7}$$

where $\lambda_i, i = 1,2,3,$ are the parameters to weight the effects of the energy terms. Clearly, when α =0, (7) becomes the typical ROF model for image restoration and noise removal [9].

2.2 PDE for fusion and enhancement

When $E(f)$ is minimized, we achieve an everywhere-in-focus image. The first order variation of (7) is

$$\delta E(f) = 2\lambda_1(\alpha\, div(V) - \Delta f) + 2\lambda_2(f - \sum_{i=1}^{n} c(i) I_i) + \lambda_3 \nabla \cdot \frac{\nabla f}{|\nabla f|} \tag{8}$$

By introducing time variable t, the gradient descent flow is applied for the steady solution of (8)

$$\frac{\partial f}{\partial t} = -\delta E(f) \tag{9}$$

In [10], an approach for histogram modification based on partial differential equation was proposed for image contrast enhancement. Let $f(x,y,t)$ is the evolving image by time t , it deforms according to

$$\frac{\partial f(x,y,t)}{\partial t} = (1 - \frac{1}{M} f(x,y,t)) A(\Omega) - A[(v,w) : f(v,w,t) \geq f(x,y,t) \tag{10}$$

where M is the number of gray level and $A[.]$ represents the number of pixels, Ω denotes the domain of f .

Shock filters are successful class for sharpening the edges of an image, initially proposed by Osher and Rudin [11]. To enhance the contrast of the fused image for recognition and analysis task, the inverse diffusion term and the histogram equalization term are added to the gradient descent flow and the fusion equation can be written in PDE form, as:

$$\frac{\partial f}{\partial t} = -2\lambda_1(\alpha\, div(V) - \Delta f) - 2\lambda_2(f - \sum_{i=1}^{n} c(i) I_i)$$
$$- \lambda_3 \nabla \cdot \frac{\nabla f}{|\nabla f|} + \lambda_4\,((1 - \frac{1}{M} f) A(\Omega)$$
$$- A[(v,w): f(v,w,t) \geq f(x,y,t)])$$
$$- \lambda_5 |\nabla f| * sign(\Delta f) \qquad (11)$$

where $\lambda_i, i = 4{,}5$ are the parameters to control the influences of image enhancement on the fusion result.

2.3 The discrete scheme

The PDE of fusion process is discretized in image domain. We use the forward difference to implement the first order derivatives except ∇f^k of the inverse diffusion term and the Laplace operator is simply expressed by 5-points discretization scheme

$$\Delta f^k = f^k(i+1, j) + f^k(i, j+1) + f^k(i-1, j) + f^k(i, j-1) - 4 f^k(i, j)$$

We reformulate (11) as

$$\frac{f^{k+1} - f^k}{\Delta t} = -2\lambda_1(\alpha\, div(V) - \Delta f^k)$$
$$- 2\lambda_2(f^k - \sum_{i=1}^{n} c(i) I_i)) - \lambda_3 \nabla \cdot \frac{\nabla f^k}{|\nabla f^k|}$$
$$+ \lambda_4\,((1 - \frac{1}{M} f^k) A(\Omega)$$
$$- A[(v,w): f^k(v,w,t) \geq f^k(x,y,t)])$$
$$- \lambda_5 |\nabla f^k| * sign(\Delta f^k), \qquad (12)$$

where Δt is the time increment and the inverse diffusion term ∇f^k needs to be discretized as

$$\frac{\partial f^k}{\partial x} = m(f^k(i,j) - f^k(i-1,j),\ f^k(i+1,j) - f^k(i,j))$$
$$\frac{\partial f^k}{\partial y} = m(f^k(i,j) - f^k(i,j-1),\ f^k(i,j+1) - f^k(i,j))$$
$$m(\alpha,\beta) = \begin{cases} sign(\alpha)\min(|\alpha|,|\beta|), & if \quad \alpha\beta > 0 \\ 0, & if \quad \alpha\beta \leq 0 \end{cases}$$

The initial image of (12) is set to the weighted average of the input images defined in (5) and it should be normalized. The histogram equalization term is processed before each iteration by mapping its gray value to the range between 0 and $M-1$ and afterwards using the array to calculate $A[.]$. After each iteration, f^k should be remained in the range between 0 and 1 and it would be clipped off to meet the constraint if necessary. The computational cost of (12) for an iteration is $O(N)$, being N the number of pixels in the image.

3 Experiments and analysis

We have implemented and analyzed the fusion model on several multi-focus image dataset. The parameters in the model need to be tuned for the solution stability and the specific application, but there are some heuristic rules for their selection. If we hope to make full of the geometry of the input images, we should choose a larger λ_1 and α. To keep the coherence of the fused image and the input images better, λ_2 could increase reasonably. λ_3 controls the smooth level for the noise image and λ_4, λ_5 weight the contrast of the fused image. We set λ_1 =0.3, λ_2 =0.4, λ_3 =0.001, λ_4 =0.2, λ_5 =0.1, Δt = 0.1, α = 2, a time increment Δt = 0.1 in all of the experiments. The convergence condition is set to the root-mean-squared difference less than 0.1%. In all cases, no more than 15 iterations, the equation reaches a steady solution.

We have compared our model with Laplacian, wavelet and contrast-based fusion approaches. Daubechies symmetric spline DBSS(2,2) is used for 4 decomposition levels of input images [12]. Figure 1 and 2 (a)-(b) are the input images. Figure 1 and 2 (c)-(f) show the results with Laplacian, DWT, contrast and our model respectively. The images obtained by our model are perceptually good because of the contrast enhancement. The ringing artifacts have appeared in the results by DWT.

To achieve the quantitative measures, we choose the following indexes to evaluate the fusion performance: entropy, average gradient magnitude, mean structural similarity index (MSSIM) [13], and normalized mutual information (MI) [14].

The structural similarity index of the signal x and y is defined as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where C_1, C_2 are the constant and

$$C_1 = (K_1 L)^2, K_1 << 1; C_2 = (K_2 L)^2, K_2 << 1$$

L is the dynamic range of the pixel values. μ_x denotes the mean of x, and σ_x, σ_{xy} denote the variance of x and the covariance of x and y. SSIM is an index of the local similarity.

MSSIM is defined to evaluate the overall quality:

$$MSSIM(x,y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j, y_j)$$

x_j and y_j are the contents at the j-th local window; and

M is the number of local windows.

In this paper, MSSIM measures the global similarity between the fused image and the input images:

$$MSSIM(I_M, I) = \sum_{i=1}^{n} \frac{1}{n} MSSIM(I_i, I)$$

Where I_i and I are the input images and the fused image.

Mutual information is defined as

$$MI(I_M, I) = \sum_{i=1}^{n} M(I_i, I)$$

where

$$M(I_i, I) = \frac{H_{I_i} + H_I}{H_{I_i I}}$$

H_{I_i}, H_I are the marginal entropy for I_i and I, $H_{I_i I}$ is the joint entropy. MI can be interpreted as the index of dependence between the input images I_i and the fused image I. A larger MI means the better fusion resemblance [14].

The first experiment is to compare the different fusion approach for multi-focus image of *clock*. As shown in Table 1, our model achieves the best score in entropy, average gradient magnitude and a little lower MSSIM, MI partly because the histogram equalization leads to a little structural change. The fused image has a higher spatial quality and attains a good trade-off between improving the visualization effects and preserving image coherence.

The second example is shown in Figure 2 and Table 2, which is for multi-focus image of *disk*. The fused image is everywhere-in-focus and it contains the focus region in both input images. The quantitative result is similar to the first experiment.

Our model can be extended to the multi-modalities image fusion, as seen in Figure 3. The MRI image has a higher imaging quality for the soft tissue while the CT image achieves a better diagnosis value for the bone structure. The result is shown in Figure 3 and Table 3. The fused images based on the contrast model and our model have an ideal perceptual quality which contains the rich tissue texture and the protrudent bone structure.

4 Conclusion

A fusion model for multi-focus images has been proposed which combines the geometry, coherence and noise removal of input images into a variational framework. For the purpose of enhancement visualization, the PDE of the fusion process is modified by introducing the terms of the histogram equalization and the inverse diffusion. The performance of the fusion model has been tested on multi-focus and multi-modalities images. Compared with the conventional fusion approaches, our model can enhance the visual contrast while improving the spatial quality and avoiding the artifacts at the edges. Our model provides a good alternative for multi-focus image fusion.

For further research, we would study the influences of the different parameters on the fusion results and how we can choose them reasonably. As for the fusion of the multi-modalities images, we would consider the local histogram equalization for contrast enhancement. Another direction is how MRA technique can be integrated into the fusion process to achieve a better performance.

Acknowledgements

This work was supported by Natural Science Foundation of Shanghai (09ZR1413700) and PLA University of Science and Technology (2009QX04).

References

[1] M. Gonzáles-Audícana, J. L. Saleta, R. G. Catalán, and R. García, "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," IEEE Trans. Geosci. Remote Sens., Vol. 42, No. 6, pp. 1291- 1299, 2004.

[2] J. Nùñez, X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," IEEE Trans. Geosci. Remote Sens., vol. 37, no. 3, pp. 1204-1211, 1999.

[3] M. Choi, R. Y. Kim, M.-R. Nam, and H. O. Kim, "Fusion of multispectral and panchromatic satellite images using the curvelet transform,"IEEE Trans. Geosci. Remote Sens. Lett., vol. 2, no. 2, pp. 136-140, 2005.

[4] Shutao Li, James Tin-Yau Kwok, Ivor Wai-Hung Tsang, and Yaonan Wang, "Fusing Images With Different Focuses Using Support Vector Machines", IEEE Trans. Neural Networks, Vol. 15, No. 6, pp. 1555-1561, 2004.

[5] Socolinsky D A, Wolff L B, "Multispectral image visualization through first-order fusion," IEEE Trans. Image Processing, vol. 11, no. 8, pp. 923-931, 2002.

[6] Gemma Piella, "Image fusion for enhanced visualization: a variational approach," International Journal of Computer Vision, Vol. 83, no. 1, pp. 1-11, 2009.

[7] M. Kumar, S. Dass, et al, "A total variation-based algorithm for pixel-level image fusion," IEEE Trans. Image Processing, vol. 18, no. 9, pp. 2137-2143, 2009.

[8] Wei-Wei Wang, Peng-Lang Shui, and Xiang-Chu Feng, "Variational models for fusion and denoising of multifocus images," IEEE Signal Processing Letters, Vol. 15, pp. 65-68, 2008.

[9] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variational based noise removal algorithms,” Phys. D, vol. 60, pp. 259–268, 1992.

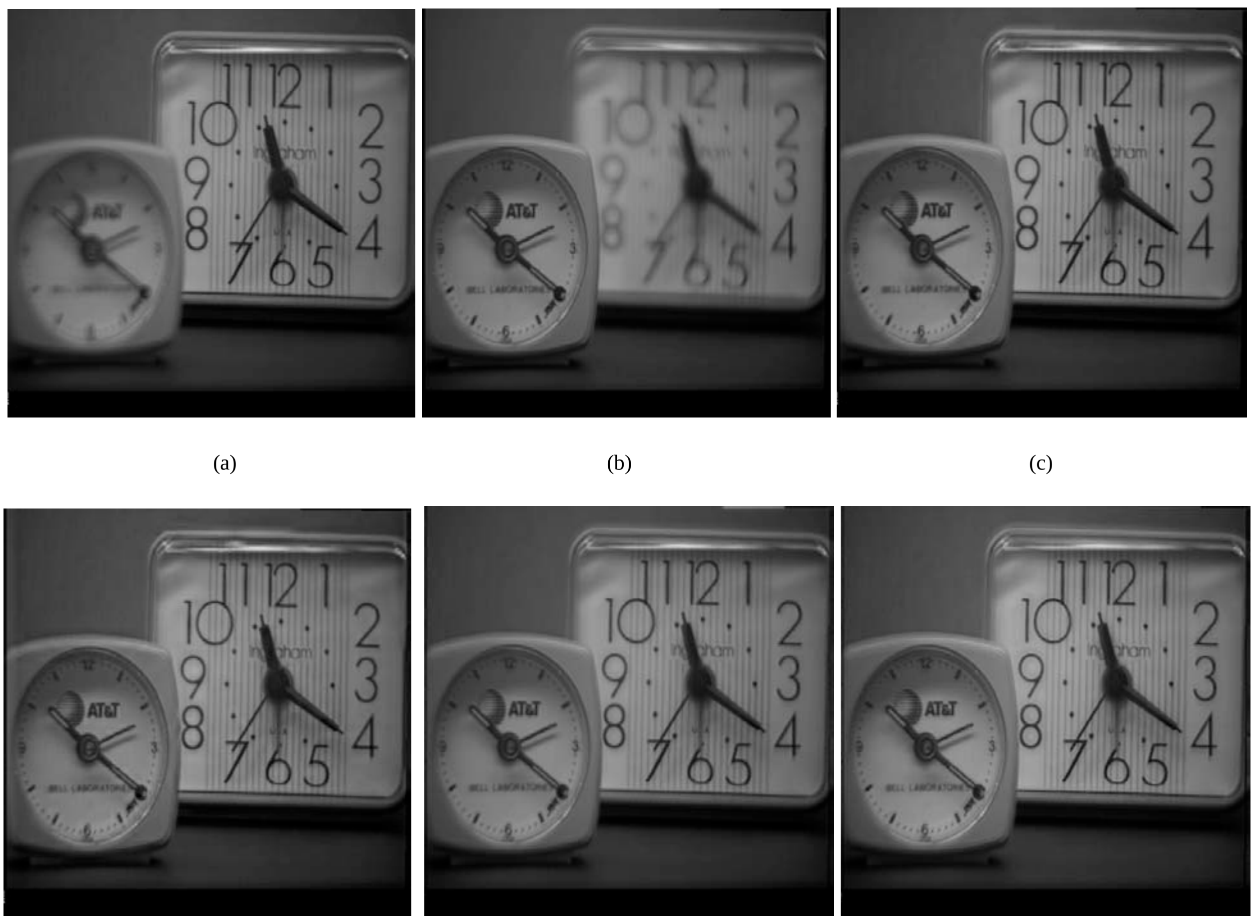
[10] Guillermo Sapiro, Vicent Caselles, “Histogram modification via differential equations,” Journal of Differential Equations, Vol 135, No. 2, pp 238-268, 1997.

[11] S. Osher, L. Rudin, “Feature-oriented image enhancement with shock filters,” SIAM Journal on Numerical Analysis, vol. 27, no.3, pp.919-940, 1990.

[12] Rockinger, Oliver: "Image sequence fusion using a shift invariant wavelet transform", in: Proc. IEEE Intl. Conference on Image Processing, pp. III-288-291, 1997.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error measurement to structural similarity”, IEEE Transactions on Image Processing, vol. 13, no. 4, Apr. 2004.

[14] Yufeng Zheng, “Multi-scale fusion algorithm comparisons: Pyramid, DWT and Iterative DWT”, 12th International Conference on Information Fusion, USA, July 6-9, pp. 1060-1067, 2009.



(a) (b) (c)

(d) (e) (f)

Figure 1. Fusion results for multi-focus image of *clock*. (a)-(b) input images; (c) fused image using Laplacian Pyramid; (d) fused image using DWT; (e) fused image using contrast model; (f) fused image using our model

Table 1. Fusion quality analysis for multi-focus image of *clock*

Fusion model	Entropy	Average Gradient	MSSIM	MI
Laplacian-based	7.0628	2.9264	0.9065	**2.3277**
DWT-based	7.0584	3.0049	0.9001	2.2431
Contrast-based	7.0060	2.5349	**0.9142**	2.2857
Our model-based	**7.1970**	**3.1891**	0.8932	2.2879

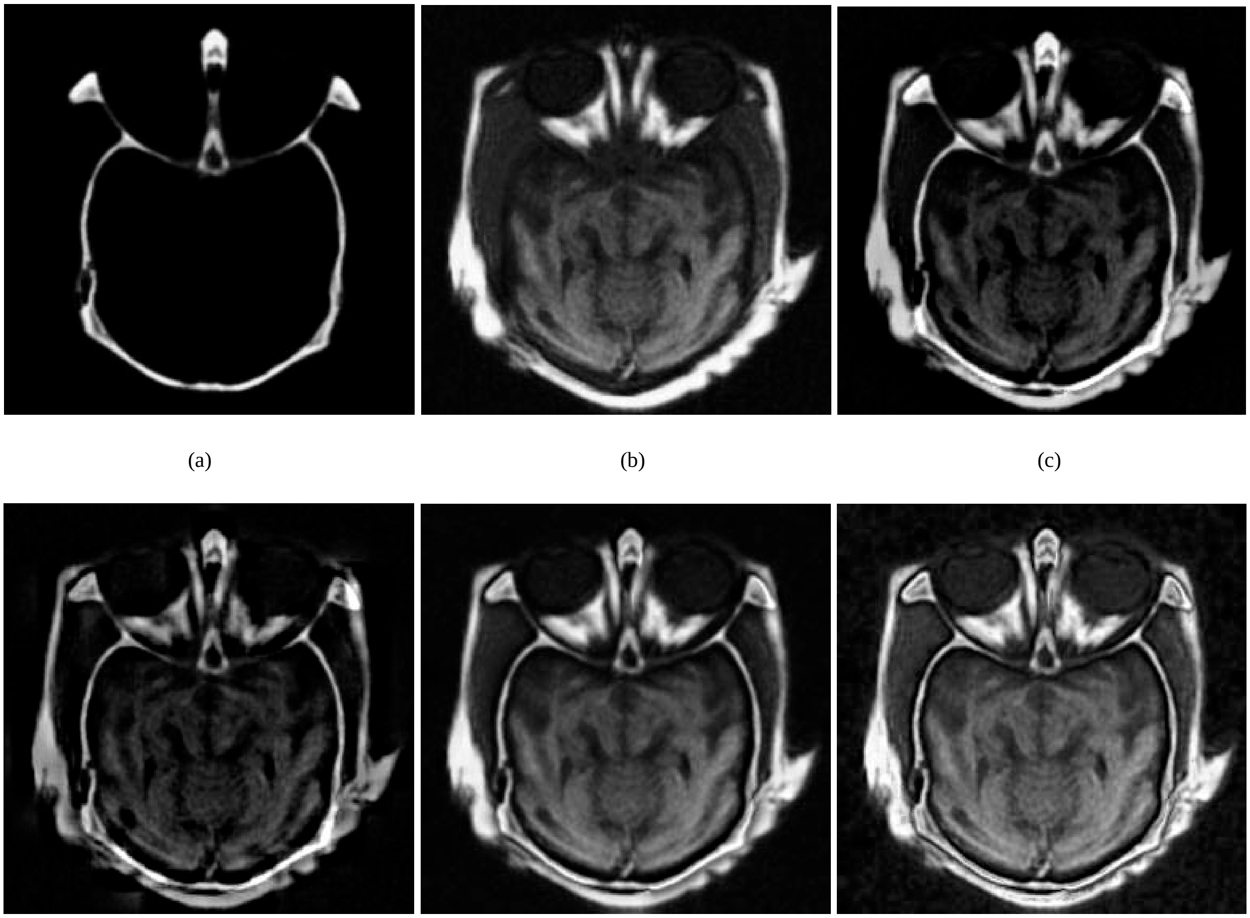
(a) (b) (c)

(d) (e) (f)

Figure 2. Fusion results for multi-focus image of *disk*. (a)-(b) input images; (c) fused image using Laplacian Pyramid; (d) fused image using DWT; (e) fused image using contrast model; (f) fused image using our model

Table 2. Fusion quality analysis for multi-focus image of *disk*

Fusion model	Entropy	Average Gradient	MSSIM	MI
Laplacian-based	7.3673	4.8081	**0.8528**	**2.1624**
DWT-based	7.3717	4.9490	0.8378	2.0981
Contrast-based	7.2727	4.0592	0.8497	2.1122
Our model-based	**7.4802**	**5.2752**	0.8269	2.1219



(a) (b) (c)

(d) (e) (f)

Figure 3. Fusion results for multimodalities image of *MRI and CT*. (a)-(b) input images; (c) fused image using Laplacian Pyramid; (d) fused image using DWT; (e) fused image using contrast model; (f) fused image using our model

Table 3. Fusion quality analysis for multi-modalities image of *MRI and CT*

Fusion model	**Entropy**	**Average Gradient**	**MSSIM**	**MI**
Laplacian-based	6.0977	6.6179	0.4310	2.2611
DWT-based	6.1512	6.3462	0.4052	2.2095
Contrast-based	6.7529	6.6502	**0.4824**	2.2945
Our model-based	**7.3443**	**8.5922**	0.4362	**2.3040**