This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2016 conference.

Multifocus Image Fusion Using Discrete Wavelet Transform And Sparse Representation

Aishwarya N
SSN College of Engineering
Chennai, India

Abirami S
SSN College of Engineering
Chennai, India

Amutha R
SSN College of Engineering
Chennai, India

***Abstract*—Multifocus image fusion increases the depth of field of a sensor by combining the images of same scene with different focus settings. So, extracting the useful image information of different source images plays a crucial role in image fusion process. In this paper, a novel fusion algorithm based on Discrete Wavelet Transform (DWT) and Sparse Representation (SR) is proposed. Initially, DWT is applied to extract the low frequency components and high frequency components of source images. High frequency components are merged using SR based fusion approach and low frequency components are combined using variance as activity level measurement. Finally, inverse DWT is performed on the fused coefficients to get the fused image. Experimental results demonstrate the effectiveness of proposed method in terms of visual perception and quantitative analysis.**

***Index Terms*—Multifocus image fusion, discrete wavelet transform, sparse representation, variance.**

I. INTRODUCTION

Due to the limited depth of focus of optical lenses it is not possible to get an image that contains all relevant objects in focus. However, for accurately interpreting and analyzing images, it is desired to obtain images with every object in focus [1]. Multifocus image fusion is an effective technique to solve this problem by combining two or more images of the same scene taken with different focus settings into a single all-in-focus image with extended depth of field. The resultant fused image is very useful for human or machine perception. The multi-focus image fusion has been applied in various applications such as remote sensing, medicine, machine vision, automatic change detection, biometrics etc [2]. In the past decade, many image fusion algorithms have been developed which are categorized into two main groups: spatial domain based methods and frequency domain based methods [3–10]. The spatial domain methods operate directly on image pixels or image regions which may lead to undesirable artifacts and may produce spatial distortion in the fused image. The frequency domain fusion algorithms extract the coefficients of source images with respect to some localized bases and these extracted coefficients are fused according to some criteria and then transformed back to fused image. DWT is a widely used multi resolution fusion technique as it can provide better spatial and spectral resolution than other traditional multi resolution methods [5]. In DWT based fusion methods, the "max-absolute" rule is used to select the fused coefficients from source images. But this method is sensitive to noise and artifacts [11]. In transform based fusion algorithms [2] [12] a simple "averaging rule" is adopted to fuse the low frequency coefficients. Since it is a well known fact that the low frequency components contain detailed information of source images, a sharpness measure which selects the focused pixels from source images has to be employed for better visual clarity. Motivated by the above key issues, a new multi-focus image fusion method is proposed to fuse the DWT coefficients. Initially, DWT is applied to decompose the source images into low frequency components and high frequency components. Since the high frequency components reflect the structural information of source images, sparse representation based fusion rule is applied to get the fused high frequency coefficients. Low frequency components are segmented into fixed sized blocks and a sharpness focus measure is computed for each corresponding blocks of input images. The block which has maximum sharpness index is absorbed into the fused low frequency coefficients. Finally, inverse DWT is applied to obtain the fused image. The rest of this paper is organized as follows. In Section 2, the signal sparse representation theory, is briefly reviewed. Section 3 describes the proposed fusion strategy based on sparse representation. The performance metrics used to evaluate the fusion algorithm is briefly explained in Section 4. Experimental results and objective evaluations are demonstrated in Section 5. Finally, Section 6 concludes the paper.

II. SPARSE REPRESENTATION THEORY

A.Sparse Representation Theory

Sparse representations have recently drawn much interest in vision, signal and image processing [14, 15]. It represents a signal by a linear combination of "few" atoms from an over-complete dictionary. Overcompleteness and sparsity are major characteristics of sparse representation. Overcompleteness means that the number of atoms in a dictionary is greater than the signal dimension. In sparse representation modeling, an input signal $y \in R^n$ can be represented by

$$y = D\alpha \tag{1}$$

where the vector $\alpha \in R^K$ contains the representation coefficients of the signal y and the dictionary matrix $D \in R^{m \times K} (K > n)$ contains K prototype signals referred as atoms. The sparsest solution with the fewest number of nonzero coefficients can be found by solving the following optimization problem

978-1-4673-9338-6/16/$31.00 ©2016 IEEE

$$\min_{\alpha} \|\alpha\|_0 \ \ s.t. \ \|y - D\alpha\|_2 \le \varepsilon \tag{2}$$

where $\|.\|_0$ is the l_0 semi-norm that counts the number of nonzero entries in a vector and $\varepsilon > 0$ is the error tolerance. The above optimization problem is a NP-hard problem and can be solved approximately by greedy approaches such as Orthogonal Matching Pursuit (OMP) algorithm [16].

Constructing a proper dictionary is a key issue in sparse representation. Generally, there are two ways to obtain a dictionary. The first category is using analytical models such as Discrete Cosine Transform (DCT) and Curvelet Transform (CVT). The second category is applying machine learning technique to obtain the dictionary from a large number of training image patches. The dictionary learning model can be represented by,

$$\min_{D,\{\alpha_i\}_{i=1}^{M}} \sum_i \|\alpha_i\|_0 \ \ s.t. \ \|y_i - D\alpha_i\|_2 \le \varepsilon, \ \ i \in \{1.....M\} \tag{3}$$

The above minimization problem can be solved using K-SVD [17].

III. Proposed Fusion Scheme

The schematic diagram of the proposed fusion framework is shown in Fig. 1. First, DWT is applied to the source images $\{I_A, I_B\}$ to obtain their low-pass bands $\{L_A, L_B\}$ and high-pass bands $\{H_A, H_B\}$. Image variance [18] is used as a saliency measure to select the focused coefficients in the approximation bands. High frequency sub bands reflect the structural information of source images in horizontal, vertical and diagonal directions. Hence, SR based fusion rule is applied to obtain the fused high frequency bands.

A. Low Pass Fusion

1. Divide the low frequency bands L_A and L_B into non overlapping patches of size 8x8. Suppose assume that there are T patches in L_A and L_B denotes as $\{p_{LA}^i\}_{i=1}^{T}$ and $\{p_{LB}^i\}_{i=1}^{T}$ respectively.
2. Calculate the image variance [18] for each corresponding patch of L_A and L_B denoted as var_{LA}^i and var_{LB}^i.
3. The fused patch p_{LF}^i is obtained by,

$$p_{LF}^i = \begin{cases} p_{LA}^i & if \ \text{var}_{LA}^i > \text{var}_{LB}^i \\ p_{LB}^i & otherwise \end{cases} \tag{4}$$

4. Iterate the above process for all image patches to obtain fused low pass band.

B. High Pass Fusion

1. Divide the high frequency bands H_A and H_B into image patches of size 8x8 from upper left to lower right with a step length of one pixel using sliding window technique. Suppose that there are T patches denotes as $\{p_{HA}^i\}_{i=1}^{T}$ and $\{p_{HB}^i\}_{i=1}^{T}$ in H_A and H_B respectively.
2. Rearrange the image patches $\{p_{HA}^i, p_{HB}^i\}$ into column vectors $\{v_{HA}^i, v_{HB}^i\}$.
3. Calculate the sparse coefficient vectors $\{\alpha_{HA}^i, \alpha_{HB}^i\}$ of $\{v_{HA}^i, v_{HB}^i\}$ using OMP,

$$\alpha_{HA}^i = \arg\min \|\alpha\|_0 s.t. \|v_{HA}^i - D\alpha\|_2 \le \varepsilon \tag{5}$$

$$\alpha_{HB}^i = \arg\min \|\alpha\|_0 s.t. \|v_{HB}^i - D\alpha\|_2 \le \varepsilon \tag{6}$$

where D is the dictionary.

4. Merge α_{HA}^i and α_{HB}^i with the "max-L1" rule to obtain the fused sparse vector. It is given by,

$$\alpha_{HF}^i = \begin{cases} \alpha_{HA}^i & if \ \|\alpha_{HA}^i\|_1 > \|\alpha_{HB}^i\|_1 \\ \alpha_{HB}^i & otherwise \end{cases} \tag{7}$$

The fused vector is calculated by,

$$v_{HF}^i = D\alpha_{HF}^i \tag{8}$$

5. Iterate the above process for all the image patches to obtain all the fused vectors $\{v_{HF}^i\}_{i=1}^{T}$. For each v_{HF}^i, reshape it into a patch and then merge all the patches to get the final fused high pass coefficients. Perform inverse DWT over the fused sub bands to reconstruct the final fused image.

IV. Performance Metrics

Fusion quality evaluation is essential to analyze the fusion result. There are two kinds of evaluation methods. One is a subjective evaluation method, which involves making conclusions relying on the judgment of human eye and human brain. Other one is an objective assessment method, where the evaluation of the fusion image depends on some objective parameters from the fusion image. Generally standard deviation, spatial frequency, entropy [18,19], quality metrics $Q^{AB/F}$ and Q_w [17,18] reflect the quality of the fused image. These fusion metrics are described in the following sections.

(1) Entropy (H) is considered as one of the vital image quality index to evaluate the information content in the fused image. It is defined as,

$$H = -\sum_{i=0}^{255} p_F(i1) \log_2 p_F(i1) \tag{9}$$

where $p_F(i1)$ stands for probability of intensity value $i1$ in image F Higher the value of the entropy indicates that the fused image has better information content.

(2) Standard Deviation (σ) is considered as one of the best metrics for contrast value measurement of the fused image. The image has a better contrast if the standard deviation value is higher. It is given by,

$$\sigma = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (F(i,j) - \mu)^2} \tag{10}$$

where μ - mean of the gray scale F.

(3) Spatial Frequency (SF) reflects the amount of edge information preserved in the fused image. It is defined as,

$$SF = \sqrt{RF^2 + CF^2} \tag{11}$$

where RF and CF indicates the row frequency and column frequency of the fused image F.

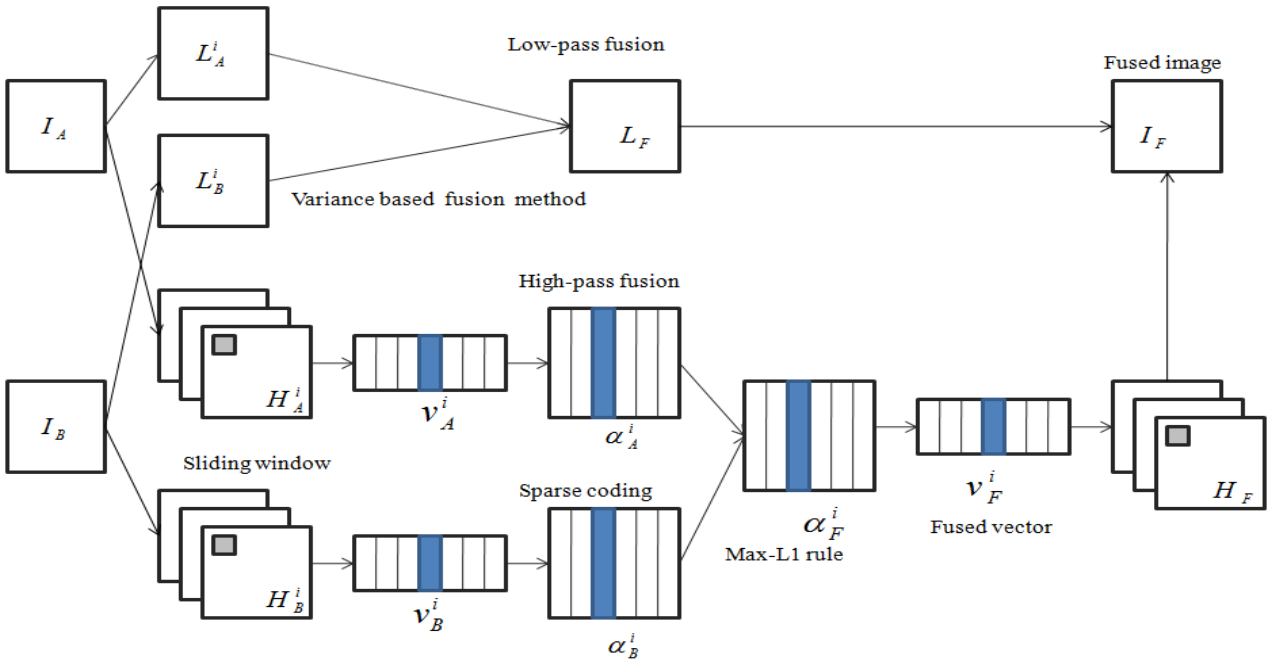


Fig. 1. Schematic diagram of the proposed fusion scheme

RF and *CF* are calculated as,

$$RF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}[F(i,j)-F(i,j-1)]^2} \tag{12}$$

$$CF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}[F(i,j)-F(i-1,j)]^2} \tag{13}$$

(4) Petrovic's metric $Q^{AB/F}$ [20] evaluates the relative amount of edge information that is transferred from the input images to the composite image. It is calculated by,

$$Q^{AB/F} = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} Q^{AF}(n,m)\,w^{A}(n,m) + Q^{BF}(n,m)\,w^{B}(n,m)}{\sum_{i=1}^{M}\sum_{j=1}^{N}(w^{A}(i,j)+w^{B}(i,j))} \tag{14}$$

where $Q^{AF}(n,m) = Q_g^{AF}(n,m)Q_\alpha^{AF}(n,m)$. Edge strength and orientation preservation values at pixel (n,m) are denoted by $Q_g^{AF}(n,m)$ and $Q_\alpha^{AF}(n,m)$. The weighting factors $w^A(n,m)$ and $w^B(n,m)$ indicates the significance of Q^{AF} and Q^{BF}. The definition of Q^{BF} is the same as that of Q^{AF}. The range of $Q^{AB/F}$ is between 0 and 1.

(5) The universal image quality index (UIQI) [21] based fusion metric proposed by Piella and Heijmans [22]. Q_w is defined as,

$$Q_w = \sum_{w\in W} C(w)\lambda(w)Q_0(A,F/w) + (1-\lambda(w))Q_0(B,F/w) \tag{15}$$

where $Q_0(A,F/w)$ and $Q_0(B,F/w)$ are calculated using the method in [21] in a local sliding window *w*.

The saliency weight is calculated by

$$\lambda(w) = \frac{s(A/w)}{s(A/w)+s(B/w)} \tag{16}$$

where the saliency measure $s(A/w)$ and $s(B/w)$ are calculated with the variance of *A* and *B* in window *w*, respectively. $C(w)$ is the normalized salience of *w* among all the local windows. $C(w)$ is obtained by,

$$C(w) = \frac{\max(s(A/w),s(B/w))}{\sum_{w'\in W}\max(s(A,w'),s(B,w'))} \tag{17}$$

A larger value of Q_w generally indicates a better fused result.

V. Experimental Results and Analysis

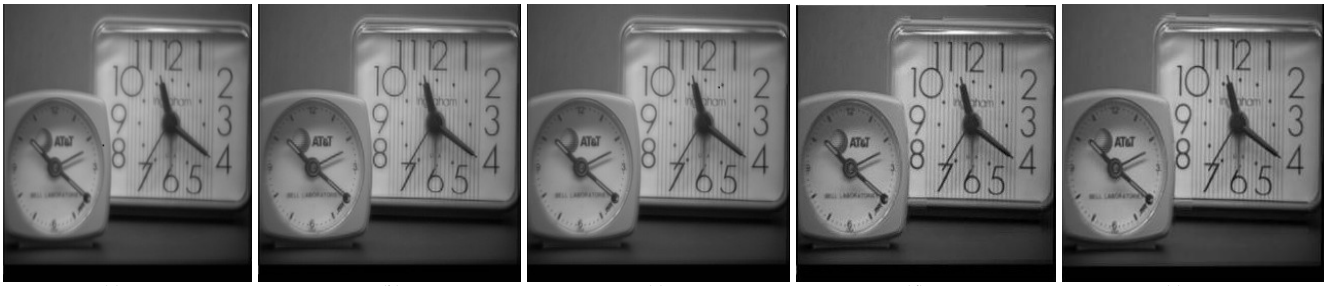
The performance of the proposed fusion method is tested on various multi-focus image pairs of size 256×256 as shown in Fig. 2. The proposed method is compared with Ratio pyramid (RP) [13], DWT, Shift invariant DWT (SIDWT) [23, 24]. Both fixed DCT dictionary and learned dictionary using K-SVD is used to verify the effectiveness of proposed method. The error tolerance ε is set to 0.1. The size of over-complete dictionary is set to 256 and the iteration number of K-SVD is fixed to 10. Figure 3 and 4 shows the simulation results of "clock" and "Lab" image pairs. It is clear that the proposed method has better contrast and good image quality in the fused image than existing methods.

The objective evaluation indices of "Book", "Leopard" and "Bottle" for the proposed method and transform based methods such as RP, DWT, SIDWT are compared and given in Table 1. Tables 2 compares the fusion metrics of "clock", "pepsi", "disk" and "lab" images for the proposed method and transform based methods.

Fig. 2. Multi-focus source image pairs.

As indicated in the Table 1 and Table 2, DWT-SR fusion method using DCT dictionary gives better values in terms of *H*, *SF* and $Q^{AB/F}$ compared to learned dictionary using K-SVD. This indicates that better spatial details are preserved in the proposed method using fixed DCT dictionary.

DWT-SR using K-SVD gives better contrast than fixed DCT dictionary. Also, the values of Q_w is almost same for both fixed and learned dictionaries. Hence, it is evident that DWT-SR using fixed DCT dictionary is more suitable for multifocus image fusion.

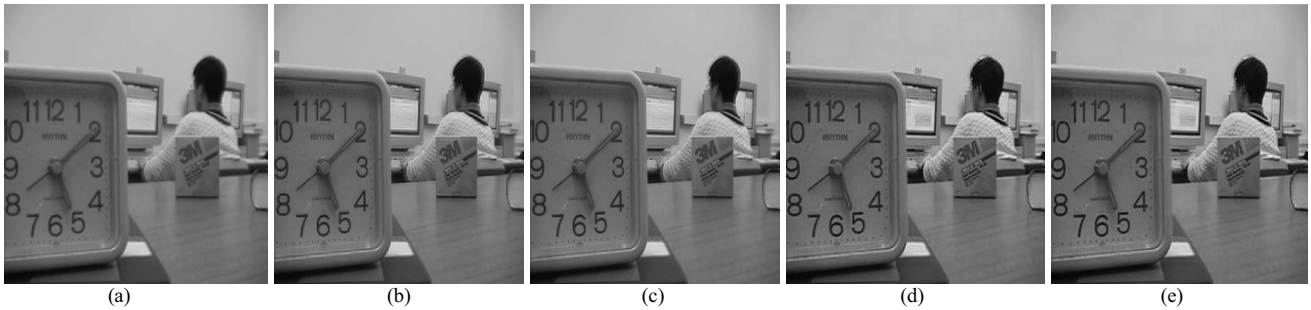


(a) (b) (c) (d) (e)

Fig. 3. Fusion results of "Clock" image. (a) RP (b) DWT (c) SIDWT (d) DWT-SR using K-SVD (e) DWT-SR using DCT.

TABLE I. OBJECTIVE ASSESSMENT OF "BOOK", "LEOPARD","BOTTLE" IMAGES FOR DIFFERENT METHODS

Source images	Methods		Performance Metrics				
			H	*SD*	*SF*	$Q^{AB/F}$	Q_w
Book	RP	[13]	7.3007	58.6275	22.0523	0.8257	0.9577
	DWT	[23]	7.3359	58.9935	25.3364	0.8431	0.9517
	SIDWT	[24]	7.3183	58.9027	24.6492	0.8100	0.9623
	DWT-SR using K-SVD	Proposed	7.3546	59.0699	25.8041	0.8090	**0.9665**
	DWT-SR using DCT	Proposed	**7.3563**	**59.0894**	**25.9533**	**0.8479**	0.9510
Leopard	RP	[13]	7.3596	63.6397	18.0795	0.8408	0.8990
	DWT	[23]	**7.3643**	64.0238	21.0585	0.8300	0.9216
	SIDWT	[24]	7.3632	64.0671	20.9479	**0.8566**	0.9281
	DWT-SR using K-SVD	Proposed	7.3633	**65.1985**	24.0808	0.8428	**0.9309**
	DWT-SR using DCT	Proposed	7.3633	65.1916	**24.1144**	0.8429	0.9307
Bottle	RP	[13]	7.5412	64.9429	26.1966	0.5082	0.6838
	DWT	[23]	7.5469	66.9005	33.5189	0.5660	0.7960
	SIDWT	[24]	7.6313	66.9652	32.6635	0.6439	0.8173
	DWT-SR using K-SVD	Proposed	7.5656	**70.0331**	38.8042	0.6773	**0.8727**
	DWT-SR using DCT	Proposed	**7.6669**	69.9132	**39.2995**	**0.6825**	0.8715



(a) (b) (c) (d) (e)

Fig. 4. Fusion results of "Lab" image. (a) RP (b) DWT (c) SIDWT (d) DWT-SR using K-SVD (e) DWT-SR using DCT.

TABLE II. OBJECTIVE ASSESSMENT OF "CLOCK", "PEPSI", "LAB", "DISK" IMAGES FOR DIFFERENT METHODS

Source images	**Methods**		**Performance Metrics**				
			H	*SD*	*SF*	$Q^{AB/F}$	Q_w
Clock	RP	[13]	7.2810	49.0859	10.5859	0.6621	0.7708
	DWT	[23]	7.3205	49.5632	13.3972	0.6726	0.8325
	SIDWT	[24]	7.2910	49.5041	13.2212	0.7107	0.8205
	DWT-SR using K-SVD	Proposed	7.3196	**51.1425**	16.2569	0.7210	**0.8957**
	DWT-SR using DCT	Proposed	**7.3216**	51.1267	**16.2723**	**0.7243**	0.8955
Pepsi	RP	[13]	7.0065	43.6136	9.7679	0.6786	0.8086
	DWT	[23]	7.0145	43.9916	13.976	0.7155	0.8806
	SIDWT	[24]	7.0135	43.9511	13.3629	**0.7481**	0.8857
	DWT-SR using K-SVD	Proposed	7.0317	**44.5915**	15.4655	0.7430	**0.9061**
	DWT-SR using DCT	Proposed	**7.0334**	44.5885	**15.4840**	0.7445	0.9060
Lab	RP	[13]	6.9415	45.1196	11.7439	0.7081	0.8448
	DWT	[23]	6.9712	45.6053	14.8926	0.6978	0.8727
	SIDWT	[24]	6.9735	45.6607	14.7338	0.7387	0.8829
	DWT-SR using K-SVD	Proposed	7.0118	**47.0138**	17.4985	0.7390	**0.8840**
	DWT-SR using DCT	Proposed	**7.0139**	47.0118	**17.5121**	**0.7415**	**0.8840**
Disk	RP	[13]	7.1734	43.7213	13.4043	0.6702	0.8188
	DWT	[23]	7.2131	44.3410	17.5614	0.6670	0.8592
	SIDWT	[24]	7.2114	44.3916	17.2669	0.7166	0.8691
	DWT-SR using K-SVD	Proposed	7.2670	**45.8147**	20.9018	0.7239	**0.8864**
	DWT-SR using DCT	Proposed	**7.2673**	45.8056	**20.9444**	**0.7242**	0.8804

VI. CONCLUSION

In this paper, we proposed a novel multifocus image fusion algorithm using DWT and sparse representation. Initially, the low frequency components and high frequency components are separated using DWT. Image variance is used as a focus measure to fuse the low frequency sub bands. SR based fusion rule is applied to obtain the fused high frequency coefficients. Finally, inverse DWT is applied to get the fused image. According to the fusion results, it was observed that the proposed method can well preserve the focused regions of source images with reduced artifacts. Furthermore, the proposed method using fixed DCT dictionary performs well as compared to learned dictionary in terms of objective evaluations.

ACKNOWLEDGMENT

We would like to thank Yu Liu for providing the source images [25]. We also thank Rockinger for sharing the Matlab package [26] used in our experiments.

REFERENCES

[1] Wan, C. Zhu, Z. Qin, "Multifocus image fusion based on robust principal component analysis", *Pattern Recog. Lett*., vol. 34, pp. 1001–1008, 2013.

[2] S. Li, B. Yang, J. Hu, "Performance comparison of different multi-resolution transforms for image fusion", *Inf. Fusion*, vol. 12, pp. 74–84, 2011.

[3] S. Li, X. Kang, J. Hu, B. Yang, "Image matting for fusion of multi-focus images in dynamic scenes", *Inf. Fusion*, vol. 14, pp. 147–162, 2013.

[4] Z. Liu, K. Tsukada, K. Hanasaki, Y.K. Ho, Y.P. Dai, "Image fusion by using steerable pyramid", *Pattern Recog. Lett*., vol. 22, pp. 929–939, 2001.

[5] J. Tian, L. Chen, "Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure", *Signal Process*., vol. 92, pp. 2137–2146, 2012.

[6] J.J. Lewis, R.J. O'Callaghan, S.G. Nikolov, D.R. Bull, N. Canagarajah, "Pixel and region based image fusion with complex wavelets", *Inf. Fusion*, vol. 8, pp. 119–130, 2007.

[7] B. Yang, S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit", *Inf. Fusion*, vol. 13, pp. 10–19, 2012.

[8] B. Yang, S. Li, "Multifocus image fusion and restoration with sparse representation", *IEEE Trans. Instrum. Meas*., vol. 59, pp. 884–892, 2010.

[9] Y. Asnath Victy Phamila, R. Amutha, "Discrete Cosine Transform based fusion of multi-focus images for Visual Sensor Network", *Signal Process*., vol. 95, pp. 161-170, 2014.

[10] N. Aishwarya, Y. Asnath Victy Phamila, R. Amutha, "Multi-focus image fusion using multi structure top hat transform and image variance", *in Proc. of IEEE Int. Conf. Commun. Signal Process*., pp. 686-689, 2013.

[11] Yong Yang, "A Novel DWT Based Multi-focus Image Fusion Method", *Procedia Engineering*, vol. 24, pp. 177-181, 2011.

[12] Nirmala Paramanandham, Kishore Rajendiran, Deepika Narayanan, Indu Vadhani.S, Mrinalini Anand, "An efficient Multi transform based Fusion for Multi Focus Images", *IEEE Int. Conf. Commun. Signal Process*., pp. 984-988, 2015.

[13] A. Toet, "Image fusion by a ratio of low pass pyramid", *Pattern Recog. Lett*., vol. 4, 1989, pp. 245–253.

[14] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, S. Yan, "Sparse representation for computer vision and pattern recognition", *Proc. IEEE*, vol. 98, pp. 1031–1044, 2010.

[15] M. Elad, M. Figueiredo, Y. Ma, "On the role of sparse and redundant representations in image processing", *Proc. IEEE*, vol. 98, pp. 972–982, 2010.

[16] M. Aharon, M. Elad, A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation", *IEEE Trans. Signal Process*., vol. 54, pp. 4311–4322, 2006.

[17] M. Elad, M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries", *IEEE Trans. Image Process*., vol. 15, pp. 3736-3745, 2006.

[18] Wei Huang, Zhongliang Jing, "Evaluation of focus measures in multi-focus image fusion", *Pattern Recog. Lett*., vol. 28, pp. 493-500, 2007.

[19] P. Jagalingam, Arkal Vittal Hegdeb, "A Review of Quality Metrics for Fused Image", *Aquatic Procedia*, vol. 4, pp. 133 – 142, 2015.

[20] C.S Xydeas and V. Petrovic, "Objective Image Fusion Performance Measure", *Electron. Lett*., vol. 36, pp. 308-309, 2000.

[21] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Trans. Image Process.,* vol. 13, pp. 600–612, 2004.

[22] G. Piella, H. Heijmans, "A new quality metric for image fusion", *in Proc. Int. Conf. Image Process*., pp. 173–176, 2003.

[23] Hui Li, B. Manjunath, S.K Mitra, "Multi sensor image fusion using the wavelet transform", *Graph. Models Image Process*., vol. 3, pp. 235-245, 1995.

[24] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform*", in Proc. IEEE Int. Conf. Image Process*., vol. 3, pp. 288-291, 1997.

[25] Source images available at http://home.ustc.edu.cn/~liuyu1/.

[26] http://www.metapix.de/toolbox.html