

# MULTI-FOCUS PIXEL-BASED IMAGE FUSION IN DUAL DOMAIN

Zhiyu CHEN

Niigata University  
Graduate School of Science and Technology  
8050 2-no-cho, Ikarashi, Nishi-ku,  
Niigata, 950-2181, JAPAN  
chenzhiyu@telecom0.eng.niigata-u.ac.jp

Shogo MURAMATSU

Niigata University  
Dept. of Electrical and Electronic Eng.  
8050 2-no-cho, Ikarashi, Nishi-ku,  
Niigata, 950-2181, JAPAN  
shogo@eng.niigata-u.ac.jp

## ABSTRACT

Redundant-transform-based image fusion approaches require a lot of memory and computations. This paper proposes an effective and efficient multi-focus image fusion technique in dual domain using a redundant transform. The proposed scheme captures high-frequency informations, e.g. edges and slant textures, of images efficiently, and reduce the computational cost. The proposed scheme extracts the high-frequency information of images with multiple directional lapped orthogonal transforms (M-DirLOTs) through the following procedure: (1) reconstruct the detail image using high-pass subbands, (2) execute a fusion operation in spatial domain through joint measurement and (3) improve the performance by mathematical morphology processing. The proposed method overcomes some disadvantages of traditional transform-based and spatial-based fusion techniques. Experimental results show that the proposed method is able to significantly improve the fusion performance.

**Index Terms**— Dual domain, Image fusion, M-DirLOTs, Slant texture, Joint measurement, Mathematical morphology

## 1. INTRODUCTION

It is difficult to clearly acquire all objects in the same scene, because the focused range of visible imaging system is limited. A well-focused image is comparatively clear while a defocused image is blur. Image fusion is a scheme to improve the quality of information from a set of images by combining relevant information from multiple images into a single image. The fused image will be more informative than any of the original images. Obtaining a focused image is an essential task for human perception and machine vision. Image fusion is a good approach to acquire a well-focused image for providing reliable and accurate information by combining multi-sensor data [1]. It is basically realized by combining well-focused clear parts of multiple source images.

In the pixel-based fusion, the pixels are fused by selecting either value on corresponding position or weighted average of multiple pixels. Attributions for the fusion commonly include the average value, standard deviation and energy etc. So far, various approaches of multi-focused image fusion have been developed [2] – [5]. Some approaches select pixels from clear parts of source images in the spatial domain or feature domain to compose a sharp image [2], [3], [4]. These approaches, however, require edge and texture detection and the detection has great influence on the quality of image fusion. Generally, fusion algorithms take the single pixel or pixels in a local re-

gion into account. This approach, however, could lead image degradation to the result, e.g. reduced contrast and blocking artifacts, and hardly depends on the adopted algorithm.

Another approach of image fusion combines coefficients of a multi-scale transform under a premise that the detailed information of image is distributed in high frequency subbands. This transform-based approach is called coefficient-based fusion. Compared with the pixel-based method, the coefficient-based one can preserve more detail information of the source images. The basic steps of the procedure are as follows: Firstly, analyze every image using a discrete wavelet transform (DWT) decomposition. Secondly, fuse individual subband, where different fusion algorithms are used. Thirdly, reconstruct a picture using an inverse DWT.

The discrete cosine transform (DCT), Curvelet and Contourlet can also be used as the transform. Images are effectively decomposed into approximation and detail coefficients. However, these transforms have some disadvantages. The traditional separable transforms have limited directional characteristics. Curvelet has a question how to construct a tight curvelet-like transform in discrete domain. Contourlet causes pseudo-Gibbs phenomena easily. In order to overcome these problems, some new constructions such as dual-tree complex wavelet transform (DT-CWT), non-subsampled contourlet transform (NSCT) and non-subsampled shearlet transform (NSST) were proposed [6], [7], [8]. 2-D DT-CWT has six subbands that give directional information of an image, where the angles are set to  $\pm 15, \pm 45, \pm 75$  degrees. The construction is based on real 1-D two-channel filter banks, and thus the design and implementation are not complicated. However, the structure is somehow restrictive. NSCT and NSST can capture two-dimensional geometrical structure much more effectively than traditional multi-scale transforms. There, however, remains a disadvantage that NSCT and NSST cannot satisfy orthogonality, and the computational complexities and redundancy are high.

For more effective representation of images, we proposed a new transform, M-DirLOTs, which is a union of directional filter banks that can represent slant contours, textures and gradation with few coefficients [9]. M-DirLOTs can represent edges and other singularities along trend surfaces efficiently [5], [10]. Remind that the performances of spatial domain fusions were restricted by an adopted detail information detection algorithm. As well, traditional fusion methods in redundant transform domain need high computational costs. The computational cost is proportional to the redundancy.

In order to improve the fusion performance, we propose an effective image fusion method based on M-DirLOTs and joint measurement in dual domain. Although fusion of two images is illustrated in this paper, it is not hard to be extended for multiple images by

This work was supported in part by JSPS KAKENHI (No.26420347).

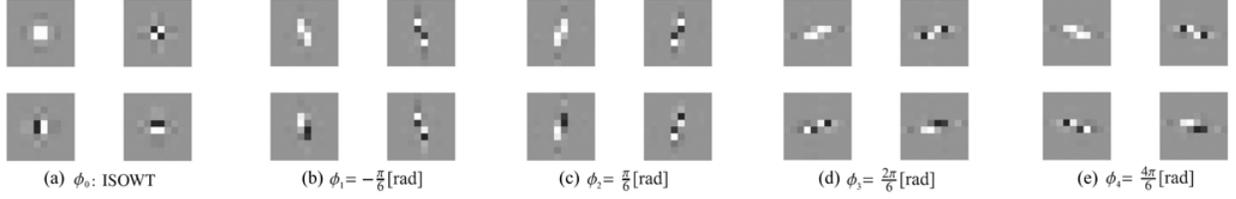


Fig. 1: Examples of atomic images of M-DirLOTs.

analogy. The following is the structure of this paper. A brief introduction of M-DirLOTs is given in Section II. In Section III, we will propose an efficient image fusion method. The experimental results and discussion are presented in Section IV in order to verify the significance of the proposed method. Finally, Section V concludes this paper.

## 2. REVIEW OF MULTIPLE DIRLOTS

M-DirLOTs provide multi-scale and multi-directional decomposition. We use M-DirLOTs to analyze an image, which can acquire geometrical features of images and provide significant high-frequency information.

A single DirLOT is able to simultaneously satisfy the following three properties: orthogonality, symmetry, and overlapping property with a non-separable basis. This transform can be constructed with a lattice structure [11]. In addition, DirLOT can also satisfy the fixed-critically-subsampling, real-valued, and compact-support properties. Furthermore, it can hold the trend vanishing moments (TVMs) for any direction. The directional property works well for slant textures and edges.

Aiming at expressing the oblique textures and edges in various directions better, we define a dictionary  $\mathbf{D}$  by multiple DirLOTs as

$$\mathbf{D} = [\Psi_{\phi_0}^T \Psi_{\phi_1}^T \Psi_{\phi_2}^T \Psi_{\phi_3}^T \dots \Psi_{\phi_{R-1}}^T]^T,$$

where  $\Psi_{\phi_0}$  is a nondirectional isotropic symmetric orthonormal DWT (ISOWT) with the classical two-order vanishing moments (VMs), and  $\Psi_{\phi_k}$  is a directional anisotropic symmetric orthonormal wavelet transforms constructed by DirLOTs with the two-order TVMs for the direction  $\phi_k$ .  $R$  corresponds to the redundancy. Fig.1 shows examples of single-level atomic images in  $\mathbf{D}$ , which constructs a normalized tight frame and satisfy

$$\mathbf{D}^T \mathbf{D} = \sum_{k=0}^{R-1} \Psi_{\phi_k}^T \Psi_{\phi_k} = \mathbf{R} \mathbf{I}.$$

A heuristic approach takes the average of results obtained by independent reconstruction with  $\Psi_{\phi_k}$  for  $k = 0, 1, \dots, R-1$ . It is simply realized by

$$\hat{\mathbf{x}}_A = \frac{1}{R} \mathbf{D}^T \mathbf{c}_A = \frac{1}{R} \sum_{k=0}^{R-1} \Psi_{\phi_k}^T \Theta(\Psi_{\phi_k} \mathbf{x}_A),$$

where  $\mathbf{x}_A$  and  $\hat{\mathbf{x}}_A$  are vector representations of input image  $A$  and reconstructed image  $\hat{A}$ , respectively.  $\Theta(\cdot)$  is an operation for the coefficients. In this paper, M-DirLOTs of polyphase order  $N_y = N_x = 4$  were adopted. The atom size is  $10 \times 10$ . The TVM angles  $\phi_1, \phi_2, \phi_3$  and  $\phi_4$  were set to  $-\frac{\pi}{6}, \frac{\pi}{6}, \frac{2\pi}{6}$  and  $\frac{4\pi}{6}$ , respectively.

## 3. PROPOSED IMAGE FUSION METHOD

In [5], a fusion method based on M-DirLOTs and sum-modified-Laplacian was introduced. The fusion processing was implemented in M-DirLOTs domain. When the redundancy of M-DirLOTs is set to  $R$ , the fusing attributions are calculated for  $R$  times. In order to reduce the computational cost and obtain effective high-frequency information, we propose to use M-DirLOTs to analyze input images ( $A$  and  $B$ ), and reconstruct detail images ( $A^H$  and  $B^H$ ) with high-frequency subbands. Consequently, the approximation images  $A^L = A - A^H$  and  $B^L = B - B^H$  are computed. The fusion operation will be performed in spatial domain. The computational cost for fusion of detail images in Sec3.1.2 can be significantly reduced.

### 3.1. Spatial Domain Fusion based on Joint Measurement

A classical fusion rule takes an average of low-frequency coefficients and selects absolute maximum of high-frequency coefficients in DWT domain [12]. We propose to apply this DWT domain rule to the spatial domain. That is, the fusion rules are directory used for the approximation and detail images instead of subband coefficients.

#### 3.1.1. Fusion of Approximation Images

The fusion rule takes average of approximation images. The fusion rule among the pixels is represented by

$$F^L = \frac{1}{2}(A^L + B^L), \quad (1)$$

where  $F^L$  denotes an approximation of the fusion image.

#### 3.1.2. Fusion of Detail Images

Salient features of a given image, such as edges, appear as large absolute values of high-frequency informations. Thus, we adopt a fusion rule to extract maximum absolute value of the corresponding pixel of high-frequency informations. In [13], spatial frequency (denoted as  $S$ ) was used as a sharpness measurement which evaluates the variation of pixel values. It can be represented by

$$S_A^H = \sqrt{(S_A^{W,H})^2 + (S_A^{E,H})^2}, \quad (2)$$

where  $\sqrt{\cdot}$  and  $(\cdot)^2$  denote an element-wise square root and squared operation,  $S_A^{W,H}$  and  $S_A^{E,H}$  are the row frequency

$$S_A^{W,H} = \sqrt{\frac{1}{MN} \sum_{m=-P}^P \sum_{n=-Q+1}^Q [A^H(m, n) - A^H(m, n-1)]^2}$$

and column frequency

$$S_A^{E,H} = \sqrt{\frac{1}{MN} \sum_{m=-P+1}^P \sum_{n=-Q}^Q [A^H(m, n) - A^H(m-1, n)]^2},$$

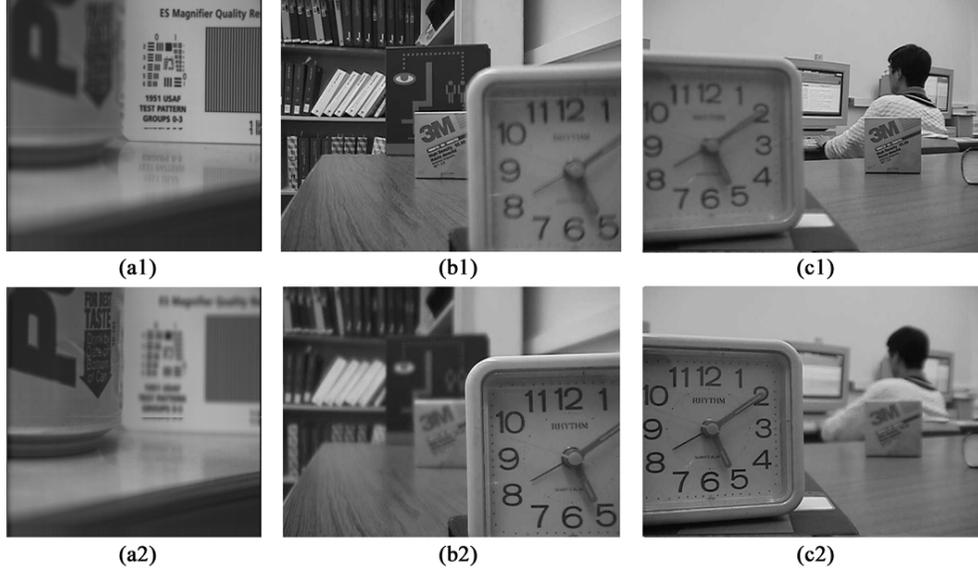


Fig. 2: Multi-focus source images.

respectively. The parameters  $P$  and  $Q$  set the window size  $(2P + 1) \times (2Q + 1)$ , which are used to compute the  $S^W$  and  $S^E$ . In this paper, the size is set to  $(2P + 1) \times (2Q + 1) = 3 \times 3$ . In the same manner,  $S_B^H$  is calculated.

Respecting the above mentioned facts, two metrics in (1) and (2) have different features. Hence, we propose a combination metric to improve the fusion performance. Considering the human visual contrast (e.g. sensitivity to local contrast change, edges, and directional features), the local luminance contrast was developed [14]. Meanwhile, Gaussian low-pass filter (denoted as  $G$ ) of size  $3 \times 3$  is used to provide good characterization of the high-frequency information. We define, the fusion-map-based joint measurement as follows:

$$\mathbf{map}(i, j) = \begin{cases} 1, & \text{if } G\left(\frac{S_A^H \cdot \text{abs}(A^H)}{(\bar{A}^L)^{2(1+\alpha)}}\right)(i, j) \geq G\left(\frac{S_B^H \cdot \text{abs}(B^H)}{(\bar{B}^L)^{2(1+\alpha)}}\right)(i, j) \\ 0, & \text{otherwise} \end{cases},$$

where  $\bar{A}^L$  and  $\bar{B}^L$  are the mean coefficients of approximation images for each pixel in  $3 \times 3$  windows.  $\alpha$  is a visual constant, which is set by perceptual experiment, ranging from 0.6 to 0.7 [15]. The standard deviation of  $G$  is set to 5 from experience. The detail image of fused image is constructed by

$$F^H = \mathbf{map} \cdot A^H + (\mathbf{1} - \mathbf{map}) \cdot B^H.$$

### 3.2. Refinement of Focused Region Detection

In [16], a detection of focused regions was presented. The procedure is as follows:

1. Fuse  $F^L$  and  $F^H$  to get an initial fused image.
2. Calculate the RMSE (root mean square error) of each pixel within a local area ( $7 \times 7$ ) between the source images and the initial fused image.
3. Compare  $\text{RMSE}_A(i, j)$  and  $\text{RMSE}_B(i, j)$  to determine fusion map  $\mathbf{Z}$  (logical matrix), where '1' in  $\mathbf{Z}$  indicates the pixel at position  $(i, j)$  in image A is fused, i.e.,  $\text{RMSE}_A(i, j) \leq \text{RMSE}_B(i, j)$ , and '0'

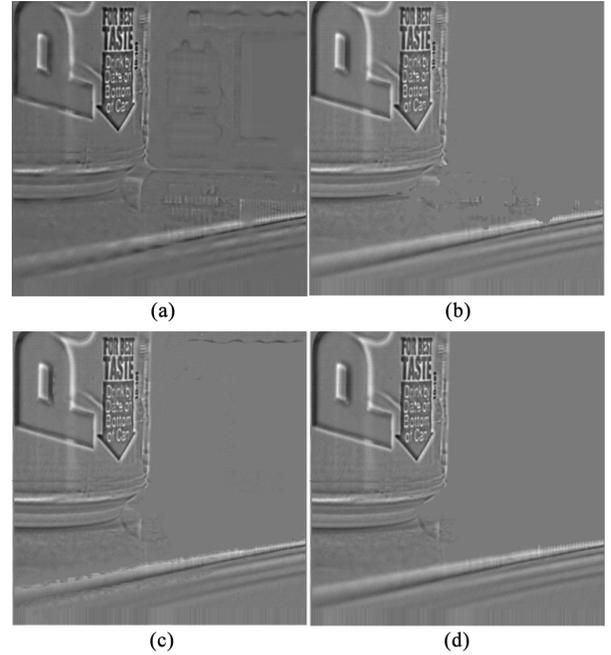
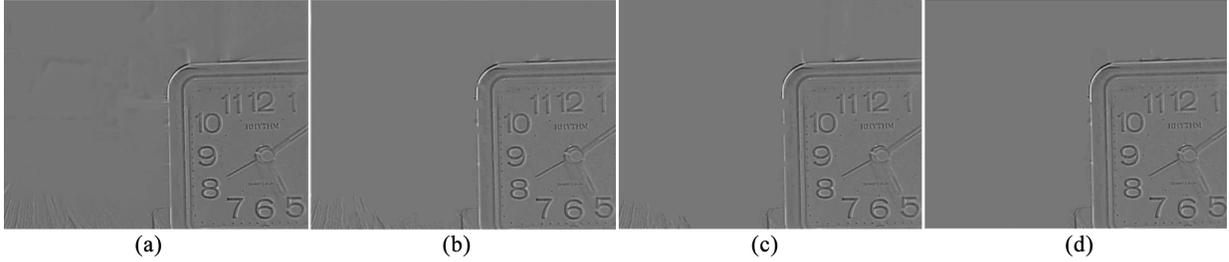


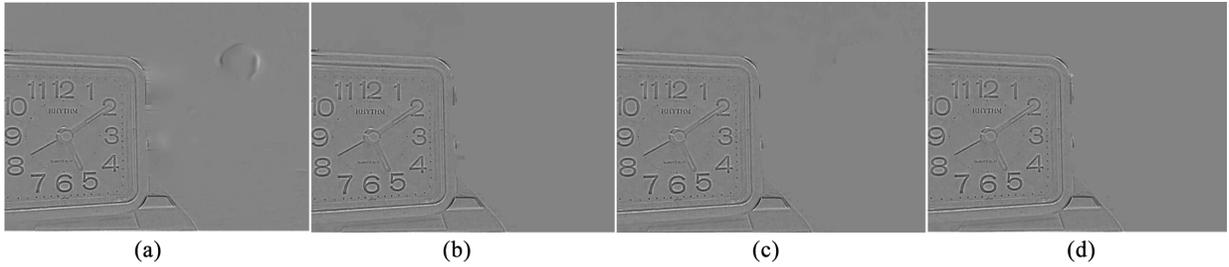
Fig. 3: "Pepsi" fused images of different methods. (a)-(d) are the difference images between fused images and Fig.2(a1), where Li's method, Bai's method, Liu's method and proposed method, respectively.

indicates the pixel in B is fused.

4. Morphological opening ( $\mathbf{Z} \circ \mathbf{B}$ ) and closing ( $\mathbf{Z} \bullet \mathbf{B}$ ) with small structural element are used to optimize  $\mathbf{Z}$  (thin connections, thin protrusions, narrow breaks, fill long thin gulfs). Meanwhile, a threshold  $T_{th}$  is set to remove the holes smaller than the threshold. The struc-



**Fig. 4:** “Disk” fused images of different methods. (a)-(d) are the difference images between fused images and Fig.2(b1), where Li’s method, Bai’s method, Liu’s method and proposed method, respectively.



**Fig. 5:** “Lab” fused images of different methods. (a)-(d) are the difference images between fused images and Fig.2(c1), where Li’s method, Bai’s method, Liu’s method and proposed method, respectively.

**Table 1:** Comparison of MI and  $Q^{AB/F}$  among four methods

Method	MI			$Q^{AB/F}$		
	Pepsi(512 × 512)	Disk(480 × 640)	Lab(480 × 640)	Pepsi(512 × 512)	Disk(480 × 640)	Lab(480 × 640)
Li’s method[2]	7.3762	7.0630	7.9109	0.7940	0.7243	0.7505
Bai’s method[3]	8.8769	<b>8.3003</b>	8.6841	0.7920	0.7385	0.7589
Liu’s method[4]	8.6289	8.2165	8.5201	0.7885	0.7364	0.7585
Proposal	<b>8.9189</b>	8.2885	<b>8.8292</b>	<b>0.7943</b>	<b>0.7388</b>	<b>0.7591</b>

tural element  $\mathbf{B}$  is a  $5 \times 5$  matrix with logical 1’s and  $T_{th}$  is set according to the experimental results. The resulting logical matrix denoted as  $\mathbf{Z}'$ . Morphological operations are again performed to smooth object contours [17].

5. The final image  $F$  is restored by map  $\mathbf{Z}'$ .

The final fused image can be described by

$$F = \mathbf{Z}' \cdot A + (\mathbf{1} - \mathbf{Z}') \cdot B.$$

#### 4. EXPERIMENTAL RESULTS

In order to confirm the effectiveness of the proposed image fusion method, some experiments were conducted. In these experiments, the number of hierarchical levels is set to five. “Pepsi”, “Disk” and “Lab” were used. Fig.2 shows multi-focus source images. The Li’s[2], Bai’s[3] and Liu’s[4] methods were used as references, and three algorithms were implemented using MATLAB programs provided by their authors.

We use mutual information (MI) and  $Q^{AB/F}$  as information measures for evaluating image fusion performance were used. MI represents how much of the information in the fused image was obtained from the input images [18].  $Q^{AB/F}$  uses Sobel edge detection to measure how much edge information in the fused image can be obtained

from the source images [19].

Figs.3-5 show the experimental results. It can be seen that the proposed method shows better quality for edges and slant textures compared with the results of Li’s, Bai’s and Liu’s method. Table 1 compares the MI performances and  $Q^{AB/F}$  metric among four methods. We can see that the proposed method outperforms the other methods except the case that the MI of “Disk”.

The experimental results imply that proposed method are not only able to reproduce the structure appropriately, but also generate good results. By introducing directional decomposition into the transform, M-DirLOTs can efficiently obtain diagonal components present in natural images. The directional property works well for edges and slant textures.

#### 5. CONCLUSIONS

Image fusion methods based on DWT for multi-focus images were reviewed. It was pointed out that the representation of geometric structures is relatively insufficient by using traditional transform. In order to solve this problem, M-DirLOTs was introduced to improve the image fusion quality for edges and textures. Meanwhile, the new framework was applied to reduce computational cost. Experimental results show the imaging fused performance was improved, and their effectiveness has been verified.

## 6. REFERENCES

- [1] A. A. Goshtasby, S. Nikolov and Guest editorial, "Image fusion: Advances in the state of the art," *Information Fusion*, vol. 8, pp. 114-118, Apr. 2007.
- [2] Shutao Li, Xudong Kang and Jianwen Hu, "Image Fusion with Guided Filtering," *IEEE Trans. on Image Process.*, vol. 22, no. 7, pp. 2864-2875, Jul. 2013.
- [3] Xiangzhi Bai, Yu Zhang, Fugen Zhou, Bindang Xue, "Quadtree-based multi-focus image fusion using a weighted focus-measure," *Information Fusion*, vol. 22, pp. 105-118, Mar. 2015.
- [4] Yu Liu, Shuping Liu, Zengfu Wang, "Multi-focus image fusion with dense SIFT," *Information Fusion*, vol. 23, pp. 139-155, May. 2015.
- [5] Zhiyu Chen and Shogo Muramatsu, "Multi-focus Image Fusion based on Multiple Directional LOTs," *IEICE Trans. on Fundamentals*, vol.E98-A, no.11, pp. 2360-2365, Nov. 2015.
- [6] Selesnick, I. W., Baraniuk, R. G. and Kingsbury, N.C., "The dual tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no.6, pp. 123-151, Nov. 2005.
- [7] L.da Cunha, Jianping Zhou and Minh N. Do, "The non-subsampled contourlet transform: theory, design, and applications," *IEEE Trans. on Image Processing*, vol. 15, no. 10, pp. 3089-3101, Oct. 2006.
- [8] Glenn Easley, Demetrio Labate and Wang-Q Limc, "Sparse directional image representations using the discrete shearlet transform," *Appl. Comput.Harmon. Anal.*, vol. 25, no. 1, pp. 25-46, Jul. 2008.
- [9] Shogo Muramatsu and Dandan Han, "Image Denoising with Union of Directional Orthonormal DWTs," *IEEE Proc. ICASSP 2012*, pp.1089-1092, Mar. 2012.
- [10] Zhiyu Chen and Shogo Muramatsu, "SURE-LET Poisson Denoising with Multiple Directional LOTs," *IEICE Trans. on Fundamentals*, vol.E98-A, no.8, pp. 1820-1828, Aug. 2015.
- [11] S. Muramatsu, Dandan Han, T. Kobayashi, and H. Kikuchi, "Directional Lapped Orthogonal Transform: Theory and Design," *IEEE Trans. on Image Process.*, vol. 21, no. 5, May 2012.
- [12] Zhong Zhang and Rick S. Blum, "A Categorization of Multiscale-decomposition-based Image Fusion Schemes with a Performance Study for a Digital Camera Application," *Proc. of the IEEE*, vol. 87, no. 8, pp.1315-1326, Aug. 1999.
- [13] A.M. Eskicioglu and P.S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Commun.*, vol. 43 no. 12, pp. 2959-2965, Dec. 1995.
- [14] A. Toet, L. J. van Ruyven and J. M. Valetton, "Merging Thermal And Visual Images By A Contrast Pyramid," *Optical Engineering*, vol. 28, no. 7, pp. 789-792, Jul. 1989.
- [15] A. B. Watson, "Efficiency of a model human image code," *J. Opt. Soc. Am. A*, vol. 4, no. 12, pp. 2401-2417, Dec. 1987.
- [16] Huafeng Lib, Yi Chaia and Zhaofei Li, "Multi-focus image fusion based on nonsubsampling contourlet transform and focused regions detection," *Optik*, vol. 124, no. 1, pp. 40-51, Jan. 2013.
- [17] T. Stathaki, "Image fusion algorithms and applications," *Academic Press is an imprint of Elsevier*, 2008.
- [18] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137-154, Sep. 1997.
- [19] V. Petrovic and C. Xydeas, "Objective Image Fusion Performance Characterisation," *Proceedings of ICCV'05*, vol. 2, pp. 1866-1871, Oct. 2005.