Multi-Focus Image Fusion Based on Spatial Frequency in Discrete Cosine Transform Domain

Liu Cao, Longxu Jin, Hongjiang Tao, Guoning Li, Zhuang Zhuang, and Yanfu Zhang.

***Abstract*—Multi-focus image fusion in wireless visual sensor networks (WVSN) is a process of fusing two or more images to obtain a new one which contains a more accurate description of the scene than any of the individual source images. In this letter, we propose an efficient algorithm to fuse multi-focus images or videos using discrete cosine transform (DCT) based standards in WVSN. The spatial frequencies of the corresponding blocks from source images are calculated as the contrast criteria, and the blocks with the larger spatial frequencies compose the DCT presentation of the output image. Experiments on plenty of pairs of multi-focus images coded in Joint Photographic Experts Group (JPEG) standard are conducted to evaluate the fusion performance. The results show that our fusion method improves the quality of the output image visually and outperforms the previous DCT based techniques and the state-of-art methods in terms of the objective evaluation.**

***Index Terms*—Discrete cosine transform, image fusion, multi-focus, spatial frequency.**

I. INTRODUCTION

AS A result of the limited depth of focus in optical lenses, it is difficult to describe the complex situation with a single image accurately [1]. In wireless visual sensor networks, multiple sensors are applied to obtain images of the same scene, and a centralized fusion centre combines source images from multiple sensors into a single image, which is more suitable for human visual and machine perception [2]. Then, the fused image will be transmitted to an upper node.

So far, a lot of researches have concentrated on image fusion performed in the spatial domain [3]–[5]. Methods based on multi-scale transform such as discrete wavelet transform (DWT) [3], shift invariant discrete wavelet transform (SIDWT) [4], and non-subsampled contourlet transform (NSCT) [5] are popular. However, most of the image fusion approaches based on multi-scale transform are complex and time-consuming, which limits their applications for wireless visual sensor networks equipped with constrained resources.

Manuscript received June 18, 2014; revised August 19, 2014; accepted August 21, 2014. Date of publication September 04, 2014; date of current version September 12, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tolga Tasdizen.

L. Cao, Z. Zhuang, and Y. Zhang are with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033 China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: caol223@mail.ustc.edu.cn; 691117262@qq.com; 839020811@qq.com).

L. Jin, H. Tao, and G. Li are with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033 China (e-mail: jinlx@ciomp.ac.cn; hjtao@yahoo.com.cn; Lgning@netease.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2354534

In WVSN, images are compressed before transmission to the other nodes. When the source images are saved or transmitted in DCT based standards, the methods applied in DCT domain will reduce computation complexity considerably [6]. Recently, several image fusion techniques in DCT domain have been proposed. Tang *et al.* [7] proposed two methods in DCT domain, namely, $\mathrm{DCT} + \mathrm{Average}$ and $\mathrm{DCT} + \mathrm{Contrast}$. But these methods suffer some undesirable side effects like blurring or blocking artifacts which degrade the image quality. The algorithm proposed in [8] called $\mathrm{DCT} + \mathrm{AC}$- Max leads to mistakes in selecting right JPEG coded blocks because the number of higher valued AC coefficients is an invalid criterion when the most of the AC coefficients are quantized to zeros during the quantization. In another approach [9], variance is considered as a contrast criterion of fusion. However, experiment results in [10] show that variance provides worse performance than other focus measures.

In this letter, a general image fusion technique in DCT domain is proposed. Here, the image blocks with high spatial frequencies are absorbed to the fused image. A consistency verification procedure is followed to increase the quality of output image. Experimental results, performed on several databases which are coded in JPEG format, indicate our method improves the quality of the fused image considerably.

The rest of this letter is organized as follows: Section II demonstrates the basic concepts of our algorithm. Section III describes the proposed approach of image fusion. Section IV analyzes the experimental results, followed by conclusions in Section V.

II. DCT BLOCKS ANALYSIS

The discrete cosine transform (DCT) is one of the most widely used transform in image compression applications [11]. Several commercial standards widely used such as JPEG still image coding standard [12], Motion-JPEG, MPEG and the H263 video coding standards [13] are based on DCT.

Using vector processing, the output matrix F of a two-dimensional 8×8 DCT for an input matrix f is given by:

$$F = C.f.C^t \tag{1}$$

where C is an orthogonal matrix consisting of the cosine coefficients and C^t are the transpose coefficients.

$$C^{-1} = C^t \tag{2}$$

The inverse DCT (IDCT) is also defined as:

$$f = C^t.F.C \tag{3}$$

According to [10]:

$$trace(ff^t) = trace(FF^t) \tag{4}$$

where $trace(x)$ stands for the trace of x.

The row frequency (RF) and column frequency (CF) of an 8×8 image block are given by:

$$RF^2 = \frac{1}{8 \times 8} \sum_{x=0}^{7} \sum_{y=1}^{7} (f(x,y) - f(x,y-1))^2 \tag{5}$$

$$CF^2 = \frac{1}{8 \times 8} \sum_{x=1}^{7} \sum_{y=0}^{7} (f(x,y) - f(x-1,y))^2 \tag{6}$$

The total spatial frequency (SF) of an 8×8 block in the spatial domain is calculated as:

$$SF^2 = RF^2 + CF^2 \tag{7}$$

After a small amount of calculation, we can calculate the spatial frequency of the block from the AC coefficients in the DCT domain.

We denote Δx and Δy as the difference matrixes of rows and columns respectively:

$$\Delta x = \begin{pmatrix} f(0,1) - f(0,0) & \cdots & f(0,7) - f(0,6) & 0 \\ \vdots & \ddots & \vdots & \vdots \\ f(7,1) - f(7,0) & \cdots & f(7,7) - f(7,6) & 0 \end{pmatrix}$$

$$\Delta y = \begin{pmatrix} f(1,0) - f(0,0) & \cdots & f(1,7) - f(0,7) \\ \vdots & \ddots & \vdots \\ f(7,0) - f(6,0) & \cdots & f(7,7) - f(6,7) \\ 0 & \cdots & 0 \end{pmatrix}.$$

It's clear that:

$$\Delta x = fb = C^t F C C^t B C = C^t F B C, \tag{8}$$

$$\Delta y = b^t f = (C^t B C)^t C^t F C = C^t B^t F C, \tag{9}$$

$$\text{where} \quad b = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

and B is the DCT presentation of b.

From (3), (8), and (9), we can find FB and $B^t F$ are the DCT presentations of Δx and Δy, respectively. Then, we get:

$$\begin{aligned} RF^2 &= \frac{1}{8 \times 8} \sum_{x=0}^{7} \sum_{y=0}^{7} \Delta x^2(x,y) = \frac{1}{8 \times 8} trace(\Delta x (\Delta x)^t) \\ &= \frac{1}{8 \times 8} trace(FB(FB)^t) = \frac{1}{8 \times 8} trace(FBB^t F^t) \end{aligned} \tag{10}$$

$$\begin{aligned} CF^2 &= \frac{1}{8 \times 8} \sum_{x=0}^{7} \sum_{y=0}^{7} \Delta y^2(x,y) = \frac{1}{8 \times 8} trace((\Delta y)^t \Delta y) \\ &= \frac{1}{8 \times 8} trace((B^t F)^t B^t F) = \frac{1}{8 \times 8} trace(F^t B B^t F) \end{aligned} \tag{11}$$

Let D be the product of B and B^t. We can find D is a diagonal matrix shown at the bottom of the page. Then, we will have:

$$\begin{aligned} SF^2 &= RF^2 + CF^2 = \frac{1}{8 \times 8} [trace(DF^t F) + trace(DFF^t)] \\ &= \frac{1}{8 \times 8} \sum_{u=0}^{7} \sum_{v=0}^{7} [(D(u,u) + D(v,v)) \times F^2(u,v)] \\ &= \frac{1}{8 \times 8} \sum_{u=0}^{7} \sum_{v=0}^{7} [E(u,v) \times F^2(u,v)] \end{aligned} \tag{12}$$

where (see the second matrix at the top of the next page. In conclusion, the spatial frequency of an 8×8 block of pixels can be accurately calculated by the weighted sum of squares of AC coefficients in the DCT block.

III. Proposed Method

The spatial frequency, which had its beginning with the study of the human visual system, indicates the overall active level of an image [14]. It is difficult to completely comprehend the human visual system with current physiologic means, but the spatial frequency supplies an effective contrast criterion for image fusion [15]. It is shown in Section II that the calculation of spatial frequency in DCT domain is simple. Hence, we can use the spatial frequency value as the contrast measure of the 8×8 blocks of the source images.

Fig. 1 shows the schematic diagram of the proposed multi-focus image fusion method. For simplicity, we only consider two source images A and B, but the method can be extended for more than two source images. The fusion process consists of the following steps:

1) Decode and de-quantize the source images, and then divide them into blocks of size 8×8. Denote the block pair at location (i, j) by $A_{i,j}$ and $B_{i,j}$ respectively.

$$D = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1522 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5858 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.2346 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.0000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2.7654 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3.4142 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3.8478 \end{pmatrix}$$

$$E = \begin{pmatrix} 0 & 0.1522 & 0.5858 & 1.2346 & 2.0000 & 2.7654 & 3.4142 & 3.8478 \\ 0.1522 & 0.3045 & 0.7380 & 1.3869 & 2.1522 & 2.9176 & 3.5665 & 4.0000 \\ 0.5858 & 0.7380 & 1.1716 & 1.8204 & 2.5858 & 3.3512 & 4.0000 & 4.4335 \\ 1.2346 & 1.3869 & 1.8204 & 2.4693 & 3.2346 & 4.0000 & 4.6488 & 5.0824 \\ 2.0000 & 2.1522 & 2.5858 & 3.2346 & 4.0000 & 4.7654 & 5.4142 & 5.8478 \\ 2.7654 & 2.9176 & 3.3512 & 4.0000 & 4.7654 & 5.5307 & 6.1796 & 6.6131 \\ 3.4142 & 3.5665 & 4.0000 & 4.6488 & 5.4142 & 6.1796 & 6.8284 & 7.2620 \\ 3.8478 & 4.0000 & 4.4335 & 5.0824 & 5.8478 & 6.6131 & 7.2620 & 7.6955 \end{pmatrix}$$

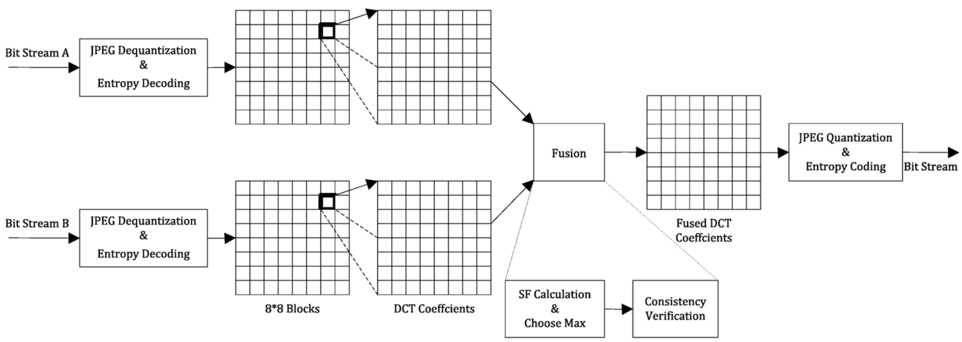

Fig. 1. Schematic diagram for fusing images coded in JPEG format.

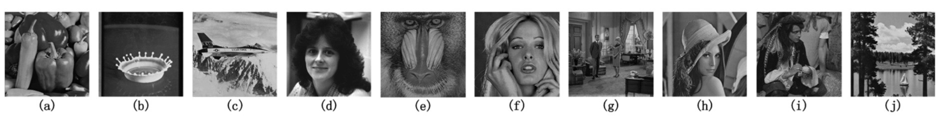

Fig. 2. Images used for simulations.

2) Compute the spatial frequency of each block by (12), and denote the results of $A_{i,j}$ and $B_{i,j}$ by $SFA_{i,j}$ and $SFB_{i,j}$, respectively.
3) Compare the spatial frequencies of two corresponding blocks to decide which should be used to construct the fused image. Create a decision map W to record the feature comparison results according to a selection rule:

$$W_{i,j} = \begin{cases} 1 & SFA_{i,j} > SFB_{i,j} + T \\ -1 & SFA_{i,j} < SFB_{i,j} - T \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Here, T is a user-defined threshold.
4) Apply a consistency verification process to improve quality of the output image. Use a 3×3 majority filter [3] to obtain a refined decision map R:

$$R_{i,j} = \sum_{x=i-1}^{i+1} \sum_{y=j-1}^{j+1} w_{x,y}. \tag{14}$$

Then, obtain the DCT representation of the fused image F based on R as:

$$F_{i,j} = \begin{cases} A_{i,j} & R_{i,j} > 0 \\ B_{i,j} & R_{i,j} < 0 \\ (A_{i,j} + B_{i,j})/2 & R_{i,j} = 0 \end{cases}. \tag{15}$$

5) Quantize the resulting DCT coefficients with a standard quantization table in the standard JPEG coder [12] and then use entropy coding to produce the output bit stream.

IV. Experimental Results and Analysis

The simulations of the fusion methods have been conducted with an Intel i5 4570 processor with 4 GB RAM. For the wavelet based methods, the DWT with DBSS (2, 2) and the SIDWT for Haar basis with three levels of decomposition are applied. For the NSCT, we use 2, 4, 8 directions in the scales from coarser to finer. For the proposed method, we obtain the results with the threshold of 2. All the images in the simulations are converted to JPEG files.

TABLE I
Objective Evaluation of Image Fusion

Method	*SSIM*	*RMSE*	*T(μs)*
DCT + aver	0.9555	7.213	**10.99**
DCT + Contrast	0.9792	5.103	511.0
DCT + AC-Max	0.9227	9.326	24.70
DCT + Variance	0.9866	4.541	13.65
DCT + Variance + CV	0.9884	4.260	25.23
DWT	0.9833	4.721	-
SIDWT	0.9871	5.032	-
NSCT	0.9894	4.068	-
DCT + SF(proposed)	0.9888	4.220	13.96
DCT + SF + CV(proposed)	**0.9902**	**4.037**	25.54

In this section, we compare the performance of our technique with the existing image fusion methods in the DCT domain, like $\mathrm{DCT+Average}$ [7], $\mathrm{DCT+Contrast}$ [7], $\mathrm{DCT+AC-Max}$ [8], $\mathrm{DCT+Variance}$ [6], and $\mathrm{DCT+Variance+CV}$ [6]. The multi-scale based fusion such as DWT [3], SIDWT [4], and NSCT [5] are treated as state-of-the-art approaches.

In the first experiment, the performance of the proposed fusion method is demonstrated by fusing 30 pairs of blurred images which are generated by filtering the standard grayscale images shown in Fig. 2 with averaging filter of different radiuses (5, 7, and 9 pixels). In each of these pairs, complementary regions of the source images are blurred. The standard grayscale images are taken as ground truth images.

The root mean square error (RMSE) and structural similarity measure (SSIM) [16] are used for objective evaluation in

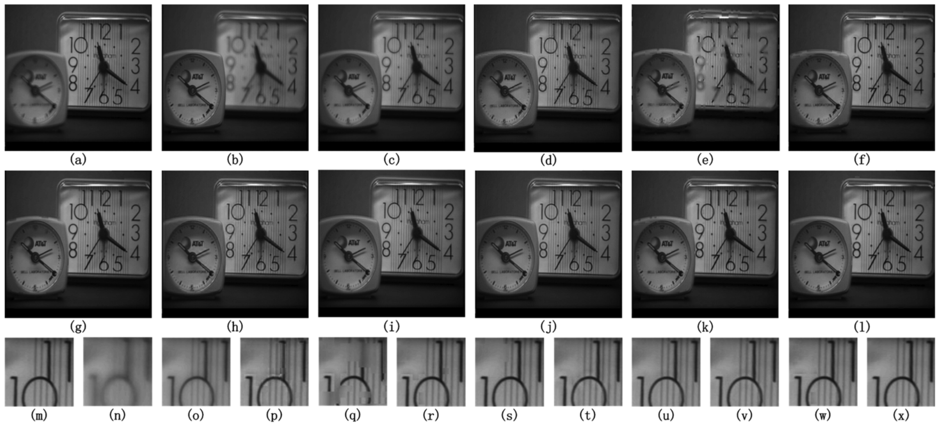

Fig. 3. Source images "Clock" and the fusion results. (a) The first source image with focus on the right. (b) The second source image with focus on the left. (c) $\mathrm{DCT} + $ Average result. (d) $\mathrm{DCT} + $ Contrast result. (e) $\mathrm{DCT} + \mathrm{AC} - $ Max result. (f) $\mathrm{DCT} + $ Variance result. (g) $\mathrm{DCT} + \mathrm{Variance} + $ CV result. (h) DWT result. (i) SIDWT result. (j) NSCT result. (k) Result of the proposed $\mathrm{DCT} + \mathrm{SF}$ algorithm. (h) Result of the proposed algorithm with consistency verification $\mathrm{DCT} + \mathrm{SF} + \mathrm{CV}$. (m)-(x) are the local magnified version of (a)-(l), respectively.

TABLE II
OBJECTIVE EVALUATION OF THE IMAGE FUSION

Method	Clock			Pepsi			Books		
	LMI	Q_w	*FMI*	*LMI*	Q_w	*FMI*	*LMI*	Q_w	*FMI*
DCT + aver	0.8509	0.8540	0.9054	0.8851	0.8750	0.9095	0.8804	0.8266	0.9012
DCT + Contrast	0.8645	0.8990	0.8992	0.8957	0.9271	0.9146	0.8969	0.9176	0.9050
DCT + AC-Max	0.8642	0.7003	0.8991	0.8214	0.6932	0.6932	0.8315	0.6710	0.9067
DCT + Variance	0.8698	0.9057	0.9142	0.9257	0.9465	0.9244	0.9337	0.9026	0.9134
DCT + Variance + CV	0.8702	0.9160	0.9150	0.9268	0.9534	0.9249	0.9339	0.9115	0.9145
DWT	0.8580	0.9056	0.9097	0.8889	0.9487	0.9191	0.8977	0.9107	0.9114
SIDWT	0.8589	0.9016	0.9125	0.8936	0.9485	0.9195	0.8992	0.9179	0.9123
NSCT	0.8602	0.9114	0.9137	0.8950	0.9525	0.9216	0.9021	0.9209	0.9137
DCT + SF(proposed)	0.8702	0.9198	0.9151	0.9271	0.9535	0.9251	0.9355	0.9162	0.9148
DCT + SF + CV(proposed)	**0.8721**	**0.9210**	**0.9157**	**0.9296**	**0.9551**	**0.9253**	**0.9356**	**0.9297**	**0.9157**

the first experiment. RMSE is the cumulative squared error between the fused and the referenced image. SSIM is used to evaluate salient information that has been transferred into the fused image [16]. Table I lists the average RMSE values and SSIM values of 30 experimental images. The average periods of the run-time for fusion of every 8×8 block of images in the DCT based methods are also given in Table I. The best results are shown in bold fonts. Obviously, the proposed approach without a CV performs better than the other DCT based algorithms and DWT based methods. Furthermore, the result of the proposed method with a CV even outperforms that of the NSCT based algorithm, at the cost of a little complexity.

The second experiment is carried out on sets of non-referenced multi-focused images from online resources [17]. Due to the lack of space in this letter, only the results of "Clock" are shown in Fig. 3. We also conducted experiments on "Pepsi" and "Book", and similar subjective results were obtained. By carefully observing the fusion results, we can see clearly that the method $\mathrm{DCT} + \mathrm{Average}$ suffers undesirable blurring effects (Fig. 3(c)) and the method $\mathrm{DCT} + \mathrm{Contrast}$ results in blocking artifacts (Fig. 3(d)) In addition, the method $\mathrm{DCT} + \mathrm{AC} - \mathrm{Max}$ leads to the error selection of the best blocks distinctly. It is also obvious that the DWT based method exhibits undesirable ringing artifacts round figures. Moreover, it can be easily found in the magnified images corresponding to $\mathrm{DCT} + \mathrm{Variance}$ and $\mathrm{DCT} + \mathrm{Variance} + \mathrm{CV}$, respectively in Figs. 3(p) and 3(q), the variance based algorithms bring about erroneous selection of some blocks from the blurred image. Since the results of the proposed method and the NSCT based method can't be visually comparable, we use some state-of-the-art performance metrics such as localised mutual information (LMI) [18], Piella metric ($\mathrm{Q_w}$) [19], and feature mutual information (FMI) [20] for further comparison. These metrics estimate the transfer of local structures from source images into the fused image. The higher the values of these metrics, the better are the quality of the fused image. The performance analyses of three well-known images "Clock", "Pepsi" and "Book" are shown in Tables II. It is observed from Table II that the performance of the proposed technique is superior to the conventional techniques in terms of the above metrics.

Based on the above analysis, we can see that the proposed method is effective and it outperforms the traditional image fusion approaches in terms of both subjective evaluation and objective evaluation.

V. CONCLUSIONS

In this letter, a new approach based on spatial frequency for fusion of multi-focus images has been proposed in the DCT domain instead of the spatial domain. We evaluate the performance of the proposed method with various evaluation metrics and it is found that the performance of fusion in the DCT domain is superior to that of conventional approaches based on DCT and the state-of-the-art methods including DWT, SIDWT, and NSCT, in terms of visual quality and quantitative parameters. Moreover, the proposed method is simple to implement and computationally efficient when the source images are coded in JPEG format, especially in wireless visual sensor networks.

References

[1] B. K. Shreyamsha Kumar, M. N. S. Swamy, and M. Omair Ahmad, "Multiresolution DCT decomposition for multifocus image fusion," in *2013 26th Annu. IEEE Can. Conf. Electrical and Computer Engineering (CCECE)*, 2013, IEEE.

[2] D. Egfin Nirmala, R. K. Vignesh, and V. Vaidehi, "Multimodal image fusion in visual sensor networks," in *2013 IEEE Int. Conf. Electronics, Computing and Communication Technologies (CONECCT)*, 2013, IEEE.

[3] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57.3, pp. 235–245, 1995.

[4] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform," in *1997 Proc. Int. Conf. Image Processing*, 1997, vol. 3, IEEE.

[5] Q. Zhang and B.-L. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," *Signal Process.*, vol. 89.7, pp. 1334–1346, 2009.

[6] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "Multi-focus image fusion for visual sensor networks in DCT domain," *Comput. Electr. Eng.*, vol. 37.5, pp. 789–797, 2011.

[7] T. Jinshan, "A contrast based image fusion technique in the DCT domain," *Digital Signal Process.*, vol. 14.3, pp. 218–226, 2004.

[8] Y. Phamila and R. Amutha, "Discrete cosine transform based fusion of multi-focus images for visual sensor networks," *Signal Process.*, pp. 161–170, 1995.

[9] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "Real-time fusion of multi-focus images for visual sensor networks," in *6th IEEE Iranian Machine Vision and Image Processing (MVIP)*, 2010.

[10] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Patt. Recognit. Lett.*, vol. 28.4, pp. 493–500, 2007.

[11] S. Djamel and B. Mouldi, "Image compression via embedded coder in the transform domain," *Asian J. Inf. Tech*, vol. 5.6, pp. 633–639, 2006.

[12] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34.4, pp. 30–44, 1991.

[13] I. E. Richardson*, Video Codec Design*. Hoboken, NJ, USA: Wiley, 2002.

[14] S. Li and B. Yang, "Multifocus image fusion using region segmentation and spatial frequency," *Image Vis. Comput.*, vol. 26.7, pp. 971–979, 2008.

[15] S. Li, J. T. Kwok, and Y. Wang, "Combination of images with diverse focuses using the spatial frequency," *Inf. Fusion*, vol. 2.3, pp. 169–176, 2001.

[16] Z. Wang *et al.*, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13.4, pp. 600–612, 2004.

[17] http://imagefusion.org/.

[18] M. Hossny *et al.*, "Image fusion performance metric based on mutual information and entropy driven quadtree decomposition," *Electron. Lett.*, vol. 46.18, pp. 1266–1268, 2010.

[19] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. 2003 Int. Conf. Image Processing, ICIP*, 2003, vol. 3, IEEE.

[20] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "A non-reference image fusion metric based on mutual information of image features," *Comput. Elect. Eng.*, vol. 37.5, pp. 744–756, 2011.