DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-198

Image fusion method for a single sensor based multispectral filter array containing a near infra-red channel

Seung Hoon Jee and Moon Gi Kang; Institute of BioMed-IT, Energy-IT and Smart-IT Technology (BEST); City, Yonsei University 134 Shinchon-Dong, Seodaemun-Gu, Seoul, South Korea

Abstract

This paper proposes an image fusion method for a single sensor based RGB+NIR (near infrared) MFA (multi-spectral filter array) sensor system. Unlike conventional color filter arrays, Bayer patterns, MFA sensors can receive not only visible band information but also NIR band information by using a single sensor system, which does not need to be registered. The main reasons for using MFA sensors are to increase object discrimination in the target images and improve the brightness of visible band images in extremely low light conditions by using NIR bands. As described in this paper, the fusion method consists of resolution fusion and color fusion, both of which are based on the texture decomposition method. In the experimental results, we compared the visible band, which is the input image for the fusion method, and fused the output image with the NIR band. Fusion results showed higher object discrimination and produced less noise than input visible images.

Introduction

Multispectral imaging systems have emerged as visible bands for handling the limitations of visible band information. By exploiting the fine spectral differences between various natural and artificial materials of interest, these systems can support improved detection and classification capabilities relative to visible bands and other spectral information. By using multispectral information, the image fusion method is useful for making images visible to the human eye. The image fusion method maximizes resolution. It follows the colors of visible bands by fusing the visible band information and the invisible band information with the multispectral information. Ideally, the goal of multispectral image fusion is to represent all the visual information contained in multiple source images into a single fused image without introducing artifacts or information loss.

To obtain spectral information, multi-sensor based multispectral imaging systems have been developed. When using these systems, however, resolution can be limited when producing multispectral information because of differences among the sensors such as physical location, pan, tilt, and zoom. A MFA (multispectral filter array) sensor has been developed to solve these limitations. These sensors can receive light on the same lens and they are appropriate for fusion since they do not require registration. In addition, they produce the same optical character through the same lens. As one sensor receives the multispectral information, fusion becomes easier than when using multisensory systems.

A disadvantage of multispectral sensing, however, is that the spatial resolution is often coarser when viewed through the same lens than when using only visible band sensors. Therefore, enhanced spectral fidelity often comes at the expense of spatial fidelity, and fine resolution can be lost. A wide variety of fusion algorithms have been studied to find a solution to this problem. For example, this paper suggests a texture decomposition based fusion algorithm, which takes into account the differences observed in previous studies. The proposed algorithm is based on total variation (TV) minimization inspired by [2]. The most widely known form was introduced in image processing by [3] for image segmentation and later by [1] for noise removal through the optimization of a cost function. Unlike pre-designed filter methods, the suggested method can be used adaptively with any image.

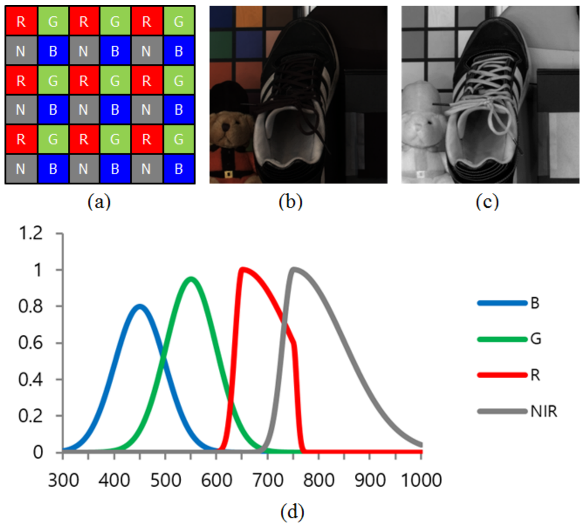


Figure 1. Characteristics of MFA sensor. (a) MFA sensor pattern. (b) RGB color image from MFA. (c) NIR band image from MFA. (d) Spectral response of MFA containing the NIR

Because of strong edges in the images, resolution maximization was easier. As the brightness increased by strong reflectance of invisible bands in the objects, excess artifacts were suppressed and texture information was stably maximized along the strong edges. To merge the decomposed information, we used fusion rules suitable for texture decomposition.

Fusion Method

Maximized resolution and accurate color representation can be expected when using MFA sensor information that merges NIR information and RGB information. However, it is difficult to maximize resolution, improve sensitivity, and represent accurate color in the RGB domain. If the fusion that improves sensitivity by using NIR information in the RGB domain is performed in each channel, it is extremely difficult to maintain the unique hue and saturation. A sense of color of the entire image depends on the hue. The color composition changes since hue is one of the main properties of color. Thus, to achieve effective resolution and color, we had to separate the color information and resolution information.

In the experiments, we used a YUV converter based framework to separate the resolution (or luminance) domain and the color

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-198

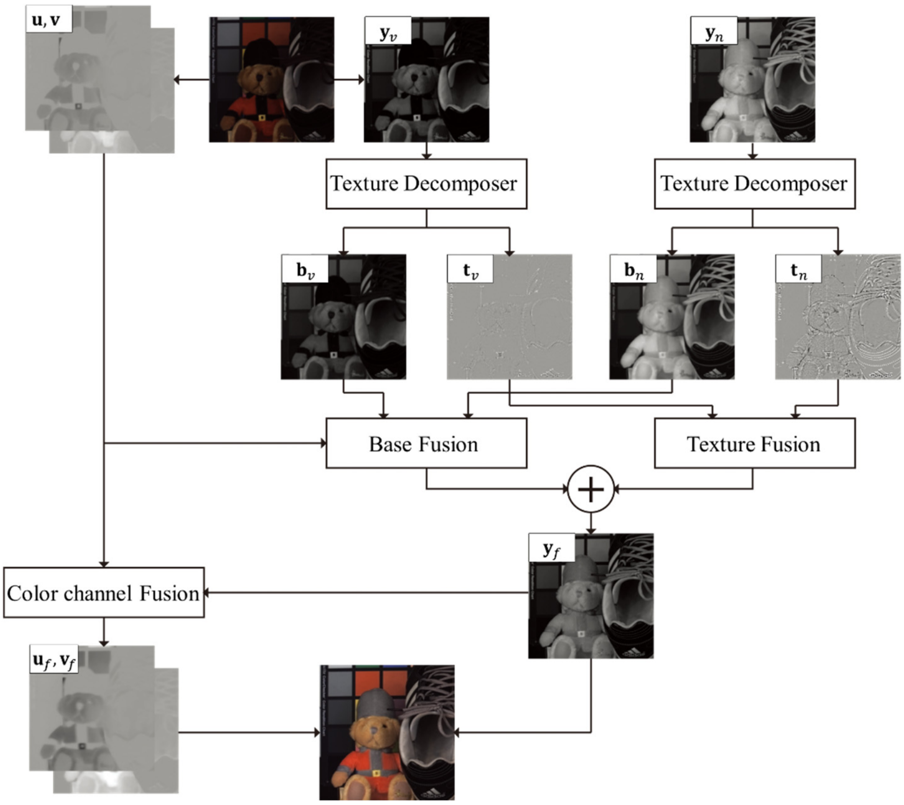


Figure 2. Block diagram for the proposed fusion method

information domains. The separated resolution information performed the color information and individually optimized the algorithm to maximize the image resolution. Likewise, an optimized color representation algorithm was performed by using color channel information in which the resolution information was already separated. The Y channel presents the luminance information of the images. This is because the human vision system recognizes this as image resolution. Through the luminance information fusion of the NIR band and the Y channel information of the visible band, we were able to maximize resolution and improve image sensitivity. By performing individual fusion, the unique hue of the images was maintained.

Regarding the fusion of multispectral images, the fusion of the luminance information was a significant factor. This is because the luminance information in the human visual system controls the image resolution. Fusion algorithms usually start from a dividing characteristic basis. Each characteristic basis performs fusion by using the fusion rule that each algorithm suggests. In general methods, the characteristic basis is decomposed by sub-bands produced by wavelet transform techniques. These conventional methods decompose image luminance information and make it suitable to cut-off frequencies obtained by pre-designed filters. In a more advanced way, filter phases also can be diversely designed and separate frequency information can be obtained in various ways. The separated sub-band maximized information appears in images obtained by performing fusion according to each individual fusion rule. However, there are several limitations to this approach. First, a predesigned filter library should be used. Designing filters that reflect image character is not an easy process. Moreover, manufacturing a filter library for each separate image is an extremely time consuming process. Second, segmentation between strong edges and detailed parts is difficult. NIR reflectance presents different material reflection characters from the visible bands. If each object reflection is different, there are several fusion issues in the strong edges of the object edges. When high-frequency areas are divided by using a sub-band decomposer, there is a trade-off between obtaining detailed information and finding strong edges. In addition, by using the maximum selection method with high frequency information to stress detail, artifacts can occur along the strong edges (object edges). On the other hand, a loss of detailed information can occur when reducing the high-frequency energy in order to stabilize the strong edges.

In order to solve this issue, this paper presents a texture-decomposer method that is different from the sub-band decomposer. This texture decomposer method was used to separate the strong edges and the detailed texture information. The textures within the images contained mainly high frequency information such as oscillated patter and detail. The information remaining after the texture was separated from each image is called base information and this showed the structural information of each image, which was used to separate the object edges. Natural image sensitivity depends on this base information fusion which controls the entire image brightness. At strong edges, image texture information fusion maximizes detail information without leaving any artifacts. Moreover, it can be stably used without a particular filter library and it also can be used in every image regardless of image character, again without any particular filter library. The individual fusion-rule was used to merge the component separated texture information and

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-198

Figure 3. Texture decomposition results. (a) original input image. (b) base information image. (c) texture information image.

the base information. Finding the average value is the simplest way of fusing the base channel information and fusing data at a particular ratio such as a one to one ratio between the visible bands and the NIR bands. Thus, in this paper, we propose a fusion ratio parameter. This parameter can control the reflection ratio of the NIR information in the fusion results. Thus, we obtained output images at a diverse fusion ratio by using the proposed algorithm.

Texture decomposition with variation minimization:

It is necessary to separate texture and base information before using the luminance channel of the input visible band image and the luminance channel information fusion of the NIR band. This paper describes total variation minimization based texture decomposition [2]. We have an observed image which is a luminance channel of the input visible band or the NIR band.

Given $\mathbf{y}$ finding another image $\mathbf{b}$ is a simplification of $\mathbf{y}$. In general, $\mathbf{b}$ is an image formed by homogeneous regions and sharp boundaries. Most models assume that

$$\mathbf{f}(h,v) = \mathbf{b}(h,v) + \mathbf{t}(h,v)\ , \tag{1}$$

where $\mathbf{t}$ is a significant piece of small scale repeated detail (texture). Texture can be defined as a repeated pattern of small scale details. This type of pattern can be modeled by oscillatory functions taking both positive and negative values, with a zero mean value. The textured component $\mathbf{t}$ is completely represented using two functions $(\mathbf{g_1}, \mathbf{g_2})$. Let $\mathbf{y}: \boldsymbol{R^2} \to \boldsymbol{R}$ be a given image. In this paper, the observed image $\mathbf{y}$ is the texture version of the true image $\mathbf{b}$, and $\mathbf{b}$ would be a sketchy approximation of $\mathbf{y}$. [2] proposed cost function; $\mathrm{E}(\mathbf{b})$, which is

$$\mathbf{E}(\mathbf{b}) = \int |\nabla \mathbf{b}| + \lambda \| \mathbf{t} \|_{*}, \mathbf{y} = \mathbf{b} + \mathbf{t}\ , \tag{2}$$

where λ>0 is a tuning parameter. The second term in the energy is a fidelity term used to remove small details while retaining essential features and sharp edges. And then, [2] defines $\mathbf{t}$, which is

$$\mathbf{t}(h,v) = \partial_h \mathbf{g_1} + \partial_v \mathbf{g_2}\ , \tag{3}$$

Then [3] is the proposed cost function

$$\min_{\mathbf{b},\mathbf{g_1},\mathbf{g_2}} \{\mathbf{G_p}(\mathbf{b},\mathbf{g_1},\mathbf{g_2}) = \int |\nabla \mathbf{b}| + \lambda \int |\mathbf{y} - \mathbf{b} - \partial_v \mathbf{g}_1 - \partial_h \mathbf{g_2}|^2\, dhdv + \mu\Big[\int (\mathbf{g}_1^2 + \mathbf{g}_2^2)^p\, dhdv\Big]^{\frac{1}{p}}\}\ , \tag{4}$$

Figure 4. Experiments on real image. (a), (d) RGB input image from MFA. (b), (e) NIR input image from MFA. (c), (f) fused output image with texture decomposition

where $\lambda, \mu > 0$ are tuning parameters, and $p \to \infty$.

Local saturation adaptive base fusion:

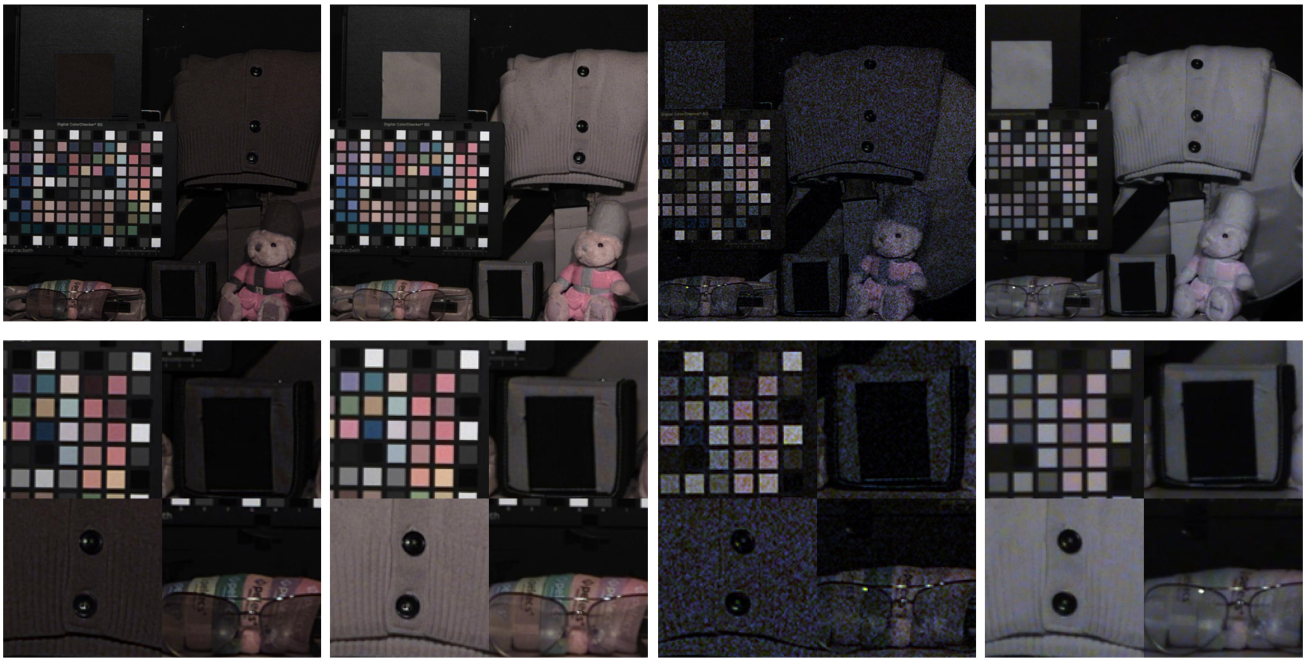
Base separated by texture decomposer suggests fusion algorithm. Each band of the visible bands and the NIR bands was represented as $\mathbf{b}_v, \mathbf{b}_n$. The global fusion ratio parameter was represented as $w_g (0 \le w_g \le 1)$. The global fusion ratio parameter can control the sensitivity of the entire image. If w_g gets higher, the ratio of $\mathbf{b}_n$ increases while w_g gets lower ratio of $\mathbf{b}_n$ decreases. w_g is a user selection parameter and it is able to control brightness according to a user's choice. The point of using the base fusion algorithm is to obtain a stable result depending on the alteration of the global fusion ratio parameter. The more the base of the NIR band is reflected, the greater the increase of the image sensitivity. However, a wash out effect can occur in object areas where saturation is high. In general, saturation enhancement can be used when wash out effects occur. However, increasing the luminance can cause an overflow of a particular channel in the RGB domain. In this case, saturation enhancement is not effective. This paper suggests a local saturation adaptive base fusion algorithm in order to solve this problem. This algorithm decreases the local base fusion ratio in areas with high saturation energy. Because we assumed that the highly saturated regions were marked bright, enhancing the sensitivity by fusing the NIR bands was unnecessary. The saturation energy can be described as

$$\mathbf{S}(h,v) = \sqrt{\mathbf{u}(h,v) + \mathbf{v}(h,v)}\ , \tag{5}$$

where $\mathbf{S}(h,v)$ denotes saturation energy h,v at the pixel location. Through the estimated saturation energy, the proposed base fusion strategy can be described as

$$w_b(h,v) = w_g \cdot \mathbf{DF}(\mathbf{S}(h,v)) \tag{6}$$

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-198



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

Figure 5. Experiments on real image in general illumination and low illumination. (a) RGB input image in general illumination (120Lux) by incandescent. (b) fused output image in general illumination (120Lux) by incandescent. (c) RGB input image in low illumination (1Lux) by incandescent. (d) fused output image in general illumination (1Lux) by incandescent.

$$\mathbf{DF}(x) = \begin{cases} -\dfrac{1}{\tau^2}x^2 + 1 & 0 \le x \le \tau \\ 0 & \tau \le x \end{cases}, \tag{7}$$

$$\mathbf{b}_f(h,v) = w_b(h,v) \cdot \mathbf{b}_n(h,v) + [1 - w_b(h,v)] \cdot \mathbf{b}_v(h,v), \tag{8}$$

where $\mathbf{b}_f(h,v)$ denotes the fused base for the output images, $w_b(h,v)$ denotes the pixel-based weighting parameter for the NIR band base information, denotes the global fusion ratio parameter and τ is the tuning parameter. $\mathbf{DF}(x)$ is a second order function designed as a decreasing function. The $w_b(h,v)$ value decreases when saturation energy $\mathbf{S}(h,v)$ increases and $w_b(h,v)$ becomes zero when $\mathbf{S}(h,v)$ is bigger than τ.

This paper proposes a fusion algorithm of texture separated by a texture decomposer. Because much of the texture was already separated from the strong edges, we obtained the maximum resolution by using texture. Generally, texture fusion can solve the focus issue at each band. When images are well-focused in either the NIR or visible band, the texture of each well-focused image can be used by fusion. Moreover, through fusion, texture fusion can naturally present texture which is invisible in the visible bands, but not in the NIR bands. Base fusion can prevent artifacts at the object edges or in the strong edge regions. In addition, texture fusion can maximize the texture information individually. Each visible band and NIR band represents texture as $\mathbf{t}_v, \mathbf{t}_n$. Texture fusion can be described as

$$\mathbf{t}_f = \frac{(\mathbf{t}_n(h,v))^3 + (\mathbf{t}_n(h,v))^3}{(\mathbf{t}_n(h,v))^2 + (\mathbf{t}_n(h,v))^2 + \alpha}, \tag{9}$$

where $\mathbf{t}_f(h,v)$ denotes the fused texture for the output images and where α denotes a very small parameter used to avoid dividing by zero. Through $(\mathbf{t}_n(h,v))^2$ and $(\mathbf{t}_n(h,v))^2$, the fusion ratio reflected at the output texture was determined. The bigger absolute value, the higher weight it gained. Therefore, the texture information was maximized.

Experimental Results

Two types of characteristics, the visible band image (input) and the result of the fusion method image (output), were compared to verify the performance of our proposed fusion method. The experiment was conducted under various light intensities – 500Lux (high illumination), 120Lux (normal illumination) and 1Lux (extremely low illumination) - under 3000K (incandescent light) condition, and results are presented in figures 4 and 5. Especially, more extreme environments were used with incandescent light due to its inequality of light energy that the RGB and each channel received since the incandescent light contained most of the frequency of the light source concentrated in the near infrared side.

The overall procedure for the RGBN patterns was composed of the following two parts: color interpolation for each channel and RGB+NIR fusion. For the RGBN pattern, the results of the color interpolation based DLMMSE [4] and the proposed fusion method were compared respectively. Although there are lots of color interpolation methods other then the DLMMSE, the DLMMSE method was used in these experiments, because it is widely known in the color interpolation literature.

As shown in figures 4-5, the visual results of the proposed method show less false color and fewer grid artifacts in the edges and details than the input images. In addition, the fused results increased object discrimination in Figure 4 (d, e, f) and Figure 5. As shown in Figure 5 (b, d), furthermore, because of using NIR, which has very high sensitivity in extremely low light conditions, the fused

DOI: 10.2352/ISSN.2470-1173.2016.15.IPAS-198

results show less noise and fewer color artifacts than the input visible band images. For comparison of color expression with visible band images, in Figure 5, despite the merging of the NIR information, the color expression of the output images was quite accurate. Especially in Figure 5 (c, d), the fused images show less color noise and seem to have improved.

Conclusion

This paper proposed an image fusion algorithm through a RGB+NIR MFA (multispectral filter array) sensor system. Generally, NIR band information can improve object discrimination in visible band images. Moreover, the sensitivity of image sensors can be improved by using NIR band information in low illumination environments, where the proposed fusion algorithm can supplement the visible band information by reflecting the NIR band information, and improve the overall brightness of the images.

Acknowledgement

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No. 2015R1A2A1A14000912).

References

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

[2] Y. Meyer, Oscillating Patterns in Image Processing and Non-linear Evolution Equations. Providence, RI: American Mathematical Society, 2001.

[3] L. A. Vese and S. J. Osher, “ Modeling textures with total variation minimization and oscillating patterns in image processing”, J. sci. comput., vol. 19, no. 1-3, Dec. 2003

[4] L. Zhang, X. Wu, Color demosaicking via directional linear minimum mean square-error estimation, IEEE Trans. Image Process. 14 (12) (2005) 2167–2178.

[5] C. Pohl and J. L. Can Genderen, “Multisensor image fusion in remote sensing: Concepts, methods, and applications,” *Int. J. Remote Sensors*, vol. 19, no. 5, pp. 823-854, 1998.

[6] P. J. Burt and E. H. Adelson, “The Laplacian pyramid as a compact image code,” *IEEE Trans. Commun.*, vol, 31, no. 4, pp. 532-540, Apr. 1983.

[7] A. Toet, “Image fusion by a ratio of low pass pyramid,” *Pattern Recognit Lett.*, vol, 9, no. 4, pp. 245-253, 1989.

[8] A. Toet, L. J. van Ruyven, and J. M. Valeton, “Merging thermal and visual images by a contrast pyramid,” *Opt. Eng.*, vol. 28, no. 7, pp.789-792, Jul. 1989.

[9] S. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674-693, Jul. 1989.

[10] V. S. Petrovix and C. S. Xydeas, “Gradient-based multiresolut-ion image fusion,” *IEEE Trans. Image Process.*, vol. 13, no, 2, pp. 228-237, Feb. 2004.

[11] T. Huntsberger and B. Jawerth, “Wavelet based sensor fusion.’ *Proc. SPIE*, vol. 2059, pp. 488-498, Aug. 1993

[12] M. Choi, R. Y. Kim, M.-R. Nam, and H. O. Kim, “Fusion of multispectral and panchromatic satellite images using the curvelet transform,” *IEEE Geosci. Remote Sensing Lett.,* vol. 2, no. 2, pp. 136–140, Apr. 2005.

[13] C. Deng, H. Cao, C. Cao, and S. Wang, “Multisensor image fusion using fast discrete curvelet transform,” *Proc. SPIE*, vol. 6790, pp. 679004-1–679004-9, Nov. 2007.

[14] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, “Remote sensing image fusion using the curvelet transform,” *Inf. Fusion*, vol. 8, no. 2, pp. 143–156, Apr. 2007.

[15] M. N. Do and M. Vetterli, “The contourlet transform: An efficient directional multiresolution image representation,” *IEEE Trans. Image Process.,* vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[16] M. Qiguang and W. Baoshu, “The contourlet transform for image fusion,” *Proc. SPIE,* vol. 6242, pp. 62420Z-1–62420Z-8, Apr. 2006.

[17] L. Tang and Z. Zhao, “Multiresolution image fusion based on the wavelet-based contourlet transform,” *in Proc. IEEE 10th Int. Conf. Inf. Fusion*, Jul. 2007, pp. 1–6.

[18] S. Yang, M. Wang, L. Jiao, R. Wu, and Z. Wang, “Image fusion based on a new contourlet packet,” *Inf. Fusion*, vol. 11, no. 2, pp. 78–84, Apr. 2010.

Author Biography

Seung Hoon Jee received his BS. degree in electronics engineering in 2010 from Yonsei University, Seoul, Korea. He is currently pursuing a Ph.D. at the Department of Electrical and Electronic Engineering at Yonsei University. His current research interests include image and video fusion in multispectral system, filtering, restoration, and enhancement.

Moon Gi Kang received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1986 and 1988, respectively, and the Ph.D. degree in electrical engineering from Northwestern University, Evanston, IL, USA, in 1994. He was an Assistant Professor with the University of Minnesota, Duluth, MN, USA, from 1994 to 1997, and since 1997, he has been with the Department of Electronic Engineering, Yonsei University, Seoul, where he is currently a Professor. His current research interests include image and video filtering, restoration, enhancement, and superresolution reconstruction. He has authored more than 150 Technical articles in his areas of expertise. He served as the Editorial Board Member for the IEEE Signal Processing Magazine, the Editor of SPIE Milestone Series Volume (CCD and CMOS imagers), and the Guest Editor of the IEEE Signal Processing Magazine Special Issue on Superresolution Image Reconstruction (May 2003). He has served in the technical program and steering committees of several international conferences. He has also served as the Associate Editor of the EURASIP Journal of Advances in Signal Processing. He was a recipient of 2006 and 2007 Yonsei Outstanding Research Achievement Awards (Technology Transfer), the 2002 HaeDong foundation Best Paper Award, and the 2000, 2009, and 2010 Awards of Teaching Excellence at Yonsei University.