World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:8, 2022

Multi-Sensor Image Fusion for Visible and Infrared Thermal Images

Amit Kr. Happy

Open Science Index, Computer and Information Engineering Vol:16, No:8, 2022 publications.waset.org/10012628.pdf

Abstract**—**This paper is motivated by the importance of multi-sensor image fusion with specific focus on Infrared (IR) and Visible image (VI) fusion for various applications including military reconnaissance. Image fusion can be defined as the process of combining two or more source images into a single composite image with extended information content that improves visual perception or feature extraction. These images can be from different modalities like Visible camera & IR Thermal Imager. While visible images are captured by reflected radiations in the visible spectrum, the thermal images are formed from thermal radiation (IR) that may be reflected or self-emitted. A digital color camera captures the visible source image and a thermal IR camera acquires the thermal source image. In this paper, some image fusion algorithms based upon Multi-Scale Transform (MST) and region-based selection rule with consistency verification have been proposed and presented. This research includes implementation of the proposed image fusion algorithm in MATLAB along with a comparative analysis to decide the optimum number of levels for MST and the coefficient fusion rule. The results are presented, and several commonly used evaluation metrics are used to assess the suggested method's validity. Experiments show that the proposed approach is capable of producing good fusion results. While deploying our image fusion algorithm approaches, we observe several challenges from the popular image fusion methods. While high computational cost and complex processing steps of image fusion algorithms provide accurate fused results, but they also make it hard to become deployed in system and applications that require real-time operation, high flexibility and low computation ability. So, the methods presented in this paper offer good results with minimum time complexity.

Keywords**—**Image fusion, IR thermal imager, multi-sensor, Multi-Scale Transform.

I. Introduction

IN this age of information explosion, sensor techniques are being developed rapidly. The various applications require comprehensive information about a certain scenario for enhanced understanding of different conditions. Sensors of the same type acquire information from only one aspect and are thus unable to provide all required information. As a result, the fusion of imagery from different sensors is a subject of research that has gained importance in recent years in the scientific community. Fig. 1 shows the block level scheme of a general multi-sensor image fusion procedure.

For an image fusion system, the input source images can be acquired from either different type of imaging sensors (different modalities) or a sensor whose optical parameters can be modified, and the output i.e., fused image will be more suitable for human or machine perception than any individual source image. Images of different types, such as visible, IR, computed tomography (CT), and magnetic resonance imaging (MRI), are good source images for fusion. Among the combinations of these types, IR and visible image fusion is of greater significance. Firstly, because their signals come from different modalities, thereby providing scene information from different aspects; i.e., visible images capture reflected light, whereas IR images capture thermal radiation. Therefore, this combination is more informative than that of single-modality signals. Secondly, IR and visible images present characteristics that are inherent in nearly all objects [6]. Furthermore, this fusion is very important for night vision applications including military surveillance. Visible images typically have high spatial resolution and considerable detail which are suitable for human visual perception. However, these images can be easily influenced by severe environmental conditions, such as poor illumination, fog, and other effects of bad weather whereas, IR images, which depict the thermal radiation of objects, are resistant to these disturbances but typically have low resolution and poor texture information.[3]

Image fusion can be done at the level of pixel, feature or at decision level. This paper specifically addresses the problem of pixel-level fusion. It is desired that no information should be lost in the image fusion process. Not only the structure, but also the origin of the details from the different image modalities should be clearly represented in the fused image in some applications. Preferably the resulting image should also have a natural appearance so that it can be readily interpreted. So, the keys to an excellent fusion method are effective image information extraction and appropriate fusion principles, that allow useful information to be extracted from source images and integrated in the fused image without introducing any artifact in the process.

II. Related Knowledge

A. Image Fusion Methods

IR and VI fusion methods can be categorized as pixel-based fusion methods and region-based fusion methods [11]. Pixel-based fusion methods are most popular because their implementation is at the lowest physical level and also there are minimum artifacts in the fused image. In these methods contributing pixels can be selected by the measurements of the source image pixels or the transformed coefficients [12]. The conventional pixel-based IR and VI fusion methods are transform domain based methods whose performance mainly

Amit Kr. Happy is with Electronics and Electrical Communication Engg. Dept., IIT Kharagpur, India (e-mail: amitkhappy1@gmail.com).

depends on the image transform selected and coefficient fusion strategies [14]. Some of the transform domain-based IR and VI fusion algorithms are pyramids, wavelet [16], contourlet and so on. Moreover, the filter-based image decomposition methods are becoming more and more common in this field [4], [12], [15]. Region-based image fusion is another approach widely used in IR and VI fusion, and the key steps in this approach are to get the regional salient information of the source images. The region-based image fusion methods can extract the salient regions and create a saliency map by image segmentation or other saliency techniques [11], [13] and then fuse the source images.

Pixel-level image fusion is widely used in remote sensing [7], [8], medical imaging [10], and computer vision [9]. Although it is not possible to design a universal method applicable to all scenarios due to the diversity of source images, but majority of the image fusion methods have three main stages as shown in Fig. 2, i.e., image transform, fusion of the transform coefficients, and corresponding inverse transform. In addition to the signal transform scheme, the other key factor affecting fusion results is the fusion strategy [1]. The fusion strategy is the scheme that determines the generation of the fused image from the coefficients or pixels of the source images.

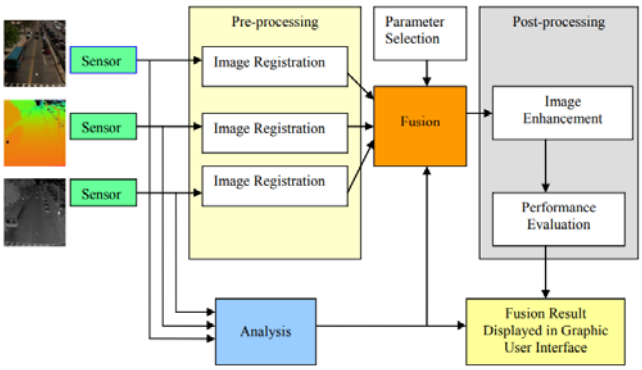

Fig. 1 Block Scheme of a general multi-sensor image fusion procedure

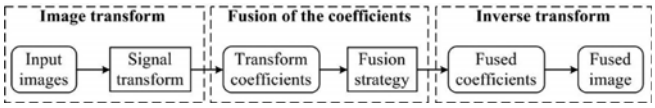

Fig. 2 The summary of the main stages for a generic pixel-level image fusion method

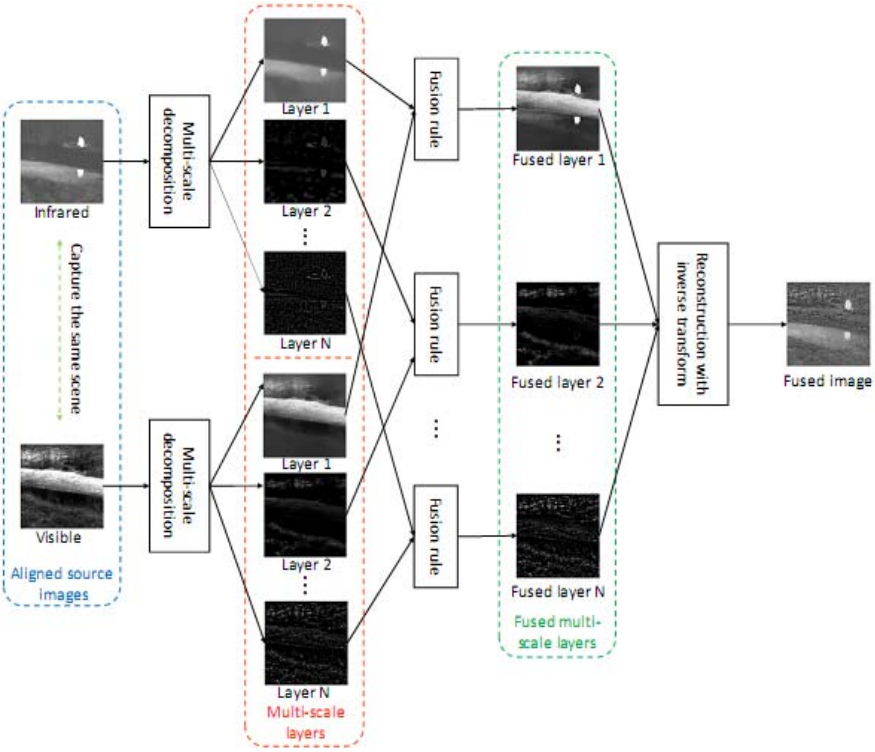

Fig. 3 Schematic diagram for the basic MST based image fusion scheme

B. MST Based Methods

MST has proved to be a very useful tool for image fusion and other image processing applications. Also, several studies [1], [14] have demonstrated that MSTs are consistent with human visual characteristics, and this property can enable fused images to have good visual effect. Fig. 3 shows the schematic diagram of a general multi-scale decomposition based image fusion scheme. First, the MST is used to obtain the multiscale representations of the source images, in which the image features are represented in a joint space-frequency domain. Then, the corresponding layers of MST are fused together as per a specific fusion rule to obtain multi-scale representation of the fused image. Finally, inverse MST is applied on the fused representation to get the final fused image. The key in MST-based fusion schemes lies in the selection of the transforms and

Open Science Index, Computer and Information Engineering Vol:16, No:8, 2022 publications.waset.org/10012628.pdf

Open Science Index, Computer and Information Engineering Vol:16, No:8, 2022 publications.waset.org/10012628.pdf

fusion rules. Next, we review the techniques in this category on the basis of these two aspects. The most commonly used multi-scale decomposition methods for image fusion are the pyramid and wavelet transform, such as the Laplacian pyramid, discrete wavelet transform (DWT), and stationary wavelet decomposition.

C. Strategies Used for Fusion of Multi-Scale Representations

In order to enhance the fusion quality, another direction that can be explored in muti-scale decomposition-based fusion is using effective fusion strategies. Some classical fusion schemes are like choose-max and weighted-average based coefficient combining method, window and region-based consistency verification, etc. In recent years, many novel fusion schemes have been developed to get better results [3], [8]. But as compared to the development of decomposition methods, fusion rules have not received enough attention in early years. However, now it's coming up that advanced fusion rules can address many shortcomings of the decomposition level, and thus, provide better fusion performances. Although the traditional fusion rules like the widely used choose-max and averaging methods are quite simple but these rules may introduce visual artifacts when the images are not perfectly registered or contain noise. These issues can be effectively improved by making use of the strong correlation between neighboring pixels and the dependency between the coefficients of different scales.

First, the source images are divided into low-frequency subbands and a series of high-frequency subbands in various scales and orientations. The highest saliency value is then chosen at each position in the transformed subbands to generate the fused subbands. Finally, the inverse transform is applied to the fused subbands to produce the fused image. The coefficients combining should integrate the visual information contained in all source images into the fused image without introduction of distortion or loss of information. However, this goal is almost impossible. A more practical and effective image fusion rule for the multi-resolution-based methods is adopted in this paper. The low-frequency coefficients are fused by the average method, meaning the fused coefficient is the average of the corresponding coefficients of the source images. The high-frequency coefficients are generally fused by the approach of choosing absolute maximum which can be formulated as:

$$D_F(p)=\begin{cases} D_{I1}(p) & A_{I1}(p) > A_{I2}(p) \\ D_{I2}(p) & \text{Otherwise} \end{cases}$$

But this technique does not always give the best result. So, in this paper some modification to this technique is introduced and implemented.

D. The Modified Feature Selection Algorithm

The pixel-by-pixel maximum selection criteria may not be the best suited strategy for fusion because the useful features in the images are generally larger than one pixel. Therefore, in the fusion scheme it is proposed to use an area-based selection rule. The images are first decomposed into a gradient pyramid. we use the maximum absolute value within the window as an activity measure associated with the center pixel. In this way a high activity value indicates the presence of a dominant feature in the local area. A binary decision map of the same size of the MST is then created to record the selection results based on a maximum selection rule. This binary map is subject to a consistency verification. Specifically, if the center pixel value comes from image A while the majority of the surrounding pixel values come from image B, the center pixel value is switched to that of image B. In the implementation, a majority filter (which outputs 1 if the count of 1's outnumbers the count of 0's, and outputs 0 otherwise) is applied to the binary decision map; the map is then negated, and is followed by the application of a majority filter. The resulting map is negated again. A fused image is finally obtained based on the new binary decision map. A schematic diagram of the proposed feature selection rule is shown in Fig. 4. This selection scheme helps to ensure that the dominant features are incorporated as complete as possible into the new images. Thus, consistency verification [16] is based on the idea that a composite multiscale coefficient is unlikely to be generated in a completely different manner from all its neighbors.

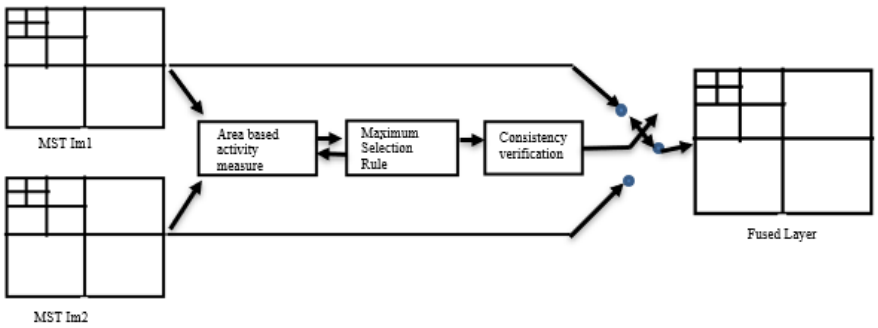

Fig. 4 The modified feature selection scheme

Open Science Index, Computer and Information Engineering Vol:16, No:8, 2022 publications.waset.org/10012628.pdf

III. Implementation

The steps of implantation are described as follows:

Step0. Source images IR and VI are given.

Step1. The IR and VI are decomposed by MST Laplacian Pyramid (LP) to get several corresponding low frequency sub-images and high frequency sub-images sets, and LP uses 2/4/6/8 layers of decomposition and makes a comparative study on their performance.

Step2. The low-frequency coefficients are fused by the average method.

Step3. The pixel-by-pixel maximum selection rule is used for high frequency layers and fusion results are saved for comparison.

Step4. Again, an area-based selection rule with consistency verification is used for high frequency layers and a comparison is done with the previous results.

Step5. The fused image is reconstructed from the fused layer coefficients by inverse MST.

Step6. Performance evaluation metrics are calculated for each result and a comparison is presented in tabular form.

IV. Results & Performance Evaluation

Several groups of often-used IR and VI are taken to test the validity of the proposed method. Also, some frequently-used image fusion methods are selected to compare them with the proposed method. The comparison methods are: morphological difference pyramid (MDP); ratio pyramid (RP); contrast pyramid (CP); wavelet transform (WP); dual-tree complex wavelet transform (DTWT); nonsubsampled contourlet transform (NSCT); nonsubsampled shearlet transform (NSST). The results of these methods are taken from [5] for comparison of the obtained results.

A. Evaluation Index System

In order to evaluate the performance of different image fusion methods, the frequently-used image fusion assessment indexes are adopted as the evaluation index system in this paper, and they are mean value (MV), edge based on similarity measure (Q^{ABF}), mutual information (MI), standard deviation (SD), space frequency (SF) and entropy (EN). The higher the assessment index values are, the better quality of the fused image is. Details of indexes can be found in [5].

B. Experimental Results

The 1st pair of IR and VI is shown in Figs. 7 (a) and (b) whose resolution is 496 x 632; and it is the street scene at night, named Bristol Queen's Road image which is one of the most often-used test images. The fused images generated by different methods are shown in Figs. 7 (c)-(i) [5]. Result generated by proposed method is shown in Fig. 7 (j). It can be seen that the proposed method does well in extracting the key features of the source images. There are some artifacts in Figs. 7 (c)-(e). The brightness of the image generated by the proposed method is better than others; and the visible details in VI and the IR areas in IR of the proposed method are more prominent than the others. It shows that the proposed method achieves better result than the contrastive image fusion methods. Table I lists the fusion quality indexes of all experimental fusion methods for the first pair of images. There, it shows that the fused image generated by the proposed method contains more information. The MV, QABF, MI, SD and EN values of the proposed method are much better than most of other methods. The SF value of the proposed method is very close to the others. The EN values of all methods are very close; however, the value of the proposed method is slightly larger than others. According to the above analysis, we can conclude that the proposed method is better than the competitors.

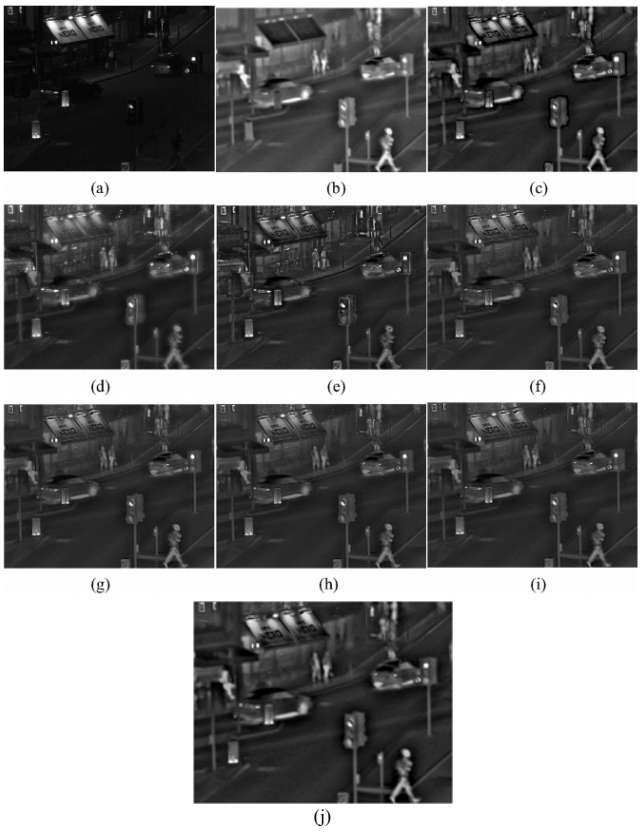

Fig. 7 Source images and fused images using different methods for the first pair of IR and VI: (a) VI, (b) IR, (c) MDP, (d) RP, (e) CP, (f) WP, (g) DTWT, (h) NSCT, (i) NSST, (j) Proposed method

Fig. 5 & Table II show the fusion quality indexes of the results with different decomposition levels (2, 4, 6 & 8) for the first pair of IR and VI images. It is observed that the best result is obtained when the decomposition levels are 6 in the LP. The fusion quality indexes of the results with level 8 are also close but on some indexes the performance is deteriorating like QABF and SF. Moreover, the execution time increases with number of levels 8 without much gain in fusion quality.

Fig. 6 & Table III show the fusion quality indexes of the results with two different fusion rules for high frequency layers: (i) Choose max. criteria in which the maximum of the two coefficients at the corresponding level is selected as the coefficient in the fused layer and (ii) Consistency Check criteria which is an area-based selection rule explained in detail in Section III *C*. This comparison clearly shows that the second

method is giving much better results than the first one.

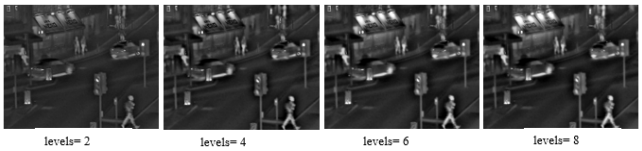

Fig. 5 Fused images using different no. of decomposition levels for the first pair of IR and VI

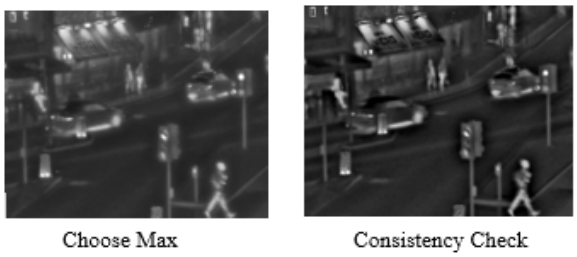

Fig. 6 Fused images using different Fusion rules for the first pair of IR and VI

C. Some More Results

Fig. 8 shows some more fusion results, where first row shows the IR images, second row shows the visible band images and the third row shows their fusion result.

TABLE I
FUSION QUALITY INDEXES WITH DIFFERENT METHODS FOR THE ABOVE PAIR OF VI AND IR IMAGES [5]

	Mv	Q^{ABF}	SD	SF	EN	MI
MDP	50.2136	0.5727	1.7662	31.9488	13.5463	6.4539
RP	**59.7712**	0.2875	1.8827	27.5253	11.070	6.1610
CP	49.924 4	0.3376	1.3681	25.5886	**16.2967**	6.104
WT	51.8918	0.4926	1.8888	23.2813	12.3594	6.036
DTWT	51.8995	0.5053	1.9604	22.9947	12.2322	6.0108
NSCT	51.9115	0.6003	1.7971	26.2640	12.4976	6.1957
NSST	51.8981	0.5174	1.9883	22.9760	12.2422	6.0032
Proposed	54.3574	**0.6675**	**2.6977**	**37.8057**	12.7704	**6.7120**

TABLE II
FUSION QUALITY INDEXES WITH DIFFERENT DECOMPOSITION LEVELS FOR THE ABOVE PAIR OF VI AND IR IMAGES

Level	*Mv*	Q^{ABF}	SD	SF	EN	MI
2	51.9746	0.5817	2.0697	23.9123	12.2017	6.0625
4	52.5067	0.6516	2.0711	31.5615	12.725	6.4729
6	54.3574	**0.6675**	2.6977	37.8057	**12.7704**	6.7120
8	**54.7917**	0.6667	**2.7655**	**39.4327**	12.7027	**6.7401**

TABLE III
FUSION QUALITY INDEXES WITH DIFFERENT FUSION RULE FOR THE ABOVE PAIR OF VI AND IR IMAGES

Fusion Rule (high freq.)	*Mv*	Q^{ABF}	MI	SD	SF	EN
Choose Max	**63.9899**	0.4144	1.8961	30.6090	8.241	6.1669
Consistency Check	54.3574	**0.6675**	**2.6977**	**37.8057**	**12.7704**	**6.7120**

Fig. 8 1st Row: IR Images; 2nd Row: Visual Images (VI); 3rd Row: Fused Images

V. CONCLUSION

In this paper, the image fusion performance of six multi-resolution transforms with different filters and different numbers of decomposition levels are compared. The experimental results show that the optimum number of decomposition levels is six and the proposed method gives the best result among all the compared methods. The number of decomposition levels chosen is a trade-off between spatial detail capture and susceptibility to noise and transform errors.

Open Science Index, Computer and Information Engineering Vol:16, No:8, 2022 publications.waset.org/10012628.pdf

When the number of decomposition levels is more, one coefficient in coarse resolutions corresponds to a larger group of pixels in the fused image. Therefore, an error in coarse resolutions has a great effect on final fused image. Some errors inevitably occur in the process of fusion, producing some artificial distortion. Large decomposition levels give rise to fusion methods that are sensitive to noise. Moreover, large decomposition levels consume more time and have higher memory requirements. When the number of decomposition levels is too small, spatial details cannot be captured well. The comparison in Table III clearly shows that the method of consistency check for fusion of higher layers gives better results than the fusion rule of selecting maximum coefficient at just pixel level. Running time of Proposed fusion method on CPU intel Pentium(R)N3710 (@1.60 GHz, RAM 4G, MATLAB 2020a) is 0.0571 sec (average of 10 readings)

Though several methods have been successful in fusing IR and VI images, there still exist many challenges in image fusion resulting from image noise, moving targets, resolution difference between images, computational complexity, imperfect environmental conditions and limitations of the imaging hardware. Therefore, it is expected that new researches and practical applications utilizing image fusion will continue to grow in future [2].

REFERENCES

[1] Liu, Y.; Liu, S. & Wang, Z. A general framework for image fusion based on multi-scale transform and sparse representation *Information Fusion, Elsevier BV,* 2015*, 24*, 147-164

[2] Liu, Y.; Chen, X.; Wang, Z.; Wang, Z. J.; Ward, R. K. & Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects *Information Fusion, Elsevier BV,* 2018*, 42*, 158-173

[3] Ma, J.; Ma, Y. & Li, C. Infrared and visible image fusion methods and applications: A survey *Information Fusion, Elsevier BV,* 2019*, 45*,

[4] Yaochen Liu Lili Dong, Y. J. & Xu, W. Infrared and Visible Image Fusion through Details Preservation

[5] Jin, X.; Jiang, Q.; Yao, S.; Zhou, D.; Nie, R.; Lee, S.-J. & He, K.Infrared and visual image fusion method based on discrete cosine transform and local spatial frequency in discrete stationary wavelet transform domain *Infrared Physics & Technology, Elsevier BV,* 2018*, 88*, 1-12

[6] Theses, M. & School, G.University of Tennessee, Knoxville Trace: Tennessee Research and Creative Exchange

[7] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. Licciardi, R. Restaino, L. Wald, A critical comparison among pansharpening algorithms, IEEE Trans. Geosci. Remote Sensing 53 (5) (2015) 2565–2586.

[8] A.A. Goshtasby, S. Nikolov, Image fusion: Advances in the state of the art, Inf. Fus. 8 (2) (2007) 114–118.

[9] A.P. James, B.V. Dasarathy, Medical image fusion: A survey of the state of the art, Inf. Fus. 19 (1) (2014) 4–19.

[10] J.G. Han, E.J. Pauwels, P.D. Zeeuw, Fast saliency-aware multi-modality image fusion, Neurocomputing 2013 (111) (2013) 70–80.

[11] W.W. Kong, B.H. Wang, Y. Lei, Technique for infrared and visible image fusion based on non-subsampled shearlet transform and spiking cortical model, Infrared Phys. Technol. 2015 (71) (2015) 87–98.

[12] D.P. Bavirisetti, R. Dhuli, Two-scale image fusion of visible and infrared images using saliency detection, Infrared Phys. Technol. 2016 (76) (2016)

[13] G. Pajares, J.M. de la Cruz, A wavelet-based image fusion tutorial, Pattern Recognit. 37 (9) (2004) 1855–1872.

[14] Z. Zhou, B. Wang, S. Li, M. Dong, Perceptual fusion of infrared and visible im- ages through a hybrid multi-scale decomposition with gaussian and bilateral filters, Inf. Fus. 30 (1) (2016) 15–26.

[15] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, IEEE Trans. Image Process. 22 (7) (2013) 2864–2875.

[16] H. Li, B. Manjunath, S. Mitra, Multisensor image fusion using the wavelet transform, Graph. Models Image Process. 57 (3) (1995) 235–245.

[17] Toet, Alexander (2014): TNO Image Fusion Dataset. figshare. Dataset. https://doi.org/10.6084/m9.figshare.1008029.v1

Open Science Index, Computer and Information Engineering Vol:16, No:8, 2022 publications.waset.org/10012628.pdf