The New Segmentation and Fuzzy Logic based Multi-Sensor Image Fusion

Jamal Saeedi
Electrical Engineering Department
Amirkabir University of Technology
Tehran, Iran.
jamal_saeidi@aut.ac.ir

Karim Faez
Electrical Engineering Department
Amirkabir University of Technology
Tehran, Iran.
kfaez@aut.ac.ir

Abstract—**In this paper, we propose a new multi-resolution fusion algorithm for spatially registered multi-sensor images fusion. First, we use a modified watershed algorithm to produce a region map of source images. This segmentation is then used to generate region-based decision map, which is obtained based on local texture features in the dual-tree discrete wavelet transform domain. The region-based decision map is very accurate for selecting high frequency coefficients between the source images compare to the pixel-based decision map, but very important details of source images is missed in the final fused image. In this study, we propose an algorithm based on fuzzy logic, which uses both pixel and region-based decision map based on a dissimilarity measure of source images for fusion rule. Also a region-based activity measure of each source image is used for low frequency fusion rule instead of simple averaging. Our new method provides improved subjective and objectives results compared to the previously pixel and region-based fusion methods.**

Keywords-image fusion;dual-tree discrete wavelet transform; watershed segmentation; multi-sensor images.

I. INTRODUCTION

Image fusion is described as the task of enhancing perception by integration multiple images into a composite image, which is more appropriate for the purposes of human visual perception and computer processing tasks such as segmentation, feature extraction and target recognition. Important applications of the fusion of images include medical imaging, microscopic imaging [1], remote sensing [2], computer vision, and robotics [3].

In this paper we concentrate on visible and infrared images fusion. The main reason for combining visible and infrared (IR) sensors is that a fused image, constructed by combination of features, enables improved detection and instantly recognizable localization of a target (in the IR image) with respect to its background (in the visible image).

Fusion techniques include the simplest method of pixel averaging to more complicated methods such as principal component analysis [4], and multi-resolution (MR) fusion [5]. MR image fusion is a biologically-inspired method which fuses images at different spatial resolutions. The key step in MR image fusion is the coefficient combination step or fusion rules, specifically, the process of merging the wavelet coefficients in an appropriate way in order to obtain the best quality in the fused image. Some general alternative for construction a fusion rule are demonstrated in Figure 1.The activity level measurement attempts to determine the quality of each source image. Grouping and combining methods are also used to obtain the combined multi-scale representation of the fused image. A consistency verification procedure is then performed which incorporates the idea that a composite MR decomposition coefficient is not likely to be generated in a completely different manner from all its neighbors [6].

Several pixel-based fusion method exist in the literature [7, 8 and 9], wherein each individual coefficient of the MR decomposition (or possibly the coefficients in a small fixed neighborhood) is treated more or less independently. It is followed by many region-based fusion methods to combine objects instead of pixels [10, 11, 12 and 13].

For fusion of IR and visible images, it could be more meaningful to combine objects rather than pixels, because there are some regions or objects in the IR image that are not visible in the visual image, and vice versa. Additional advantage of region-based approach is that the fusion process becomes more robust and, moreover, it may help to avoid some of the well-known disadvantages of pixel-based methods, such as blurring effects, high sensitivity to noise and mis-registration.

Commonly, there are two ways for region-based fusion method: maximum selection (MS) and weighted average (WA). In the MS scheme, a region from one of the source images is selected based on an activity measurement, and other corresponding regions in the remaining images are neglected. In the WA scheme, first, activity measures of the same regions in the different source images are obtained, and then the resultant coefficients for reconstruction are calculated from these measures via a weighted average of the coefficients in the same regions of the different sources.

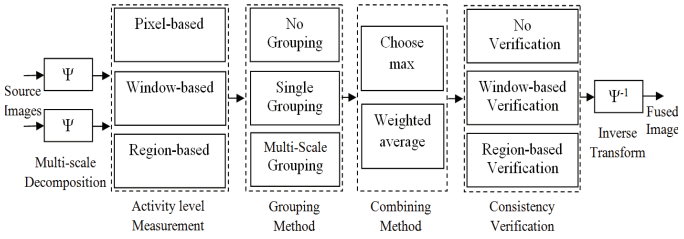


Figure 1: A general framework for multi-resolution image fusion [6].

978-1-4244-4698-8/09/$25.00 ©2009 IEEE

These fusion rules ignore some useful information and sensitive to noise. Selective operation made the fused coefficients completely dependent on the coefficients with larger activity in the image regions and ignores the other corresponding coefficients from other images. In the weighted average scheme, the weights were computed by a linear function which cannot describe the uncertainty of each source image contributions.

In this paper, we propose a new MR fusion algorithm, which combines pixel and region-based method using fuzzy rules. The basic idea is to compute a segmentation map based on all source images and so using this segmentation to obtain region-based decision map from pixel-based decision map, which is obtained through the measurement activities of wavelet coefficients. Also a region-based activity measure of each source image is used for low frequency fusion rule instead of simple averaging, because equally weighing the input approximation sub-bands leads to the problem of contrast reduction. For testing our new fusion algorithm with other wavelet based fusion algorithm, we use dual-tree discrete wavelet transform (DT-DWT), which introduces limited redundancy and allows the transform providing approximate shift invariance and directionally selective filters while preserving the usual properties of perfect reconstruction and computational efficiency [14], [15].

The paper is structured as follows: In section II an image segmentation algorithm using watershed transform is described. In Section III our proposed fuzzy fusion rule is presented. Section IV gives various results and comparisons. Finally, we conclude with a brief summary in section V.

II. Image Segmentation Algorithm

In this section, we describe an algorithm based on watershed segmentation for multivalued images segmentation, especially for IR and visible images.

A. Waterhsed Segmentation

In the immersion-based watershed algorithm of [16], the gradient magnitude of an image is calculated. That image is considered as a topographic relief where the brightness value of each pixel corresponds to a physical elevation. The technique can simply be described by figuring that holes are pierced in each local minimum of the topographic relief. In the sequel, the surface is slowly immersed into a 'lake', by that filling all the catchment basins, starting from the basin that is associated with the global minimum. As soon as two catchment basins tend to merge, a dam is built. The procedure results in a partitioning of the image in many catchment basins of which the borders define the watersheds.

For multivalued images, a single-valued approach can be adopted by segmenting each band, separately, but jointly segmented images work better for image fusion. This is because the segmentation map will contain a smaller number of regions to represent all features in the both IR and visible images.

Here we use the following method to obtain a gradient image from IR and visible images. First a Gaussian derivative function is used to generate gradient magnitude from source images as:

$$G_S(x,y) = \sqrt{(S(x,y) * G'_x)^2 + (S(x,y) * G'_y)^2} \qquad (1)$$

where $S(x,y)$ is the gray scale image, G'_x and G'_y are the Gaussian partial derivative filters in the x and y directions, and $*$ denotes convolution.

It is followed by morphological opening. Opening is based on a $\min(.)$ followed by an $\max(.)$ operation on a local neighborhood around the pixels (denotes by $\tilde{G}_S$). Then a joint gradient image (JG) is obtained using:

$$JG(x,y) = \frac{1}{2}\left(\frac{\tilde{G}_A(x,y)}{\max_{x,y}\left(\tilde{G}_A(x,y)\right)} + \frac{\tilde{G}_B(x,y)}{\max_{x,y}\left(\tilde{G}_B(x,y)\right)} \right) \qquad (2)$$

where A and B are the IR and visible images.

We use the joint gradient image to obtain segmentation map via watershed algorithm.

B. Marker Selection

The watershed algorithm is known to suffer from the over segmentation problem due to noise and other local irregularities of the gradient image. A solution is to limit the number of regional minima. Use markers to specify the only allowed regional minima. First, we use a binary version of the JG image as the external marker (em), and its edges as the internal marker (im). Then we use following formula to modify the JG image as:

$$JG' = imimposemin\ (JG\ , em\ \&\ im\ ) \qquad (3)$$

where *imimposemin* modifies the joint gradient image using morphological reconstruction so it only has regional minima wherever is $(em \,\&\, im)$ nonzero [17], and & is logical AND operator.

Now, the watershed algorithm applied to the modified JG image. Thus, at the end of it we have a joint region map for both IR and visible images (see Figure 2).

III. The New Image Fusion Algorithm

In this section, we explain proposed image fusion algorithm. Figure 3 shows the block diagram of our new method.

Figure 2: Segmentation of IR and visible images. From left to right: IR image, visible image, the joint gradient image and segmentation map.

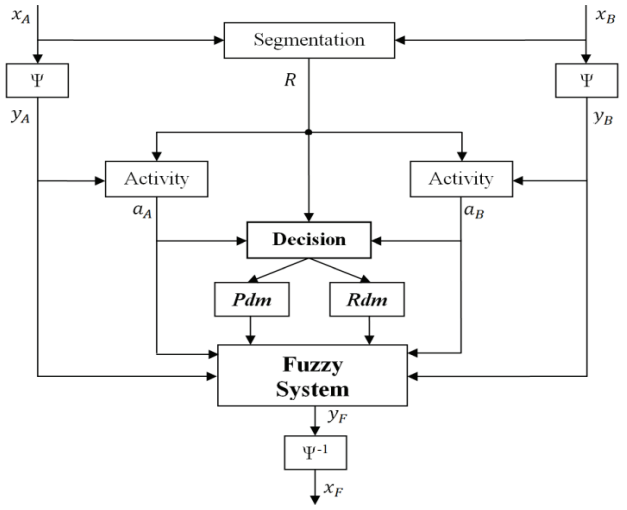


Figure 3: The new segmentation and fuzzy logic based image fusion with two input sources $\boldsymbol{x_A}$ and $\boldsymbol{x_B}$.

First, a pixel-based decision map is obtained through the activities measurement of wavelet coefficients. Having the segmentation map, which is described in the previous section, a region-based decision map is generated from pixel-based decision map. Then we use a dissimilarity measure of wavelet coefficients to combine pixel and region-based fusion rules.

A. Notations

In this paper, the DT-DWT of an image $x = x^0$ is denoted by y and it is assumed in the different scales to be of the form:

$$y = \left[y^1, y^2, ..., y^L, x^L \right] \tag{4}$$

Here, x^L represents the approximation sub-bands at the last level of the DT-DWT, while y^l represent the details sub-bands at level l. Also y^l is composed of $D = 6$ directional sub-bands, i.e. $y^l = \{y^l(.|1), y^l(.|d), ..., y(.|D)\}$. We use the coordinates (i, j) (or the shorthand notation of $(.)$) to index the spatial position of the coefficients.

B. Generation of pixel and Region-based decision map

The activity level measurement is used to judge the quality of a given part of each source image in the transform domain. Activity level measurement can be categorized into three classes: Pixel-based, window-based and region-based.

First, we use two texture features as window-based activities for generating pixel-based decision map in each level of decomposition. The first feature calculates standard deviation, and the second one calculates energy of the high-frequency coefficients in the small neighborhood. We calculate the two features using following formula:

$$a_1^l(n \mid d) = \sum_{\Delta n \in w^l(d)} w^l(\Delta n \mid d) \times \left(y_s^l(n + \Delta n \mid d) - \overline{y_s^l} \right)^2 \tag{5}$$

$$a_2^l(n \mid d) = \sum_{\Delta n \in w^l(d)} w^l(\Delta n \mid d) \times \left| y_s^l(n + \Delta n \mid d) \right| \tag{6}$$

where $w^l(d)$ is a fix window at level l and direction d, and $w^l(.|d)$ are the window's weights, which is obtained by Gaussian function.

Having the two features, pixel-based decision map (Pdm) in each level of decomposition is obtained using:

$$dm^l(.|d) = \begin{cases} 1 & \text{if } a_{1A}^l(.|d) > a_{1B}^l(.|d) \text{ and } a_{2A}^l(.|d) > a_{2B}^l(.|d) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Then

$$Pdm^l(.) = \begin{cases} 1 & \text{if } \left(\frac{1}{D} \sum_{d=1}^{D} dm^l(.|d) \right) > \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

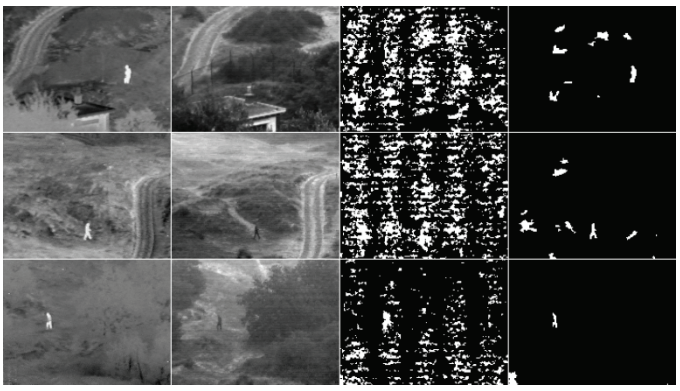
In fact, we use inter-scale dependency of wavelet coefficients in the different orientations (in the DT-DWT at orientations $\pm 15^\circ, \pm 45^\circ, \text{and } \pm 75^\circ$) to obtain a confident Pdm for each level.

Region-based activity measurement also used to obtain the decision map for selecting wavelet coefficients between different source images, but it is very sensitive to the large coefficients in the region and is not very accurate [10], [11]. Here, we use the segmentation map to obtain the region-based decision map (Rdm) based on Pdm and consistency verification:

$$Rdm^l(R) = \begin{cases} 1 & \text{if } \left(\sum_{n \in R} Pdm^l(n) \right) > \frac{|R|}{2} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where $|R|$ is number of pixels in the region.

Figure 4 shows Pdm and Rdm obtained from different IR and visible images.



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

Figure 4: (a) Original IR images, (b) Original visible images, (c) pixel-based decision maps and (d) region-based decision maps at first level.

C. Fuzzy Fusion Rule for High Frequency Sub-bands

Having the pixel and region-based decision maps, we have two strategies for fusion rule:

$$y_1^l(.\mid d) = Pdm^l(.)\times y_A^l(.\mid d) + \left(1 - Pdm^l(.)\right)\times y_B^l(.\mid d) \qquad (10)$$

And,

$$y_2^l(.\mid d) = Rdm^l(.)\times y_A^l(.\mid d) + \left(1 - Rdm^l(.)\right)\times y_B^l(.\mid d) \qquad (11)$$

As it can be seen in Figure 4, the Rdm is very accurate for selecting high frequency coefficients between the source images compare to the Pdm, but very important details of source images is missed in the final fused image. Also in some regions of the source images, we cannot make a good decision for selecting wavelet coefficients, because there is not enough difference between extracted features. As for this reason, we define a new rule using following formula:

$$y_3^l(.\mid d) = \begin{cases} \dfrac{y_A^l(.\mid d) + y_B^l(.\mid d)}{2} & sign\left(y_A^l(.\mid d)\right) = sign\left(y_B^l(.\mid d)\right) \\ \max\left(\left|y_A^l(.\mid d)\right|, \left|y_B^l(.\mid d)\right|\right) & otherwise \end{cases} \qquad (12)$$

We want to design a good fusion rule with combining these three fusion rules to integrate as much information as possible into the fused image. We define a dissimilarity measure for this purpose. The dissimilarity measure (Dis) is intended to quantify the degree of 'dissimilarity' between the source images $y_A^l(.)$ and $y_B^l(.)$. In the following expression, this measure is defined as:

$$f_1^{\,l}(i,j) = \frac{1}{D}\sum_{d=1}^{D}\left|a_{1A}^l(i,j\mid d) - a_{1B}^l(i,j\mid d)\right| \qquad (13)$$

$$f_2^{\,l}(i,j) = \frac{1}{D}\sum_{d=1}^{D}\left|a_{2A}^l(i,j\mid d) - a_{2B}^l(i,j\mid d)\right| \qquad (14)$$

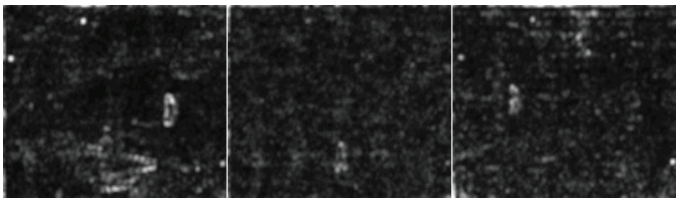
Then

$$Dis^{\,l}(i,j) = \left[\sin\left(\frac{\pi \times f_1^{\,l}(i,j)\times f_2^{\,l}(i,j)}{2\times T}\right)\right]^{1/2} \qquad (15)$$

where $a_1^l(.)$ and $a_2^l(.)$ are obtained from equations (2) and (3), and $T = \max_{i,j}(f_1^l(i,j)\times f_2^l(i,j))$.

Figure 5 shows the Dis measure for the IR and visible images. By analyzing the Dis measure, we can determine where the source images differ and to what extent, and use this information to combine fusion rules. First we define the following linguistic rules for fusion rule:

IF the Dis measure at a given position is high (i.e. the sources are distinctly different at that position) **THEN** we use the first fusion rule.



(a) (b) (c)

Figure 5: Dissimilarity measure at first level of decomposition for the images (a) "UN camp", (b) "Dune" and (c) "Trees".

IF the Dis measure at a given position is medium (i.e. the sources are different at that position) **THEN** we use the second fusion rule.

IF the Dis measure at a given position is low (i.e. the sources are similar at that position) **THEN** we use the third fusion rule.

Then, for constructing standard rules from Linguistic rules, we define two thresholds T_1 and T_2:

$$y_F^{\,l}(.\mid d) = \begin{cases} y_3^{\,l}(.\mid d) & Dis^{\,l}(.) < T_1 \\ y_2^{\,l}(.\mid d) & T_1 < Dis^{\,l}(.) < T_2 \\ y_1^{\,l}(.\mid d) & Dis^{\,l}(.) > T_2 \end{cases} \qquad (16)$$

We obtained $0.05 \le T_1 \le 0.1$ and $0.15 \le T_2 \le 0.3$ using test images and try and error. Also for improving object contrast in the final fused image, we use the following formula to modify wavelet coefficients in the transform domain:

$$\tilde{y}_F^l(.\mid d) = \left(1 + Dis^l(.)\right)\times y_F^l(.\mid d) \qquad (17)$$

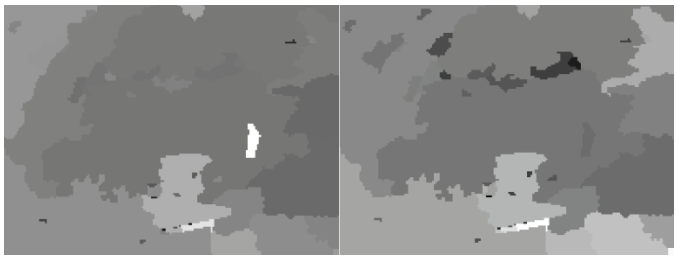
D. Fusion Rule for Low Frequency Sub-bands

Commonly, averaging is a popular method to fuse approximation or low frequency sub-bands of the source images [10], [11]. However, equally weighing the input approximation sub-bands leads to the problem of contrast reduction, because IR and visible images have different contrast conditions. In this study, a weighted averaging method uses for low frequency fusion rule. We use a region-based activity measurement (the normalized Shannon entropy) of high frequency coefficients in the first level of decomposition as the priority (pr) of the region, for weighted combination:

$$pr(R) = \frac{1}{D}\sum_{d=1}^{D}\frac{1}{|R|}\sum_{n\in R}\left|y^1(n\mid d)\right|\times \log\left|y^1(n\mid d)\right| \qquad (18)$$

Then

$$x_F^L(.) = \frac{\hat{pr}_A(.)\times x_A^L(.) + \hat{pr}_B(.)\times x_B^L(.)}{\hat{pr}_A(.) + \hat{pr}_B(.)} \qquad (19)$$



(a) (b)

Figure 6: Priorities generated using region-based activity, (a) for the IR images and (b) for the visible images.

where $\hat{p}r_A$ and $\hat{p}r_B$ are obtained via down-sampling from the pr_A and pr_B with respect to level of decomposition, and $0 \times \log(0) = 0$.

A pair of priority for the IR and visible images is shown in Figure 6.

IV. Experimental Results

The proposed image fusion method was tested against several state-of-the-art image fusion methods consist the simple averaging, the Dual-tree discrete wavelet transform with maximum selection rule [18] and Lewis's region-based algorithm [10], in two modalities: IR and visible.

The images used in experiments are surveillance images from TNO Human Factors, publicly available at the Image Fusion web site [20]. Image sequence ''UN Camp'' consists of 32 images (32 visible and 32 IR images) and image sequences ''Trees'' and ''Dune'' contain 19 images.

Also we apply contrast stretching to enhance the representation of the source image. Other image processing techniques like histogram equalization can be used. Figures 7, 8 and 9 show the results of different fusion methods.

Visual (subjective) comparison between methods indicates that our method is better than pervious image fusion algorithms. For example, in Figure 7, it is clear that the fence details from the visible image and the person details from the IR image are far better transferred into the fused image in the proposed method than in the other pixel-based methods. Also in the fused image using Lewis's region-based method, some details such as the contours of trees and the bright points, are not transferred in to the fused image.

In addition, it can be seen in Figures 7, 8 and 9 the person is brighter in the proposed method and images contrast is far better compare to other methods.

Two metrics are considered in this paper, which do not require ground truth images: The Xydeas and Petrovic $Q^{AB/F}$ metric, proposed in [21], considers the amount of edge information transferred from the input images to the fused images using a Sobel edge detector to calculate the strength and orientation information at each pixel in both source and the fused images, and the Entropy metric, which measures the information content in an image. An image with high information content will have high entropy.

Table 1 shows the objective fusion results from different methods. It should be mentioned that the obtained results is different with results listed in previous publications, because

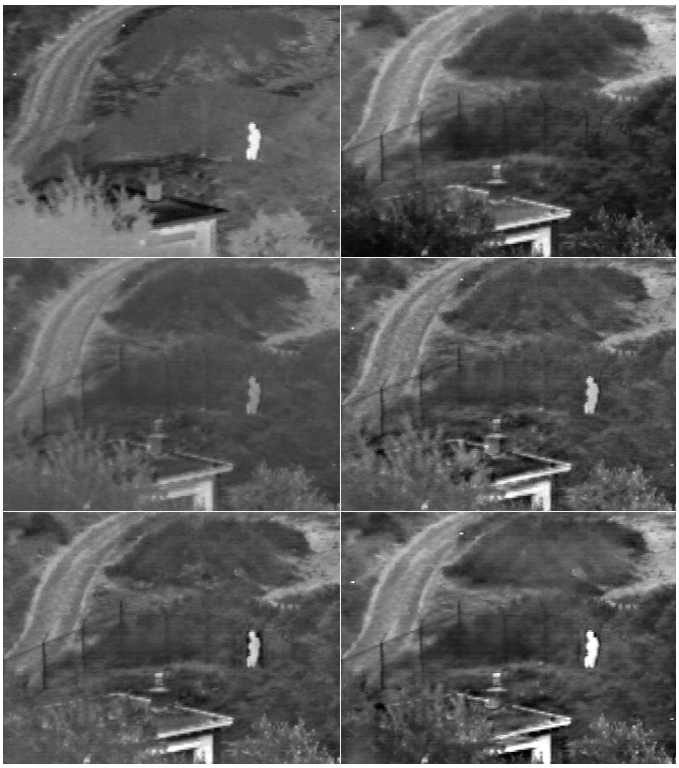


Figure 7: Subjective fusion results of "Un camp" images. Original IR image, original visible image (top-row), fused images using averaging, DT-DWT and MS rule (middle-row), Lewis's method and proposed method (bottom-row).

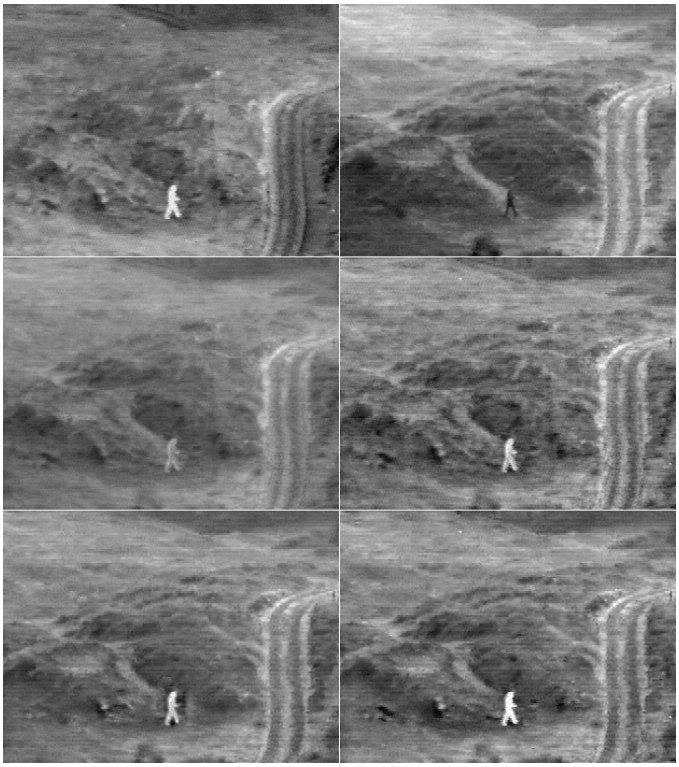


Figure 8: Subjective fusion results of "Dune" images. Original IR image, original visible image (top-row), fused images using averaging, DT-DWT and MS rule (middle-row), Lewis's method and proposed method (bottom-row).

Figure 9: Subjective fusion results of "Trees" images. Original IR image, original visible image (top-row), fused images using averaging, DT-DWT and MS rule (middle-row), Lewis's method and proposed method (bottom-row).

we apply contrast stretching to enhance the representation of the source images. The metric's values confirm the subjective assessment, that our proposed image fusion algorithm generally integrates more details from the visible and IR images into the fused images.

V. CONCLUSION

In this paper, we have presented a new wavelet based multi-sensor image fusion method using watershed segmentation and fuzzy rules. Proposing a new fusion rule for merging wavelet coefficients, which is the second step in the wavelet based image fusion, is the main novelty of this paper. Also this new method used DT-DWT for finer frequency decomposition and shift invariant property compared to discrete wavelet transform. The experimental results demonstrated that the proposed method outperforms the standard fusion methods in the fusion of IR and visible images.

Table 1: Performance of Image Fusion methods

Metric	Method	Un camp	Dune	Trees
Petrovic	Averaging	0.332	0.338	0.351
	DT-CWT (MS)	0.439	0.409	0.470
	Lewis [10]	0.503	0.502	0.553
	Proposed	**0.540**	**0.525**	**0.602**
Entropy	Averaging	6.29	6.75	6.47
	DT-CWT (MS)	6.75	6.90	6.58
	Lewis [10]	6.66	6.88	6.54
	Proposed	**6.92**	**7.08**	**6.79**

REFERENCES

[1] Garg, S., Ushah Kiran, K. Mohan, R. Tiwary, "Multilevel Medical Image Fusion using Segmented Image by Level Set Evolution with Region Competition," *27th Annual International Conference of the Engineering in Medicine and Biology Society*, 17-18 Jan, pp. 7680-7683, 2006.

[2] Xu-Hong Yang, Zhong-Liang Jing, Gang Liu, Li Zhen Hua, "Fusion of multi-spectral and panchromatic images using fuzzy rule," *Communications in Nonlinear Science and Numerical Simulation 12*, pp. 1334–1350, 2007.

[3] Kam, M., Zhu, X., and Kalata, P, "Sensor fusion for mobile robot navigation," *Proceedings of the IEEE 85*, pp. 108-119, January 1997

[4] Kumar, S. Senthil, Muttan, S, "PCA-based image fusion," Proceedings of the SPIE, Volume 6233, pp. 62331T, 2006.

[5] Rui. Zhang Ke, Li Yan-Jun, "An Image Fusion Algorithm Using Wavelet Transform," *ACTA ELECTRONICA SINICA*, Vol.32, No.5, pp. 750-75, 2004.

[6] R. S. Blum and Zheng Liu, Multi-Sensor Image Fusion and Its Applications, *Taylor & Francis Group*, New York, 2006.

[7] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany, pp. 173–182, May 1993.

[8] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," Graphical Models and Image Processing, vol. 57, no. 3, pp. 235–245, May 1995.

[9] Koren, A. Laine, and F. Taylor, "Image fusion using steerable dyadic wavelet transforms," in Proceedings of the IEEE International Conference on Image Processing, Washington D.C., pp. 232–235, October 1995.

[10] J. Lewis, R. O'Callaghan, S. Nikolov, D. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," *Information Fusion*, vol. 8, no. 2, pp. 119–130, Apr. 2007.

[11] G. Piella, H. Heijmans, "Multiresolution image fusion guided by a multimodal segmentation," in: Proceedings of Advanced Concepts of Intelligent Systems, Ghent, Belgium, pp. 1-8, September 2002.

[12] N. Cvejic, D. Bull, and N. Canagarajah, "Region-Based Multimodal Image Fusion Using ICA Bases," IEEE SENSORS JOURNAL, VOL. 7, NO. 5, pp. 743-75, 1MAY 2007.

[13] T.Wan, N. Canagarajah, and A. Achim, "Segmentation-Driven Image Fusion Based on Alpha-Stable Modeling of Wavelet Coefficients," IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 11, NO. 4, pp. 624-633 JUNE 2009.

[14] N.G. Kingsbury, "The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters," in *Proc. 8th IEEE DSP Workshop*, Utah, Aug. 9–12, 1998.

[15] Ivan W. Selesnick, Richard G. Baraniuk, and Nick G. Kingsbury, "The Dual-Tree Complex Wavelet Transform," *IEEE Signal Processing Magazine*, pp. 124-152, 2005.

[16] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Machine Intel*, vol. 13, pp. 583–593, 1991.

[17] Rafael C. Gonzalez, Richard E. Woods, *digital image processing*, Tom Robbins, Upper Saddle River, New Jersey, 2002.

[18] Sun. Wei., Wang. Ke, "A Multi-Focus Image Fusion Algorithm with DT-CWT," *International Conference on Computational Intelligence and Security*, pp. 147-151, 2007.

[19] P. Soille, *Morphological Image Analysis, Principles and Applications*. Berlin, Germany: Springer-Verlag, 1999.

[20] http://imagefusion.org

[21] V. Petrovic, C. Xydeas, "On the effects of sensor noise in pixel-level image fusion performance," in: Proceedings of the Third International Conference on Image Fusion, vol. 2, pp. 14–19, 2000.