

PAPER • OPEN ACCESS

Image enhancement using thermal-visible fusion for human detection

To cite this article: Ezrinda Mohd Zaihidee *et al* 2017 *J. Phys.: Conf. Ser.* **890** 012038

View the [article online](#) for updates and enhancements.

Related content

- [Advanced Digital Imaging Laboratory Using MATLAB® \(Second edition\) : Methods of image enhancement](#)
L P Yaroslavsky
- [Study on Efficiency of Fusion Techniques for IKONOS Images](#)
Yanmei Liu, Haiyang Yu, Yang Guijun *et al.*
- [Remote sensing image fusion in the context of Digital Earth](#)
C Pohl

Image enhancement using thermal-visible fusion for human detection

Ezrinda Mohd Zaihidee^{1,2}, Kamarul Hawari Ghazali¹ and Mohd Zuki Saleh²

¹Faculty of Electrical & Electronics Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang

²Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, 26300 Gambang, Pahang

E-mail: ezrindamz@gmail.com

Abstract. An increased interest in detecting human beings in video surveillance system has emerged in recent years. Multisensory image fusion deserves more research attention due to the capability to improve the visual interpretability of an image. This study proposed fusion techniques for human detection based on multiscale transform using grayscale visual light and infrared images. The samples for this study were taken from online dataset. Both images captured by the two sensors were decomposed into high and low frequency coefficients using Stationary Wavelet Transform (SWT). Hence, the appropriate fusion rule was used to merge the coefficients and finally, the final fused image was obtained by using inverse SWT. From the qualitative and quantitative results, the proposed method is more superior than the two other methods in terms of enhancement of the target region and preservation of details information of the image.

1. Introduction

The fusion of thermal and visible images become more popular in image enhancement where it combines the advantages in thermal and visible images to produce a higher quality of image. A good quality of image is very important for the systems to detect and monitor the presence of a human, hence give alarms for suspicious activities [1]. Close Circuit Television (CCTV) or visible cameras are commonly used for monitoring systems. These cameras captured images with high resolution and provide details of the scene but they only works well with good lighting conditions [2]. The characteristics of visible camera are totally different with those of thermal camera. An image captured by a thermal camera has low contrast, is very sensitive to temperature changes and lack of background details but it can detect an infrared radiation from different objects in dark environment [3]. Therefore, the fusion image of visible and thermal cameras will overcome the weaknesses in both cameras and give better information about the image.

Extensive works have been done on thermal-visible fusion for indoor or outdoor human detection. It is a challenging task because both cameras have different modalities with different fields of view. Some of the challenges are in handling different data modalities, data imperfection, data alignment etc [4]. Multiscale transform at pixel level become more popular in thermal-visible fusion. It converts the raw input images to a more convenient representation and has low complexity compared to fusion at feature or decision level.

Curvelet [5] and contourlet transform [6] are not translation invariant due to upsampling and downsampling in the transformation process. Besides, the edge of the fused images is not smooth and



it may affect the detection of a human. Therefore, Cunha et al [7] proposed a Non-Subsampled Contourlet Transform (NSCT) that is more suitable and can overcome the drawback in contourlet transform. Then, several researchers applied and enhanced this method for human detection [8,9]. The results are slightly better than traditional NSCT but some of the images still have low contrast. Other than that, saliency analysis [10], combination of saliency analysis and non-subsampled Shearlet transform [11], gradient transfer [12], Gaussian [13] and sparse representation [14] can also be used. Some of these methods produced good results but with low computational speed.

This paper proposed another fusion approach using Stationary Wavelet Transform (SWT). The wavelet transform is widely used in digital signal processing and the concept of SWT are introduced in the context of image fusion in Section 2. Section 3 briefly presents the experiments. Section 4 provides results and discussion based on qualitative and qualitative analysis. Finally, Section 5 concludes the paper.

2. Fusion Techniques

It is very important to preserve as much as information in input images to get a better quality of fused image. In SWT, the input images of $m \times n$ pixels, thermal image, I_{IR} and visible image, I_{VS} will be decomposed into approximation, A , vertical details, VD , horizontal details, HD and diagonal details, DD components respectively as shown in figure 1. All of these components represent lowpass and highpass filter for each decomposition level. In SWT, it modifies the filters at each level by padding them with zeroes to remain the coefficients during the decomposition process. Then, the coefficients for approximations, A_{IR} and A_{VS} were merged using maximum absolute value fusion rule because it contains main energy of the image. For merging the highpass coefficients or details components, the appropriate fusion rule was chosen to produce the best fused image. Finally, fused image, I_F was obtained using inverse SWT.

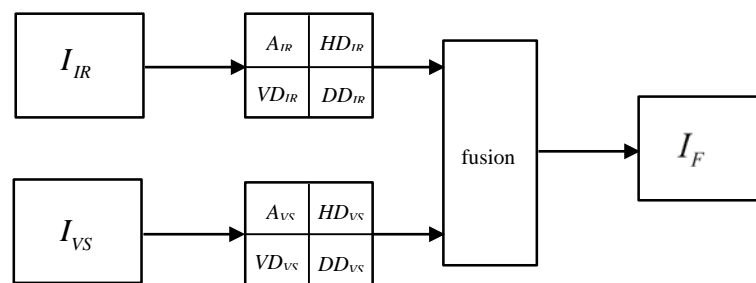


Figure 1. SWT fusion scheme.

3. Experiments

3.1 Dataset and experimental setting

Thermal-visible nighttime imagery dataset were obtained from online dataset [15,16] where it consists of three different scenarios. All the images have the same size, 256×256 . These images captured a human in outdoor environment. The experiments were performed using MATLAB R2014a with 2.30Ghz Intel® Core™ i3-2350M CPU and 8GB of main memory.

3.2 Qualitative Analysis

Instead of objective assessment, all fused images are also evaluated by visual interpretation. A group of 20 people needs to answer two questions (overall quality of the image and human physical appearance) based on the given input and fused images for three different scenarios. The viewer can rate the given question based on the absolute and relative measure as shown in table 1. This evaluation is simple and easy to analyze, however it depends largely on the observer's experience, personal preference and viewing condition [17].

Table 1. Qualitative Measurement.

GRADE	ABSOLUTE MEASURE	RELATIVE MEASURE
1	Very poor	Missing information, very blurred image
2	Fair	Average, good image but not so clear
3	Good	Very clear image

3.3 Quantitative Analysis

These images have been measured quantitatively using Image Quality Index [18] without a reference image and also based on fusion factor [19]. The chosen quality matrices for objective assessment are as follows:

(a) Standard Deviation (SD)

It measures the contrast in the fused image. The standard deviation of an $M \times N$ image is given by

$$SD = \sqrt{\frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N (f(n,m) - \mu)^2} \quad (1)$$

where M, N represent the number of rows and columns respectively, $f(n,m)$ is the pixel value of fused image and μ is mean value of the image. The higher value of SD indicates higher contrast of fused image.

(b) Information Entropy (IE)

IE measures the richness of information in fused image. IE [20] is defined as:

$$IE = - \sum_{i=0}^{L-1} \left(\frac{N_i}{N} \right) \log \left(\frac{N_i}{N} \right) \quad (2)$$

where L is a grey level of fused image, N_i is the pixel number of the grey value i and N represents the total pixel number of the image. The larger value of IE implies that fused image has more abundant information.

(c) Fusion Factor (FF)

For two input images A, B and fused image F , FF is defined as:

$$FF = I_{AF} + I_{BF} \quad (3)$$

where I_{AF} and I_{BF} are the similarity information between input images and fused image. Higher value of FF indicates that the image has better quality.

4. Results and discussion

The results of the fused images are divided into two parts; qualitative and quantitative evaluation. Two methods have been chosen to be compared with the proposed method. The chosen methods are Weighted Averaging Method (AVG) and Discrete Wavelet Transform (DWT). The first two images in figure 2, 3 and 4 represent grayscale visible light and thermal images and also known as input images. In the thermal images, the thermal radiation of an object in the scene was detected, while the grayscale visible light images provide details of the background information.

4.1 Assessment Results for Image A

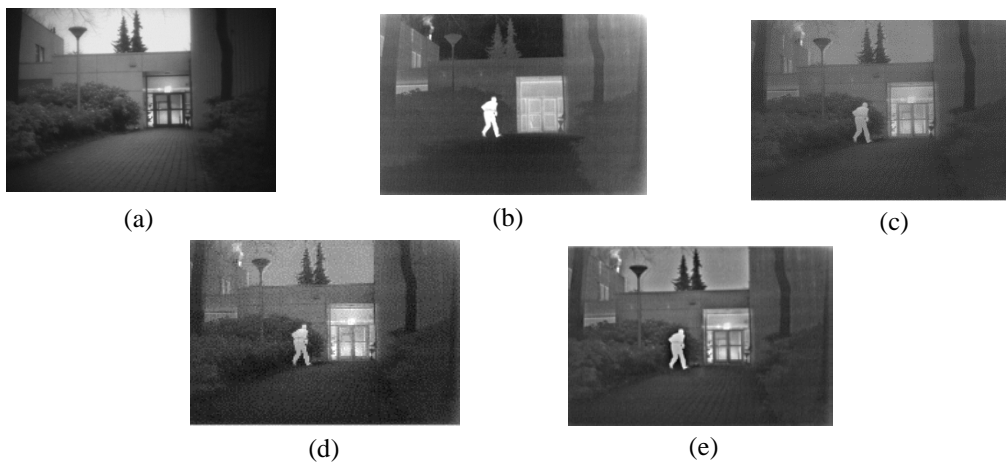


Figure 2. Image A (a) visible image (b) thermal image (c) AVG (d) DWT (e) proposed SWT.

Table 2. Image Quality Index for Image A.

METHOD	EVALUATION METRIC		
	IE	SD	FF
AVG	6.2449	6.2131	0.8032
DWT	5.9025	7.0300	1.2337
PROPOSED METHOD	7.1012	7.5611	1.4565

Figure 2 (a) and (b) show an elderly walking towards the door. Although both cameras captured the same scene but there is a big difference in terms of the details of the image. In figure 2(c), the fused image is not clear and information on the background details is missing while in figure 2(d) there is a presence of noise and the edge is not well preserved as in figure 2(e). The results in table 2 for IE and SD also show that less information can be obtained from the fused image using AVG and DWT methods. Besides, among 20 respondents, 75% of them agreed that the fused image by SWT provide the best quality of image.

4.2 Assessment Results for Image B

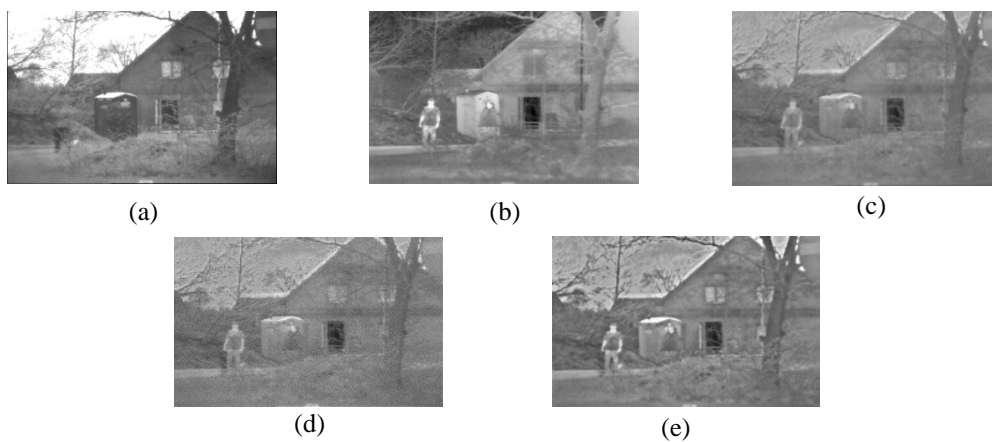


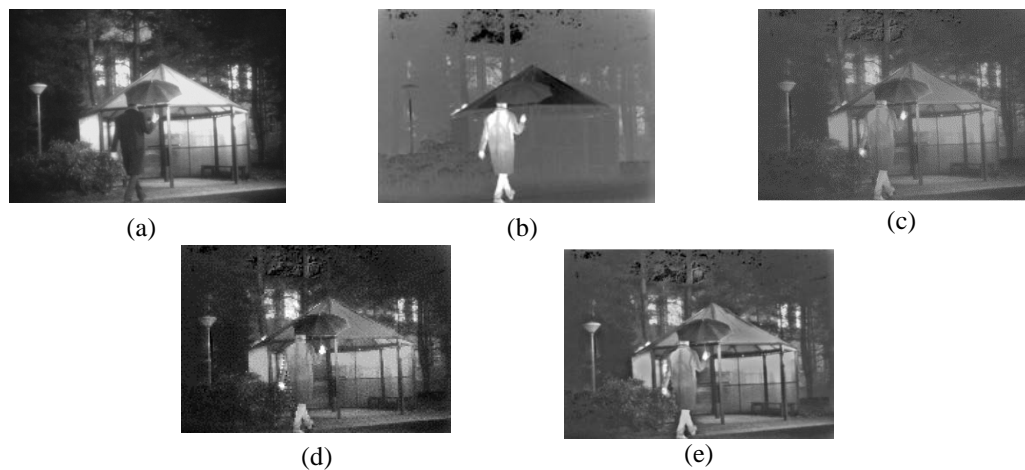
Figure 3. Image B (a) visible image (b) thermal image (c) AVG (d) DWT (e) proposed SWT.

Table 3. Image Quality Index for Image B.

METHOD	EVALUATION METRIC		
	IE	SD	FF
AVG	5.4233	6.0021	0.9030
DWT	5.9128	6.4540	0.9004
PROPOSED METHOD	6.3400	7.2151	1.2309

As in figure 3, image B shows two person standing side by side in front of the house. The comparison between these three images (figure 3(c), (d) and (e)) clearly shows that the fused image by SWT has high contrast and all the details are very clear. From the results in table 3, the proposed method has the highest values for IE, SD and FF. From the questionnaires, 18 out of 20 respondents stated that the fused image by SWT is clearer and can detect human better than other two methods.

4.3 Assessment Results for Image C

**Figure 4.** Image C (a) visible image (b) thermal image (c) AVG (d) DWT (e) proposed SWT.**Table 4.** Image Quality Index for Image C.

METHOD	EVALUATION METRIC		
	IE	SD	FF
AVG	6.2122	5.4247	8.8840
DWT	7.7394	7.6001	1.2426
PROPOSED METHOD	8.0021	7.2130	1.4560

As in figure 4(c), (d) and (e), the fused images show a person walking with an umbrella. Although it preserved the thermal image characteristic, it is still lack of background details in figure 4(c) as well as low contrast as indicated in table 4 by the value of IE and SD. The fused image by DWT is clear and quite similar with the input images where the FF value is above 1.0 however, some noise is still present compared to the proposed method. 80% of the respondents agreed that figure 4(e) is the best image in terms of human physical appearance and background details.

5. Conclusion

This paper proposed a shift invariant wavelet transform known as SWT at pixel level fusion where the two source images were decomposed using SWT, then the appropriate fusion rule were chosen for low and high frequency components. Finally, the fused image was obtained using inverse SWT. The proposed method was compared with two state-of-the-art methods and the experimental results clearly

indicate that SWT provide better quality of images in terms of both objective and subjective evaluations.

References

- [1] Lau R Y K, Pramanik M I and Chowdhury M K H 2016 *Proc. of the 20th Pacific Asia Conference on Information Systems (Chiayi)* vol 1 (Taipei City: Association for Information Systems)
- [2] Ibrahim S 2016 *Communication Science Technology* **1** p 7
- [3] Budzan S and Golik R W Y Z 2015 *Measurement Automation Monitoring* vol 61 (Gliwice: PAK Publishing House) pp 187-90
- [4] Khaleghi B, Khamis A and Karray F O 2011 *Information Fusion* **14** pp 28-44
- [5] Quan S, Qian W, Guo J and Zhao H 2014 *2nd Int. Conf. on Systems and Informatics (Shanghai)* vol 2 (Shanghai: Institute of Electrical and Electronics Engineers) pp 828-32
- [6] Yu Z, Yan L, Han N and Liu J 2015 *Cybernetics and Information Technologies* **15** pp 116-25
- [7] Cunha A L, Zhou J and Do M N 2006 *Image Processing* **15** pp 3089-101
- [8] Lu F and Lei X 2015 *Int. Conf. on Image and Graphics (Tianjin)* vol 3, ed Zhang Y J (Berlin, Springer) pp 297-306
- [9] Yucheng L and Yubin L 2014 *Scientific Research and Essays* **9** pp 374-79
- [10] Zhao J, Gao X, Feng H and Wang D 2016 *Infrared Physics & Technology* **76** pp 295-302
- [11] Baohua Z, Xiaoqi L, Haiquan P and Ying Z 2015 *Infrared Physics & Technology* **73** pp 286-97
- [12] Ma J, Chen C, Li C and Huang J 2016 *Information Fusion* **31** pp 100-9
- [13] Lee S and Lee D 2016 *Mathematical Problems in Engineering* **2016** pp 1-9
- [14] Liu Y, Liu S and Wang Z 2014 *Information Fusion* **24** pp 147-64
- [15] Toet A, Hogervorst M A, Son V and Dijk J 2011 *Int. Journal of Image and Data Fusion* **2** pp 287-308
- [16] Toet A, IJspeert J K, Waxman A M and Aguilar M 1997 *Displays* **18** pp 85-95
- [17] Zhang Y 2008 *The Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Beijing)* vol XXXVII, ed Jun C, Jie J and Genderen J V (Gottingen, Copernicus Publications) pp 1101-9
- [18] Jagalingam P and Hegde A V 2015 *Aquatic Procedia* **4** pp 133-42
- [19] Singh R and Khare A 2014 *Information Fusion* **19** pp 49-60
- [20] Verdu S 1998 *IEEE Transactions on Information Theory* **44** pp 2057-78