A Hybrid Thermal-Visible Fusion for Outdoor Human Detection

Ezrinda Mohd Zaihidee[1], Kamarul Hawari Ghazali[1], Jinchang Ren[2] and Mohd Zuki Salleh[3]
[1]*Faculty of Electrical & Electronics Engineering, Universiti Malaysia Pahang.*
[2]*Department of Electronic and Electrical Engineering, University of Strathclyde.*
[3]*Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang.*
ezrindamz@gmail.com

***Abstract*—Multisensory image fusion can be used to improve the visual interpretability of an image for further processing task. A hybrid thermal-visible image fusion is proposed in this paper to detect the target and produce an output image that had all the information from both sensors. The thermal target region was extracted using Niblack algorithm and some morphological operators. Then, the source images were decomposed at pixel level using Stationary Wavelet Transform (SWT). The appropriate fusion rule were chosen for low and high frequency components and finally the fused image was obtained from inverse SWT. Results show that the proposed method achieve better to include figure of merit than the other three methods.**

***Index Terms*—Multisensory Image Fusion; Niblack Algorithm; Stationary Wavelet Transform; Thermal-Visible Image Fusion.**

I. INTRODUCTION

Image fusion systems are widely used in medical diagnosis, remote sensing, surveillance, navigation guidance and agriculture. Multisensory image fusion is a technique that combine two or more images that capture the same scene but with different imaging conditions (sensor types, weather conditions, time, polarization modes etc.) to obtain an image with highly comprehensive and quality for further processing tasks [1], [2]. The concept of information fusion in military defense systems was proposed in a research report of the American military aspect in 1972 [3]. From the concept being proposed until today, numerous studies on new approaches and methods have been developed extensively to meet current demand.

Nowadays, there is a high demand in detecting and tracking human beings in surveillance systems. Among items monitored are abnormal behavior, human gait characterization and person identification [4]. These items required continuous monitor. There are many Closed Circuit Television (CCTV) being used for monitoring any suspicious activities but the cameras work well in controlled condition only and not able to work as a real time alarm system without human control [5]. Visible camera more accord to human visual characteristics and relatively high contrast ratio. In other words, it is not easy to make observation under low lighting conditions and in an uncontrolled outdoor environment [6]. To overcome this short comings, thermal camera was proposed for video surveillance system. It capture thermal radiation from different objects. Hence, it has obvious target information regardless of lighting conditions, haze or fog. However, it is difficult to identify the actual activity due to lack of details in the captured image. The low value of signal-to-noise ratio and contrast level [7] affect the quality of an image. Thus, combining both images will result in better awareness. Thermal-visible image fusion will provide better images as each image contains complementary information, different contrast level and resolution although they depict the same region at the same time [8].

In recent years, the research on thermal-visible image fusion pays particular attention to detect human at pixel level based on multiscale transforms (MST) where it consists three general steps. Firstly, decompose the input images into a multiscale transform domain. Then, merge the low and high frequency coefficients in each layer using appropriate fusion strategies. Finally, reconstruct the fused image at every layer to get the final image for further processing [9].

There are many techniques in image fusion using MST either in continuous or discrete domain. Curvelet transform in continuous domain is difficult to sample, therefore the edges of the final fused image are not smooth [10]. The contourlet transform in discrete domain [11], [12] that inspired by curvelet transform produce a better result in terms of image smoothness. However, it is not translation invariant due to upsampling dan downsampling. Then, many researchers tend to introduce other techniques in thermal-visible fusion for human detection such as non-subsampled countourlet transform [13], [14], saliency extraction with multiscale decomposition [15], [16], gradient transfer fusion [17], MST with sparse representation [18] and others.

This paper proposed a hybrid thermal-visible fusion for human detection based on thermal target extraction and Stationary Wavelet Transform (SWT). The target region in input thermal image was extracted using Niblack algorithm and some morphological operators. At the same time, both input images was decomposed into low and high frequency coefficient using SWT. Then, appropriate fusion rule was chosen to integrate the coefficients. Finally, a fused image was obtained using inverse SWT and mapping with the target region.

In the sequel, the fusion techniques are presented in Section II. In Section III the experimental setting for this research is described in detail. Then, Sections IV and V discuss the results and concludes the paper, respectively.

II. FUSION TECHNIQUES

For thermal-visible fusion, all the target regions that are captured by thermal camera were extracted for further processing using local thresholding technique and some morphological operators. At the same time, the raw thermal image and visible image were fused using Stationary Wavelet

Transform using maximum local energy for low frequency coefficients and absolute maximum value for high frequency coefficients. The research design of the proposed method is illustrated in Figure 1.

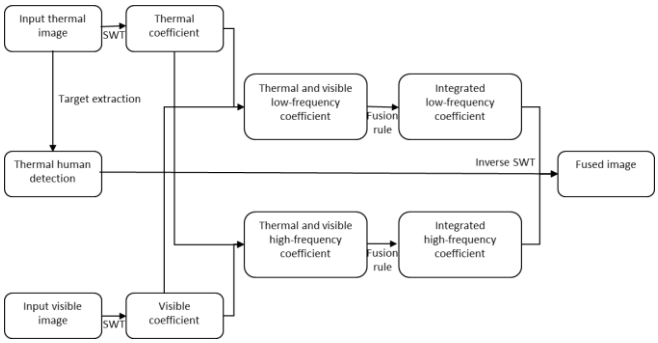


Figure 1: Research design

A. *Detection of the Thermal Target Region*

The presence of noise in thermal images influences the quality of the image. Histogram equalization was used to enhance the image contrast. Then, local thresholding technique using Niblack algorithm was chosen for thermal segmentation due to its capability to detect the targeted region although in low quality image. It calculates a threshold in 25 × 25 rectangular window at each pixel as shown in Equation (1). Experiments show that the best value of p is 0.6.

$$T_{Niblack} = m + p.s \quad (1)$$

where:

$m(x, y)$: Mean value of the pixels
$s(x, y)$: Standard deviation of the pixels
p : Correction coefficient

Morphological operators were used at the final stage of the detection. The basic effect of dilation expands the foreground while the erosion operation does the opposite. It shrinks the foreground by removing the pixels of the foreground boundaries. Hence, it results in better detection for further processing.

B. *Fusion Rules Based on Stationary Wavelet Transform*

In 1990's researchers found that Wavelet Transform (WT) can be a powerful solution as high pass filter in image processing applications. It is more useful than Fourier Transform in terms of frequencies in time and space. In multi sensor image fusion, it extract relevant information from the input images based on frequency domain to produce a single image with higher quality for further processing.

There are two main groups of WT, continuous and discrete. Discrete Wavelet Transform (DWT) is commonly used because it is more practical to apply in image fusion. However, due to lack of translation invariant property, a new kind of WT known as Stationary Wavelet Transform (SWT) is shown in Figure 2. In SWT, instead of repetition process of upsampling and downsampling like DTW, it upsamples the filter coefficients by the factor of in the jth level of the algorithm. That means if the signal is shifted at each level of decomposition, the coefficient will remain. Therefore, it will improve the resolution of edge details in the three sub images; horizontal, vertical and diagonal. The characteristic of translation invariant of SWT make it more suitable for image fusion applications [19], [20].

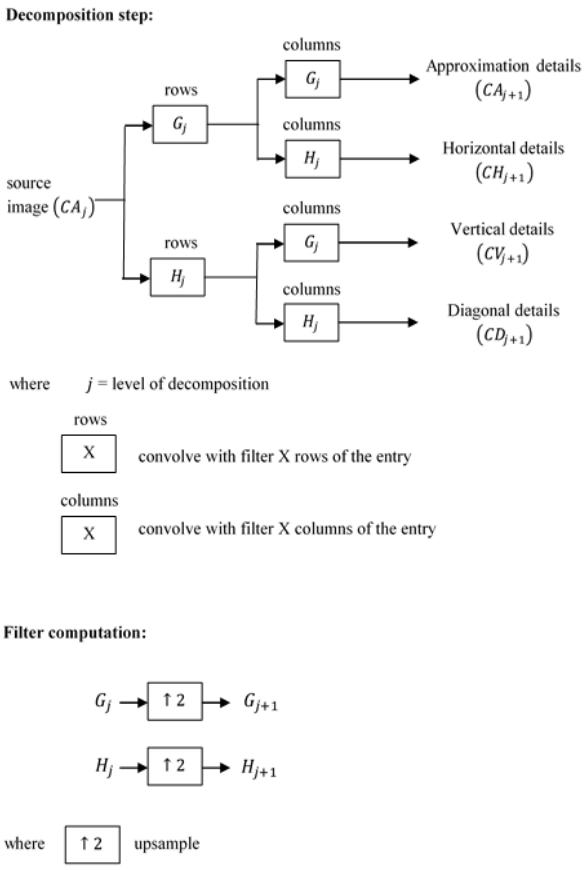


Figure 2: Block diagram of SWT

III. EXPERIMENTS

In this experiment, all visible and infrared images are taken from publicly available dataset that have been perfectly processed by image registration. Qualitative and quantitative analyses of fusion results using Image Quality Index [21] are presented in this paper. Besides, several popular image fusion methods are taken as comparison with the proposed method.

A. *Dataset*

Three sets of image fusion from TNO dataset [22] have been chosen as a source image to test the proposed algorithm. All the images are set to the same sizes, 256 × 256 and already correctly registered so that there is no misalignment issue in this research that will affect the experimental results [23]. These images captured a human in outdoor environment and are widely used in thermal-visible image fusion related researches.

B. *Experimental Settings*

The experiments were performed using MATLAB R2014a with 2.30Ghz Intel® Core™ i3-2350M CPU and 8GB of main memory.

C. *Qualitative Evaluation of the Fusion Method*

For subjective assessment, all fused images are evaluated using human visual to analyze the quality of fused images by comparing with source images in terms of texture, spatial details, edge of object, etc. This evaluation is simple but it depends on the observer's experience and viewing conditions

[24].

D. Quantitative Evaluation of the Fusion Method

The effects of fused images are evaluated using several quality metrics that does not involve reference images because there are no ground truth images for the chosen dataset. The metrics for performance evaluation of fused image are as follows:

a. Information Entropy (IE)

IE measure the information richness in fused image. The value of *IE* represents the level of information that it contains. Based on Shannon's principle of information theory [25], *IE* 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 : Grey level of fused image
N_i : Pixel number of the grey value i
N : Total pixel number of the image

b. Structural Similarity Index (SSIM)

SSIM describes the similarity between source image and fused image, where it compare the local patterns of pixel intensities. The value is within the range -1 to 1 where the value of 1 shows that both images are similar.

$$SSIM(s,f) = \frac{(2\mu_s\mu_f C_1)(2\sigma_{sf} + C_2)}{(\mu_s{}^2 + \mu_f{}^2 + C_1)(\sigma_s{}^2 + \sigma_f{}^2 + C_2)} \quad (3)$$

where:

μ : Mean intensity of an image
σ : Standard deviation of an image
C : Constant
s : Source image
f : Fused image

c. Average Gradient (AG)

AG indicates the clarity and the details of the image based on the edge of the image. It also known as image sharpness. The larger value of AG, the richer information of the image. For the fused image of size, the AG is defined as:

$$AG = \frac{1}{(M-1)(N-1)} \cdot \sum_{x=1}^{M-1}\sum_{y=1}^{N-1}\frac{1}{4}\sqrt{\left(\frac{\partial F(x,y)}{\partial x}\right)^2 + \left(\frac{\partial F(x,y)}{\partial y}\right)^2} \quad (4)$$

where:

(x,y) : Coordinate of fused image, F
$\frac{\partial F}{\partial x}$: Horizontal gradient value
$\frac{\partial F}{\partial y}$: Vertical gradient value

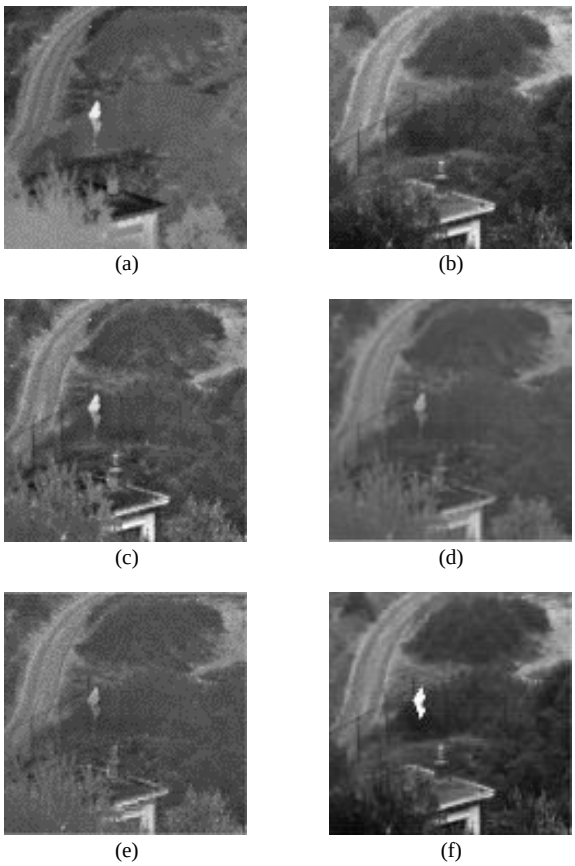
IV. Result and Discussions

The results of the fused images are divided into two parts; qualitative and quantitative evaluation. There are four methods have been chosen as a comparison with the proposed method. The methods are Laplacian Pyramid (LP), Weighted Averaging Method (AVG), DWT and proposed method based on SWT.

The first two images in Figures 1, 2 and 3 represent thermal and gray-scale visible light images and also known as input images. The thermal images detect the thermal radiation of an object in the scene, while the gray-scale visible light images provide details of the background information.

A. Assessment Results for United Nations Camp Images

Figure 3(c-f) represent image fusion using several methods. It shows that the intruder tried to climb the fence at the Camp of United Nations. However, the contrast level in Figure 3(d and e) are very low and the image are not very clear. In Figure 3(c), the image of an intruder is clear but the details of the scene are not obvious. Through overall comparison, image fusion using the proposed method as revealed in Figure 3(f) give the best visual effect with higher contrast and details are obvious. Furthermore, the quantitative results as shown in Table 1 also indicate that the proposed method provides the best fused image in terms of information richness, similarity images and image sharpness.



(a) (b) (c) (d) (e) (f)

Figure 3: Qualitative comparison for United Nations camp images (a) thermal image, (b) visible image, (c) LP, (d) AVG, (e) DWT, and (f) proposed method

Table 1
Quantitative Comparison for United Nations Camp Images

Method	Evaluation Metric		
	IE	AG	SSIM
LP	6.6498	7.3786	0.7787
AVG	6.2449	5.5154	0.5615
DWT	6.4024	7.7701	0.6726
Proposed Method	**7.2125**	**8.7560**	**0.9498**

B. Assessment Results for Tree Images

Figure 4 capture a person walking in the forest and there are a different between the image fusions in Figure 4 (c-f). The background image especially the texture of the trees in Figure 4(c) is not obvious and clarity while Figure 4 (d and e) are not clear at all. The details of fused images including the person are not clear. The overall information is the best in Figure 4(f). It has higher contrast and the edge profiles of the background are clearer. In addition, the quantitative comparison as in Table 2 shows the highest values of all the three quality metrics for the proposed method compare with other three methods.

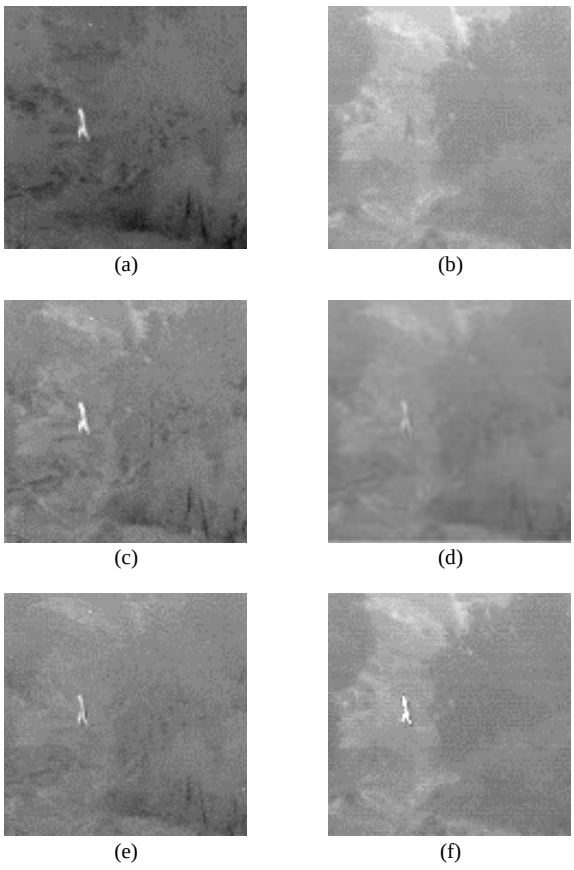


Figure 4: Qualitative Comparison for Tree images (a) thermal image, (b) visible image, (c) LP, (d) AVG, (e) DWT, and (f) proposed method

Table 2
Quantitative Comparison for Tree Images

Method	Evaluation Metric		
	IE	AG	SSIM
LP	6.0894	7.7503	0.7162
AVG	5.9267	4.4875	0.6408
DWT	6.4241	5.0550	0.6581
Proposed Method	**6.9308**	**7.9033**	**0.8630**

C. Assessment Results for Dune Images

Figure 5 illustrated a person walking on a sand dune. Although both input images as shown in Figure 5 (a and b) are vague, image enhancement using fusion method still provide the better image. In Figure 5(c), it preserve the characteristics of the thermal image but the background details are not clear. As shown in Figure 5 (d and e), both fused images are similar basically with lack of information. There are low contrast and the texture of the sand dune are not clear. Obviously, the proposed method as in Figure 5(f) provides the best image. The image shows higher contrast and more richness information compare with the other fused images as shown in Table 3 where the value of IE is 7.0244 and the AG value is 7.1507. Higher SSIM value indicates that the final fused image of the proposed method absorbs as much as possible the characteristics of the two input images.

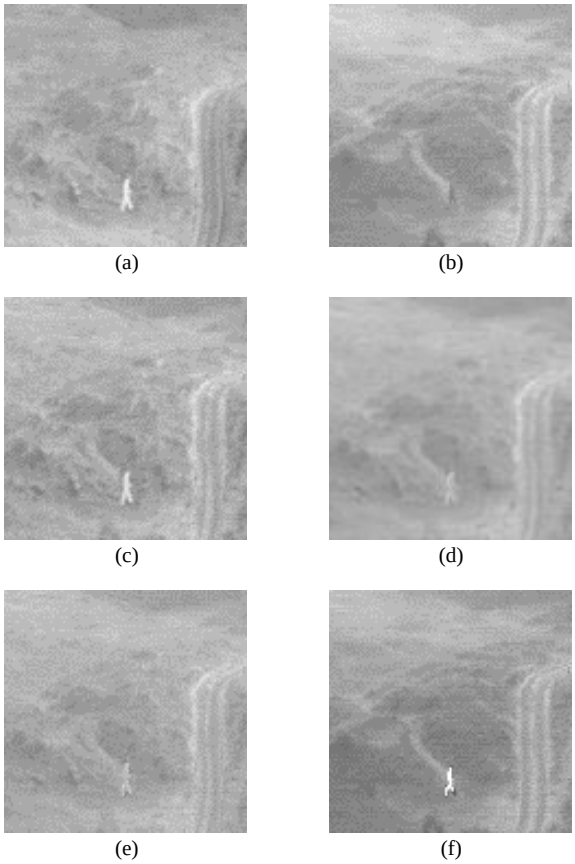


Figure 5: Qualitative Comparison for Dune images (a) thermal image, (b) visible image, (c) LP, (d) AVG, (e) DWT, and (f) proposed method

Table 3
Quantitative Comparison for Dune Images

Method	Evaluation Metric		
	IE	AG	SSIM
LP	6.3874	6.6566	0.8271
AVG	5.7542	3.9729	0.7717
DWT	5.3325	5.4370	0.7903
Proposed Method	**7.0244**	**7.1507**	**0.8633**

V. CONCLUSION

This paper proposed a hybrid thermal-visible fusion for image enhancement in an outdoor environment using SWT with thermal target extraction. The edge sharpness require improvements for the input thermal image before the target extraction due to the present of noise in the image. Then, Niblack algorithm and some morphological operators was chosen to extract the target region. SWT with appropriate

fusion rule is used to fuse both input thermal and visible images as well as the thermal target region. Experimental results prove that the proposed hybrid method provides the fused image with higher quality than other methods in terms of qualitative and quantitative evaluations.

REFERENCES

[1] M. Talha and R. Stolkin, "Particle filter tracking of camouflaged targets by adaptive fusion of thermal and visible spectra camera data," IEEE Sens. J., vol. 14, no. 1, pp. 159–166, 2014.

[2] Y. Jiang and M. Wang, "Image fusion with morphological component analysis," Inf. Fusion, vol. 18, no. 1, pp. 107–118, 2014.

[3] Z. Yan and L. Yan, "Some New Concepts and Key Techniques in Multi-Sensor Image Fusion," Sciencetimes.Com.Cn, pp. 1–19, 2010.

[4] D. Frejlichowski, K. Gosciewska, P. Forczmanski, and R. Hofman, "Application of foreground object patterns analysis for event detection in an innovative video surveillance system," Pattern Anal. Appl., vol. 18, no. 3, pp. 473–484, 2015.

[5] T. Sikandar and K. H. Ghazali, "A review on human motion detection techniques for ATM-CCTV surveillance system," Int. J. Comput. Commun. Instrum. Enggineering, vol. 3, no. 2, pp. 213–217, 2016.

[6] L. Snidaro, I. Visentini, and G. L. Foresti, "Fusing multiple video sensors for surveillance," ACM Trans. Multimed. Comput. Commun. Appl., vol. 8, no. 1, pp. 1–18, 2012.

[7] S. Budzan and R. W. Y. Ż. Golik, "Remarks on noise removal in infrared images," Measurement Automation Monitoring, vol. 61, no. 6, Poland, pp. 187–190, Jun-2015.

[8] W. K. K. and R. P. K. T.N. Dat, G.H. Hyung, "Person recognition system based on a combination of body images from visible light and thermal cameras," Sensors, vol. 17, no. 3, pp. 1–29, 2017.

[9] H. Li, L. Liu, W. Huang, and C. Yue, "An improved fusion algorithm for infrared and visible images based on multi-scale transform," Infrared Phys. Technol., vol. 74, pp. 28–37, 2016.

[10] S. Quan, W. Qian, J. Guo, and H. Zhao, "Visible and infrared image fusion based on curvelet transform," in 2nd International Conference on Systems and Informatics (ICSAI), 2014, pp. 828–832.

[11] Z. Yu, L. Yan, N. Han, and J. Liu, "image fusion algorithm based on contourlet transform and PCNN for detecting obstacles in forests," Cybern. Inf. Technol., vol. 15, no. 1, pp. 116–125, 2015.

[12] J. Liao, J. Huang, Z. Wang, H. Wang, L. Yang, B. Yi, and L. Ren, "Fusion of visible image and infrared image based on Contourlet transform and improved spatial frequence," in Proceedings of 2nd International Conference on Information Technology and Electronic Commerce, ICITEC 2014, 2014, pp. 322–325.

[13] X. Lu, Fei; Lei, "A novel fusion method of infrared and visible images based on non-subsampled Contourlet transform," in Image and Graphics, vol. 9217, Y.-J. Zhang, Ed. Tianjin: Springer International Publishing, 2015, pp. 297–306.

[14] L. Yucheng and L. Yubin, "A novel fusion algorithm for infrared image and visible light image based on non-subsampled Contourlet transform," Sci. Res. Essays, vol. 9, no. 9, pp. 374–379, 2014.

[15] J. Zhao, X. Gao, Y. Chen, H. Feng, and D. Wang, "Multi-window visual saliency extraction for fusion of visible and infrared images," Infrared Phys. Technol., vol. 76, pp. 295–302, 2016.

[16] G. Cui, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition," Opt. Commun., vol. 341, pp. 199–209, 2015.

[17] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," Inf. Fusion, vol. 31, pp. 100–109, 2016.

[18] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," Inf. Fusion, 2015.

[19] H. Qayyum, M. Majid, S. M. Anwar, and B. Khan, "Facial expression recognition using stationary wavelet transform features," Hindawi, vol. 2017, no. 1, pp. 1–9, 2017.

[20] U. Bhatt, A. Singh, H. S. Bhadauria, and M. Kumar, "Image super resolution based on discrete and stationary wavelet transform using canny edge extraction and non local mean," in International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1–5.

[21] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," in International Conference on Water Resources, Coastal and Ocean Engineering (ICWRCOE 2015), 2015, no. 4, pp. 133–142.

[22] A. Toet, J. K. IJspeert, A. M. Waxman, and M. Aguilar, "Fusion of visible and thermal imagery improves situational awareness," Displays, vol. 18, no. 2, pp. 85–95, 1997.

[23] J. Dou, Q. Qin, Z. Tu, X. Peng, and Y. Li, "Infrared and visible image registration based on SIFT and Sparse Representation," in Control and Decision Conference (CCDC), 2016, pp. 5420–5424.

[24] Y. Zhang, "Methods for image fusion quality assessment - A review, comparison and analysis," in The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2008, vol. XXXVII, pp. 1101–1109.

[25] S. Verdü, "Fifty years of Shannon theory," IEEE Trans. Inf. Theory, vol. 44, no. 6, pp. 2057–2078, 1998.G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.