A Novel Image Fusion Scheme using Wavelet Transform for Concealed Weapon Detection

Hanan A. Hosni Mahmoud

Department of Computer Science, College of Computer and Information Science
Princess Nourah bint Abdulrahman University, Riyadh, 11671 KSA
Dept. of Computer and Systems Engineering, Faculty of Engineering, University of Alexandria, Egypt

Abstract—The aim of this paper is to detect concealed weapons, especially in high security places like in airports, train stations and places with large crowds, where concealed weapons are not allowed. We aim to specify suspicious person who may have a concealed weapon. In this paper, an Image Fusion technique using pixel alignment and discrete wavelet transform is proposed. It is mainly utilized for Concealed Weapon Detection. Image fusion can be defined as extracting information from two or images into a single image to enhance the detection. Image fusion allows detecting concealed weapons underneath a person's clothing with imaging sensors such as Infrared imaging or Passive Millimeter Wave sensors. A data fusion scheme for simpler sensors based on correlation coefficients is proposed and utilized. We proposed an image fusion scheme that utilizes fusion dependency rules using wavelet (WT) and inverse wavelet transform (IWT). The fusion rule is to select the coefficient with the highest correlation rate. The higher the correlation the stronger of the co-existed feature. Experimental results shows the superiority of the proposed algorithm both in quality and real time requirement. The proposed algorithm has a real time response time that is less than other comparable algorithms by 40%. At the same time it retains higher quality as shown in the experimental results. It outperforms other algorithms by superior PSNR of more than 10% of the comparable algorithms in average.

Keywords—*Concealed weapon detection; image fusion; pixel alignment; wave sensors*

I. INTRODUCTION

Video surveillance systems acquire a video stream from the scene under monitoring from several sensors distributed across the area of interest. The analysis of the video stream begins with the detection of moving objects, and then recognition of the detected object is performed in order to classify it [1]-[3]. Then the object trajectory is identified in order to analyze the object's behavior or activities. Fig. 1 and Fig. 2 illustrate the processing flow in a visual surveillance system. The processing flow gives the first glance of the core challenges such as object identification. CCTV provides the visual sensors used for this type of systems; however, the CCTV cameras offer low resolution and low frame rate as well as varying quality due to environmental conditions such as changes in illumination. Tracking is another core challenge because of coordination required between different cameras [4]-[5].

A. Processing of Visual Surveillance

1) Object detection: Optical flow and background subtraction are usually utilized in object detection. Optical flow is computationally complex [6]. The main disadvantage of background subtraction technique is that a fixed background is required [7]-[8]. In outdoor environments, the high variability of environmental conditions requires robust adaptive background models that are computationally more expensive. In [9] the introduced a new technique to detect moving objects in MPEG videos utilizing modified optical flow and background subtraction algorithms.

2) Object classification: Object classification is a tough task in such low-resolution images. Surveillance footage has a quite poor resolution, resulting in an object of interest spanning in only a few pixels in each frame [10]-[11]. Two approaches have been used to solve this problem, histogram and model-based techniques. Color histogram of frames are calculated in the first approach. The model-based approach employs apriori geometrical knowledge of the objects of interest. The apriori knowledge can be constructed based on the appearance and features of the object. In [12] the authors used Convolutional Neural Network to analyze Low Resolution Thermal Images based on Embedded Platform.

3) Tracking of detected objects: Tracking introduces several challenges to the system, such as tracking in different lighting conditions. Shadow detection must be performed in order to avoid tracking a shadow instead of the real moving object. Another challenge is to track a person in a cluttered and dynamic background. The techniques employed for tracking can be classified in three groups: filtering techniques, statistical models, and multi-agents systems [13]-[14]. Filtering techniques such as Kalman filtering have been successfully employed in surveillance systems. Object tracking system utilizing Kernalized filters and Kalman predictive estimates was introduced in [15]. A Hidden Markov Model has been also used for tracking purposes; however, offline training data is required. Another technique used for tracking is Multi-agent systems, which is a well suited framework.
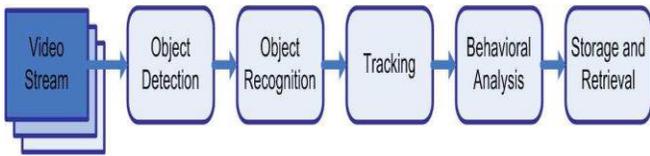
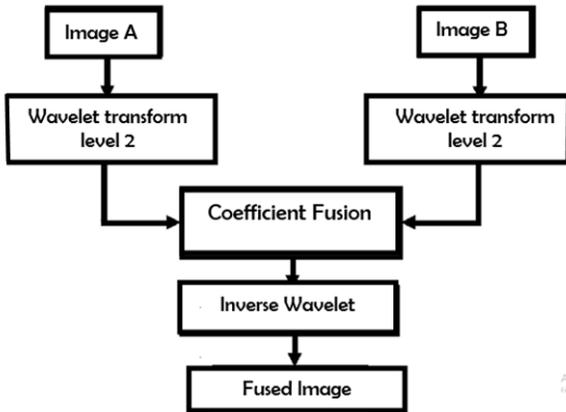Fig. 1. Classical Processing Flow in Visual Surveillance Systems.

Fig. 2. The Proposed Image Fusion Scheme FMWFR.

4) Behavioral analysis: Behavioral analysis aims to describe the activity that is taking place in the area under monitoring. This is a classification problem for the feature data provided by the previous stages during some period of time. In [16], they introduced predicting unrest in social events using hidden Markov models.

B. Image Fusion Schemes

The processing of multiple images into a single image is defined as image fusion. It is very important the image fusion is done without employing final image distortion or decreasing entropy [11]. Image fusion methods vary and based on many techniques. Wavelet transform includes various frequency sub-bands. In [14], the authors introduced Multi-sensor image fusion utilizing empirical wavelet transform, showing high performance through experimentation of images came from different sensors. Normalized convolution framework through a multi-frame super resolution of digitized videos is introduced in [16]. Panchromatic Fusion utilizing Empirical Wavelet technique was introduced in [7]. In [13], they implemented image fusion algorithm in the wavelet domain and built a high performance FPGA to implement the proposed algorithm. Also, the authors in [15] proposed an Image Fusion Implementation. In [15], the authors proposed several image fusion Techniques that were based on discrete wavelet transform. Approaches other than wavelet transform are the techniques proposed in [7]. They proposed a dictionary entropy image fusion framework that was based on deep learning. In [8] the classification process was based on wavelet image fusion was introduced. In [9], texture features were used to perform image fusion.

II. Methodology

We propose an image fusion scheme that fuses images at the pixel level using a wavelet transform of the source images.

Image decomposition by wavelet transform at different levels is shown in Fig. 3. The fusion rule is described by the dependency of the wavelet coefficient. The dependency metric will identify the stronger pattern.

A. The Proposed Image Fusion Scheme with Salient Feature Extraction FMWFR

We employed linear dependency metric to be performed on the pixels of the two block of dimension n × n. If the two blocks are unrelated then those pixels are linearly independent, and no relevant feature in the window. According to this metric, if the two blocks are related, then the pixels are linearly dependent, and pattern exists. We utilized Wronskian determinant formula to assess this metric. The problem is that background information of any two images are highly correlated. Therefore, in this paper we accommodate a simple preprocessing step, where background information in an image is detected from the frames of videos containing this image. Then background information is converted to all black pixels. For image B the background will be converted to all white pixels. Thus, no similarities of salient feature will be determined based on background similarities.

The LL band of the wavelet transform (WT) preserves the related features of the original image. LH, HL, HH bands hold information about the prominent feature of the original image. Fusion rule diagram is depicted in Fig. 4. The Wronskian determinant of the prominent features on both images deploys the dependency as follows:

$$E_{MWFR(A,B)}(m,n) = \sum_{y=m-1}^{m+1} \sum_{x=n-1}^{m+1} \left[\frac{[WT_A(y,x)]}{[WT_B(y,x)]}\right]^2 - \left[\frac{[WT_A(y,x)]}{[WT_B(y,x)]}\right] \tag{1}$$

Where, WTA(y,x) is defined as the WT coefficient of image A with the position (y,x) in the 2D image.

Linear dependency is proven in equation 1. The similarity of the features in images A and B is detected by an increased value of EAB. EAB equals zero in the case of no features coexist in both mages A and B. The fusion algorithm determines the existence of the feature in image A which are more significant that the feature exists in B as depicted below:

$$F_{MWFR(A,B)}(m,n) = \begin{cases} WT_A(y,x) \ if \ E_{MWFR(A,B)}(m,n) \geq F_{MWFR(B,A)}(m,n) \\ WT_B(y,x) \ if \ E_{MWFR(A,B)}(m,n) < F_{MWFR(B,A)}(m,n) \end{cases} \tag{2}$$
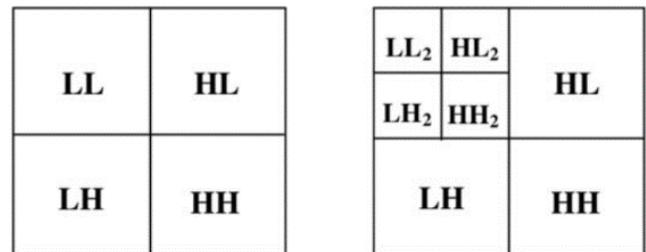

Fig. 3. Image Decomposition by Wavelet Transform at different Levels.
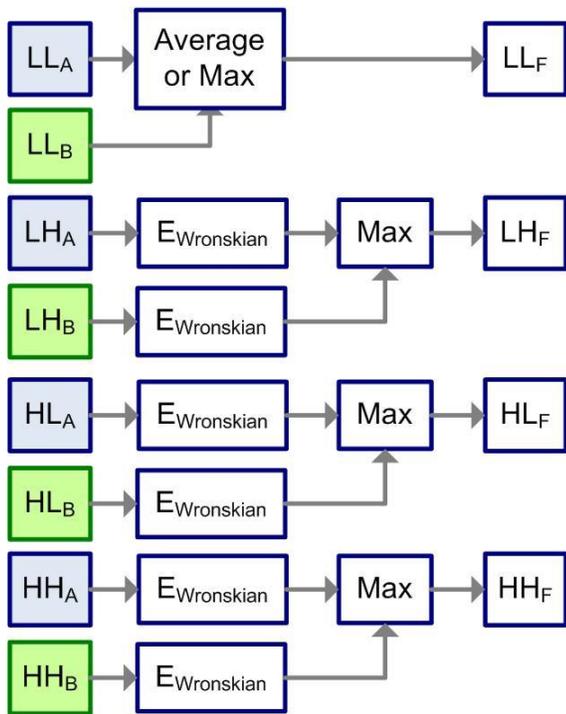
Fig. 4. Fusion Rule Diagram.

The main steps of the proposed algorithm of fusion of several bands of two source images is depicted in Fig. 5, and Fig. 6. The algorithms are illustrated as follows:

Algorithm 1. Fusion (Input: Image A; Image B, Output: Fused Image F)

Start

Preprocessing Phase:

a. Register the source images: source image *A and* source image *B.* from videos V_A and V_B
b. Detect background information in image *A* from the frames of videos containing this image.
c. Convert background information to all black pixels.
d. Detect background information in image *B* from the frames of videos containing this image.
e. Convert background information to all white pixels.

The Fusion Phase FMWFR

1. Resample images A and B in such a way that the pixels of the A and B are aligned.
2. Perform the discrete wavelet transform of A and B.
3. Calculate $W_A(y,x)$ (The WT coefficient of the original image *A* at position *y,x* in the spatial fixation).
4. Calculate $W_B(y,x)$ (The WT coefficient of the original image *B* at position *y,x* in the spatial fixation).
5. Calculate the Wronskian determinant of A and B using equation 1 as follows:

$$E_{MWFR_{AB}}(i,j) = \sum_{r=i-1}^{i+1}\sum_{c=j-1}^{j+1}\left[\frac{W_A(r,c)}{W_B(r,c)}\right]^2 - \frac{W_A(r,c)}{W_B(r,c)}$$

The algorithm performs the fusion decision as follows:

If $(E_{AB} = 0)$
Then
The feature *f* does not exist in B
Elseif (E_{AB} is greater value)
Then
{There is a feature *f* exists in in A and B
Call Fusion rule algorithm}

Fusion rule algorithm

1. $$F_{MWFR}(i,j) = \begin{cases} W_A(i,j) & if \quad E_{MWFR_{AB}}(i,j) \geq E_{MWFR_{BA}}(i,j) \\ W_B(i,j) & if \quad E_{MWFR_{AB}}(i,j) < E_{MWFR_{BA}}(i,j) \end{cases}$$

2. Join WT of A and B spatially to get *AB*.
3. Apply inverse WT on *AB*.
4. final image is fused from A and B.

End

Algorithm 2. Fusion rule algorithm (Input: wavelet transform of image A; wavelet transform of image B, Output : LL_F, LH, HL, HH of the fused image F).

Start

1. *Compute LL_F= Average(LL_A, LL_B);*
2. *Compute LH_F= Max($E_{Wronskian}(LH_A)$, $E_{Wronskian}(LH_B)$);*
3. *Compute HL_F= Max($E_{Wronskian}(HL_A)$, $E_{Wronskian}(HL_B)$);*
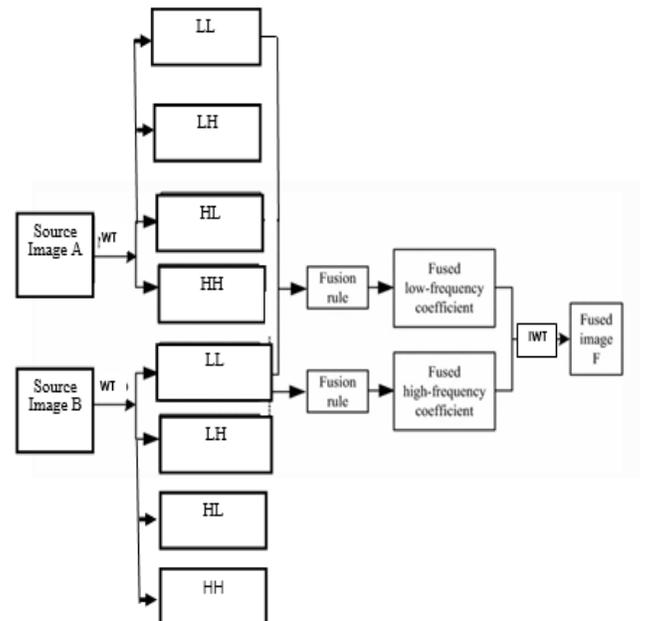4. *Compute HH_F= Max($E_{Wronskian}(HH_A)$, $E_{Wronskian}(HH_B)$);*

End

Fig. 5. Fusion of Several Bands of Two Source Images.
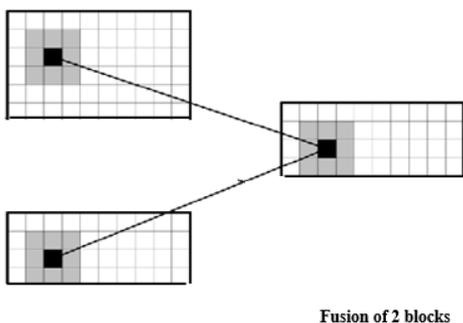
Fusion of 2 blocks

Fig. 6.   Fusion of Two Blocks from Two Source Images.

III.   EXPERIMENTAL RESULTS

Simulation using test images that are extracted from [10] with permission are shown in Fig. 7. The results with the previous schemes as well as the proposed schemes are shown in Fig. 8 and Fig. 9 for images a and e respectively, the rest of the results are omitted due to space limit. The test images contained images with different exposition or focusing on different objects. Fig. 7 contains images coming from different sensors including Infrared, visible, MMW, and MRI.   The methods tested are a combination of the rules explained above and are described in Table I.

Image Fusion at real time is very crucial to detect threats such as concealed weapon. Therefore, the evaluation criteria of any fusion algorithm should include fusion time. Image fusion

scheme retains as much information as possible from the sources while introducing as few inconsistencies as possible. Therefore, another evaluation criteria should be done in terms of information, quality, and spectral efficiency. Qualitative measures by showing before and after images are shown in Fig. 7, 8 and 9. The assessment of the quality of the algorithm are performed with the help of reference image. Peak Signal to Noise Ratio (PSNR) are usually employed to compare a distorted image with a distortion-free image. PSNR establishes spectral differences within the fused image compared to reference image. The spatial criterion determined the spatial details in terms of maximizing the correlation between them. We compared all the algorithms in table.

In respect of fusion time and spectral quality through PSNR, the comparison is shown in Fig. 10 and 11. The results for Mutual Information show that WFR-MS: WFR using MS for the LL bands achieves the best PSNR, while the WFR-Ave achieves comparable PSNR as WFR-MS but wilt better fusion time as shown in Fig. 11. The spectral quality of fused images can be measured by several metrics such as Spectral angle mapper (SAM), Relative average spectral error (RASE). The Cross correlation (CC). In Table II, we are comparing the algorithms in Table I for images (a) and (b) in the first row of Fig. 9. Table III shows the same metrics calculating the averages of SAM, RASE and CC for all images in the first rows of Fig. 8 and 9. From the two Tables II and III, it shows the superiority of our proposed techniques WFR-MSWR, WFR-MS and WFR-Ave.

TABLE. I.   THE METHODS TESTED WHICH ARE COMBINATION OF DIFFERENT RULES

Method	Description
DMR: The Decision Map Rule	DMR consists of Energy feature extraction Energy for all bands except LL with a maximum criterion. The LL band is fused using the average rule [15].
FBR: The Feature Based Rule	FBR employs the expected value of features for extraction of the feature and the selection of the coefficient is based on the maximum value of the feature. The LL band is fused by averaging the coefficients using the average rule [16].
Ave : The Average Rule	Ave does not extract any feature and the fusion rule consists of averaging the wavelet coefficients of the two source images [12].
MSWR: The Maximum Square Window Rule	MSWR uses the extraction rule and the fusion rule F_{Max} [13].
WFR-MSWR: The proposed algorithm case # 1	WFR- MSWR uses the Wronskian Fusion Rule combined with the proposed feature extraction FMWFR for all the bands except LL. The LL band was fused using maximum square window rule.
WFR-MS: The proposed algorithm case # 2	The Wronskian fusion rule uses the proposed feature extraction FMWFR for all the bands except LL. The LL band was fused by the maximum square rule.
WFR-Ave: The proposed algorithm case # 3	The Wronskian fusion rule uses the proposed feature extraction FMWFR for all the bands except LL. The LL band was fused by averaging.

TABLE. II.   SPECTRAL METRICS COMPARISON FOR INDIVIDUAL IMAGES

	Spectral Metrics					
	Image a			*Image b*		
	RASE	*SAM*	*CC*	*RASE*	*SAM*	*CC*
DMR: The Decision Map Rule	58.36	0.176	0.87	56.48	0.158	0.91
FBR: The Feature Based Rule	55.38	0.211	0.78	56.01	0.252	0.81
Ave : The Average Rule	52.85	0.243	0.78	53.05	0.201	0.73
MSWR: The Maximum Square Window Rule	64.54	0.066	0.95	65.34	0.076	0.89
WFR-MSWR: The proposed algorithm case # 1	72.58	0.058	0.98	73. 88	0.058	0.97
WFR-MS: The proposed algorithm case # 2	72.57	0.250	0.97	72.46	0.276	0.95
WFR-Ave: The proposed algorithm case # 3	72.69	0.118	0.92	72.91	0.126	0.93

TABLE. III.   AVERAGE SPECTRAL METRICS

| Spectral Metrics			
	Average RASE	*Average SAM*	*Average CC*
DMR: The Decision Map Rule	44.547	0.159	0.87
FBR: The Feature Based Rule	43.352	0.282	0.78
Ave : The Average Rule	43.801	0.337	0.78
MSWR: The Maximum Square Window Rule	61.388	0.200	0.95
WFR-MSWR: The proposed algorithm case # 1	65.593	0.250	0.98
WFR-MS: The proposed algorithm case # 2	63.891	0.351	0.97
WFR-Ave: The proposed algorithm case # 3	64.520	0.206	0.92
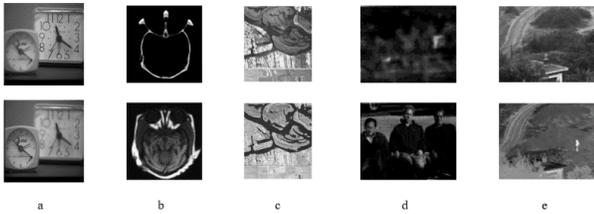
Fig. 7.   Test Images from different Sensor Types or with different Exposition or Focus [10].
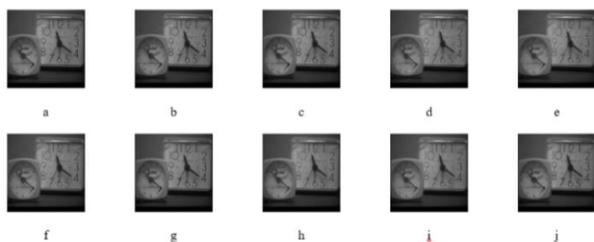

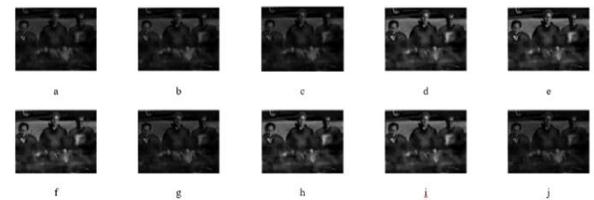Fig. 8.   Images Obtained after Fusing Two Images Focused in different Objects.

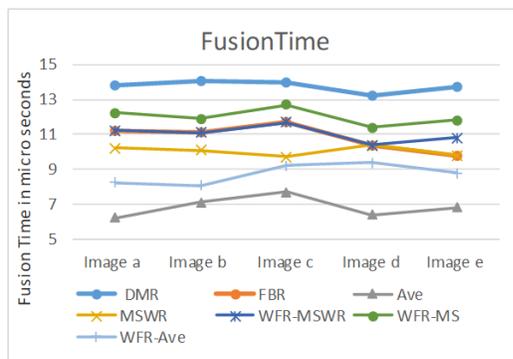Fig. 9.   Images Obtained after Fusing a MMW Image and a Visible Image.

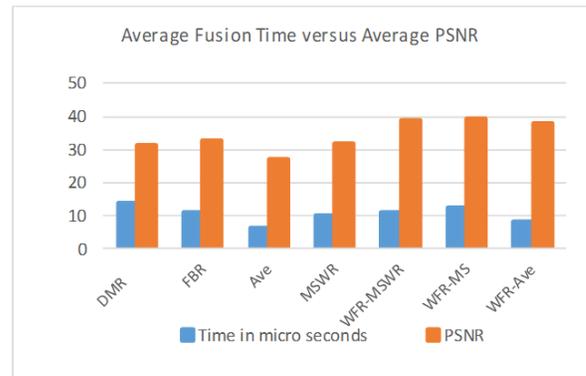Fig. 10. Fusion Time for Images Obtained after Fusing a MMW Image and a Visible Image.

Fig. 11. Average Fusion Time Versus Average PSNR.

IV. CONCLUSION

Security systems emphasize on safety measures, one of them is the detection of concealed weapons. In the proposed architecture, the fusion of infrared images and visual information offers concealed weapon detection for secure-sensitive areas. Some of the results were shown earlier. In addition, the use of infrared technology for detection of concealed weapon raises privacy issues. In order to protect the privacy of the people crossing the scene under monitoring, a conditional image fusion can be done. First, a search for a concealed weapon should be applied to the infrared image alone. If the scene contains a suspected weapon, then fusion should be performed to identify the person carrying it. In the proposed technique, we employed linear dependency metric to be performed on the pixels of the two block of dimension $n \times n$. If the two blocks are unrelated then those pixels are linearly independent, and no relevant feature in the window. According to this metric, if the two blocks are related, then the pixels are linearly dependent, and pattern exists. We utilized Wronskian determinant formula to assess this metric. Spatial and spectral metrics are applied to test images comparing the proposed algorithm with other well-known algorithms in the literature, all metrics show the superiority of our algorithm in its different cases. Also, Image Fusion at real time is very crucial to detect threats such as concealed weapon. Therefore, evaluation criteria of any fusion algorithm should include fusion time. We presented experimental results showing the superiority of the proposed algorithm w.r.t fusion time.

R<small>EFERENCES</small>

[1] Wei Liu, Wei Chen, "Recent Advancements in Empirical Wavelet Transform and Its Applications," Access IEEE, vol. 7, pp. 103770-103780, 2019.

[2] K J A S Sundar, V. Vaithiyanathan, "Design and Analysis of Fusion Algorithm for MultiFrame Super-Resolution Image Reconstruction using Framelet," Defence Science Journal, vol. 65, no. 4, pp. 292-299, 2015.

[3] K. J. A. Sundar, M. Jahnavi and K. Lakshmisaritha, "Multi-sensor image fusion based on empirical wavelet transform," 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, 2017, pp. 93-97.

[4] X. Lu and C. Xu, "Novel Gaussian mixture model background subtraction method for detecting moving objects," 2018 IEEE International Conference of Safety Produce Informatization (IICSPI), Chongqing, China, 2018, pp. 6-10.

[5] Y. Zhang, "Generalized Wronskian Solutions for a Non-isospectral MKdV Equation," 2010 International Conference on Computing, Control and Industrial Engineering, Wuhan, 2010, pp. 65-67.

[6] Y. Wu, C. Chang and Y. Tao, "Closed-Circuit Television-Enabled Service: A Review of Security and Privacy Issues," 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops, Japan, 2013, pp. 306-309.

[7] Y. -L. Wu, Y. -H Tao, . C-J. Chang, and C. -W. Chang, Dilemma of CCTV adoption between security and privacy: The Taiwanese context, International Conference on Business and Information Conference, Sapporo, Japan, 2012.

[8] F. Conche, and M. Tight, Use of CCTV to determine road accident factors in urban areas, Accident Analysis and Prevention, vol. 38, 2006 pp. 1197-1207.

[9] S. Manchanda and S. Sharma, "Identifying moving objects in a video using modified background subtraction and optical flow method,"2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 129-133.

[10] Xu, Tuzhi, "Multisensor Concealed Weapon Detection Using the Image Fusion Approach" (2016). Electronic Theses and Dissertations.

[11] Gianmarco Cerutti, Bojan Milosevic, Elisabetta Farella, "Outdoor people detection in low resolution thermal images", 2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech), pp. 1-6, 2018.

[12] G. Cerutti, R. Prasad and E. Farella, "Convolutional Neural Network on Embedded Platform for People Presence Detection in Low Resolution Thermal Images," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 7610-7614.

[13] M. Kivanc Mihcak, I. Kozintsev and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258), Phoenix, AZ, USA, 1999, pp. 3253-3256 vol.6.

[14] D. Casasent, R. Patnaik, "Analysis of kernel distortion invariant filters", Proceedings of SPIE, vol. 6764, no. 2007, pp. 1.

[15] J. Luo and D. Chen, "Mapping Rules Based Data Mining for Effective Decision Support Application," 2008 International Seminar on Business and Information Management, Wuhan, 2008, pp. 506-509.

[16] A. Rao and K. Shah, "An optimized rule based approach to extract relevant features for sentiment mining," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 2330-2336.