



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 2, February 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Review on Object Recognition and Tracking using Machine Learning

Meghna Kanaujia¹, Prof. Suresh S. Gawande²

M. Tech. Scholar, Department of Electronics and Communication, RKDF College of Engineering, Bhabha University, Bhopal, India¹

Guide, Department of Electronics and Communication, RKDF College of Engineering, Bhabha University, Bhopal, India²

ABSTRACT: The growth in amount of video content has led to development of number of applications like video summarization, extraction of useful content from videos, development of automatic video surveillance systems etc. One common use analysis for such applications is detection and tracking of moving objects from videos. This study is reliable hand gesture recognition system using the machine learning technique. The first step, carried out pre-processing, to separate the foreground and background. Then the foreground is transformed using the discrete wavelet transform (DWT) to take the most significant subband. The wavelet transform is a well-known signal analysis method in several engineering disciplines. In image processing and pattern recognition, the wavelet transform is used in many applications for image coding as well as feature extraction purposes. It can be used to describe a given object shape by wavelet descriptors (WD). The last step is image classification with machine learning approach.

KEYWORDS: - Discrete Wavelet Transform, Machine Learning, Object Recognition

I. INTRODUCTION

Many computer vision applications require object detection and tracking in video sequences as a vital step for target detection, target monitoring and activity analysis of target in videos. Since video based systems are relatively inexpensive and can capture a huge amount of desired information, there is an immediate need for automated video understanding methods in which object detection and tracking can be done without involvement of human operators. This growth has led to great demand on real time video processing. Object detection and tracking objects are among the most challenging task to accomplish for determining meaningful events and suspicious activities from video sequence. The general framework for object detection and tracking in videos is shown in Figure 1.1. The framework is initialized by feeding video imagery. The process of object detection begins with object segmentation followed by extraction of object from video sequence. Detection of object is followed by labeling and tracking of that object. Labeling is important to differentiate and track an object over series of frames. Preprocessing operations can be applied before or after an object is detected. Preprocessing before object extraction will target to enhance quality of video sequences by suppressing of noise or redundant information to enable more accurate and precise detections. It is different from other perspective where it is targeted to enhance boundaries or shapes of detected object with help of morphological operations.

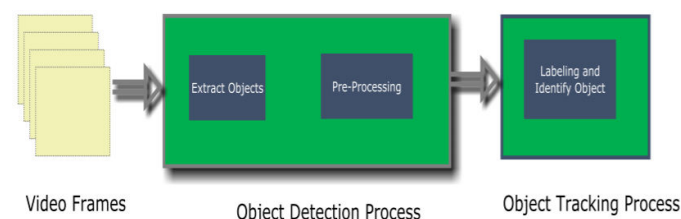


Figure 1: Framework of object detection and tracking in videos

Object Detection

The interpretation and analysis of video streams is an active research field and one of the important tasks in this field is object detection. Object detection is the process of locating the occurrence of object using number of techniques like background subtraction, feature extraction, statistical methods etc. Figure 1.2 shows some example of object detection in videos. Colored boxes on the figures are used to highlight the detected object on to frame. Applications like video event detection need to detect and track objects firstly, followed by recognizing what is going on around those tracked objects.



Figure 2: Examples of object detection in videos

II. LITERATURE REVIEW

Jaehoon Lee et al. [1], have propose three-dimensional (3D) visualization of objects in heavy scattering media by using peplography and wavelet transform. Conventional haze removal techniques can remove the light haze in the image by using various image processing algorithms or machine learning techniques. However, they may not provide a clear image under heavy scattering media. On the other hand, peplography can visualize the object by detecting ballistic object photons from heavy scattering media. Then, 3D image can be generated by integral imaging. However, it may not visualize 3D object information accurately because of the noise photons from scattering media. Therefore, the image quality for 3D object visualization may be degraded. To solve this problem, we use the discrete wavelet transform in peplography. It can detect the object photon signals from the scattering media and enhance 3D image contrast ratio by using a specific coefficient threshold technique. To prove our method, we carry out optical experiments and compare results with the conventional haze removal method and peplography by using various image quality metrics such as correlation, structural similarity, and peak signal-to-noise ratio.

J. Lee et al. [2], present a computational volumetric reconstruction method for three-dimensional (3D) photon counting imaging with enhanced visual quality when low-resolution elemental images are used under photon-starved conditions. In conventional photon counting imaging with low-resolution elemental images, it may be difficult to estimate the 3D scene correctly because of a lack of scene information. In addition, the reconstructed 3D images may be blurred because volumetric computational reconstruction has an averaging effect. In contrast, with our method, the pixels of the elemental image rearrangement technique and a Bayesian approach are used as the reconstruction and estimation methods, respectively. Therefore, our method can enhance the visual quality and estimation accuracy of the reconstructed 3D images because it does not have an averaging effect and uses prior information about the 3D scene. To validate our technique, we performed optical experiments and demonstrated the reconstruction results.

J. Lee et al. [3], have propose three-dimensional (3D) photon counting double random phase encryption (DRPE) with enhanced visual quality and security level. Conventional 3D photon counting DRPE can quickly encrypt the data by using the 4f optical system and random phase masks with enhanced security because the 3D photon counting technique is used. 3D photon counting DRPE extracts photons from the encrypted data by using statistical methods such as the Poisson random process, and the visual quality of the decrypted data can be enhanced through the 3D reconstruction process. However, it still has a problem to visualize the data when we extract extremely a few photons. To improve the security and the visual quality of the decrypted data, we propose the random amplitude reconstruction process in the

encryption stage. Our proposed method reconstructs the amplitude of the encrypted data at a random depth. Thus, the shifting pixel value and depth information can be another important key for decryption through the random process. Therefore, it can effectively decrypt data securely, and the visual quality of the decrypted data can be enhanced. Finally, through the random reconstruction process in the decryption stage, our proposed method can simultaneously enhance the security and visual quality. To verify our proposed method, we carry out the simulation and optical experiment.

J. Lee et al. [4], propose an effective noise reduction method for photon counting imaging using a discrete wavelet transform. Conventional 2D photon counting imaging was used to visualize the object under dark conditions using statistical methods, such as the Poisson random process. The photons in the scene were estimated using a statistical method. However, photons which disturb the visualization and decrease the image quality may occur in the background where there is no object. Although median filters are used to reduce the noise, the noise in the scene remains. To remove the noise effectively, our proposed method uses the discrete wavelet transform, which removes the noise in the scene using a specific thresholding method that utilizes photon counting imaging characteristics. We conducted an optical experiment to demonstrate the denoising performance of the proposed method.

M. Hupfel et al. [5], fluorescence microscopy images are inevitably contaminated by background intensity contributions. Fluorescence from out-of-focus planes and scattered light are important sources of slowly varying, low spatial frequency background, whereas background varying from pixel to pixel (high frequency noise) is introduced by the detection system. Here we present a powerful, easy-to-use software, wavelet-based background and noise subtraction (WBNS), which effectively removes both of these components. To assess its performance, we apply WBNS to synthetic images and compare the results quantitatively with the ground truth and with images processed by other background removal algorithms. We further evaluate WBNS on real images taken with a light-sheet microscope and a super-resolution stimulated emission depletion microscope. For both cases, we compare the WBNS algorithm with hardware-based background removal techniques and present a quantitative assessment of the results. WBNS shows an excellent performance in all these applications and significantly enhances the visual appearance of fluorescence images. Moreover, it may serve as a pre-processing step for further quantitative analysis.

Sunit Vaidya et al. [6], our work provides a solution to people who are suffering from partial blindness due to diseases of the eye or due to accidents. Unlike people who are born blind, they depend a lot on other people for doing their day to day activities. Our project is a boon to such people as our end product; a smart walking stick detects the trained objects that are used daily by the person and intimates the person using audio messages. This project changes the visual world into an audio world by informing the visually impaired of the objects in their environment. These people can use this particular prototype for self -navigating their way. A YOLO (You look only once) algorithm is used in our project. Real-time objects in an image are detected with their names represented on a bounding box and these names are converted to speech signals. The conversion to audio signals is done by using an e-Speak tool which forms Google's Text to Speech (gTTS) system. The prototype consists of several modules. The Raspberry Pi camera module takes the image and transfers it to the Raspberry Pi desktop. Then, real-time object detection is carried out by using YOLO network. The detected image is converted to speech by using the gTTS module and the audio result is provided to the user through a headset. For achieving the required portability, the battery backup is being used.

Ganbayar Batchuluun et al. [7], recently, thermal cameras are being widely used in various fields, such as intelligent surveillance, biometrics, and health monitoring. However, the high cost of the thermal cameras poses a challenge in terms of purchase. Additionally, thermal images have an issue pertaining to blurring caused by object movement, camera movement, and camera focus settings. There have been very few studies on image restoration centered around thermal images to address such problems. Moreover, it is important to increase the processing speed of image restoration methods to jointly conduct with methods such as action recognition and object tracking that use temporal information from thermal videos. However, no study has been conducted on simultaneously performing super-resolution reconstruction and deblurring using thermal images. Furthermore, existing studies on object detection using thermal images have errors owing to the incapability in distinguishing reflections on the surrounding ground or wall due to the heat radiated from the object. To address such issues, this study proposes a deep learning-based thermal image restoration method that simultaneously performs super-resolution reconstruction and deblurring. According to recent development of deep learning, generative adversarial network (GAN)-based methods which have ability to preserve texture details in images, and yield sharper and more plausible textures than classical feed forward encoders show success in image-to-image translation tasks. Considering the advantages of GAN, we propose a deblur-SRRGAN for thermal image reconstruction. In addition, we propose a light-weighted Mask R-CNN for object detection in the

reconstructed thermal image. For the input, we employ an image processing method that converts 1-channel thermal images (often used in the existing studies) into 3-channel images.

G. Batchuluun et al. [8], a high-resolution thermal camera is very expensive and is thus difficult to be used. Furthermore, thermal images become blurred in various cases of object motion, camera shaking, and camera defocusing. To solve these problems, a previous super-resolution restoration (SRR) technique converting a thermal image acquired by a low-resolution camera into a high-resolution one, and a thermal image deblurring method have been researched. However, existing studies were performed based on 1-channel (grayscale) images. In addition, a large-sized and whole image has been used in the existing thermal image deblurring methods, which causes lower deblurring performance. In this study, we propose novel SRR and deblurring methods. The proposed deblurring method is conducted based on small region images. The proposed methods are also conducted using 3-channel (color) thermal images and generative adversarial networks. In addition, the performances of this method are compared in various color spaces (RGB, Gray, HLS, HSV, Lab, Luv, XYZ, YCrCb), image sizes, and thermal databases. Through experiments using self-collected databases and open databases, it was confirmed that the proposed methods show better performance than the state-of-the-art methods.

II. VIDEO RECONSTRUCTION

Essentially, the true aim of salient object detection is to find objects that are distinctive from the image background. It needs to calculate the contrast between the objects and the image background and then select those with high contrast as salient objects. However, the local and global contrast based methods do not identify which regions form the image background. They blindly assume the neighboring regions or the entire image to be the background and then calculate the contrast between each location and the assumed background. As their assumed background may not be the real one, the determined contrast also becomes incorrect, which in turn reduces the performance of saliency detection. To overcome these problems, a few emerging methods [7], [8] using background priors were proposed based on the idea of modeling the property of background first and thereby separating salient objects from the background. Specially, Wei et al. [9] exploited the boundary and connectivity priors about the background in natural images and detected saliency based on the geodesic distance. Considering that the salient object may be partially cropped on the boundary, this paper adopts an existing saliency detection method [10] to compute the saliency of boundary patches and generates weights for the virtual background nodes. However, in some challenging images, where [11] could not calculate the saliency of boundary patches precisely, the method of [12] is difficult to obtain satisfactory results. Yang et al. [13] modeled saliency detection as a manifold ranking problem and proposed a two-stage scheme for graph labeling. They represent the image as a close-loop graph with superpixels as nodes. In saliency detection, they first use the nodes on the image boundary as background seeds to rank other nodes in the graph. Then, in the second stage, they select the salient nodes from the detection results of the first stage and use them to refine the saliency of other nodes in the graph. On the assumption that the image boundary is mostly background, these methods result in a background template. As a result, the contrast between salient object and background can be precisely obtained. By incorporating background priors into traditional contrast-based methods, they show improved results in saliency detection. However, existing background prior-based methods still have certain limitations.

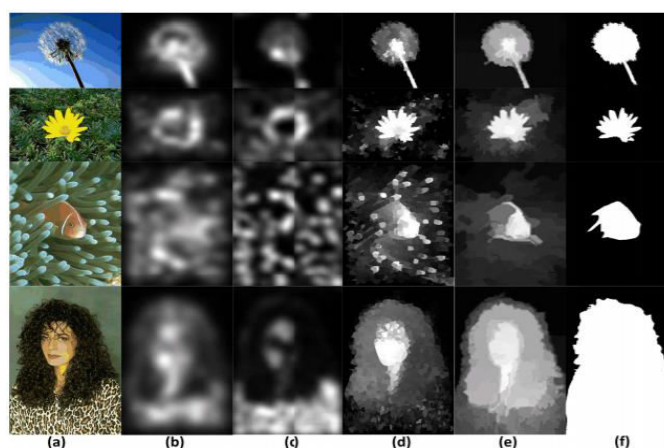


Fig. 3: Some examples of saliency detection for Video Collection

In this paper, the SDAE is used to model image background. Rather than adopting hand-designed features as used in previous works [13], [14], the deep-structured SDAE is employed to learn more powerful representation directly from the raw image data in an unsupervised way, which also enables capturing the latent pattern of the input data hierarchically. It thus helps to deal with the second scenario (shown in the second row of Fig. 3) where the background regions share latent patterns. Then, the measure of contrast between the salient objects and the background is formulated as the reconstruction residuals in the deep-structured SDAE. Unlike the previous works, which mainly focused on the way to calculate the similarity or distinctiveness between a certain image patch and the image boundary, this paper pays more attention to modeling the background regions.

III. METHODOLOGY

Machine Learning is a subset of Artificial Intelligence concerned with “teaching” computers how to act without being explicitly programmed for every possible scenario. The central concept in Machine Learning is developing algorithms that can self-learn by training on a massive number of inputs. Machine learning algorithms are used in various applications, such as email filtering and computer vision, where it is difficult or infeasible to develop conventional algorithms to perform the needed tasks [4]. Machine learning enables the analysis of vast amounts of information. While it usually delivers faster, more precise results to identify profitable prospects or dangerous risks, it may also require additional time and assets to train it appropriately. Merging machine learning with AI and perceptive technologies can make it even more effective in processing vast volumes of information. Machine learning is closely associated with computational statistics, which focuses on making predictions using computers. Machine learning approaches are conventionally divided into three broad categories, namely Supervised Learning, Unsupervised Learning & Semi-supervised Learning, depending on the nature of the “signal” or “feedback” available to the learning system.

Face anti-spoofing (FAS) has lately attracted increasing attention due to its vital role in securing face recognition systems from presentation attacks (PAs). As more and more realistic PAs with novel types spring up, traditional FAS methods based on handcrafted features become unreliable due to their limited representation capacity. With the emergence of large-scale academic datasets in the recent decade, machine learning based FAS achieve remarkable performance and dominate this area.

Supervised Learning

A model is trained through a process of learning in which predictions must be made and corrected if those predictions are wrong. The training process continues until a desired degree of accuracy is reached on the training data. Input data is called training data and has a known spam / not-spam label or result at one time.

Unsupervised Learning

By deducting the structures present in the input data, a model is prepared. This may be for general rules to be extracted. It may be through a mathematical process that redundancy can be systematically reduced, or similar data can be organized. There is no labeling of input data, and there is no known result.

Semi-Supervised Learning

Semi-supervised learning fell between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data). There is a desired problem of prediction, but the model needs to learn the structures and make predictions to organize the data. Input data is a combination of instances that are marked and unlabeled.

4.1 Discrete Wavelet Transform

The concept of discrete wavelet transform is introduced here (Graps, 1995). Through the discrete wavelet transform analysis, the whole input value should be analyzed and further given to the process. The input data should be in the form of time domain. The DWT is used to change the input data from a time domain into a frequency domain. The frequency domains have many features and by the use of frequency domain it can easily identify the audio signal. From the above steps details of DWT is known and now the continuous wavelet transform is studied. Here, the discrete wavelet transform is transmitted by a continuous wavelet transform and so a discussion on the continuous wavelet transform and its features has to be made. The continuous wavelet transform is a shifted version of the wavelet function ψ , and the main function of CWT is used to divide a continuous wavelet functions into a small wavelet.

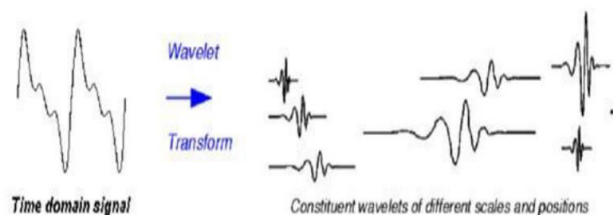


Fig. 4: Waveform represented of DWT

The model used in [5] to implement the tree structure of Direct Wavelet Packet Transform (DWPT) is based on the filtering process. Figure 1 depicted a complete 3-level Direct WPT. In this figure G and H is the high pass and low pass filter respectively.

Computation period is the wide variety of the input cycles for one time produces output samples. In trendy, the computation duration is $M=$ for a j -level DWPT. The duration of the 3-stage computation is eight. discern 1, The Sub band Coding set of rules for example, assume that the authentic sign $X[n]$ has N - pattern points, spanning a frequency band of 0 to π rad/s. At the first decomposition level, the sign exceeded through the high pass and low bypass filters, followed through subsampling via 2. The output of the high pass filter out has $N/2$ - sample factors (as a result half of the time resolution) however it handiest spans the frequencies $\pi/2$ to π rad/s (consequently double the frequency decision).

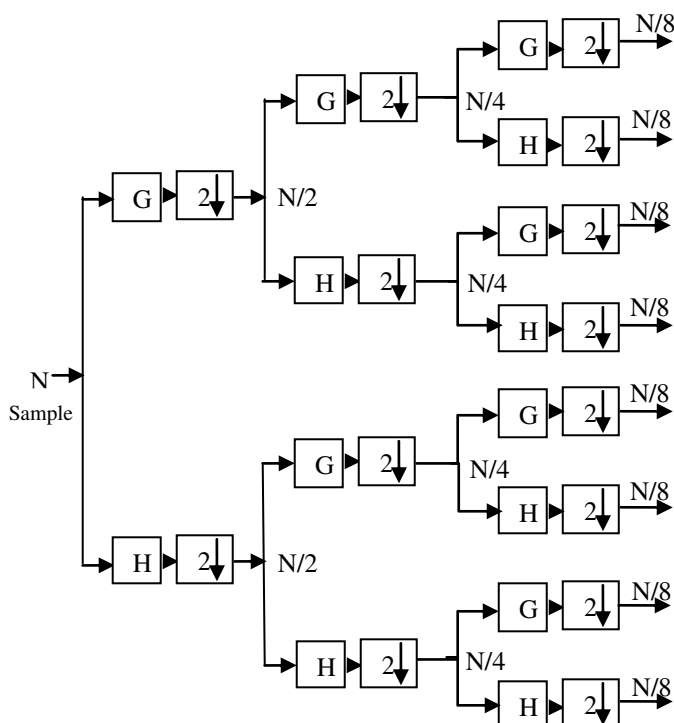


Fig. 5: 3- Levels for DWPT. Where G, H are the high-pass and low-pass filter coefficient.

The output of the low-bypass filer additionally has $N/2$ - sample points, but it spans the opposite half of the frequency band, frequencies from 0 to $\pi/2$ rad/s. again low and excessive-pass filter output passed via the identical low skip and excessive pass filters for similarly decomposition.

The output of the second one low skip clear out followed by way of sub sampling has $N/4$ samples spanning a frequency band of zero to $\pi/4$ rad/s, and the output of the second high bypass filter out observed by using sub sampling has $N/4$ samples spanning a frequency band of $\pi/4$ to $\pi/2$ rad/s. the second one excessive bypass filtered sign constitutes the second stage of DWPT coefficients. This sign has half of the time resolution, but two times the

frequency decision of the first level sign. This process continues till two samples are left. For this specific instance there would be three ranges of decomposition, each having half of the range of samples of the preceding degree. From the above steps details of DWT is known and now the continuous wavelet transform is studied. Here, the discrete wavelet transform is transmitted by a continuous wavelet transform and so a discussion on the continuous wavelet transform and its features has to be made. The continuous wavelet transform is a shifted version of the wavelet function ψ , and the main function of CWT is used to divide a continuous wavelet functions into a small wavelet.

The DWT of the original signal is then obtained by concatenating all coefficients starting from the last level of decomposition (remaining two samples, in this case). The DWT will then have the same number of coefficients as the original signal.

IV. CONCLUSION

Object detection and tracking is one of the challenging research tasks of computer vision aimed at detecting moving objects from video sequence. It is followed by predicting the path of moving object for the duration of its presence in video frame sequences. This study has provided a comprehensive review of the state-of-the-art methods on object detection and tracking with focus on soft computing based approaches. The number of WD needed to recognize a given object increases according to the complexity of the object shapes and must be set according to the given application. It is possible in some cases to use only the components of the approximation signal in order to recognize an unknown object using the minimum distance method, but generally the use of the detail signal will include detail information about small contour changes between the compared objects. The starting point on the contour has a large influence on the recognition process because the values of the WD depend strongly on it.

REFERENCES

- [1] Jaehoon Lee, Myungjin Cho and Min-Chul Lee, "3D Visualization of Objects in Heavy Scattering Media by using Wavelet Peplography", IEEE Access 2022.
- [2] J. Lee, M.-C. Lee, and M. Cho, "Three-dimensional photon counting imaging with enhanced visual quality," *J. Inf. Commun. Conver. Eng.*, vol. 19, no. 3, pp. 180_187, 2021.
- [3] J. Lee, M. Cho, and M.-C. Lee, "Three-dimensional photon counting optical encryption with enhanced visual quality and security level," *IEEE Access*, vol. 9, pp. 128862_128869, 2021.
- [4] J. Lee, M. Kurosaki, M. Cho, and M.-C. Lee, "Noise reduction for photon counting imaging using discrete wavelet transform," *J. Inf. Commun. Conver. Eng.*, vol. 19, no. 4, pp. 276_283, 2021.
- [5] M. Hupfel, A. Y. Kobitski, W. Zhang, and G. U. Nienhaus, "Wavelet- based background and noise subtraction for _uorescence microscopy images," *Biomed. Opt. Exp.*, vol. 12, no. 2, pp. 969_980, 2021.
- [6] Sunit Vaidya, Naisha Shah, Niti Shah and Prof. Radha Shankarmani, "Real-Time Object Detection for Visually Challenged People", International Conference on Intelligent Computing and Control Systems (ICICCS 2020).
- [7] Ganbayar Batchuluun, Jin Kyu Kang, Dat Tien Nguyen, Tuyen Danh Pham, Muhammad Arsalan And Kang Ryoung Park, "Deep Learning-based Thermal Image Reconstruction and Object Detection", IEEE Access, 2020.
- [8] G. Batchuluun, Y. W. Lee, D. T. Nguyen, T. D. Pham, and K. R. Park, "Thermal Image Reconstruction Using Deep Learning", IEEE Access, vol. 8, pp. 126839-126858, 2020.
- [9] G. Batchuluun, H. S. Yoon, D. T. Nguyen, T. D. Pham, and K. R. Park, "A Study on the Elimination of Thermal Reflections," IEEE Access, vol. 7, pp. 174597–174611, 2019.
- [10] D. Bianli, R. Xiaokang, and R. Jie, "CNN-based Image Super-Resolution and Deblurring," in Proc. Association for Computing Machinery, Wuhan, China, 29–31 Oct. 2019, pp. 70–74.
- [11] X. Zhang, F. Wang, H. Dong, and Y. Guo, "A Deep Encoder-Decoder Networks for Joint Deblurring and Super-Resolution," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, Canada, 15–20 Apr. 2018, pp. 1448–1452.
- [12] F. Albluwi, V.A. Krylov, and R. Dahyot, "Image Deblurring and Super-Resolution Using Deep Convolutional Networks," in Proc. IEEE 28th International Workshop on Machine Learning for Signal Processing, Aalborg, Denmark, 17–20 Sept. 2018, pp.1–6.
- [13] B. Oswald-Tranta, "Temperature Reconstruction of Infrared Images with Motion Deblurring," *J. Sens. Sens. Syst.*, vol. 7, pp. 13–20, 2018.
- [14] H. Choi and J. Jeong, "Despeckling images using a preprocessing _filter and discrete wavelet transform-based noise reduction techniques," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3131_3139, Apr. 2018.



- [15] Y. Matsushita, H. Kawasaki, S. Ono, and K. Ikeuchi, "Simultaneous Deblur and Super-Resolution Technique for Video Sequence Captured by Hand-Held Video Camera," in Proc. IEEE International Conference on Image Processing, Paris, France, 27–30 Oct. 2014, pp. 4562–4566.
- [16] H. Linyang, L. Gang, and L. Jinghong, "Joint Motion Deblurring and Superresolution from Single Blurry Image," Math. Probl. Eng., vol. 2015, pp. 1–10, 2015.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details