Enhanced Denoising Algorithm for Improving Quality of Objects in Frames for Video Surveillance

P.Vijayakumar[1], A.V.Senthilkumar[2]

[1]*Head, Department of Computer Applications, Sri Jayendra Saraswathy Maha Vidyalaya College of Arts and Science, Coimbatore, Tamilnadu, India.*
[2]*Director of MCA, Hindusthan College of Arts and Science, Coimbatore, Tamilnadu, India.*

Abstract: In the 21st century's security awareness scenario, video surveillance plays an important role in improving security to people and property. Fully automatic surveillance systems are much sought after by both public and government organizations. These systems perform surveillance in four steps, namely, object detection, tracking, classification and interpretation. After object detection, more often, the detected objects have noise and insignificant extra objects; the presence of which effects the performance of the subsequent steps like tracking and classification. In order to solve this issue, this paper presents an algorithm that performs noise detection, noise removal with a post processing procedure using enhanced morphological operations to remove extra objects. This algorithm considers three types of noise, namely, Gaussian, Salt & Pepper and Speckle, which are detected using a trained support vector machine classifier. After detection, the noises are removed using appropriate noise removal methods. Experimental results show that the efficiency of the proposed enhancement algorithm in terms of noise detection accuracy and enhancement quality is high and can be used to improve the operations of automatic video surveillance systems.

Keywords: Denoising, Morphological Operations, Noise Detection, Object Enhancement, Video Surveillance System

I.   Introduction

Increasing crime rate and terrorist acts worldwide have increased the demand for sophisticated video surveillance systems that can be used to detect abnormal behaviours and thus improve security to people and property. Video surveillance systems are currently undergoing a transition from traditional analog solutions to digital solutions. These systems are increasingly becoming a part of daily life and are used in airports, banks, offices, hospitals, traffic points and even at residential apartments. It also provides high level security like border monitoring and transport environment.

The usage of video surveillance system in India increased only after 2009. Market research report conducted to analyze the state of video surveillance revealed Asia and within Asia, India, as one of the biggest market for video surveillance in the world [1]. It has been estimated that the usage of video surveillance has grown to 24.5% in 2010 to massed revenue of $135 million. IP (Internet Protocol) video surveillance system alone has contributed to around 28% of the overall value of surveillance industry [2]. The reason of such a tremendous growth can be attributed to the volatile security environment of the country and according to [3], "increasing terrorist activities and attacks have created strong demand for advanced safety and security solutions. As a result, the CCTV (Closed Circuit Television) market in India is anticipated to grow at a CAGR (Compound Annual Growing Rate) of more than 30% between 2010 and 2013".

Apart from industry reports, according to [4], increased number of CCTV usage has been envisaged in kinder gardens, schools, supermarkets and retail stores have been noted along with banking and transport sectors, prompted mostly after the incidence of 9/11 terrorism. Delhi International Airport is reported to have the single largest IP video surveillance system in Asia having more than 3700 IP surveillance cameras with monitoring screen having the capacity of displaying up to 25 multiple camera images. The state of the CCTV industry and video surveillance equipment is changing rapidly as new technologies force security professionals worldwide into the domain of Information Technologies (IT) and vice versa.

This wide usage has directly motivated several researchers to design and develop sophisticated automated video surveillance systems that can mimic the activity of human brain in identifying events in moving objects [5]. However, due to the advanced technological breakthroughs in video capturing devices, increase in the number of cameras used, dynamic nature of the objects (human, vehicles, animals) in the video, has made this process a very challenging field of research. The main motivation behind the system proposed in this paper is the increasing demand for high and efficient surveillance systems that avoids continuous manual monitoring and cumbersome process of searching historical recordings for identifying past events.

A typical Automatic Video Surveillance (AVS) System consists of four steps, namely, object detection, enhancement, tracking, classification and event understanding [6]. The three subtasks, object detection, tracking

and classification, converts the raw video image sequences into intelligent knowledge that can be used for event understanding [7]. The performance of video surveillance systems are generally by three factors, namely, noise, illumination and lighting variations and shadows, that degrade the quality of objects in a video scene. These issues arise due to imperfect acquisition, transmission errors, recorded time (day or night), lighting and sensitivity of the recording device. The presence of noise, illumination and lighting variations and shadows has direct impact on the functioning of object tracking and classification. This necessitates the need to implement enhancement techniques in video surveillance systems to work with a common aim of making the system robust against changes in illumination, noise and shadow. This paper focuses on the design and development of enhancement techniques through the use of noise algorithm, so as to increase the efficiency of the subsequent steps of AVS. The rest of the paper is organized as follows: Section 2 presents a brief literature review of the current status of noise removal algorithms. Section 3 presents the image enhancement procedure while Section 4 presents the results of performance evaluation. Section 5 concludes the work with future research directions.

II. Literature Study

Video surveillance systems rely on the ability to detect moving objects in the video sequence which is a relevant information mining step in a wide range of computer vision applications. Many research works have been conducted for video object detection and tracking. Image and video quality enhancement is a longstanding area of the research. Image and video signals are often contaminated by noise during acquisition and transmission and noise is a dominant factor that degrades image/video quality. Low-end camera market is growing rapidly (digital cameras, web-cams, cell phones etc.) and there is a need now more than ever for fast, effective and reliable image and video enhancement technologies to improve their output. Even high-end and professional equipment (surveillance cameras, medical devices) have to cope with image degradation and noise corruption (especially in extreme conditions). Nowadays, practically every image-capturing device incorporates some sort of noise removal technology. Video denoising is highly desirable, not only for improving perceptual quality, but also for increasing compression speed and effectiveness and facilitating transmission bandwidth reduction. In [8], ELT Method was introduced for multiple moving object segmentation in video surveillance. In [9], proposed a novel Framework for Specific Object Tracking in the Multiple Moving Objects Environment. In [10], MELT method was introduced for moving object segmentation in video sequences. In [11], IELT technique initiates the process of video object segmentation, object tracking and finally object detection. In [12], TFVODT framework initiates and improves the video Object detection accuracy.

In [13], Bayesian Particle Filter-based Median Enhanced Laplacian Thresholding (BPF-MELT) is introduced. The Bayes Sequential Estimation is constructed based on the posterior and prior function. To represent a comprehensive representation, Bayes Estimation algorithm discriminately trains the high quality video files with aiming at reducing the mean square tracking error. The Color Histogram-based Particle Filter measures the color histogram value using the likelihood function. The Histogram-based Particle Filter algorithm in BPF-MELT framework improves the object detection accuracy. In [14], Improved Threshold Filtered Video Object Detection and Tracking (ITFVODT) framework is designed for efficient video object segmentation based on their features like shape, texture, intensity and tracking of moving objects. The main goal of EMFVD method is to segment the video object based their features like size, color, texture, shape, intensity and to efficiently detect the moving object in video sequence using threshold filtering technique [15].

According to the definition provided by [16], video surveillance is used to monitor the behavior, activities and other changing information, usually of objects like people or vehicle, for the purpose of offering protection in geographical space. Almost all of the image denoising techniques can be categorized into spatial domain and transform domain technique. Spatial filters employ low pass filtering techniques on group of pixels and works on the assumption that noise occupies the high frequency domain. They remove noise efficiently but at the cost of blurring and removal of edges. Spatial filters can further be classified into non-linear and linear filters. Examples of non-linear filters include median [17], rank conditioned rank selection [18] and relaxed median [19]. Examples of linear filters include wiener filtering and mean filter.

Apart from these, transform domain based denoising techniques also exist. These can be either data adaptive or non-adaptive techniques. Under non-adaptive transform methods, spatial frequency filtering and wavelet based methods are normally used. The transform domain filtering methods can be subdivided according to the choice of the basis function. The basis function can be classified as data adaptive and non-adaptive. Spatial-frequency filtering refers to the use of low pass filters using Fast Fourier Transform (FFT) and removes noise by using a frequency domain filter and adapting a cut-off frequency when the noise components are decorrelated from the useful signal in the frequency domain [20]. These methods are time consuming and depend on the cut-off frequency and the filter function behavior. Furthermore, they may produce artificial frequencies in the processed image.

Filtering operations in the wavelet domain can further be subdivided again into linear and nonlinear methods. Linear filters such as Wiener filter in the wavelet domain yield optimal results when the signal

corruption can be modeled as a Gaussian process and the accuracy criterion is the Mean Square Error (MSE) [21]. Under non-linear wavelet filtering approaches, threshold based filtering is more prominent [22]. Use of anisotropic diffusion for denoising image has also received a great deal of attention because of its impressive performance in preserving edge sharpness and suppressing noise. They work well with both Gaussian and Poisson noise.

According to [23], noise in an image can be detected at a high noise level. Using this information [24] introduced a two-phase method for removing noise using Adaptive Center-Weighted Median Filter (ACWMF) and variation method. This method successfully removed the noise, but was sensitive to the size and shape of the filter window. Bigger size window resulted in over smoothening of the image and small size did not remove the noise efficiently. The window size problem is a common problem that is shared by many filtering algorithms. Solutions to these weaknesses in noise removal algorithms were provided by [25]. They proposed a method based on Interval-Valued Fuzzy Sets (IVFS) entropy application to denoise an image. The method combined image histogram information and spatial information about pixels of different gray levels by using an IVFS multi-thresholding technique. The advantage of this method was that it was able to remove both impulse and Guassian noise. Extended this method to remove impulse and Guassian noise present in the same image by using Iteratively Reweighted Norm (IRN) method to yield predictions of the original pixel values and compute the corresponding predicted errors and train the noise model by using an Expectation-Maximization (EM) algorithm. The proposed algorithm can effectively remove the impulse noise with a wide range of noise density and produce better results in terms of the qualitative and quantitative measures of the images.

III. Noise Removal Algorithm

The main aim of denoising algorithm is to remove object noise while retaining important signals required for tracking and classification processes to perform efficiently. It is considered a challenging task as they are irreversible and therefore has to be performed carefully and should make sure that important object regions are not erroneously removed as noise. Noise in video surveillance systems can occur due to various reasons, like bad digitization or due to small movements made by scene background such as moving trees or camera displacement due to wind load. The noise removal technique proposed in this paper consists of two main steps.

Step 1 : Detect type of noise degrading the frame and use suitable algorithm to remove the noise
Step 2 : Post Processing using morphological operations to obtain a smooth contrast adjusted image

3.1. Noise Identification and Removal

The section presents the procedure used to detect the type of noise in the input diatom image. For this purpose, a Support Vector Machine (SVM) is used. To train the network, images with Gaussian noise, impulse noise (Salt & Pepper) and speckle noise was created artificially. Some examples are given in Figure 1. Fifty images for each type of noise were created, from which four statistical features, namely, mean, standard deviation, skewness and kurotsis were extracted. The extracted features were used to train SVM using k-fold cross validation method.

| Original | Gaussian Noise | Salt & Pepper Noise | Speckle Noise |
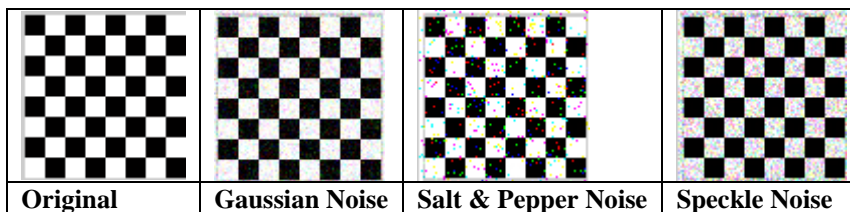
Figure 1. Noise Images Created Artificially

During denoising, the same statistical features are extracted from the input image to be denoised and the trained network is used to identify the selected type of noise. After identifying the noise, appropriate noise removal algorithm that best suits the noise type is used to denoise the object image. In this work, Gaussian noise is removed using Weiner Filter, Salt & Pepper noise is removed using weighted median filter, while speckle noise is removed using SRAD (Speckle Reduced Anisotropic Diffusion) algorithm.

3.2. Post Processing

A technique that has been proved to be efficient in these scenarios is the application of morphological operators (erosion, dilation, opening and closing) on the binary image (foreground). However, usage of morphological operators has issues like (i) Delete small regions that may be important during analysis and (ii) It introduces shape and contour distortion. To solve the above issues, a Post Processing (PP) procedure that enhances the operation of conventional Morphological Filter is used. The main aim of the enhancement is to

measure a pixel's belongingness to a structured window region around it and extract clean signals using a local threshold, using multi-scale analysis. This algorithm begins by applying morphological open and close operations on the object image enhanced using the noise removal algorithm from Step 1. One drawback of applying morphological operations directions is that it removes regions which are smaller that the Structuring Element (SE). If these removed objects belong to foreground object, then it adversely affects the performance of tracking, the next step in AVS system. In this work, a structure analysis is included to avoid this situation. This analysis identifies foreground removed objects, which are then combined with the enhanced version of the object. The process is explained in detail below.

Let the original degraded object image be O and its denoised version be E. Application of morphological open and close operations results with a smoothened and contrast adjusted image, E'. The opening operation relates to an erosion of the O using a structure element (SE) and then dilating the result using the same structure element SE. In general, opening has a smoothing effect that eliminates small (narrow) protrusions. The closing operation corresponds to a dilation of the source image followed by erosion using the same structure element. In order to avoid removing foreground small objects, a difference image, D, consisting of all the left out regions is found, that is, D = W − M and a structure analysis is performed to identify regions that belong to foreground region.

- **Structure Analysis**

For each pixel in D a structure analysis is performed to make a decision on whether the pixel belongs to foreground or background. The structuring element used for this purpose is shown in Figure 2a, where ● indicates the pixel to be analyzed. Let this pixel be termed as p_{ij}. In this method, a suitability analysis is performed on each pixel to evaluate whether they belong to background or foreground. All pixels belonging to background are removed. Suitability of p_{ij} is determined with information obtained from its 8 neighboring pixels. This is done by performing logical AND between p_{ij} and its eight neighbours (Figure 2b).

(a)   **(b)**
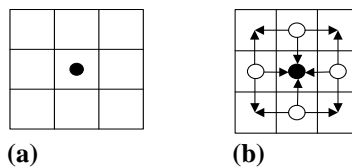
Figure 2. Structuring Element and its Neighbourhood Pixels

The basic structure of p_{ij} is looked in each of the 8 directions to make the decision of belongingness to foreground or background. The algorithm consists of three steps as listed below.

Step 1 : Let the base structure be Figure 2a

Step 2 : Perform a logical "AND" between the pixels pointed by the circle and each one of its neighborhoods (Figure 2b)

Step 3 : Increase and replace the value of the central pixel (the black pixel in Figure) by the outcome of all these "AND" operations.

Let the outcome of this procedure by R. The final enhanced foreground image with all significant objects is obtained as F' = E' + R.

IV. Experimental Results

The proposed denoising algorithm was tested with a video dataset having 20 videos collected from VIRAT and YouTube. VIRAT (Video and Image Retrieval and Analysis Tool) dataset, which is a part of video surveillance project funded by the Information Processing Technology Office (IPTO) of the Defense Advanced Research Projects Agency (DARPA). The performance of the proposed denoising algorithm is analyzed using Peak Signal to Noise Ratio (PSNR) metric. Apart from this, analysis of noise detection algorithm is shown using confusion matrix of the SVM classifier. Test frames used during discussion is given in Figure 3. Similar results were obtained for all frames in selected videos. For clarity of discussion the results presented in this paper are shown on the degraded images. During experiments Frames A and B were degraded with Gaussian noise, while C and D were respectively degraded with Salt & Pepper and Speckle noise. The degraded frames are shown in Figure 4. The confusion matrix related to the three selected noises is shown in Figure 5.
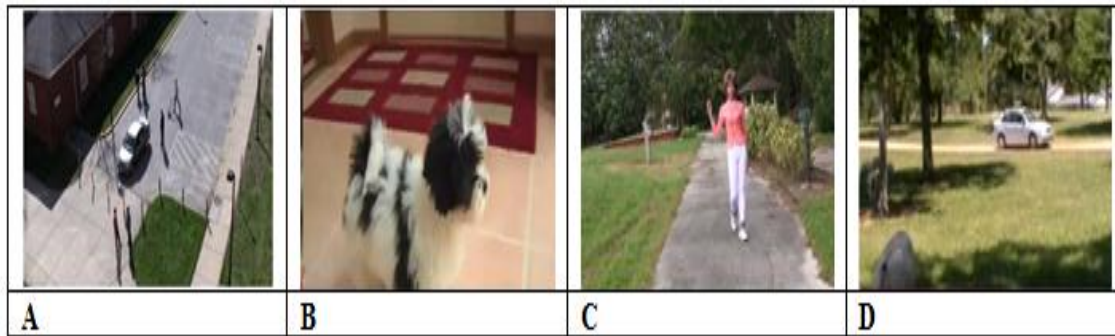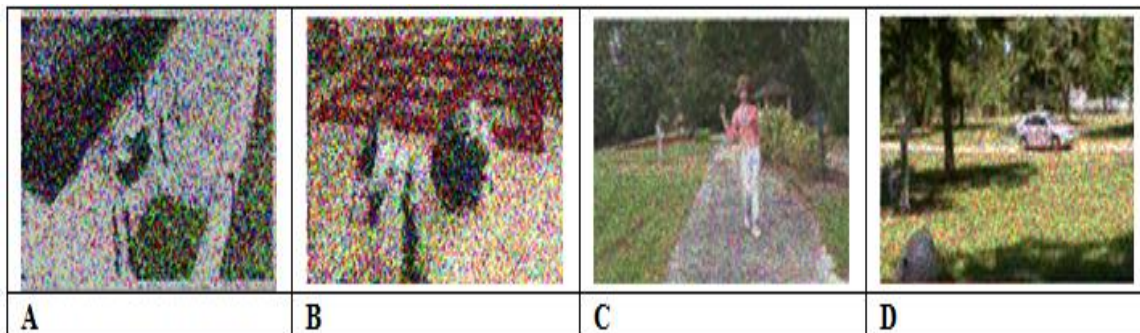
Figure 3. Test Frames Used in Experiments

Figure 4. Degraded Frames Used in Experiments

	Gaussian	Speckle	Salt &Pepper
Gaussian	98.22	1.1	0
Speckle	0	97.76	2.2
Salt &Pepper	2.6	1.5	95.44

Figure 5. Confusion Matrix of Noise Classification

In the above confusion matrix, the diagonal entries show percentage of correct classification for each class while off-diagonal entries depict mis-classification. The results show that a maximum accuracy of 98.22% was found for Gaussian noise followed by an accuracy of 97.76% and 95.44% for those with Speckle and salt & pepper noises respectively. The results thus prove that the noise detecting procedure can correctly identify the type of noise. After the correct identification of noise, the Weiner Filter (W), SRAD and Weighted Median Filters (M) are applied to remove the Gaussian, Speckle and Salt & Pepper noise respectively. The efficiency of these algorithms in terms of PSNR and speed of noise removal is presented below. As most of the existing algorithm try to enhance the objects detected only using open and close morphological operators, the results are compared with conventional morphological based denoising algorithm (CM). To analyze the effect of applying the post processing procedure which enhances the conventional CM through the use of structure analysis, the denoising algorithm with (WPP) and without (WOPP) the use of post processing algorithm is also presented. The PSNR results are presented in Table 1.

Table 1. PSNR (dB)

Frames	CM	CM with PP	W		SRAD		M	
			WOPP	WPP	WOPP	WPP	WOPP	WPP
A	26.74	30.26	36.26	38.68	37.59	37.95	32.65	34.00
B	25.96	29.42	35.87	37.58	35.91	36.74	31.76	33.50
C	28.97	31.88	37.18	39.22	36.55	38.62	33.94	35.61
D	27.13	30.77	37.01	37.95	37.80	39.46	33.14	34.80

From the table, it is clear that the proposed method of identifying noise first and then removing them using appropriate filter is successful and works better than the conventional morphological method. According to [31], a PSNR value between 30 and 40 indicates that the enhanced image is a very good match to the original image (before noise is added). In relation to this fact, it can be concluded that the proposed denoising models produce good quality enhanced images.
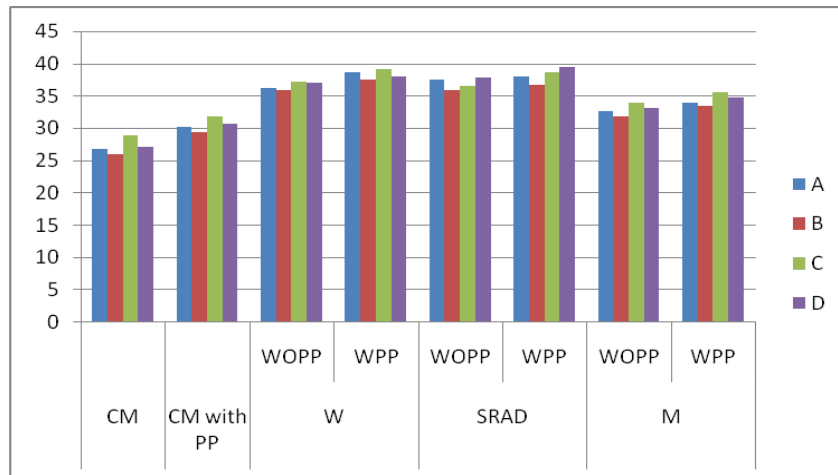
Figure 6. The visual quality of the input frame objects.

From the Figure 6, the visual quality of the input frame objects the proposed framework is improves when compared to existing methods. Further, the advantage of the inclusion of post processing procedure is also clear from the results. While comparing CM and enhanced CM, on average, the enhanced CM showed an efficiency improvement of 11.06%. Similarly, the W, SRAD and M algorithms with post processing procedure on average showed an efficiency gain of 4.6%, 3.22% and 4.7% respectively. This clearly shows that the enhanced morphological operation is advantageous and improves the visual quality of the input frame objects.

The visual comparison of the denoising algorithms is shown in Figure 7, which confirms the results obtained in Table 1. Thus, the various experimental results show that the proposed noise detection and denoising model enhances the input object frame in a positive manner and removes Gaussian, Salt & Pepper and Speckle noise successfully. The edge and image detail preserving capacity is high when compared to the traditional Weiner filter.
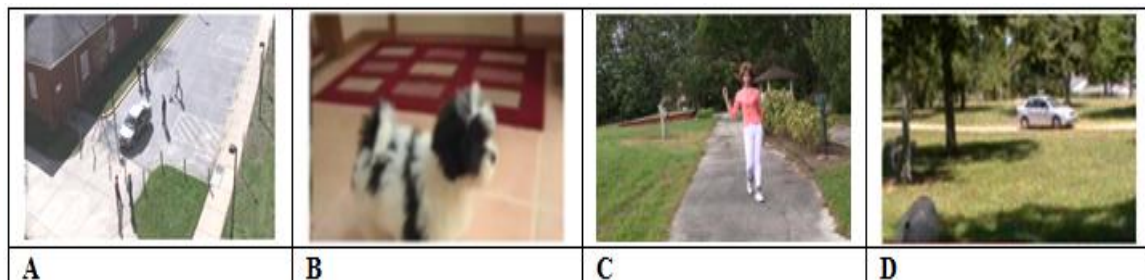

Figure 7. Visual Results of Test Frames

V. Conclusion

The state-of-the-art cameras and interest in the study of Automatic video surveillance has increased due to the increased public awareness on security. This paper presented a noise removal algorithm that can be used to remove noise, adjust contrast and remove unwanted or insignificant objects in detected objects in video frames. Three types of noises were considered, namely, Gaussian, Salt & Pepper and Speckle noise. A support vector machine based classifier was trained to detect the selected noises and were used to identify the type of noise a frame image is degraded with. After identification of the type of noise, Weiner filter, weighted median filter and SRAD filters were respectively used to remove the Gaussian, Salt & Pepper and Speckle noise. A post processing procedure based on enhanced morphological operations was included to further improve the quality of image and produce a contrast adjusted image. Experimental results showed that the proposed denoising model is effective in reducing noise in frame and enhancing the image. In future, the effect of the proposed denoising model on object tracking and classification is to be studied.

References
[1].   https://www.privacyinternational.org/reports/india/iii-surveillance-policies, Last access on February, 2014.
[2].   Jhingan,   H.   (2011)   Under   surveillance,   Express   Computer,   http://www.expresscomputer online.com/20110630/securitystrategies01.shtml, Last Access on February, 2014.
[3].   Security Concerns to Drive CCTV Demand in India (2010), http://www.rncos.com/Press_Releases/Security-Concerns-to-Drive-CCTV-Demand-in-India.htm, Last Access on February, 2014.

[4]. Norris, C., McCahill, M. and Wood, D. (2004), Editorial. The Growth of CCTV: a global perspective on the international diffusion of video surveillance in publicly accessible space, Surveillance and Society, CCTV Special (eds. Norris, McCahill and Wood), Vol. 2, No. 2/3, Pp. 110-135.

[5]. Lu, Y., Boukharouba, K., Boonært, J., Fleury, A. and Lecœuche, S. (2014) Application of an incremental SVM algorithm for on-line human recognition from video surveillance using texture and color features, Neurocomputing, Vol. 126, Pp. 132-140.

[6]. Gurwicz, Y., Yehezkel, R. and Lachover, B. (2011) Multiclass object classification for real-time video surveillance systems, Pattern Recognition Letters, Vol. 32, Pp. 805–815.

[7]. Dee, H.M. and Velastin., S.A. (2008) How close are we to solving the problem of automated visual surveillance? a review of real-world surveillance, scientific progress and evaluative mechanisms, Machine Vision and Applications, Special Issue on Video Surveillance Research in Industry and Academic, Vol. 19, Issue=5-6, Pp. 329-343.

[8]. P. Vijayakumar and A. V. Senthil Kumar "Moving Object Segmentation using Enhanced Laplacian Thresholding Method for Video Surveillance", IJCA,Volume 100, Issue 4, August 2014, Pages 13-17.

[9]. P. Vijayakumar and A. V. Senthil Kumar "A Novel Framework for Specific Object Tracking in the Multiple Moving Objects Environment", IJAER, Volume 10, Number 16 (2015)Pages 36540-36545.

[10]. P. Vijayakumar and A. V. Senthil Kumar "Moving Object Segmentation Using Median Filter Based Enhanced Laplacian Thresholding" EJSR, Volume 133, Issue Pages 415-427.

[11]. [P. Vijayakumar and A. V. Senthil Kumar"Improvising Enhanced Laplacian Thresholding Technique for Efficient Moving Object Detection in Video Surveillance" IOSRJEN, Vol. 06, Issue 01 (January. 2016), PP 42-52.

[12]. P. Vijayakumar and A. V. Senthil Kumar"Threshold based Filtering Technique for EfficientMoving Object Detection and Tracking in Video Surveillance" IJRET, Volume 05, Issue 02, Pages303-310

[13]. P.Vijayakumar and A.V.Senthil Kumar " An Effective Multiple Moving Objects Tracking using Bayesian Particle Filter-Based Median Enhanced Laplacian Thresholding", IJERT, Volume 5, Issue 7, Pages 93 - 101, July 2016.

[14]. P.Vijayakumar and A.V.Senthil Kumar "An Improved Video Object Segmentation and Tracking based on Features Using Threshold Filtering Technique", MAT Journals, Journal of Computer Science Engineering and Software Testing, Volume 2,Issue 1, Pages 1-17,2016.

[15]. P.Vijayakumar and A.V.Senthil Kumar "An Enhanced Multi Featured Video Object Detection and Tracking using Threshold Filters", IJAER, Volume 11, Number 8,Pages 5400 – 5407, 2016.

[16]. David, L. (2007) Surveillance Studies: An Overview. Cambridge: Polity Press.

[17]. Yang, R., Yin, L., Gabbouj, M., Astola, J. and Neuvo, Y. (1995) Optimal weighted median filters under structural constraints, IEEE Trans. Signal Processing, Vol. 43, Pp. 591–604.

[18]. Hardie, R.C. and Barner, K.E. (1994) Rank conditioned rank selection filters for signal restoration, IEEE Trans. Image Processing, Vol. 3, Pp.192–206.

[19]. [BenHamza, A., Luque, P., Martinez, J. and Roman, R. (1999) Removing noise and preserving details with relaxed median filters,J. Math. Imag. Vision,Vol.11,No.2,Pp.161-177.

[20]. Jain, A.K. (1989) Fundamentals of digital image processing. Prentice-Hall.

[21]. [Strela, V. (2000) Denoising via block Wiener filtering in wavelet domain, 3rd European Congress of Mathematics, Barcelona, Birkhäuser Verlag.

[22]. Bruce, A., Donoho, D. and YeGao, H. (1996) Wavelet Analysis, IEEE Spectrum, Pp.27-35.

[23]. Wang, X., Ma, X. and Grimson, W.E. (2009) Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 3, Pp. 539-555.

[24]. Zhang, J. (2010) An efficient median filter based method for removing random-valued impulse noise, Digital Signal Processing, Vol. 20, Issue 4, Pp. 1010-1018.

[25]. Bigand, A. and Colot, O. (2010) Fuzzy filter based on interval-valued fuzzy sets for image filtering, Fuzzy Sets and Systems, Vol.161, No.1, Pp. 96-117.