

# Moving Object Segmentation in Dynamic Environment by Reducing Impulsive Noise from Background Model

Satrugan Kumar  
Department of Electronics and Communication  
MANIT, Bhopal (India)

Jigyendra Sen Yadav  
Department of Electronics and Communication  
MANIT, Bhopal (India)

## ABSTRACT

A background subtraction method is a computationally inexpensive way to identify moving objects in the scene without any prior information about object and it also provides a sufficient light of information to accomplish critical task in traffic monitoring, object tracking, pattern recognition, human gait and gesture detection. However, for real time systems, the background scene is seriously affected due to changes in lightening condition, shadow cast by moving object, swaying tree, rippling water and much more, which hurdles to produce a reliable motion mask. In this concern, we focus toward the selection of the background pixel by mapping the time variance and absolute difference image in order to cope with abrupt illumination and preserve the spatial consistency. Further the local statistical properties and variance of background image are employed to reduce the local noise impulse within background candidate. Experimental results show that it can work well under static and dynamic background condition.

**Keywords:** Background subtraction, Motion detection, Time variance, Segmentation, Morphology.

## 1. INTRODUCTION

Motion detection and activity analysis of object in a video frames is an active research field in computer vision application. Motion detection [1] is a preliminary task for many real time applications [2] such as traffic surveillance, human activity recognition, people counting, public airport and shopping mall security [3]. Many computer vision algorithms are still facing some problem to cope with varying and dynamic background [4] of a video frame. Concerning real time operation, many statistical and optimization issues are generated in computer vision algorithm due to sudden or gradual illumination condition in scene, poor video quality, presence of multiple moving objects, shadow cast from object and many more[5,6].

Pixel level processing, low memory consumption and computationally inexpensive vision algorithms are often required in real-time processing to accurately define moving entities in the video sequences. If we have a optimized statistical parameters to model a reference background, an object can be detected by finding the significant intensity deviation that does not fit the initial reference background model intensity.

A background subtraction method is employed to detect the moving object, where each video frame of a video sequence is compared against a statistical defined background model to adequately explain the pixels in the current frame that differ significantly from the background .

The innovation of this work lies on the efficacy of the proposed method to generate a suitable background candidate, which can produce sufficient visual representation of moving objects on foreground mask under static and dynamic background conditions [6,7].

Recent existing methods [8,9], which are applied to detect motion in the real time environment, are statistical techniques, frame differencing (Temporal method) [5], Optical flow and Background subtraction. Amongst them background subtraction is computationally inexpensive to localize objects in the scene. Optical flow is not suited to real time due to time complexity. Frame difference method affects the aperture of an object in the scene due to unidentified interior pixels in it. Statistical methods suffer from learning mechanism and initial start-up time.

The aim of this work is to classify the points on the region of interest not belonging to background scene. In this concern, we categorize the stationary and non-stationary pixel and updated the pixel of background by estimating the noise block-wise in the initial background image.

The remaining of paper is organized into four sections. Section 2 deals with some of the related work related to background subtraction and its maintenance. Section 3 explains our proposed methodology. Section 4 explores our experimental results .Concluding remarks are discussed in section 5.

## 2. RELATED WORK

In this section, some existing state-of-the-art background subtraction methods are discussed regarding background initialization and background maintenance. Basically, recursive and non-recursive techniques are widely used in the background subtraction arts. Recursive techniques are Running Gaussian average [10], Gaussian mixture model (GMM) [11], Approximated median filtering [12].The background is estimated by updating each pixel reference background at each new frame of a video. Recursive techniques require less memory requirement but any error in the background model lingers for a longer during the extraction procedure. In Running Gaussian average method, single Gaussian distribution models each pixel of the background. The background is updated at each incoming frame by estimating the mean and variance of Gaussian

distribution. Running Gaussian average is limited to unimodal background model, hence it fails against multiple object detection. Gaussian mixture model (GMM) deals efficiently the complex background using three to five Gaussian distributions per pixel. In the GMM method, each matched pixel between incoming frame and background model is updated through mean and variance of Gaussian model. Even though there is still the problem of time complexity in GMM method. Approximated median filtering update the background model by estimating the median through the recursive filter. It updates the background model either by increasing or decreasing its intensity value to one. It is less robust against hollow space and ignores the variance of pixel. Non-recursive techniques such as Median filtering [13] and Eigen backgrounds [14] require 'n' buffer of memory to estimate the background model. Median filtering is robust to noise but less resistant to sudden illumination. The Eigen backgrounds method exploits the spatial correlation using principle component analysis of a fixed number of video frames, which has no foreground. But it fails when the background gets changed over a long time in vision application.

To deal the problem associated with dynamic background, a real time recursive algorithm using the Kalman filter and dynamic non-local mean are designed. It efficiently deals the illumination problem, but at larger computational complexity. Slow moving object [15] creates too hurdle in a detection task as it disappears in the background. In order to deal with this problem a hybrid approach is suggested in [15]. In some literature, features or edges of the image are fused with background method to generate the reliable foreground mask but the method is less resistant to hollow space and streak phenomenon[16,17].

### 3. PROPOSED WORK

This section illustrates our proposed algorithm, which is designed to model the background image at pixel level under static camera arrangement. The detail flow of this method is shown in Fig 1. In order to improve the visual and quantitative performance, we perform local impulse or speckle reduction using window with respect to the global variance of the pixel. Many of the existing methods update their background only in those pixels, which are classified as stationary or static. The pixel covered by the foreground object has been left untreated during background updation. Therefore, if once get false estimated confidence on each point belonging to foreground, there may be increase in false alarm [18,19]. Moreover, some random noise can exist on the foreground mask, if its impulsive strength is not excluded. In this concern, we have updated also those points when the foreground object covers background.

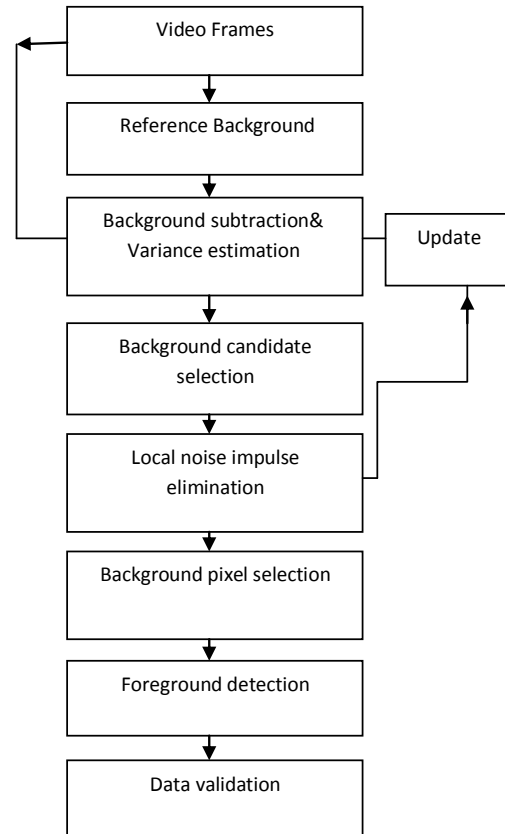


Fig 1.

In regard to the background initialization, a reference frame has taken in which no foreground pixel exists. In this work,  $B(x,y)$  tends to reference background. It is noted that a foreground pixel in incoming frame  $\{I_1, I_2, I_3, \dots, I_t\}$  deviates significantly from background frame. In that concern, the incoming frame  $I_t(x,y)$  is subtracted from reference background and the pixel that retain spatial consistency are estimated using Mahalanobis distance through absolute difference image. The absolute difference of the image is given as:

$$M_t(x,y) = (I_t(x,y) - B(x,y)) \quad (1)$$

In addition to absolute difference, the corresponding time variance is computed. Which is given as follows:

$$\nabla_t(x,y) = \text{variance}(I_t(x,y), B(x,y)) \quad (2)$$

The time variance  $\nabla_t(x,y)$  is adapted at each frame in order to cope with the lightening condition. Although, there may be some stray pixel, that can also be eliminated through the correct estimation of time variance. In fact, object should be appear in sets of connected component on the foreground mask. In that concern, we provide spatial consistency through the absolute difference image  $M_t(x,y)$ .

The background candidate selection is assessed as follows:

$$C_t(x,y) = \begin{cases} B(x,y) & \text{if } M(x,y) < T_1 \text{ and } \nabla_t(x,y) < T_2 \\ \beta * B(x,y) + (1 - \beta) * I_t(x,y) & \text{otherwise} \end{cases} \quad (3)$$

The  $T_1$  is estimated on the basis of variation of foreground pixel from the background pixel and  $T_2$  is selected using Mahalanobis distance between  $(I_t(x,y))$  and  $(B_t(x,y))$ .  $\beta$  is adaptation coefficient which is experimentally set to 0.99 in this paper. There may be the possibility of foreground pixel or

impulsive noise, if the value of  $M_t(x,y)$  and  $V_t(x,y)$  exceeds  $T_1$  and  $T_2$ . In order to treat the impulsive noise, we divide the  $C_t(x,y)$  in to block of size  $8 \times 8$  and for each block the mean and variance are calculated. The mean ' $m_t(x,y)$ ' and variance ' $v_t(x,y)$ ' of each block are computed in order to compare the pixel of  $C_t(x,y)$  with global noise strength. The global noise variance  $Z_t(x,y)$  is calculated using  $C_t(x,y)$ . In order to select the optimum pixel for a new background candidate  $J_t(x,y)$ , a comparative decision is taken as follows:

$$J_t(x,y) = \begin{cases} B(x,y) & \text{if } v_t(x,y) < Z_t(x,y) \\ \alpha * B(x,y) & \text{otherwise} \end{cases} \quad (4)$$

$J_t(x,y)$  belongs to the current background model. Finally, the absolute difference between  $J_t(x,y)$  and  $I_t(x,y)$  is computed for the binary motion mask detection.

$$R_t(x,y) = (I_t(x,y) - J_t(x,y)) \quad (5)$$

$R_t(x,y)$  belongs to the absolute difference image.

Each pixel of foreground mask is mapped into binary mask using Gaussian probability density function  $P_1$  and  $P_2$ . The binary motion mask  $D_t(x,y)$  is computed as:

$$P_1 = \frac{1}{\sqrt{2\pi\sigma_{11}^2}} \exp\left(\frac{-(R_t(x,y) - \mu_{11})^2}{2\sigma_{11}^2}\right) \quad (6)$$

$$P_2 = \frac{1}{\sqrt{2\pi\sigma_{12}^2}} \exp\left(\frac{-(R_t(x,y) - \mu_{12})^2}{2\sigma_{12}^2}\right) \quad (7)$$

$$D_t(x,y) = \begin{cases} 1 & \text{if } P_1 < P_2 \\ 0 & \text{if } P_1 > P_2 \end{cases} \quad (8)$$

$\mu_{11}$ ,  $\mu_{12}$  are means and  $\sigma_{11}^2$  and  $\sigma_{12}^2$  are the variances of  $R_t(x,y)$ . The range of  $\mu_{11}$ ,  $\mu_{12}$ ,  $\sigma_{11}^2$  and  $\sigma_{12}^2$  are calculated from  $R_t(x,y)$ . Typically these value lies between 0 to 2.5. We apply morphological open and close for one iteration operation on  $D_t(x,y)$  in order to avoid the unused pixel on the foreground.

#### 4. EXPERIMENTAL RESULTS

In this section, we have applied our proposed algorithm to some video sequences, which feature some complex and dynamic background. The basic feature of these video sequences has been mentioned in Table1.

**Table-1 Experimental datasets**

Video Title	Image Size	Sequence Type	Background Property
WS	160X128	Outdoor	Dynamic
HALL	168X144	Indoor	Dynamic
MR	160X128	Indoor	Dynamic

The WS (water sequence) video sequence has very fluctuating background due to rippling of water. Moreover, in WS sequence, the foreground and background pixel has equal

contrast and same probability of distribution in the scene in some area. The slow and slumber motion of person may create either hole inside the object or object may vanish over some time from the scene if the background is updated at high coefficient rate. Also, the rippling water may increase false positive pixels if the background is not adapted properly according to environment.

In MR video sequence, the moving curtain can hurdle to produce sufficient foreground information. Moreover, in MR Sequence, the wearing of person has changed after some frames, which matches to background pixel intensity. Therefore to handle the segmentation task by setting a single threshold could be complicated.

In HALL video sequence, the motion detection task can be complicated enough due to presence of multiple object, object staying in the scene for long time, changing illumination.

As our results show, we are able to produce sufficient visual information on the foreground mask. In WS sequence as shown **Fig 2**, the entire shape of the object has been detected even though the foreground and background pixel has same intensity distribution in some part of scene. Moreover, no ghost effect and illumination effect are seen in WS sequence. The output of the modeled background intensity is sufficient to suppress the false alarm, which in turn produces high accuracy.

In other dynamic background of MR video sequence as shown **Fig 3**, cluttered caused by moving curtain has been removed successfully from the foreground mask. Despite of the shirt of person being similar to background pixel intensity, the entire shape of the person has been detected. Indeed, the foreground mask is stable with the gradual change in illumination in the scene. Predictably, the accurate foreground mask is generated due to the stabilized signal generated through the learning of background model.

In HALL video as shown **Fig 4**, multiple people have been successfully detected in case of sudden illumination. The motion mask obtained with the proposed approach show that the correct learning and updating of all the points in the background model avoids the false detection and irrelevant points in the segmented image. We have compared the visual output through our proposed work with respect to ground truth of some sampled frames, which are shown in **Fig 2-4**. The GMM method fails when the intensity of object becomes similar to background and it is more computationally complex.

All the three video have a common feature that object retains in sleeping mode or staying condition in the scene for some time. In this situation, the ghost image will appear when background model is updated slowly. In other context of fast updating, the object will disappear in background. So, the approach to deal with blending the abandoned object into background pixel are done in an elegant way, which in turns produces sufficient visual quality to deal with many surveillance and computer vision application.

To evaluate performance quantitatively through this method, the accuracy metrics are computed at low-level pixel, where a ground truth image is compared with its corresponding foreground mask.

The accuracy metrics are precision, recall and F1 measures, which depends upon TP(True positives), FP (False positives), TN(True negatives), and FN (False negatives). TP are the

number of foreground pixels correctly detected on motion mask while FP are the number of background pixels incorrectly appeared as foreground. TN tend to number of background pixels correctly detected while FN are the number of foreground incorrectly appeared as background.

The relevant pixel on the foreground mask is measured as recall, which is given as:

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

Where, TP+FN are the number of true positive pixels in ground truth image.

The relevant pixel on the foreground mask is measured as recall, which is given as:

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

Where, TP+FP are the number of true positive pixels in ground truth image.

As precision and recall do not produce exact decision on the accuracy of detected foreground mask .In order to support the reliable decision, a parameter F1 has been proposed which is the arithmetic mean of precision and recall. The F1 score is evaluated as:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (11)$$

Sampled frame	Ground Truth	Obtained results
P*=0.9156	R*=0.9225	F1=0.9191
P*=0.9386	R*=0.9509	F1=0.9447
P*=0.9434	R*=0.9582	F1=0.9507
P*=0.9711	R*=0.9234	F1=0.9467

P\*=Precision, R\*=Recall

Fig 2. Foreground mask of WS sequence tends to Frame number 1499,1605,1615,1624 respectively.

Sampled frame	Ground Truth	Obtained results
P*=0.9884	R*=0.9203	F1=0.9531
P*=0.9764	R*=0.8208	F1=0.8919
P*=0.9935	R*=0.8671	F1=0.9260
P*=0.9942	R*=0.8334	F1=0.9067

P\*=Precision, R\*=Recall

Fig 3. Foreground mask of MR sequence tends to Frame number 22774,22847,23854,23893 respectively.

Sampled frame	Ground Truth	Obtained results
P*=0.8191	R*=0.7892	F1=0.8038
P*=0.8763	R*=0.8810	F1=0.8786
P*=0.7891	R*=0.8885	F1=0.7181
P*=0.6231	R*=0.8473	F1=0.8342

P\*=Precision, R\*=Recall

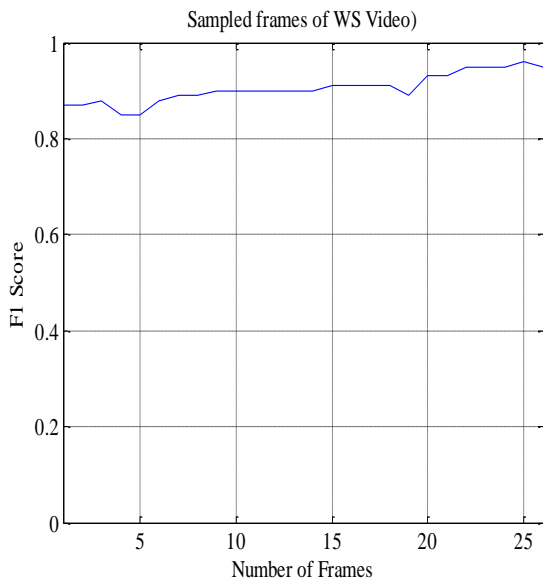
Fig 4. Foreground mask of HALL sequence tends to Frame number 1656,2289,3434,3800 respectively.

The evaluated quantitative metric ranges from 0 to 1 in which the highest attained value has the maximum accuracy. The average value of precision, recall and F1 is computed for the discussed video sequences in this paper. In this context the proposed method is compared with the GMM method and displayed in **Table 2**.

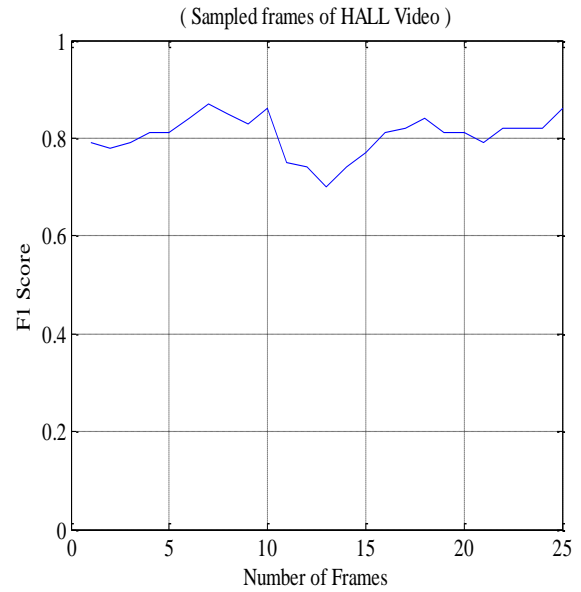
**Table 2. Average precision, Recall and F1 measure of Video sequence**

Video sequences	Quantitative measurement	Proposed Method	GMM method
WS	Precision	0.9097	0.8547
	Recall	0.9040	0.6259
	F1	0.9068	0.7226
HALL	Precision	0.7132	0.4799
	Recall	0.8594	0.6221
	F1	0.7795	0.5441
MR	Precision	0.9196	<b>0.9501</b>
	Recall	0.8678	0.6767
	F1	0.8929	0.7684

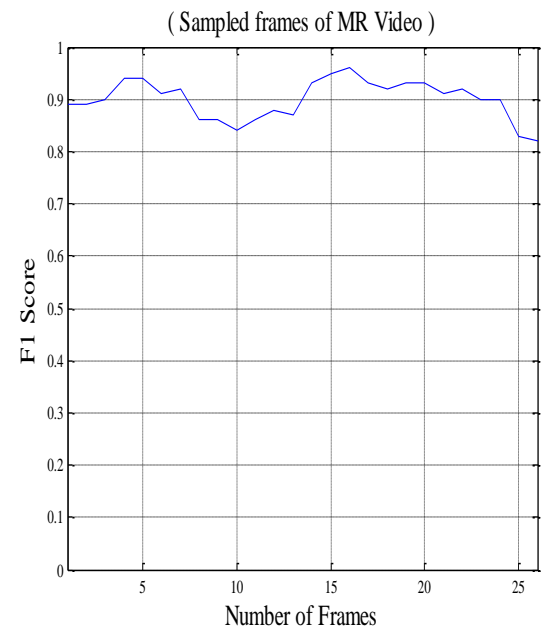
It is noted that F1 score achieved through this method for WS sequence is 18% higher than that was achieved through GMM method. In HALL monitoring video sequence, F1 score achieved through GMM method was 24% lower than our proposed scheme. Concerning the MR sequence, the F1 scores achieved through GMM were 13% lower than those obtained with this algorithm . F1 scores of sampled frames are shown in **Fig 5-7**.With regard to time complexity, GMM and this proposed algorithm are simulated using MATLAB 7.10 with 2GHz CPU. In order to process a frame of 160X128, GMM takes 0.48 sec while this algorithm takes 0.12 sec, Which shows that it executes fast when compared to GMM. The average precision,recall,F1 are calculated through GMM method by taking the optimum parameter given in [16].



**Fig 5.**



**Fig 6.**



**Fig 7.**

## 5. CONCLUSION

The experimental results obtained in different video sequence having dynamic background show the efficacy and reliability of the proposed approach. The efficient modeling of background has proven to solve the problems of a small movement of rippling water, gradual illumination, and unpredictable movement of the curtain and ghost elimination when object retains stationary in the scene. We presented a robust approach to learn all the pixel of the background, which allows our system to cope with multiple moving entities in the scene for long periods. However to calculate the precise segmentation of foreground to handle multiple object detection and noise removal, the work can be extended in future. Experimental results show the better accuracy and optimum output at foreground mask through this proposed algorithm and its also provide an important cues to many computer vision application.

## 6. REFERENCES

- [1] Cutler, R., & Davis, L. S. 2000. Robust real-time periodic motion detection, analysis, and applications. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 22(8), 781-796.
- [2] M.M. Azab, H.A. Shedeed, A.S. Hussein.2010.A new technique for background modeling and subtraction for motion detection in real-time videos.IEEE International Conference on Image Processing (ICIP), pp. 3453–3456.
- [3] Cheung, S. C. S., & Kamath, C. 2005. Robust background subtraction with foreground validation for urban traffic video. *Eurasip Journal on applied signal processing*, 2330-2340.
- [4] Patel, C. I., Garg, S., Zaveri, T., & Banerjee, A. 2014. Top-Down and Bottom-Up Cues Based Moving Object Detection for Varied Background Video Sequences. *Advances in Multimedia*, 2014.
- [5] Anderson, S. J., & Burr, D. C. 1985. Spatial and temporal selectivity of the human motion detection system. *Vision research*, 25(8), 1147-1154.
- [6] Selvanayaki, K. S., & Somasundaram, R. M. 2013. A Survey on Image Segmentation Techniques for Edge Detection. In *IJCA Proceedings on International Conference on Innovation in Communication, Information and Computing 2013* (Vol. 303, No. 2, pp. 31-34). Foundation of Computer Science (FCS).
- [7] Do, B. H., & Huang, S. C. 2011. Dynamic background modeling based on radial basis function neural networks for moving object detection. *IEEE International Conference on In Multimedia and Expo (ICME)*, (pp. 1-4).
- [8] Bouwmans, T. 2014. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11, 31-66.
- [9] L. Wang, W. Hu, and T. Tan.,2003.Recent developments in human motion analysis.*Pattern Recognition*, Vol. 36,No. 3, pp.585-601.
- [10] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland.1997Pfinder: real-time tracking of the human body. *IEEE PAMI*,9(7):780–785.
- [11] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking.1999. In *CVPR99*, volume 2, page-252.
- [12] N. McFarlane and C.1995. Schofield. Segmentation and tracking of piglets in images. *MVA*, 8:187–193.
- [13] S. Calderara,R. Melli, A. Prati, and R. Cucchiara.2006. Reliable background suppression for complex scenes. In *VSSN06*,pp-211–214.
- [14] N. Oliver, B. Rosario, and A. Pentland.2000. A bayesian computer vision system for modeling human interactions. *IEEE PAMI*,22:831–843.
- [15] Z.Zhu, Y.Wang.2012.A Hybrid Algorithm for Automatic Segmentation of Slow Moving Objects.*AEU-Int. Journal of Electronics and Communications*, vol.66, pp.249-254.
- [16] Mazoul, A., Zebbara, K., & Ansari, M. 2012. Street crossing pedestrian detection based on edge curves motion. *International Journal of Computer Applications*, 41(16), 20-24.
- [17] Rahman, F. Y. A., Hussain, A., Zaki, W. M. D. W., Zaman, H. B., & Tahir, N. M. 2013. Enhancement of background subtraction techniques using a second derivative in gradient direction filter. *Journal of Electrical and Computer Engineering*, 21.
- [18] Manzanera, A., & Richefeu, J. C. 2007. A new motion detection algorithm based on  $\Sigma-\Delta$  background estimation. *Pattern Recognition Letters*, 28(3), 320-328.
- [19] Antoine Manzanera and J. C. Richefeu .2007.A robust and computationally efficient motion detection algorithm based on  $\Sigma-\Delta$  background estimation.*ICVGIP*, 46-51 .