Estimation of Dynamic Background and Object Detection in Noisy Visual Surveillance

M.Sankari
Department of Computer Applications,
Nehru Institute of Engineering and Technology,
Coimbatore, INDIA

C. Meena
Head, Computer Centre,
Avinashilingam University,
Coimbatore, INDIA

***Abstract*—Dynamic background subtraction in noisy environment for detecting object is a challenging process in computer vision. The proposed algorithm has been used to identify moving objects from the sequence of video frames which contains dynamically changing backgrounds in the noisy atmosphere. There are many challenges in achieving a robust background subtraction algorithm in the external noisy environment. In connection with our previous work, in this paper, we have proposed a methodology to perform background subtraction from moving vehicles in traffic video sequences that combines statistical assumptions of moving objects using the previous frames in the dynamically varying noisy situation. Background image is frequently updated in order to achieve reliability of the motion detection. For that, a binary moving objects hypothesis mask is constructed to classify any group of lattices as being from a moving object based on the optimal threshold. Then, the new incoming information is integrated into the current background image using a Kalman filter. In order to improve the performance, a post-processing has been done. It has been accomplished by shadow and noise removal algorithms operating at the lattice which identifies object-level elements. The results of post-processing can be used to detect object more efficiently. Experimental results and analysis show the prominence of the proposed approach which has achieved an average of 94% accuracy in real-time acquired images.**

Keywords- Background subtraction; Background updation; Binary segmentation mask; Kalman filter; Noise removal; Shadow removal; Traffic video sequences.

I. INTRODUCTION

In visual surveillance model, estimating the dynamic background and detecting the object from the noisy environment is a computationally challenging problem. Our main target is to identify the object from the multi model background using background subtraction, shadow removal and noise removal techniques. For that we need to detect and extract the foreground object from the background image. After detecting the foreground object there may a large number of possible degradations that an image can suffer. Common degradations are blurring, motion and noise. Blurring can be caused when an object in the image is outside the cameras due to loss of depth information during the exposure. In the proposed approach after detecting the object image, converting it into its spatial frequencies, developing a point spread function (PSF) to filter the image with, and then converting the filtered result back into the spatial domain to see if blur was removed.

This can be done in several steps. At the end, an algorithm was developed for removing blur from an already blurry image with no information regarding the blurring PSF. In-class variability, occlusion, and lighting conditions also change the overall appearance of vehicles. Region along the road changes continuously while the lighting conditions depend on the time of the day and the weather. The entire process is automatic and uses computation time that scales according to the size of the input video sequence.

The remainder of the paper is organized as follows: Section II gives the overview of the related work. Section III describes the architecture and modeling of proposed methodology for background elimination and object detection. Implementation and performance are analyzed in section IV. Section V contains the concluding remarks and future work.

II. OVERVIEW OF THE RELATED WORK

Scores of research have been done in the literature in order to attain a solution to an efficient and reliable background subtraction. To detect moving objects in a dynamic scene, adaptive background subtraction techniques have been developed [1] [2] [3]. Adaptive Gaussian mixtures are commonly chosen for their analytical representation and theoretical foundations. For these reasons, they have been employed in real-time surveillance systems for background subtraction [4] [5] and object tracking [6]. For foreground analysis [7] [8], a method for foreground analysis was proposed for moving object, shadow, and ghost by combining the motion information. The computation cost is relatively expensive for real-time video surveillance systems because of the computation of optical flow. In [9], a work has presented on a novel background subtraction algorithm that is capable of detecting objects of interest while all pixels are in motion. Background subtraction technique is mostly used for motion pictures to segment the foreground object by most of the researchers [10] [11]. Liyuan Li, et al. [12] proposed foreground object detection through foreground and background classifications under bayesian framework. In addition, moving object segmentation with background suppression is affected by the problem of shadows [6] [13]. Indeed, the moving object detection do not classify shadows as belonging to foreground objects since the appearance and geometrical properties of the object can be distorted which, in turn, affects many subsequent tasks such as object classification and the assessment of moving object position. In this paper, we propose a novel simple method that exploits all these features, combining them so as to efficiently provide detection of moving objects, ghosts, and shadows. The main contribution of this proposal is the integration of knowledge of detected objects, shadows, and ghosts in the segmentation process to enhance both object segmentation and background keep posted. The resulting method proves to be accurate and reactive and, at the same time, fast and flexible in the applications.

III. PROPOSED WORK

The proposed system extracts foreground objects such as people, objects, or events of interest in variety of noisy environment. The schematic flow of the proposed algorithm is shown in Fig.1. This is an extension work of our previous method [14]. Typically, these systems consist of stationary cameras placed in the highways. These cameras are integrated with, intelligent computer systems that perform preprocessing operation from the captured video images and notify human operators or trigger control process. The objective of this real-time motion detection and tracking algorithm is to provide low-level functionality for building higher-level recognition capabilities.

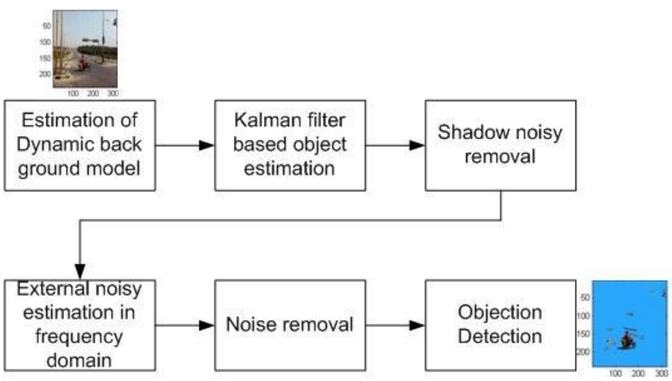

Figure 1. Schematic flow of the proposed algorithm to detect the objects in the noisy environment.

A. Preprocessing

Preprocessing is the key step and the starting point for image analysis, due to the wide diversity of resolution, image format, sampling models, and illumination techniques that are used during acquisition. In our method, preprocessing step was done by statistical method using adaptive median filter. The resultant frames are then utilized as an input for the background subtraction module. Image *I(x,y)* at time *t* is shown in Fig.2. The background image *B(x,y)* at time *t* is shown Fig.3.

Figure 2. Image at time *t*: *I (x; y; t).*

Figure 3. Image at time *t*: *B (x; y; t).*

In order to get the estimated background, we have used an adaptive median filter. Basically, impulse noise is a major artifact that affects the sequence of frame in the surveillance system. For this reason to estimate the background in the noisy environment we have proposed an adaptive median filter (AMF). The AMF can be used to enhance the quality of noisy signals, in order to achieve better forcefulness in pattern recognition and adaptive control systems. It executes on spatial processing to determine which pixels in an image have been exaggerated by impulse noise. The AMF categorizes pixels as noise by contrasting each pixel in the image to its close proximity of neighbor pixels. The size of the neighborhood and its threshold are adaptable for the robust assessment. A pixel that is dissimilar from a mainstream of its neighbors, as well as being not logically aligned with those pixels to which it is similar, is labeled as impulse noise. These noise pixels are then substituted by the median pixel value of the pixels in the neighborhood that have passed the noise labeling test [15]. The following steps were used for back ground estimation.

Step 1: Estimate the background at time t using adaptive median filter method.

Step 2: Subtract the estimated background from the input frame.

Step 3: Apply a threshold φ to the absolute difference to get the binary moving objects hypothesis mask.

Assuming that the background is more likely to appear in a scene, we can use the median of the previous n frames as the background model

$$B(x,y,t) = \Theta((I(x,y,t-i)), \quad (1)$$

$$|I(x,y,t) = \Theta(I(x,y,t-i))| > \varphi, \quad (2)$$

where $i \in \{0,1,2,\ldots,n-1\}$ and computation of $\Theta(.)$ is based on the AMF. The following are the algorithm for median filter computation. Algorithm consists of two steps. Step 1 is described for deciding whether the median of the gray values in the size of the neighborhood or not.

Step 1: $$\begin{aligned} \varpi_1 &= \Theta(\delta(x,y)) - \delta_{\min}(x,y) \\ \varpi_2 &= \Theta(\delta(x,y)) - \delta_{\max}(x,y) \end{aligned},$$

If $\varpi_1 > 0$ && $\varpi_2 < 0$ then do the Step 2 otherwise enlarge the neighborhood window size based on the maximum allowed size of the input image. Step 1 is executed until $S_w > S_{\max}$ otherwise step 2 is computed.

Step 2: $\varpi_1 = G(x,y) - \delta_{\min}(x,y)$, $\varpi_2 = G(x,y) - \delta_{\max}(x,y)$

If $\varpi_1 > 0$ && $\varpi_2 < 0$ then we have taken $G(x,y)$ as an output else $\Theta(\delta(x,y))$.

where $\delta(x,y)$ is the AMF neighborhood size. $\delta_{\min}(x,y)$ is the minimum gray level value in neighborhood $\delta(x,y)$, $\delta_{\max}(x,y)$ represents the maximum gray level value in $\delta(x,y)$, $\Theta(\delta(x,y))$ denotes the median of gray levels in $\delta(x,y)$, $G(x,y)$ is the gray level at coordinates (x,y), S_w represents window size of the current process and $S_{\max}$ is the maximum allowed size of $\delta(x,y)$.

B. Foreground Detection

In this module estimated background and foreground mask images are used as an input for further processing. Thus, we use grayscale image sequences as input. Elements of the scene and the sizes of the traffic objects (vehicles and pedestrians) are unknown. The Foreground detection is done by using accumulative difference method, which is change-detection based on subtraction of a background image. It is necessary to update the background image frequently in order to guarantee reliable object detection. The basic idea in background adaptation is to integrate the new incoming information into the current background image using a Kalman filter:

$$B_{(t+1)} = B_t + [a_1 * 1 - M_t + a_2 * M_t] D_t, \tag{3}$$

where B_t represents the background model at time t, D_t is the difference between the present frame and the background model, and M_t is the binary moving objects hypothesis mask. The gain a_1 *and* a_2 are based on an estimate of the rate of change of the background. The larger it is, the faster new changes in the scene are updated to the background frame. In our approach, $a_1 = 0.1$ *and* $a_2 = 0.01$, they are kept small and the update process based on Eq.(3) is only intended for adapting to slow changes in overall lighting.

$$M_t(x) = \begin{cases} 1, & if \ |D_t(x)| > T_t \\ 0, & Otherwise \end{cases}. \tag{4}$$

Foreground detection is started by computing a pixel based absolute difference between each incoming frame $I(x,y)$ and an adaptive background frame $B_t(x,y)$. The pixels are assumed to contain motion if the absolute differences exceed a predefined threshold level.

$$F(x,y) = \frac{|I(x,y) - B_t(x,y) - \mu|}{\sigma} > T \tag{5}$$

As a result, a binary image is formed where active pixels are labeled with a "1" and non-active ones with a "0". With the updated background image strategy using Kalman filter, we get the better foreground detection result. This is a simple, but efficient method to monitor the changes in active during a few consecutive frames. Those pixels which tend to change their activity frequently are masked out from the binary image representing the foreground detection result.

C. Shadow Removal

Shadows appear as surface features, when they are caused by the interaction between light and objects. This may lead to problems in scene understanding, object segmentation, tracking, recognition, etc. Because of the undesirable effects of shadows on image analysis, much attention was paid to the area of shadow detection and removal over the past decades and covered many specific applications such as traffic surveillance. In this paper, 8-neighborhood gray clustering method is used to define the precise shadow and remove it. The mean clustering threshold and the initial cluster seed of the gray are calculated by the following equations.

$$T_i = (1/3)\max(G(x,y) - u_i)^2, \tag{6}$$

where $G(x,y)$ is a gray value of the pixel in $I(x,y)$, u_i is the mean of $G(x,y)$. The initial seed C_i locates in the centre of $G(x,y)$.The clustering starts from the seed C_i, and the point P_i is examined in turn. If at least one point in the 8-neighborhood of P_i has been marked as a shadow region, standard deviation of P_i is calculated by the following equation.

$$\delta P_i = (P_i(x,y) - u_i)^2, \tag{7}$$

where $P_i(I(x,y))$ is the gray value of I(x,y). If $\delta P_i < T$, P_i must be shadow point; otherwise, the point need not be marked. The point P_i is checked constantly until no new point is marked. At last all the marked shadow points are removed.

D. Noise removal

In regular practice due to the camera noise and irregular object motion, there are some noise regions existed in both the object and background regions. In our method we have incorporated Gaussian noise with the acquired image and propose a solution to see how the background subtraction module would behave while the traditional background algorithms are not providing the significant results. The focus is on the background subtraction module because image noise mostly impacts the foreground extraction process. If the foreground objects are not detected well, the rest of the modules will possibly fail at their tasks.

In the proposed method, after finding the foreground object, noise is estimated and modeled using the following algorithm.

Step 1: Convert RGB image of $F(x,y)$ into gray scale image.

Step 2: Motion blurring can be estimated in a spatially linear invariant system under certain conditions. If we assume the object translates at a constant velocity *V* during the exposure time *T* with angle α, then the distortion $d = VT$ and define the point spread function (PSF) as

$$h(x,y) = \begin{cases} \frac{1}{d}, & 0 \le |x| \le d\cos\alpha, y = d\sin\alpha \\ 0 & otherwise \end{cases}, \quad (8)$$

The motion blur can be described mathematically as the result of a linear filter

$$g(x,y) = h(x,y) * F(x,y) + \eta(x,y), \quad (9)$$

where $g(x,y)$ denotes the blurred image, $f(x,y)$ denotes the original image, $\eta(x,y)$ denotes additive noise and * represents 2-D convolution.

Step 3: Dividing the Fourier transform of the PSF into the transform of the blurred image, and performing an inverse FFT, reconstruct the image without noise.

$$\Im(f(x,y)) * \Im(PSF) = \Im(g(x,y)), \quad (10)$$

$$f(x,y) = \Im^{-1}\left[\frac{\Im(g(x,y))}{\Im(PSF)}\right], \quad (11)$$

where $f(x,y)$ is the original image and $g(x,y)$ acquired blurred image. $\Im(.)$ and $\Im^{-1}(.)$ represent Fourier and inverse Fourier transform.

In order to estimate the point-spread-function (PSF) we make use of autocorrelation functions $K(n)$ of an M pixel image line l, which is defined as,

$$K(n) = \sum_{i=-M}^{M} l(i+n)l(i) \qquad n \in [-M, M], \quad (12)$$

where $l(i) = 0$ outside the image line range. The above Equation describes how pairs of pixels at particular displacements from each other are correlated. It is high where they are well correlated and low where poorly correlated. For a normal image, the $K(n)$ will be some function of distance from the origin plus random noises. But for a motion blurred image, the $K(n)$ will decline much more slowly in the direction of the blur than in other directions.

Step 4: The Noise to Signal Power Ratio was computed using the following equations:

PSNR is measured in decibels. It's defined as

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2, \quad (13)$$

Where i, j are the width and height of the frame, respectively, in pixels. The PSNR is defined as

$$PSNR = 20\log_{10}\left(\frac{MAX_I^2}{\sqrt{MSE}}\right), \quad (14)$$

Step 5: Apply Wiener filter with θ and d to deblur the image.

Step 6: The Wiener filtering is employed on the resultant Autocorrelation matrices. Wiener filtering minimizes the expected squared error between the restored and perfect images. A simplified Wiener filter is as follows:

$$\hat{F}(x,y) = \left[\frac{1}{H(x,y)}\frac{|H(x,y)|^2}{|H(x,y)|^2 + S_\eta(x,y)/S_f(x,y)}\right]G(x,y), \quad (15)$$

where $H(x,y)$ is the degradation function

$$|H(x,y)|^2 = H^*(x,y)H(x,y), \quad (16)$$

$H^*(x,y)$ is the complex conjugate of $H(x,y)$, $S_\eta(x,y) = |N(x,y)|^2$ represents PSF of the noise, $S_f(x,y) = |F(x,y)|^2$ denotes the PSF of the original image $S_\eta(x,y)/S_f(x,y)$ is called the noise to signal power ratio.

If noise power spectrum is zero, the Wiener filter reduces to a simple inverse filter. If *K* is large compared with $|H(x,y)|$, then the large value of the inverse filtering term $1/H(x,y)$ is balanced out with the small value of the second term inside the brackets.

Step 7: The Lucy-Richardson filtering on the resultant Autocorrelation matrices. Using Lucy-Richardson method the image is modeled by maximizing the likelihood function gives an equation that is satisfied when the following iteration converges

$$\hat{f}_{k+1}(x,y) = \hat{f}_k(x,y)\left[h(-x,-y) * \frac{g(x,y)}{h(x,y) * \hat{f}_k(x,y)}\right], \quad (17)$$

where * indicates the convolution. $\hat{f}$ is the estimate of the original image f at step k of the iteration. The iteration starts with $f_0 = g(x,y) * h(x,y)$ and $g(x,y)$ denotes the blurred image, $h(x,y)$ denotes the PSF.

E. Detection

After post-processing, the image is compared with the one of the original frames (usually, the first frame). If the pixels are less than certain threshold, then they are ignored. Otherwise, they are replaced by the pixels of original image. This resulting image will be consisting of the moving object ignoring the background and hence satisfying our requirement.

IV. Implementation and Performance Analysis

This system was implemented on an Intel Pentium IV 280 GHz PC. We have tested the system on image sequences on different scenarios like traffic junction intersection, highways etc. Real life traffic video sequence is used to demonstrate the vehicle tracking from traffic video sequences using the proposed framework. All the videos chosen for vehicle tracking have same light intensity and have been taken during day time. We convert the colour video frames to gray scale images.

Automatic monitoring visual surveillance system implementation needs to detect vehicles using automatic background extraction. Background subtraction is the main step for vehicle detection. Fig. 4 shows number of successive frames that are used to extract the background. Digital camera used to take shots. The camera placed over the highway directly. It shots eight frames per second.

Figure 4. Number of successive frames that are used for preprocessing.

Estimated background using Median Filter for n = 12. This is shown in Fig.5

Figure 5. Estimated background image.

After applying the threshold, φ to the absolute difference we got the binary moving objects hypothesis mask which is shown in Fig. 6. With no noise information, the Wiener and other filters do a poor job at realizing the original, non-degraded, image. However, the Lucy-Richardson filter works really well, despite having no information about the noise in the image. With noise information, the Wiener filter gives better result than other filters. Fig 7. shows the comparison of PSNR value with varies filters.

Figure 6. Binary moving objects hypothesis mask.

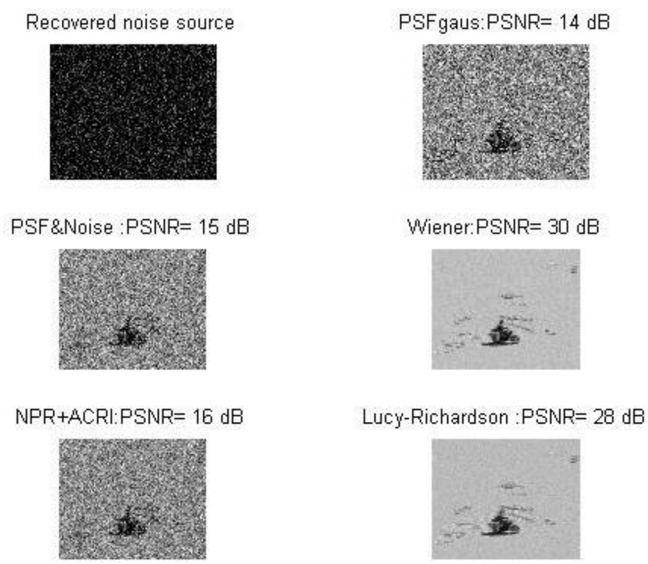

Figure 7. Comparison of PSNR value with a mixture of filters.

Table I shows the result of a mixture of filters used in the Noise removal module. Fig 8 shows the Comparative Analysis of mixture of filters using chart.

Table I. Result of Nose removal module in decibel.

Object used for analysis	Gaussian	NPR	Wiener	NPR + Autocorrelatio	L-R
car	13	13	25	13	24
bus	12	13	24	13	22
motorcycle	14	15	30	16	28
Lorry	10	25	24	23	25
Truck	14	20	26	20	24

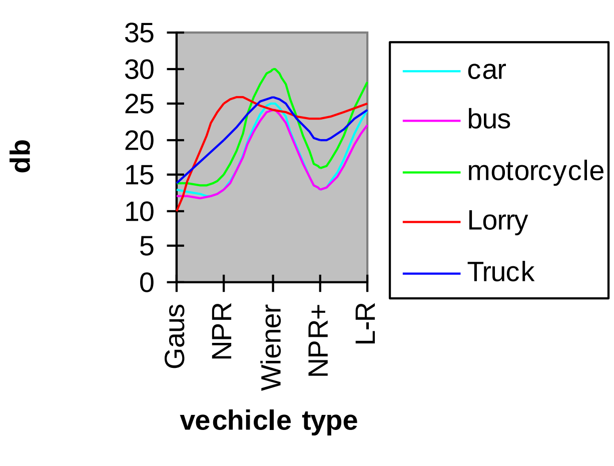

Figure 8. Comparative Analysis of a mixture of filters.

Fig.9 shows the foreground detected objects obtained after background subtraction, shadow and noise. The automatic background extraction results are very good and promising. The most effective parameters that are playing a main role for automatic background extraction are the threshold level. This threshold is used to extract the moving vehicles from the background. Matlab built-in function has been employed for the evaluation of the threshold.

Table II gives vehicle detection results. The background is subtracted from the current image then the resulted image is filtered to get moving vehicles only. By using this technique most of vehicles are detected. Moving vehicles are detected easily after background is subtracted. Performance analysis is shown in Fig.10.

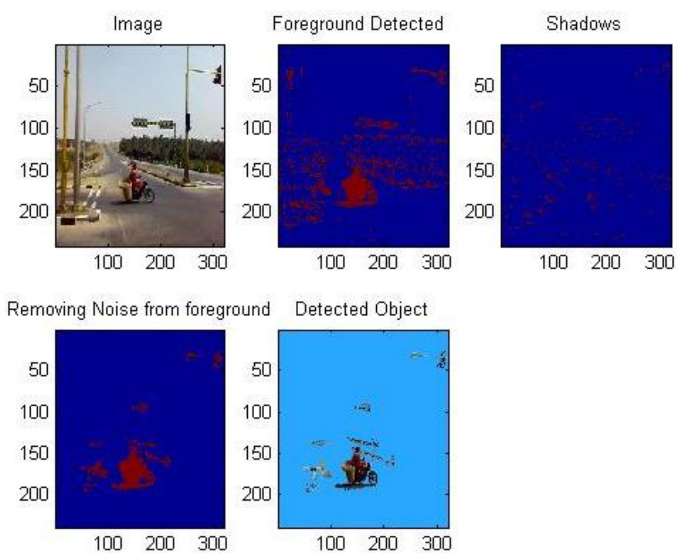

Figure 9. Sequence of steps in the foreground object detection process.

TABLE II. The results for vehicle detection after noise removal

Type of Vehicle	Actual Number of vehicles	Detected Vehicles	Rate %
Car	30	28	93
Motor cycle	30	29	97
Bus	30	27	90
Lorry	30	27	90
Truck	30	28	93

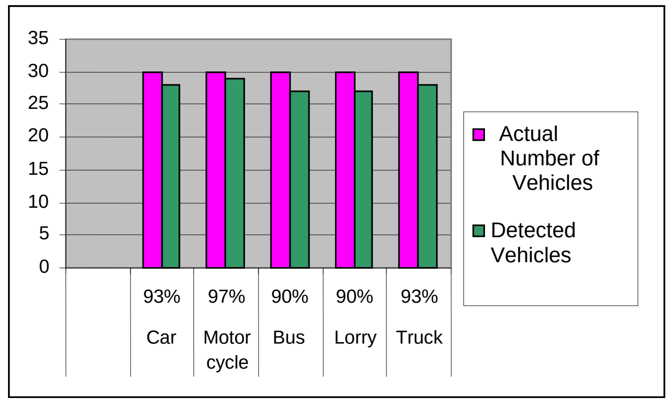

Figure.10 Performance analysis of the proposed method in detection different vehicles in the noisy traffic environment.

V. CONCLUSION

The experimental results of using this approach lead to detect moving vehicles efficiently. This algorithm has been implemented and evaluated experimentally using natural traffic images. It gives promising and effective results where an average vehicle detection rate was around 94%. In this approach the background subtraction and edge detection are used. It is mainly focused on the autocorrelation method. And then an adaptive algorithm is applied to autocorrelation. PSNR is used for the evaluation to measure the performance of the filters. It could be improved and used as a basis for automatic traffic monitoring. Failure detection resulted from occluding large vehicles with small ones and the far moving vehicles that appear as a point in the image. These difficulties could be solved in the future work.

ACKNOWLEDGMENT

Authors thank their family members and children for their continuous support and consent encouragement to do this research work successfully.

REFERENCES

[1] M. Cristani, M. Bicegi, and V. Murino, "Integrated Region- and Pixel-based Approach to Background Modeling", Proceedings of the MOTION, 2002.

[2] H. Eng, J. Wang, A. Kam, and W. Yau, "Novel Regionbased Modeling for Human Detection within High Dynamic Aquatic Environment," Proceedings on CVPR, 2004.

[3] P. KaewTraKulPong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection," In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, 2001.

[4] M. Harville, G. Gordon, and J. Woodfill, "Foreground Segmentation Using Adaptive Mixture Models in Color and Depth," Proc. ICCV Workshop Detection and Recognition of Events in Video, July 2001.

[5] C. Stauffer and W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," Proc. Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 246-252, June 1999.

[6] S.J. McKenna, Y. Raja, and S. Gong, "Object Tracking Using Adaptive Color Mixture Models," Proc. Asian Conf. Computer Vision, vol. 1, pp. 615-622, Jan. 1998.

[7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams," IEEE Trans. on PAMI, 25: (10), October 2003.

[8] Connell, J., "Detection and Tracking in the IBM PeopleVision System," IEEE ICME, June 2004.

[9] Dongxiang Zhou, Hong Zhang and Nilanjan Ray, "Texture Based Background Subtraction," *Proc. IEEE Int. Conf. on Information and Automation*, Zhangjiajie, China, pp. 601-605, 2008.

[10] A. McIvor, "Background subtraction techniques," in *Proc. Image Video Computing*, pp. 147-153, 2000.

[11] Y. Ivanov, A. Bobick, and J. Liu, "Fast lighting independent background subtraction," *Int. Journal on Comp. Vision*, vol. 37, no. 2, pp. 199–207, Jun. 2000.

[12] Liyuan Li, Weimin Huang, Irene Yu-Hua Gu and Qi Tian, "Statistical Modeling of Complex Backgrounds for Foreground Object Detection," *IEEE Transactions on Image Proc.*, vol. 13, no. 11, pp. 1459-1472, Nov 2004.

[13] A. Elgammal, D. Harwood, and L.S. Davis, "Non-Parametric Model for Background Subtraction," Proc. IEEE Int'l Conf. Computer Vision '99 FRAME-RATE Workshop, 1999.

[14] M. Sankari, Dr. C. Meena "Adaptive Background Estimation and object detection applying in Automated Visual surveillance" International Journal of Computer science and Information Security July 2010, Vol. 8 No. 4.

[15] Peng Lei, "Adaptive median filtering", Machine vision, 140.429 Digital image processing.

AUTHORS' PROFILE

Mrs. M. Sankari received her B.Sc. and M.Sc. degrees in Computer science from Bharathidasan University in 1988 and 1990, respectively. She has completed her Master of Philosophy degree in Computer science from Regional Engineering College, Trichy in 2000. Presently, she is a Assistant Professor & Head in the department of MCA at NIET. She is pursuing her Ph.D. degree in Computer science at Avinashilingam University for women, Coimbatore, India. She has published several technical papers in IEEE/IET conferences and Journal. Her field of research includes Computer vision, Pattern recognition, Analysis of algorithms, Data structure, Computer graphics and multimedia.

Dr. C. Meena received her B.Sc (Physics), and Master of Computer Applications degrees from Madurai Kamaraj University in 1987 and 1990, respectively. She has completed her Ph.D. degree in Computer science from Bharathiyar University in 2006. Presently, she is Head, Computer Centre at Avinashilingam University, Coimbatore, India. She is guiding for funded Research Project in UGC and Naval Research Board. She has presented international/national journals. She has published various technical papers in IEEE/IET conferences and national level conferences. Her field of research includes Image processing, Computer vision and Pattern recognition.