International Journal of Information Retrieval Research
Volume 12 • Issue 2

Detection of Change in Body Motion With Background Construction and Silhouette Orientation:

Background Subtraction With GMM

Rohini Mahajan, Central University of Jammu, India*
https://orcid.org/0000-0003-3577-9811

Devanand Padha, Central University of Jammu, India

ABSTRACT

Background subtraction techniques have been widely implemented and improvised to obtain a stable background model. The novelty of the proposed work is to generate a stable background model under dynamic changes in the environmental conditions where 1) an improved background subtraction algorithm is proposed based on GMM with EM algorithm for computing granulometry to run faster for the generation of a stable background model; 2) detecting the foreground by curvelet based denoising process with improvised semisoft thresholding techniques with morphological operations is done; 3) background maintenance is done by an adaptive algorithm in which the intensity values are mapped to remove the connected components with less than P pixels. The proposed scheme works for the spatio-temporal motion of the object in both spatial and temporal modes. The experimental outcome for the proposed model results in the accurate shape analysis of the object in motion thereby dipping the complexity.

KEYWORDS

Background Maintenance, Background Subtraction, Curvelet Denoising, Gaussian Mixture Model, Spatio-Temporal Analysis

I. INTRODUCTION

Tracking and assessment of the activity in a moving body is a challenging task. The evaluation of object in motion and distinguish the background and foreground, many techniques have been implemented. This includes optical flow algorithm, interframe difference and background subtraction(Kim & Jung, 2017). The optical flow algorithm proposed by researchers (Fuentes et al., 2018) requires estimation of flow vectors computing the magnitude of flow estimations and warping flow fields. But the technique cannot deal with dynamic conditions. The temporal frame differencing as per researchers (Chiu et al., 2018) exploit the consecutive frames to mine the pixel by pixel difference of the two images. The subtraction process cannot handle the fast moving objects because when the object in motion goes still or moves fast, for a few frames, background and foreground can't be distinguished. Also, thresholding the difference is a kind of exacting approach which may lead to unnoticed activity

DOI: 10.4018/IJIRR.299935 *Corresponding Author

of the interest of researcher (Anandhalli & Baligar, 2015)(Mahajan & Padha, 2018). The widespread modus operandi for distinguishing the background from the foreground is Background Subtraction (BS) which extracts low entropy based pixels of the object in motion without any prior information about the scene proving to be effectual in stationary camera arrangements and highly precise in pixel, frame as well as region level procedure (Kumar & Yadav, 2016b). The procedure of background subtraction comprise of three stages: a) initialize the background by mapping the spatio-temporal constraints and reconstruct the model based on intensity variations, b) Foreground extraction which detects and segments the moving object and c) Maintenance of background by updating of pixel variance in spatial domain along temporal constraints in the video processing of the sequences.

The techniques implemented by various researchers (Kumar & Yadav, 2016a) (Fazli et al., 2009a) (Fazli et al., 2009b) in these three stages define the accuracy and precision of the proposed algorithms. In our study, in the background modeling stage, an improved Expectation Maximization based algorithm for Gaussian mixtures is proposed to initialize the multi modal background with dynamic conditions. In the foreground detection stage, a curvelet based denoising method with adaptive normalization of the intensities values is done to minimize the intraclass variance of the pixels for the removal of isolated background pixels and filling of the isolated foreground pixels. In the background maintenance stage, the updation of pixel variance in spatial domain is done by removing all connected components (objects) that have fewer than P pixels, producing another image sequence and the convolution of image vectors in done in temporal domain. This is illustrated in Fig. 1.

Figure 1. *Tracking of the foreground moving object*

Section II contains the relevant research based on the literature surveyed, Section III contains the proposed technique for the construction and updating of the background model, Section IV contains the Experimental results and Section V contains the conclusion and future scope.

II. BACKGROUND RESEARCH AND CONTRIBUTIONS

A. Literature Review

Researchers (Wang & Adelson, 1994) have worked on temporal segmentation for tracking the activity while Wren et. al (Wren et al., 1997) models human as a set of associated blobs. Another researcher Heikkila et. al (Heikkila & Silven, 1999) proposed the use of Kalman filtering for the tracking of the object in motion. Lo et.al in his work (Lo & Velastin, 2001)suggested that the optimal removal of background pixels is done implementing a variance filter and focusing on this furthermore the researchers proved that the non-parametric kernel density with an optimal bandwidth provide better accuracy while extracting the background pixels (Okasha et al., 2006). The spatio-temporal techniques were further enhanced as when Elgammal et. al (Elgammal et al., 2002) introduced the kernelized GMM for the construction of a background model. Additionally many authors (Lee & Park, 2012) applied an adaptive threshold to extract dynamic background for which Subudhi et. al (Subudhi et al., 2011) proposed spatio temporal spatial segmentation with incorporation of Markov's Random Field (MRF) and a hybrid algorithm included SA as well as ICM for MAP estimation. To detect the change, wronskian model as per Durucan et.al (Durucan & Ebrahimi, 2001) proved to be less dependent on thresholding while Fazli et. al (Fazli et al., 2009a) stated that background reconstruction can be done with adaptive Gaussian mixture model incorporating neighborhood difference and overlapping based classification. Barnich (Barnich & Van Droogenbroeck, 2011) proposed a non-parametric model for the extraction of background initialized by a single frame. Furthermore, He et. al suggested improvisation of the ViBe (He et al., 2019) while another seminal group of workers proposed integration of Gaussian distribution for multi modal background construction and Wronskian (Subudhi et al., 2013)with ratio of current versus background frame and Panda et. al (Panda & Meher, 2013) proposed the same technique with ratio of background to current frame resulting in accurate shape detection. Another improvisation (Subudhi et al., 2013) suggested the use of linearly dependent past images and some other researchers (Subudhi et al., 2016) assigned labels based on the majority voting. Foreground extraction based on the outliers in the Gaussian distribution as well as morphological processing was proposed by Mahajan et. al (Mahajan, 2020) (Mahajan & Padha, 2020b). Chan et. al, 2018 (Chan, 2018a) suggested the background construction as well as updation in the spatio-temporal variance. Furthermore, Chan et.al, 2018 (Chan, 2018b) recommended the background initialization via calculation of LTP with Photometric features and Chen et. al (Chen et al., 2018) with Panda et.al (Panda & Meher, 2018) proposed spatio-temporal background subtraction. Authors Darwich et. al (Darwich et al., 2018) recommended fuzzification of Gaussian Mixture Construction. Naidoo et.al (Naidoo et al., 2018) uses the spatio-temporal estimation of the video sequences. Lu et. al(Lu et al., 2018) proposed improved Background subtraction and (Zheng et al., 2019) exploited low rank factorization model. Rout et. al (Rout et al., 2017) proposed the 5-frame differencing based background construction and improvised the algorithm further with GMM (Rout et al., 2018). Moreover, Mahajan et. al (Mahajan & Padha, 2020a) performed the background subtraction through thresholding and Background maintenance via selective adaptive maintenance algorithm. Lu et. al (Lu & Xu, 2018) proposed foreground detection via semi soft thresholding function for denoising of the moving object frame sequences with the wavelet based denoising which work for the horizontal, vertical, diagonal and approximate scales but not for the angular change in orientation.

To the best of the author's knowledge, the researchers still find it difficult to make a distinction between foreground and background models due to the minute difference.

B. Contributions of Research

The proposed technique has following contributions taking in consideration the research gaps of the above background literature:

- Background initialization via iteratively using EM for Gaussian mixture based multimodal background construction results in better accuracy for dynamic background.
- For Foreground Extraction, Curvelet based denoising process is proposed for adaptive normalization of the intensities values to minimize the intraclass variance of the pixels leading to removal of shadows.
- Background is updated by updation of pixel variance in spatial domain and the convolution of image vectors for handling scale and orientation.

III. PROPOSED TECHNIQUE

The proposed technique implement improvised Background Subtraction based on Spatiotemporal GMM with EM algorithm for local change detection. This promotes robust background construction and generation of a stable silhouette for the body motion to evaluate the activity.

A. Background Initialization

Background construction is the initial process for the extraction of moving objects. The background model is based on the K Gaussian distributions which outline the intensity value cluster based on mean and variance, each time a frame is processed. When the captured new frame sequence is processed based on probability distribution function of GMM, the pixels in the existing guassian distribution are compared with the overlapping intensity values of pixels of the new frame. If the new pixel matches, it is added to background model construction. To handle the complexity of the model, the iterative Expectation Minimization is used which handles the multi modal background distribution. To obtain a single gaussian background distribution, the Maximum Likelihood Estimation is done.

The initialization process begins with the calculation of the components mean and weights as per the conventional GMM model (Stauffer & Grimson, 1999). The difference of the pixel values is calculated using the frame difference method based on the standard deviation rather than mean. The guassian component for each of the pixel is computed and if the pixel matches the component, it is considered a match for the foreground else the weights, mean, standard deviation and learning rate are updated. The probabilistic distribution of the k gaussian component is calculated and matched with the incoming pixels which can be illustrated as the following set of equation (1) to (10). The initial set of equation from (1) to (6) depict the equations for conventional GMM model.

$$\left|\left(\mu_{diff\left(i,j,k\right)}\right)\right| \leq D^{*}\sigma_{\left(i,j,k\right)} \tag{1}$$

$$\mu_{diff\left(i,j,k\right)} = \mu_{t\left(i,j,k\right)} - fr_{bw_{t\left(i,j,k\right)}} \tag{2}$$

$$w_{t_{i,j,k}} = \left(1-\alpha\right)^{*} w_{t_{i,j,k}} + \alpha; \tag{3}$$

$$\mathrm{p} = \frac{\alpha}{\mathrm{w}_{\mathrm{t}\left(\mathrm{i,j,k}\right)}}; \tag{4}$$

$$\mu_{t\left(\mathrm{i,j,k}\right)} = \left(1-\mathrm{p}\right)^{*}\mu_{t\left(\mathrm{i,j,k}\right)} + \mathrm{p}^{*}\left(\mathrm{fr}_{\mathrm{bw}_{\mathrm{t}\left(\mathrm{i,j,k}\right)}}\right); \tag{5}$$

$$\tilde{\mathrm{A}}_{\mathrm{t}\left(\mathrm{i,j,k}\right)} = \sqrt{\left(1-\mathrm{p}\right)^{*}\left(\tilde{\mathrm{A}}_{\mathrm{t}\left(\mathrm{i,j,k}\right)}\right)^{2} + \mathrm{p}^{*}\left(\left(\left(\mathrm{fr}_{\mathrm{bw}_{\mathrm{t}\left(\mathrm{i,j,k}\right)}}\right)-\mu_{t\left(\mathrm{i,j,k}\right)}\right)\right)^{2}}; \tag{6}$$

In equation (1) the $\mu_{\mathrm{diff(i,j,k)}}$ is the difference of each pixel intensity in the RGB 3D matrix of the frame sequences from the mean k^{th} gaussian distribution model-based pixel intensity value represented by $\mu_{t\ \mathrm{(i,j,k)}}$, D is the positive deviation threshold and $\sigma_{\mathrm{t\ (i,j,k)}}$ represent the standard deviation of the pixels of the current frame sequence. In equation (2) $fr_{bw\_t\ (i,j,k)}$ represents the intensity of the image pixels and in equation (5) $\mu_{t\ (i,j,k)}$ represents the mean pixel intensities of the frame sequences with i as number of rows, j as number of columns and k as number of the guassian distribution while in equation (6) $\tilde{A}_{t(I,j,k)}$ represents the standard deviation of the intensities .

Background initialization includes initialization of weight, mean, variance of each Gaussian component. It is carried out with commonly used expectation maximization (EM) algorithm using pure (no foreground object present) training sequence. Based on the equation (1) if pixel is a hit for the kth gaussian distribution model then the mean $\mu_{t\ (i,j,k)}$ and standard deviation $\sigma_{t\ (i,j,k)}$ are updated for the current pixel intensities. Here in equation (3), $w_{(i,j,k)}$ is the weights associated with the guassian components, α represents the learning rate and p represents the updation rate for the background initialization where p is obtained in equation (4). But if it is a miss, then the weight is slightly decreased based on the equation (9) and (10).

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of K Gaussian component densities. The probability that the pixel has value at time t is represented in the the equation (7).

$$P_t\left(x_{ti,j,k}\right) = \sum_{k=1}^{3}\left(\mathrm{w}_{t_{(i,\ j,k)}} * f\left(x_{i,j,k} \mid \mu_{t(i,j,k),} \textstyle\sum_{t_{i,j,k}}\left(x\right)\right)\right) \tag{7}$$

$$f\left(x_{i,j,k} \mid \mu_{t(i,j,k),} \textstyle\sum_{t_{i,j,k}}\left(\ x\right)\right) = \dfrac{1}{\left(2\pi\right)^{\frac{i*j}{2}} * \left|\sum_{t_{i,j,k}}\left(\right)\right|^{\frac{1}{2}}} * \left(e^{-\left(\frac{1}{2}\right)*\left(\left(x_{i,j,k} - \mu_{t(i,j,k),}\right)^T\right)*\left(\sum_{t_{i,j,k}}^{-1} x_{i,j,k} - \mu_{t(i,j,k),}\right)}\right) \tag{8}$$

Where $\mathrm{w}_{t_{(I,\ j,k)}}, \mu_{t(I,j,k),}$ and $\sum_{t_{i,j,k}}\left(x\right)$ represents the weight, mean values and the covariance matrix for the pixel intensities at (i,j) of the k^{th} guassian distribution in the mixture model at time t and f is the guassian probability density function represented by the equation (8). There are several variants on the GMM shown in Equation (8). The covariance matrices $\sum_{t_{i,j,k}}\left(\right)$ constrained to be diagonal (Fazli et al., 2009b) . Additionally, parameters can be shared, or tied, among the Gaussian components, such as having a common covariance matrix for all components. The choice of model configuration (number of components, full or diagonal covariance matrices) is often determined by amount of data available for estimating GMM parameters and how GMM is used in a particular application.

1) Expectation Maximization Algorithm

EM algorithm is commonly used method to initialize various parameters of GMM such as weight mean, variance of each component. This algorithm is an iterative algorithm that starts with initial estimate of λ ($w_{t(i,j,k)}, \theta_{t_{i,j,k}}$) where $\theta_{t_{i,j,k}} = \sum_{t_{i,j,k}}\left(x_{i,j,k}\right), \mu_{t(I,j,k),}$ and iteratively updates until convergence is detected. In this the $w_{t(i,j,k)}$ represents the weights, $\sum_{t_{i,j,k}}\left(x_{i,j,k}\right)$ represents the co variance matrix for the pixels and $\mu_{t(I,j,k),}$ represents the mean intensities at time t. To handle this, the improved Expectation-Maximization approximation (Stückler & Behnke, 2015) is implemented with base model as GMM grounded on equation (7) and (8) by estimating the distribution of the pixel values for multi modal background while maximizing the joint distribution to obtain optimal value of pixel intensities.

EXPECTATION STEP:

The process is based on calculating the current parameter $\theta_{t_{i,j,k}} = \sum_{t_{i,j,k}} \left(x_{i,j,k}\right), \mu_{t(I,j,k),}$ and membership weights are estimated for the matrix of the frame sequences in the pixel intensities denoted by $w_{t(i,j,k)}$. The weight of distribution K at time t is updated as follows:

$$w_{t(i,j,k)} = \left(1-\alpha\right) * w_{t(i,j,k)} + \alpha * \mu\left(\frac{\left(w_{t(i,j,k)} | x_{ti,j,k}\right)}{t}\right); \text{if } t > \left(\frac{1}{\alpha}\right) \tag{9}$$

$$w_{t(i,j,k)} = \sum_{1}^{t} \mu\left(\frac{\left(w_{t(i,j,k)} | x_{ti,j,k}\right)}{t}\right); \text{ otherwise} \tag{10}$$

Where $\mu\left(\frac{\left(w_{t(i,j,k)} | x_{ti,j,k}\right)}{t}\right) = \begin{cases} 1; & if\, the\, pixel\, matches\, the\, distribution \\ 0; & otherwise \end{cases}$

MAXIMIZATION STEP:

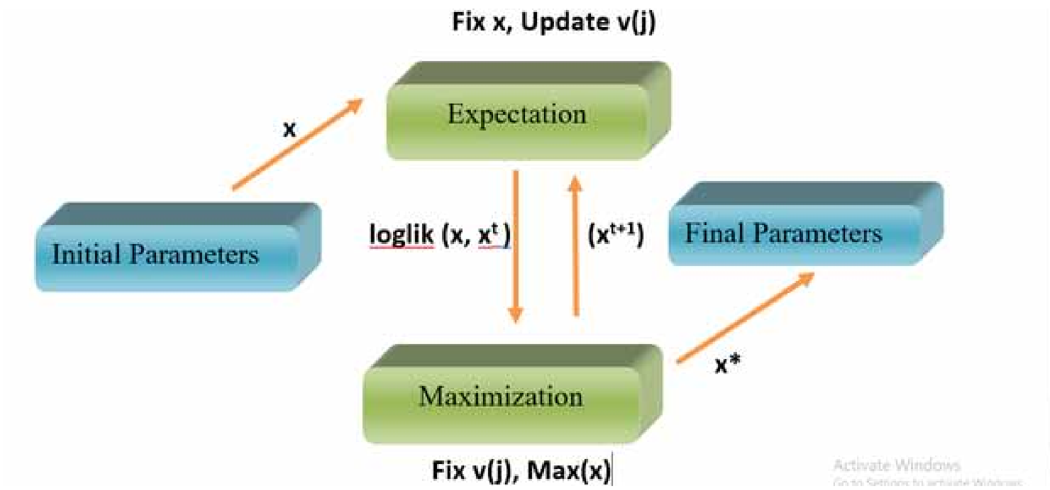
If the pixel value is within 2.5 standard deviations of one Gaussian distribution, a match is found. If this pixel value does not fit into any one of the K distributions, the distribution with the least weight is replaced by a new distribution with the current pixel value as its mean, an initially high variance, and a low weight. So, the values are now changed to 9 and 10. After computation of all new parameters, M-step is complete and it is followed by re-computation of membership weights in the E-step, then re-computation of parameters again in E-step, and parameters are updated continuously in this manner. Each pair of E and M step is considered to be one iteration.

2) Convergence of Expectation Maximization Algorithm

The EM algorithm (Fazli et al., 2009a) can be started by either initializing the algorithm with a set of initial parameters and then conducting an E-step, or by starting with a set of initial weights and then doing a first M-step. The initial parameters or weights can be chosen randomly or could be chosen via some heuristic method (such as by using k-means algorithm to cluster data first and then defining weights based on k-means memberships). Convergence is generally detected by computing value of log-likelihood after each iteration and halting when it appears not to be changing in a significant manner from one iteration to next. The log likelihood of GMM is calculated in equation 5.10 and represented in Fig. 5.3 but due to its complexity the kth mean is set to 0 which makes it an unsolvable problem. The mixture model updating is processed by checking each new pixel X against the existing K Gaussian distributions.

$$\log P_t \left(x_{ti,j,k} | \sum_{t_{i,j,k}} \left(x_{i,j,k} \right),\ \mu_{t(\mathrm{I,j,k}),} \right) = \sum_{i=1}^{N} \log \sum_{k=1}^{3} \left(\mathrm{w}_{t_{(\mathrm{i,\ j,k})}} f \left(x_{i,j,k} \mid \mu_{t(\mathrm{i,j,k}),}\ \sum_{t_{i,j,k}} \left(\right) \right) \right) \tag{11}$$

Figure 2. *Training cycle of EM based GMM*



The log likelihood of GMM is calculated as in Fig. 3 but due to its complexity the k^{th} mean is set to 0 which makes it an unsolvable problem. To handle this, the Expectation-Maximization step is defined (Stückler & Behnke, 2015) and is incorporated with GMM to estimate the distribution of the pixel values for multi modal background and maximize the joint distribution to obtain optimal value of pixel intensities.

B. Foreground Extraction

For the extraction of foreground from the voxels, the initial denoising of the frame sequences is done based on curvelet denoising process based on Bayes' Shrink Method [40].Then the soft thresholding is performed for the frame sequences by taking the median value and computing the single-level reconstructed approximation. The pixel matrix is scaled by the standard deviation and maps the values in intensity values in the registered frame sequence to the new frame sequence such that 1% of data is saturated at low and high intensities of the registered image. This increases the contrast of the background as well as foreground model and computes a global threshold that can be used to convert an intensity image to a binary image. The improvisation is done by integrating the feature interest points in the foreground moving object. The foreground extraction is based on the denoising process for the extraction of the foreground features and then thresholding the pixel values using Bayes' Shrink Method. The curvelet based denoising extracts the relevant curvelet based features to extract the movement of the object in varying scales and orientation.

C. Background Maintenance

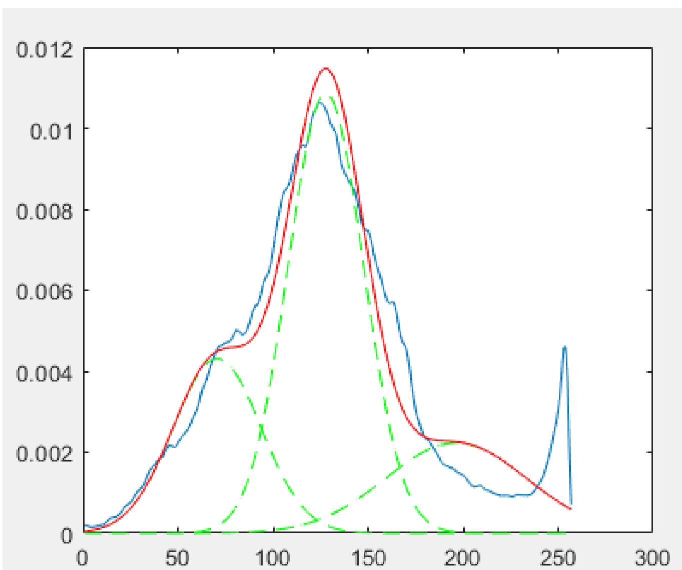
The maintenance and updation of the background model is done based on measures a set of properties for each connected component and removing pixels on the boundaries of the moving objects without allowing objects to break apart. Morphological operations with disk approximations are applied iteratively on the sequence matrix for the current frame sequence until the image no longer changes. This also requires setting a pixel to 0 if its 4-connected neighbors are all 1's, thus leaving only boundary pixels. This preserved the nature of the foreground and maintains the RGB intensities of the background pixels as illustrated in Figure 4.

IV. EXPERIMENTAL RESULTS

The experimental evaluation for the background subtraction model suggested that the proposed algorithm can deal with numerous dynamic conditions with swaying tees, moving water, flowing river, fountain and so on. The results include the experimental setup, quantitative analysis using frame sequences from three different datasets (CAVIER, PETS 2004 and ACTIVITY database) and qualitative analysis via calculating the segmentation scores for the proposed algorithm compared to the state-of-the-art technique.

A. Experimental Setup

Figure 3. *Preservation of background pixel intensities where* **X**-*axis represent the number of pixels in the frame sequences and Y-axis represents the thresholding value causing least luminance variance in the intensity values*



The database used is the well-known PETS 2004, CAVIER DATABASE and the Activity Database for background subtraction via adaptive foreground extraction and dynamic illumination conditions. The illumination condition disparity in different location and the presence of the multi modal background are the factors of misclassification of the frame sequences which may results in presence of false positives where the proposed algorithm performs well.

The calibration of camera is different for different datasets.

The machine is trained in MATLAB and the three databases are evaluated. Investigation of public spaces, recognition of doubtful activities, challenging detection/tracking scenes on water, outdoor people tracing and indoor people tracking are all included in the datasets for which the system is trained to track the body motion.

The primary features of the frame sequences are described in Table 1.

B. Quantitative Analysis

Table 1. Primary features of frame sequences

Video Title	**Properties**			
	Image Size	*Environment*	*Background*	*Illumination Condition*
Crowd	576x768x3	Outdoor	Static	Bright
Sofa	240x320x3	Indoor	Static	Dark
Fighting	256x256x3	Indoor	Static	Medium
Bus Station	240x360x3	Outdoor	Static	Bright
Canoe	240x320x3	Outdoor	Dynamic	Bright

1) Background Initialization

The background initialization process recommends the construction of a background model. Various techniques have been proposed for the construction of model and some of them have been compared with the proposed method as illustrated in Figure 4. Our system improvises the mixture of Gaussian by applying Expectation Maximization cycle for the enhancement of the background model.

Figure 4. *Performance comparison for the construction of background using various frame sequences where Column 1 represents the Crowd, Sofa, Fighting, Bus Station, Canoe frame sequences. Column 2 represents the background frame for each sequence. Column 3 represents the ground truth for each sequence. Column 4 represents the Simple Statistical Difference for each of the frame sequence. Column 5 represents the Fuzzy Background subtraction. Column 6 represents the non-parametric background subtraction. Column 7 represents the Mixture of gaussian, and Column 8 represents the proposed technique.*

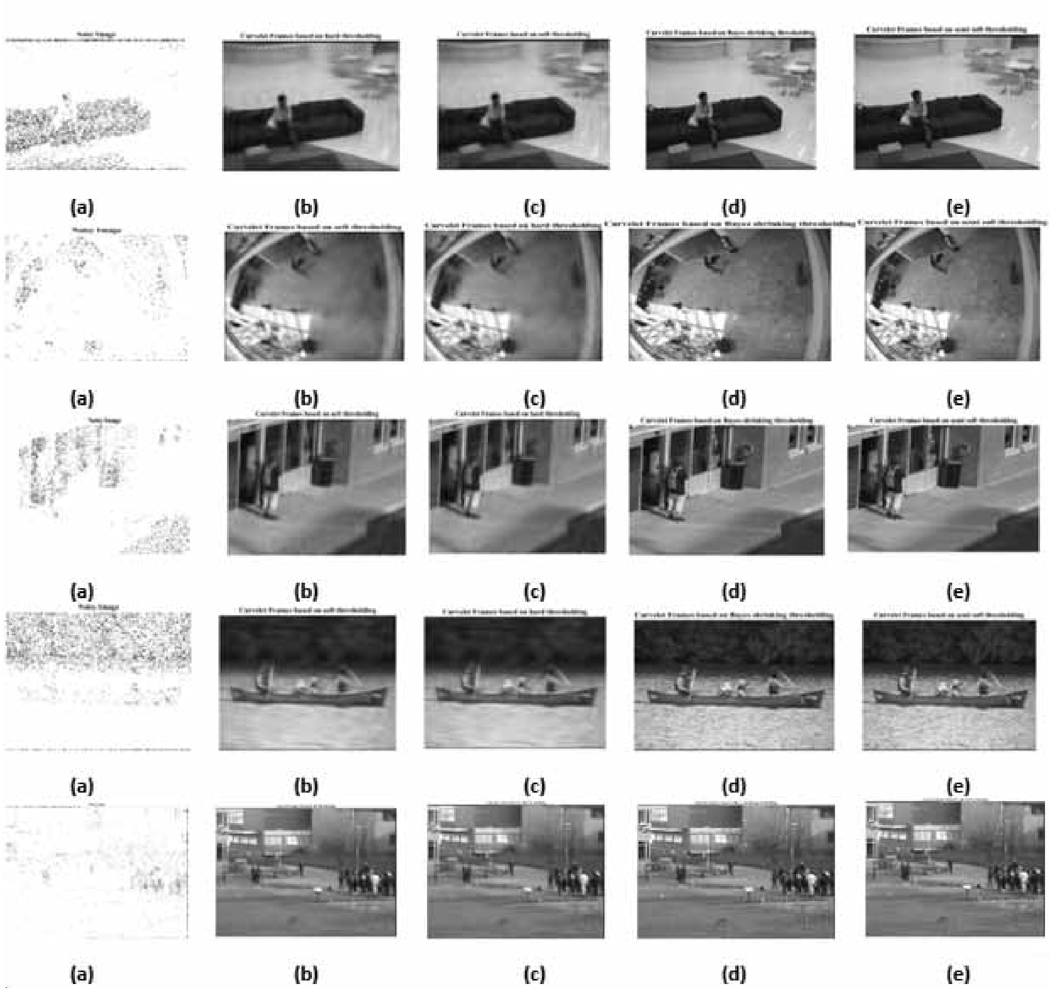
FOREGROUND EXTRACTION

The foreground extraction process is handled using the semi soft thresholding based on curvelet denoising with mapping of the pixel intensities and choose a threshold to minimize the intraclass variance. Also, spatio-temporal feature point is extracted from the frame sequences thereby removing the connected components with less than a predefined number of pixels. The process involves the encoding of local shape statistics from the foreground sections within a frame sequence extracted from the moving object video in case of a dynamic or multi modal background. The process captures large scale spatial data for the voxels in the foreground and condenses the ability for minimization of the effect of local radiance changes.

The frame sequences for accurate foreground extraction need to guarantee an intersection of at least half the frame block size to certify that an unbiased contrast normalization is performed. This is illustrated in Figure 5 and 6.

Figure 5. *Snapshots taken from the five scenes where column 1 represents the original frames, column 2 represent the anisotropic diffusion based denoising and column 3 represents the bilateral filter based denoising, column 4 represents the curvelet denoising, column 5 represents the guassian denoising with median filtering, column 6 represents wavelet based anisotropic diffusion based denoising of the foreground, column 7 represents the weiner denoising process and column 8 represents the proposed curvelet based denoising where row 1 represents the crowd in university, row 2 represents person sitting on a sofa, row 3 represents 2 people fighting, row 4 represents the person waiting on the bus station and row5 represents a moving canoe*

Figure 6. *Snapshots taken from the five scenes where columns(a) represents the noisy images, columns (b) represent the soft thresholding, column (c) represents the hard thresholding, column (d) represents the Bayes' thresholding method and columns (e) represents the proposed curvelet based semi-soft thresholding where row 1 represents person sitting on a sofa, row 2 represents 2 people fighting, row 3 represents the person waiting on the bus station, row 4 represents a moving canoe and row 5 represents the crowd in university*



2) Adaptive Background Updation

The background updation process is performed by adaptive orientation of the moving object with temporal frame sequences in consideration. The absolute mean shift difference image is generated with respect to the first background frame and the spatio temporal feature points encode local shape information from regions within an image. Then the gradient and spatial weights of the intensity values are computed to update the background model.

C. Qualitative Analysis

The background subtraction model generated the silhouette of the object in motion and detect the change by measuring the various image segmentation metrices of the system. The results are obtained by comparing the proposed background subtraction technique with the state-of-the-art techniques using segmentation parameters of accuracy, sensitivity, precision, F-measure and Mathew's Correlation coefficient. This is depicted in Table 2, 3, 4, 5 and 6 with respective graphs in the figures 7,8,9,10,11.

Table 2. Value of Image Segmentation Scores for various techniques in Frame sequence Sofa

Frame sequence	Different state-of-the-art techniques with proposed technique for SOFA				
Parameter	*SSD*	*NP*	*Fuzzy BS*	*GMM*	*Proposed*
Accuracy	44.32	47.01	45.07	89.68	92.39
Sensitivity	9.44	17.85	10.88	100	87.44
F measure	17.04	17.85	19.35	92.15	93.30
Precision	87.39	76.99	87.44	85.44	100
MCC	14.52	13.58	15.69	79.42	85.62

Table 3. Value of Image Segmentation Scores for various techniques in Frame sequence Canoe

Frame sequence	Different state-of-the-art techniques with proposed technique for CANOE				
Parameter	*SSD*	*NP*	*Fuzzy BS*	*GMM*	*Proposed*
Accuracy	85.1	86.85	85.06	88.55	98.33
Precision	92.98	88.73	91.79	79.42	94.56
Sensitivity	71.71	80.47	72.68	97.89	96.23
F measure	80.97	84.39	81.13	88.53	98.08
MCC	70.64	73.34	70.32	79.45	96.66

Table 4. Value of Image Segmentation Scores for various techniques in Frame sequence Fighting

Frame sequence	Different state-of-the-art techniques with proposed technique for FIGHTING				
Parameter	*SSD*	*NP*	*Fuzzy BS*	*GMM*	*Proposed*
Accuracy	87.096	86.19	86.76	53.52	88.54
Sensitivity	59.694	63.01	60.44	98.77	98.5
F measure	69.672	69.39	69.39	51.65	77.44
Precision	83.655	77.21	81.47	34.82	90.18
MCC	63.184	61.12	62.29	36.45	71.44

Table 5. Value of Image Segmentation Scores for various techniques in Frame sequence Crowd

Frame sequence	Different state-of-the-art techniques with proposed technique for CROWD				
Parameter	***SSD***	***NP***	***Fuzzy BS***	***GMM***	***Proposed***
Accuracy	74.32	75.53	74.88	46.76	96.55
Sensitivity	17.55	25.16	19.86	97.65	97.99
F measure	29.44	38.56	32.55	53.42	94.65
Precision	91.19	82.52	90.09	36.44	89.84
MCC	32.92	36.17	34.74	29.18	92.45

Table 6. Value of Image Segmentation Scores for various techniques in Frame sequence Bus Station

Frame sequence	Different state-of-the-art techniques with proposed technique for BUS STATION				
Parameter	***SSD***	***NP***	***Fuzzy BS***	***GMM***	***Proposed***
Accuracy	47.11	47.44	47.7	89.73	92.99
Sensitivity	12.79	16.21	14.16	99.78	91.48
F measure	22.38	26.89	24.42	92.06	93.37
Precision	89.82	78.96	88.49	85.44	90.78
MCC	18.73	14.71	19.1	79.66	80.91

Figure 7. *Value of segmentation scores for various techniques in frame sequence Sofa depicting the presence of an intermittent object*

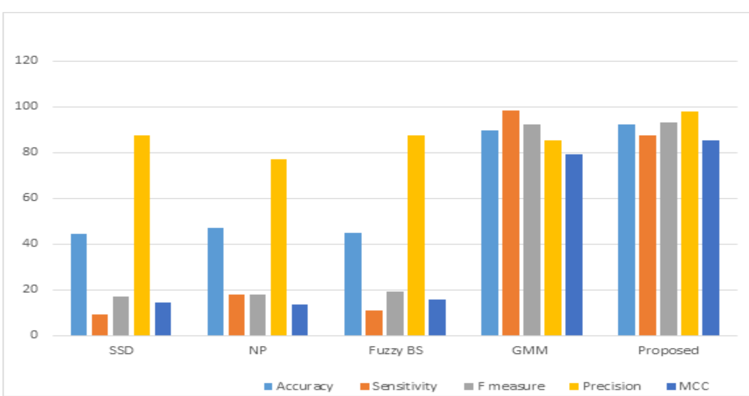


Figure 8. *Value of segmentation scores for various techniques in frame sequence Fighting depicting the presence of changing illumination conditions*

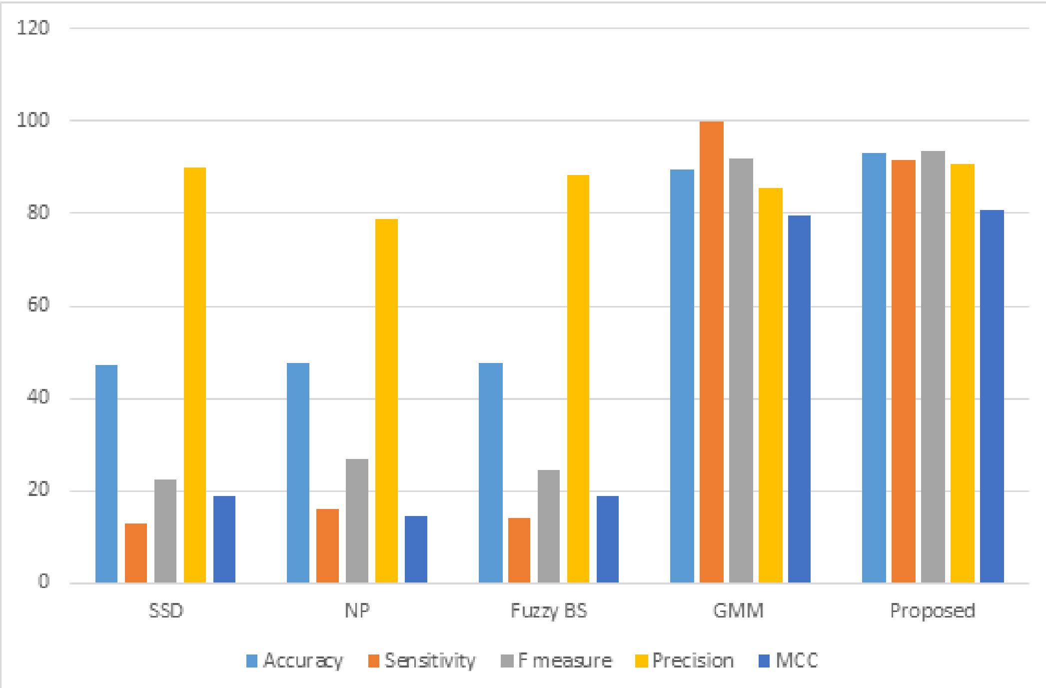


Figure 9. *Value of segmentation scores for various techniques in frame sequence Canoe depicting the presence of Dynamic Backgrounds*

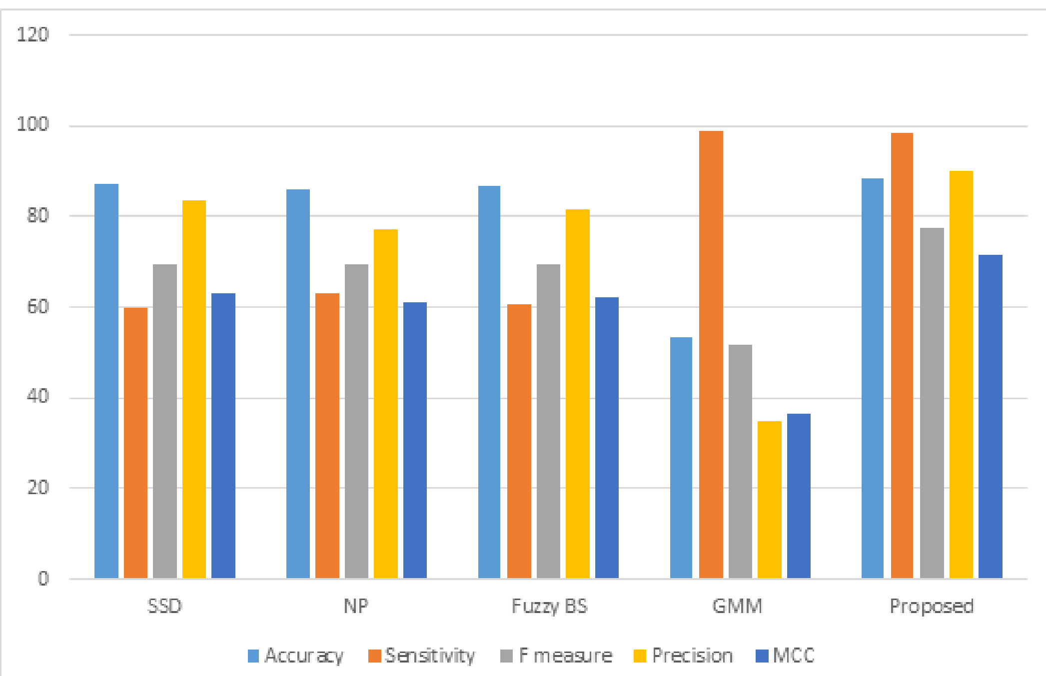


Figure 10. *Value of segmentation scores for various techniques in frame sequence Crowd depicting the presence of occlusion*

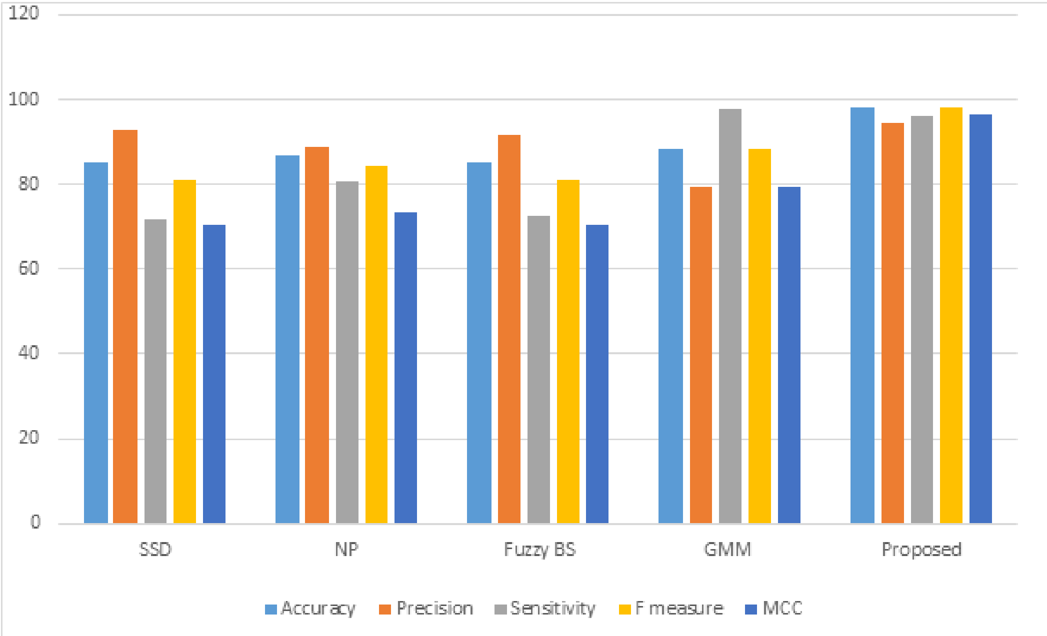
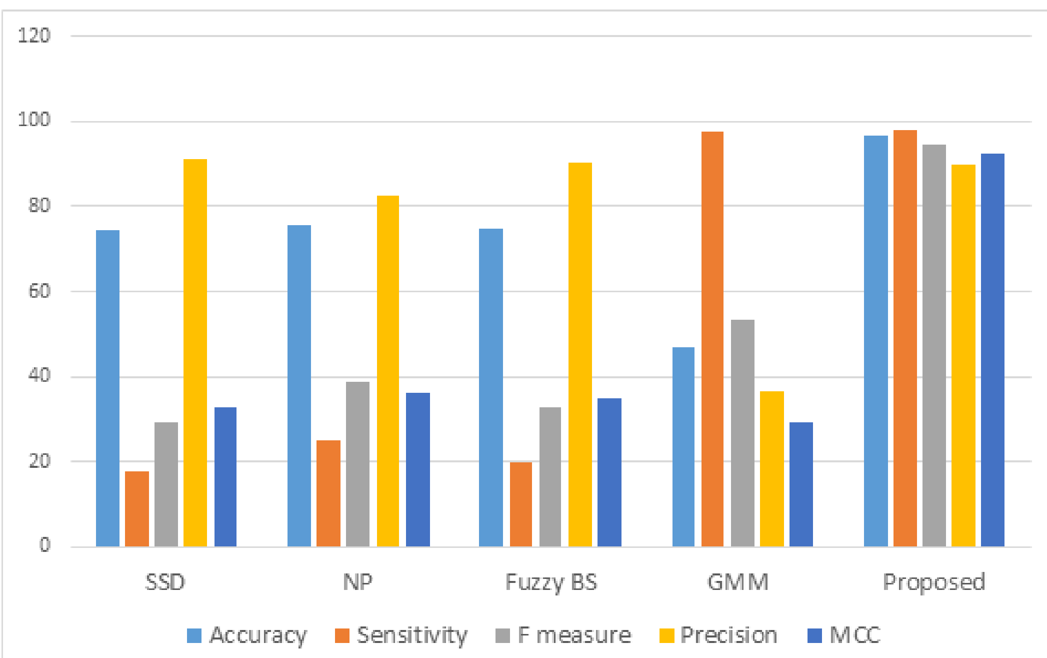


Figure 11. Value of segmentation scores for various techniques in frame sequence Bus Station depicting the presence of Shadows



V. CONCLUSION

A novel technique for the background construction based on modified GMM with Expectation Maximization Cycle for the iterative handling of multi modal background model is proposed. The foreground extraction is proposed by using the curvelet filtering based semi soft thresholding of the frame sequences leading to saturation of the model and integration of feature interest points. Finally, the background maintenance is done by setting a pixel to 0 if its 4-connected neighbors are all 1's, thus leaving only boundary pixels with the implementation of morphological operations. The experimental results proved that the accuracy of the algorithm is enhanced than the GMM and the complexity is reduced. The algorithm overcomes the research gaps of various state of the art techniques but still cannot handle the occlusions and presence of intermittent objects in the frame sequences.

Future scope of the work suggests two stream CNN based deep learning algorithms for the extraction of the foreground features which are based on spatio temporal extraction. Another aspect is to deal with the tracking of an object in motion with moving cameras or through GPS.

FUNDING AGENCY

Publisher has waived the Open Access publishing fee.

REFERENCES

Anandhalli, M., & Baligar, V. P. (2015). Improvised approach using background subtraction for vehicle detection. *Souvenir of the 2015 IEEE International Advance Computing Conference, IACC 2015,* 303–308. doi:10.1109/IADCC.2015.7154719

Barnich, O., & Van Droogenbroeck, M. (2011). ViBe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, *20*(6), 1709–1724. doi:10.1109/TIP.2010.2101613 PMID:21189241

Chan, K. L. (2018a). Detection of change in video based on local pattern and photometric features. *Computer Science Research Notes*, *2802*(May), 174–182. doi:10.24132/CSRN.2018.2802.22

Chan, K. L. (2018b). Segmentation of moving objects in image sequence based on perceptual similarity of local texture and photometric features. *Eurasip Journal on Image and Video Processing, 62*(1), 1–16. 10.1186/s13640-018-0308-4

Chen, M., Wei, X., Yang, Q., Li, Q., Wang, G., & Yang, M. H. (2018). Spatiotemporal GMM for Background Subtraction with Superpixel Hierarchy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(6), 1518–1525. doi:10.1109/TPAMI.2017.2717828 PMID:28644797

Chiu, Y. C., Liu, L. Y., & Wang, T. (2018). Automatic segmentation and summarization for videos taken with smart glasses. *Multimedia Tools and Applications*, *77*(10), 12679–12699. doi:10.1007/s11042-017-4910-8

Darwich, A., Hébert, P. A., Bigand, A., & Mohanna, Y. (2018). Background subtraction based on a new fuzzy mixture of Gaussians for moving object detection. *Journal of Imaging*, *4*(7), 1–28. doi:10.3390/jimaging4070092

Durucan, E., & Ebrahimi, T. (2001). Change detection and background extraction by linear algebra. *Proceedings of the IEEE*, *89*(10), 1368–1381. doi:10.1109/5.959336

Elgammal, A., Duraiswami, R., Harwood, D., & Davis, L. S. (2002). Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceedings of the IEEE*, *90*(7), 1151–1162. doi:10.1109/JPROC.2002.801448

Fazli, S., Pour, H. M., & Bouzari, H. (2009a). A novel GMM-based motion segmentation method for complex background. *2009 5th IEEE GCC Conference and Exhibition, GCC 2009,* 1–5. doi:10.1109/IEEEGCC.2009.5734290

Fazli, S., Pour, H. M., & Bouzari, H. (2009b). Multiple object tracking using improved GMM-based motion segmentation. *2009 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 1*, 1130–1133. doi:10.1109/ECTICON.2009.5137243

Fuentes, A., Yoon, S., & Park, D. S. (2018). Spatial multilevel optical flow architecture-based dynamic motion estimation in vehicular traffic scenarios. *Transactions on Internet and Information Systems (Seoul)*, *12*(12), 5978–5999. doi:10.3837/tiis.2018.12.020

He, W., Kim, Y. K. W., Ko, H. L., Wu, J., Li, W., & Tu, B. (2019). Local Compact Binary Count Based Nonparametric Background Modeling for Foreground Detection in Dynamic Scenes. *IEEE Access, 7*, 92329–92340. 10.1109/ACCESS.2019.2927745

Heikkila, J., & Silven, O. (1999). A real-time system for monitoring of cyclists and pedestrians. *Proceedings - 2nd IEEE International Workshop on Visual Surveillance, VS 1999*, 74–81. doi:10.1109/VS.1999.780271

Kim, W., & Jung, C. (2017). Illumination-Invariant Background Subtraction: Comparative Review, Models, and Prospects. *IEEE Access, 5*, 8369–8384. 10.1109/ACCESS.2017.2699227

Kumar, S., & Sen Yadav, J. (2016a). Segmentation of moving objects using background subtraction method in complex environments. *Wuxiandian Gongcheng*, *25*(2), 399–408. doi:10.13164/re.2016.0399

Kumar, S., & Sen Yadav, J. (2016b). Video object extraction and its tracking using background subtraction in complex environments. *Perspectives in Science*, *8*, 317–322. doi:10.1016/j.pisc.2016.04.064

Lee, J., & Park, M. (2012). An adaptive background subtraction method based on kernel density estimation. *Sensors (Switzerland)*, *12*(9), 12279–12300. doi:10.3390/s120912279

Lo, B. P. L., & Velastin, S. A. (2001). Automatic congestion detection system for underground platforms. *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, ISIMP 2001*, 158–161. doi:10.1109/ISIMP.2001.925356

Lu, X., & Xu, C. (2018). Novel Gaussian mixture model background subtraction method for detecting moving objects. *Proceedings of 2018 IEEE International Conference of Safety Produce Informatization, IICSPI 2018*, 6–10. doi:10.1109/IICSPI.2018.8690428

Lu, X., Xu, C., Wang, L., & Teng, L. (2018). Improved background subtraction method for detecting moving objects based on GMM. *IEEJ Transactions on Electrical and Electronic Engineering*, *13*(11), 1540–1550. doi:10.1002/tee.22718

Mahajan, R. (2020). Pixel-Wise Adaptive Lbp and Glcm Based Feature Extraction for Human Motion Estimation. *Journal of Critical Reviews*, *7*(9), 2344–2349.

Mahajan, R., & Padha, D. (2018). Human detection and motion tracking using machine learning techniques: A review. *PDGC 2018 - 2018 5th International Conference on Parallel, Distributed and Grid Computing*, 127–131. doi:10.1109/PDGC.2018.8745852

Mahajan, R., & Padha, D. (2020a). Multiresolution Wavelet Transform Based Anisotropic Diffusion for Removing Speckle Noise in a Real-Time Vision-Based Database. *International Journal of Multimedia Data Engineering and Management*, *11*(1), 1–14. doi:10.4018/IJMDEM.2020010101

Mahajan, R., & Padha, D. (2020b). Multi resolution Adaptive Threshold based Segmentation of real-time vision-based database for human motion estimation. *Rising Threats in Expert Applications and Solutions Proceedings of FICR-TEAS*, 1–9.

Naidoo, D., Tapamo, J. R., & Walingo, T. (2018). Human Action Recognition using Spatial-Temporal Analysis and Bag of Visual Words. *Proceedings - 14th International Conference on Signal Image Technology and Internet Based Systems, SITIS 2018,* 697–702. doi:10.1109/SITIS.2018.00112

Okasha, M., Elatrash, M., & Iqelan, M. (2006). Background Subtraction Using Nonparametric Kernel Density Estimation With Optimal Bandwidth for Visual Surveillance. *IEEE Access: Practical Innovations, Open Solutions*, 1–17.

Panda, D. K., & Meher, S. (2013). Video object segmentation based on adaptive background and Wronskian change detection model. *2013 International Conference on Signal Processing and Communication, ICSC 2013*, 219–224. doi:10.1109/ICSPCom.2013.6719786

Panda, D. K., & Meher, S. (2018). Adaptive spatio-temporal background subtraction using improved Wronskian change detection scheme in Gaussian mixture model framework. *IET Image Processing*, *12*(10), 1832–1843. doi:10.1049/iet-ipr.2017.0595

Rout, D. K., Bhat, P. G., Veerakumar, T., Subudhi, B. N., & Chaudhury, S. (2017). A novel five-frame difference scheme for local change detection in underwater video. *4th International Conference on Image Information Processing*, *17*, 290–295. doi:10.1109/ICIIP.2017.8313727

Rout, D. K., Subudhi, B. N., Veerakumar, T., & Chaudhury, S. (2018). Spatio-contextual Gaussian mixture model for local change detection in underwater video. *Expert Systems with Applications*, *97*, 117–136. doi:10.1016/j.eswa.2017.12.009

Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. *Proceedings - IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2*, 246–252. doi:10.1109/CVPR.1999.784637

Stückler, J., & Behnke, S. (2015). Efficient Dense Rigid-Body Motion Segmentation and Estimation in RGB-D Video. *International Journal of Computer Vision*, *113*(3), 233–245. doi:10.1007/s11263-014-0796-3

Subudhi, B. N., Ghosh, S., & Ghosh, A. (2013). Moving object detection using Gaussian background model and Wronskian framework. *Proceedings of the 2013 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2013*, 1775–1780. doi:10.1109/ICACCI.2013.6637450

Subudhi, B. N., Ghosh, S., Shiu, S. C. K., & Ghosh, A. (2016). Statistical feature bag based background subtraction for local change detection. *Information Sciences, 366*, 31–47. 10.1016/j.ins.2016.04.049

Subudhi, B. N., Nanda, P. K., & Ghosh, A. (2011). A change information based fast algorithm for video object detection and tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, *21*(7), 993–1004. doi:10.1109/TCSVT.2011.2133870

Wang, J. Y. A., & Adelson, E. H. (1994). Spatio-Temporal Segmentation of Video Data. *Proceedings Ofthe SPIE: Image AndVideo Processing II*, *2182*(262), 1–12.

Wren, C. R., Azarbayejani, A., Darrell, T., & Pentland, A. P. (1997). P finder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 780–785. doi:10.1109/34.598236

Zheng, A., Zou, T., Zhao, Y., Jiang, B., Tang, J., & Li, C. (2019). Background subtraction with multi-scale structured low-rank and sparse factorization. *Neurocomputing*, *328*, 113–121. doi:10.1016/j.neucom.2018.02.101

Rohini Mahajan is a Ph.D. Scholar in Department of Computer Science & IT in Central University of Jammu, Jammu.

Devanand Padha is a Professor in the Department of Computer Science & IT and Dean of School and Basic Applied Sciences at the Central University of Jammu, Jammu.