

Anisotropic Sophisticated Spatio-Temporal Contours Based Deep Neural Learned Moving Objects Detection in Video



K.N. Abdul Kader, Nihal

Abstract: Object detection in the video sequence is a significant problem to be resolved in image processing because it used different applications in video compression, video surveillance, robot technology, etc. Few research works have been designed in conventional works to discover moving objects using various machine learning techniques. However, dynamic changing background, object size variations and degradation of video frames during the object detection process remained an open issue. In order to overcome such limitations, Anisotropic Sophisticated Spatiotemporal Contours based Deep Neural Network Learning (ASSC-DNNL) practice is projected. ASSC-DNL Technique initially obtains a number of video file as input at the input layer. After acquiring the video, input layer forward it to hidden layers. Subsequently, ASSC-DNL Technique accomplishes the encoding process in the first hidden layer using Anisotropic Stacked Autoencoder (ASA). During the encoding process, ASSC-DNL practice maps each video frames pixels in input video via code. This practice results in compressed video with enhanced quality. Afterward, ASSC-DNL practice transforms compressed video into a numeral of frames in the second concealed layer. Followed by, ASSC-DNL practice carried out Teknomo–Fernandez Spatiotemporal Based Background Subtraction (TS-BS) process at the third hidden layer, in which it effectively segments the foreground images from dynamic changing background. Then, ASSC-DNL practice deep analyzes the foreground image of video frames and mines some features like shape, color, texture, intensity, and size. Finally, ASSC-DNL Technique exactly finds the moving objects in video frames according to identified features with minimal time at the output layer. Therefore, ASSC-DNL Technique obtains enhanced moving objects detection performance when compared to existing works. The simulation of ASSC-DNL practice is conducted via different metrics such as accuracy, time and false positive rate towards in detection.

Keywords: Anisotropic Stacked Autoencoder, Contours Based Object Size Estimation, Dynamic Background, Gaussian Activation Function, Video Compression, Video Frames.

Manuscript published on November 30, 2019.

* Correspondence Author

K.N. Abdul Kader*, Assistant Professor, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous), [Affiliated to Bharathidasan University], Tiruchirappalli, TN, India. (Email: akn@jmc.edu)

Nihal, Assistant Professor, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous), [Affiliated to Bharathidasan University], Tiruchirappalli, TN, India. (Email: nihal.akn@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

I. INTRODUCTION

Moving object discovery plays an imperative role in video surveillance. Object detection techniques are designed with the objective of finding objects in consecutive video frames. VSS are rising in size and ability as it is stored in dense form for reducing bandwidth and storage space. Conventional techniques decode the compressed video and utilize pixel domain methods that need a vast amount of memory space, computational complexity and also not obtain real-time concert. As a result, moving object detection in the compressed field is more viable for competent applications due to their quick delivery. During the last two decades, quantum of investigations has been proposed in moving object video detection. But, moving objects detection routine was not sufficient when considering more number of videos as input.

In addition to that, extracting background information more exactly for foreground detection is still an open challenge in case of dynamic changing background. Besides that, the false positive rate of moving object detection is still needed to be reduced by considering the loss rate caused due to degradation of video frames. Moreover, object size variation during the detection process remained unsolved. In order to address the above limitations, ASSC-DNL practice is introduced in this work with relevance of ASA, TS-BS and C-OSE and deep neural learning. Markov Random Field (MRF) model was developed in [1] to identify moving object with the help of motion vectors and quantization parameters. However, the detection accuracy of this method was poor when getting a large number of the input video. Sparse and low-rank depiction with contextual regularization (SLRC) replica was constructed in [2] for robust moving object discovery. However, the time complexity of this model was very higher.

A weighted probabilities decision method was presented in [3] to find objects from an underwater video with a sequence of frames. However, the accuracy of object detection was lower due to the degradation of video frames. A near to the ground working out moving object detection way was introduced in [4] by joint with video encoder with superior accuracy and velocity. But, false-positive rate of detection was not condensed. A simple framework was presented in [5] to identify moving objects by fusing backdrop subtraction and Convolutional Neural Networks (CNNs).



However, dynamic background during detection was not considered. A novel algorithm was designed in [6] to accurately find out moving objects in the compressed domain in bulky streams of real-time. But, the computational complexity of moving objects discovery was more.

A Variable Weighted Pipeline Filter (VWPF) was introduced in [7] to separate the true from false objects based on a spatial-temporal property with minimal false alarm and miss detection rates. However, object size variation during the detection process was not solved. A novel methodology was developed in [8] for active object discovery and monitoring systems to visually impaired community. But, the ratio of number of objects that are incorrectly detected was higher. An well-organized moving object tracking was carried out in [9] towards the application of visible image sequences under complex environment. However, object detection performance was not effective. An efficient vector-based method was introduced in [10] to discover moving object from the video. But, the amount of time required for efficient detection was more.

To address the presented issues, ASSC-DNL practice is proposed in this work. The key contribution of ASSC-DNL practice is detailed as below,

- ✓ To accurately extract background and foreground information in each video frame in case of dynamic changing background as compared to up to date works. Teknomo–Fernandez Spatiotemporal Based Background Subtraction (TS-BS) is applied in ASSC-DNL technique.

- ✓ To reduce the loss rate caused because of degradation of video frames during the compression process and thereby minimizing false-positive rate of moving object detection as compared to conventional work, Anisotropic Stacked Autoencoder is applied in ASSC-DNL practice.

- ✓ To efficiently carry out the moving objects discovery process while changing the size of the objects in consecutive video frames as compared to up to date works, Contours based Object Size Estimation (C-OSE) is applied in ASSC-DNL technique.

The respite of the paper is formulated as follows; In Section 2, ASSC-DNL practice is depicted with the aid of pictorial diagram. In Section 3, Simulation set-ups are described and experimental outcomes are depicted in Section 4. Section 5 elaborates the literature survey. Section 6 depicts the conclusion of the paper.

II. METHODOLOGY

Anisotropic Sophisticated Spatiotemporal Contours based Deep Neural Learning (ASSC-DNL) practice is submitted in order to augment the presentation of moving objects detection using video sequence with higher accuracy. The ASSC-DNL technique is deep neural networks that utilize sophisticated mathematical modeling to process input video sequence and thereby finds the moving objects. The ASSC-DNL practice is introduced by combining Anisotropic Stacked Autoencoder, Teknomo–Fernandez Spatiotemporal Based Background Subtraction (TS-BS) and Contours based Object Size Estimation concepts in deep neural learning. The ASSC-DNL practice contains an input layer (IL), hidden layers (HL), and an output layer (OL). The ASSC-DNL practice is designed to efficiently identify the moving objects when taking a large number of video file as input.

In ASSC-DNL technique, the input layer is fully linked with the output layer via adjustable, weighted links. The input layer takes a number of video file as input. Besides that, the ASSC-DNL technique uses number of hidden nodes to accurately carry out the moving objects detection process with lower time complexity. Finally, the output layer in ASSC-DNL technique produces the detection results of moving objects with enhanced accuracy. The overall diagram of ASSC-DNL practice is shown in Figure 1.

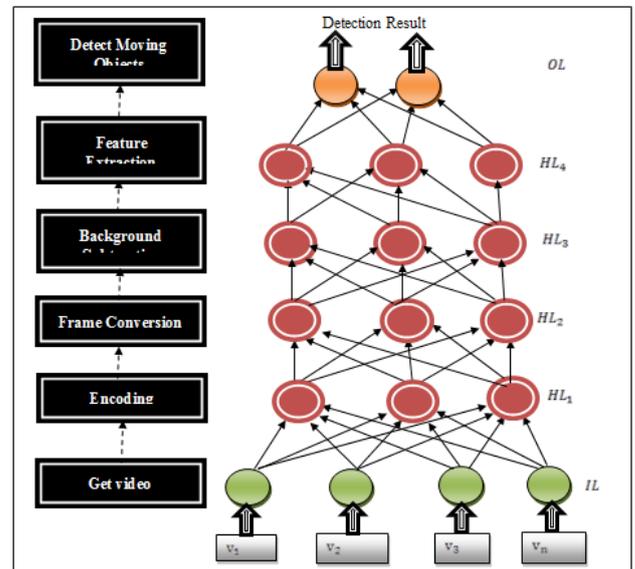


Figure 1 Architecture Diagram of ASSC-DNL Technique for Moving Objects Detection

Figure 1 shows run processes of ASSC-DNL practice for efficient moving objects discovery. As demonstrated in the above figure, input layer in ASSC-DNL practice gets a number of video file as input and then forward it to hidden layers. Then, first, hidden layer in ASSC-DNL practice performs encoding process with help of Anisotropic Stacked Autoencoder (ASA) where it maps each pixel of video frames into the code which results in compressed video. After that, second hidden layer in ASSC-DNL practice converts compressed video into a number of frames. Next, third hidden layer in ASSC-DNL practice accomplishes Teknomo–Fernandez Spatiotemporal Based Background Subtraction (TS-BS) process where it identifies the foreground and background. Followed by, fourth hidden layer in ASSC-DNL practice extracts the features such as shape, color, texture, intensity, and size from a foreground. At last, output layer accurately discovers the moving objects in video frames based on extracted features with a lower amount of time. As a result, ASSC-DNL practice achieves higher accuracy for moving objects detection compared with the predictable mechanism. The thorough processes of ASSC-DNL practice are described as;

Let us consider input dataset ‘D’ consists the number of video sequences represented as follows,

$$D = v_1, v_2, v_3, \dots, v_n \quad (1)$$

From equation (1), v_n denotes a no. of videos in dataset 'D'. Every file video contains a number of frames ' $f_1, f_2, f_3, \dots, f_n$ '. The proposed ASSC-DNL practice at first initializes the neural network with help of random weights. In ASSC-DNL practice, Input layer is created using current input video file ' $v_i(t)$ ' and weight ' a_{IH} '. Therefore, input layer is mathematically expressed as follows,

$$IL(t) = a_{IH}v_i(t) \quad (2)$$

From the above mathematical formula (2), ' $IL(t)$ ' signifies the neurons process in the input layer at time ' t '. After taking video file, input layer sent it to hidden layers. The first hidden layer ' HL_1 ' performs encoding which is significant for video applications in order to perform efficient transmission and storage. Therefore, Anisotropic Stacked Autoencoder (ASA) is applied in first hidden layer. The ASA is a compression practice which reduce the dimension of input video file via removing surplus data. On the contrary to existing works, anisotropic diffusion is combined in Stacked Denoising Autoencoder in order to increases the video compression performance and thereby avoiding degradation of video frames.

In the first hidden layer, ASA eliminates the redundant images, scenes from an input video. Therefore, ASA removes all such data to reduce the video file size. Besides that, anisotropic diffusion is applied in ASA to perform video compression with higher quality where it eliminates noise without affecting noteworthy parts of the video frame content. Thus, the proposed ASSC-DNL practice reduces degradation of video frames for performing moving objects detection with minimal false-positive rate. From that, the neuron stroke in first hidden layer ' HL_1 ' at a time ' t ' is depicted as,

$$HL_1 \leftarrow VE = \sum_{i \in H \cup I} a_{IH} v_i(t) \quad (3)$$

From the above mathematical formula (3), ' a_{IH} ' signify a weight between input and hidden layer and ' VE ' point out the encoding process of video in a first hidden layer at a time ' t '. Subsequently, the compressed video file is transmitted to second hidden layer where it accomplishes video conversion process. In ASA, encoding is described as transitions using below mathematical expression,

$$VE: v_i \rightarrow v'_i \quad (4)$$

From the above formulation (4), ' v'_i ' denotes compressed video. During the encoding process, quality of video frames are preserved with the aid of Anisotropic Diffusion using below formulation,

$$\frac{\partial v_i}{\partial t} = \text{div}(\beta(m, n, t) \nabla v_i) \quad (5)$$

$$\frac{\partial v_i}{\partial t} = \nabla \beta \cdot \nabla v_i + \beta(m, n, t) \Delta v_i \quad (6)$$

From above mathematical equation (5) and (6), ' Δ ' symbolize to the laplacian and ' ∇ ' signifies the gradient. Here, ' $\text{div}()$ ' designates the divergence operator and ' $\beta(m, n, t)$ ' denotes a diffusion coefficient that reins the rate of diffusion through the encoding procedure to conserve important parts of the video frame contents, typically edges, lines or other details. Then, the neuron action in second hidden layer ' HL_2 ' at a time ' t ' is mathematically expressed as,

$$HL_2 \leftarrow FC = f_1, f_2, f_3, \dots, f_n \in v_i \quad (7)$$

From the above mathematical representation (7), ' $f_1, f_2, f_3, \dots, f_n$ ' denote a number of video frames whereas ' FC ' indicate the video conversion process in a second hidden layer at a time ' t '. In the third hidden layer, TS-BS process find foreground from background for each input video frame. In conventional works, diverse backdrop subtraction techniques are introduced to recognize moving objects from video. However, the composite scenery of lively sight in real surveillance mission was not measured. In addition to that, computational complexity was very higher.

In order to overcome such limitations and thereby efficiently performing moving objects detection process, TS-BS Model is designed. In ASSC-DNL practice, TS-BS Model is an efficient algorithm for segmenting the background image of a given video sequence. By assuming that the background image is shown in the majority of the video, the TS-BS Model is able to produce a good background image of a video. The designed TS-BS Model increases the computational speed depending on the resolution of an image and its accuracy gained within a manageable number of frames. Furthermore, the TS-BS Model is performed for both grayscale and colored videos.

For each pixel location, the majority of pixel values in the whole video include the pixel value of the actual background image. For three frames of image sequence ' f_1 ', ' f_2 ' and ' f_3 ' are background image ' B ' is mathematically obtained as,

$$B = f_3(f_1 \oplus f_2) + f_1 f_2 \quad (8)$$

From the expression (8), three frames are selected randomly from the image series to create a background image by combining them. This yields a better background image on the contrary to existing works. In order to model backdrop for lively background scenes, the spatiotemporal model is applied in TS-BS where it initializes backdrop model from the first ' z_1 ' frames. Then, TS-BS initialize spatial model ' $S(p_i)$ ' through selecting pixel worth arbitrarily in the area of p_i for ' x ' times at each frame and ' x ' is less than 8.

$$S_1(p_i) = \{f_1(p_i), f_2(p_i), \dots, f_x(p_i)\} \quad (9)$$

$$S_2(p_i) = \{f_{x+1}(p_i), f_{x+2}(p_i), \dots, f_{2x}(p_i)\} \quad (10)$$

$$S_{z_1}(p_i) =$$

$$\{f_{(z_1-1) \times x + 1}(p_i), f_{(z_1-1) \times x + 2}(p_i), \dots, f_{z_1 \times x}(p_i)\} \quad (11)$$

Followed by, TS-BS combines the above spatial backgrounds together to generate spatiotemporal model ' $ST(p_i)$ ' using below,

$$ST(p_i) = \{S_1(p_i), S_2(p_i), \dots, S_{z_1}(p_i)\} \quad (12)$$

By using equation (11) and (12), spatial and temporal information of every frame is integrated. After that, TS-BS model compares the distance ' d ' between the present pixel ' $f(p_i)$ ' and the pixel in the backdrop ' $B(p_i)$ ' using the mathematical appearance,

$$Y(p_i) = \begin{cases} 1, & \text{if } \{d(f(p_i), ST_k(p_i)) < B(p_i)\} < \pi_{min} \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

From the equation (13), ' $ST_k(p_i)$ ' denotes the ' k^{th} ' constituent in spatiotemporal model ' $ST(p_i)$ '. Here, ' π_{min} ' represents the least no. of elements in backdrop model meeting the threshold condition.

If ‘ $Y(p_i) = 1$ ’, then TS-BS model considered that the pixel belongs to foreground. Otherwise, TS-BS model considered that the pixel belongs to background. Background changes all the time in dynamic background scenes, therefore TS-BS model update the background regularly to fit the dynamic background. From

that, the neuron activity in third hidden layer ‘ HL_3 ’ at a time ‘ t ’ is mathematically obtained as follows,

$$HL_3 \leftarrow BS = Y(p_i) \quad (14)$$

From the above mathematical depiction (14), ‘ BS ’ represent a backdrop subtraction procedure in a third hidden layer at a time ‘ t ’. After completing the backdrop subtraction process, significant features is extracted at fourth hidden layer. ASSC-DNL practice get the features i.e. shape, color, texture, intensity and size from a foreground image of video frame using below,

$$HL_4 \leftarrow OSE = Size(O_i) \quad (15)$$

From the above representation (15), ‘ HL_4 ’ denotes neuron activity in the fourth hidden layer at a time ‘ t ’ and ‘ OSE ’ signify an object size estimation process. During the feature extraction process, measuring the different object size is a considerable problem. Hence, Contours’ based Object Size Estimation (C-OSE) model is proposed. In C-OSE model uses the idea to find out the size of each object using below numerical equation,

$$Size(O_i) = \sum \sum O_i(a, b) \quad (16)$$

From the above mathematical expression (16), ‘ $Size(O_i)$ ’ denotes the measured size of each objects in video frame whereas ‘ (a, b) ’ represents the pixel position inside the object. The discovered features i.e. shape, color, texture, intensity and size in the foreground of video frame is then forward to the output layer. Accordingly, the activity of neuron ‘ i ’ can be noted with reference to time ‘ t ’,

$$OL(t) = Q(a_{HO} HL_4(t)) \quad (17)$$

From the formulation (17), ‘ $OL(t)$ ’ refers to the detection output and ‘ a_{HO} ’ indicates the load, linking the unseen and output layer and ‘ Q ’ is a Gaussian function. ASSC-DNL practice, Gaussian activation function compares a fetched feature of each video frames with pre-stored templates and thereby finds detection results. Hence, Gaussian activation function is determined mathematically as,

$$Q = \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{(f_i - \varepsilon)^2}{2v^2}} \quad (18)$$

From the above mathematical representation (18), ‘ f_i ’ refers to video frame with their extracted features such as shape, color, texture, intensity and size whereas ‘ ε ’ and ‘ μ ’ designates mean and variance value between extracted features and pre-stored templates. The output of the Gaussian activation function value in ASSC-DNL Technique is either ‘0’ or ‘1’. If the extracted features of video frame are matched with pre-stored templates, then Gaussian activation function gives result as ‘1’. Otherwise, Gaussian activation function returns ‘0’. From that, the detection result at the output layer is determined mathematically as,

$$OL(t) = \begin{cases} 1, & \text{moving objects are detected} \\ 0, & \text{moving objects are not detected} \end{cases} \quad (19)$$

From the above equation (19), ‘ $OL(t) = 1$ ’ represents that moving objects in video frames are accurately detected whereas ‘ $OL(t) = 0$ ’ indicates that moving objects are not

found. The algorithmic processes ASSC-DNL Technique is described as follows,

// Anisotropic Sophisticated Spatiotemporal Contours based Deep Neural Learning Algorithm

Input: Video Dataset ‘ $v_i = v_1, v_2, \dots, v_k$ ’, X : Termination Condition, Input Layer ‘ IL ’; Hidden Layer ‘ HL ’; Output Layer ‘ OL ’

Output: Achieve higher accuracy and minimal time complexity for moving objects detection

Step 1: Begin

Step 2: Define network with random weights

Step 3: While (‘ X ’ is attained) **do**

Step 4: For each input video ‘ v_i ’ at the input layer

Step 5: ‘ IL ’ forward input ‘ v_i ’ to hidden layers using (1)

Step 6: ‘ HL_1 ’ applies Anisotropic Stacked Autoencoder using (3)

Step 7: ‘ HL_2 ’ converts compressed video into a number of frames using (7)

Step 8: ‘ HL_3 ’ perform background subtraction using (14)

Step 9: ‘ HL_4 ’ extracts features in foreground image and sent it to OL using (15)

Step 10: ‘ OL ’ applies Gaussian activation function ‘ F ’ using (18)

Step 11: ‘ OL ’ gives detection result ‘ $c(t)$ ’ using (19)

Step 12: End for

Step 13: End while

Step 14: If ($OL(t) = 1$) **then**

Step 15: Detection of moving objects in the video frame is obtained

Step 16: Else

Step 17: Moving objects are not detected

Step 18: End If

Step 19: End

Algorithm 1 Anisotropic Sophisticated Spatiotemporal Contours based Deep Neural Learning

Algorithm 1 depicts the step by step process of ASSC-DNL Technique. As presented in the above algorithmic steps, ASSC-DNL Technique at first defines network with random weights. For each input video file obtained at the input layer is transmitted to the unseen layer. Then, the four unseen layers in ASSC-DNL procedure deeply analyzes each video frames and extracts important features i.e. shape, color, texture, intensity, and size with the application of ASA, TS-BS and C-OSE model. Next, the hidden layer forward extracted features to the output layer. Finally, the output layer in ASSC-DNL Technique utilizes Gaussian activation function where it compares the extracted features of each video frames with pre-stored templates to produce detection result. From that, ASSC-DNL method efficiently finds the moving objects with a lower amount of time complexity as compared with latest methodology.

III. SIMULATION SETTINGS

ASSC-DNL method is executed in MATLAB simulator using 168VJ Clips video dataset repository [21].



The ASSC-DNL method takes a diverse number of video frames from 168VJ Clips video dataset to perform simulation process. The performance of ASSC-DNL method is measured relevance with finding accuracy, time and false-positive rate. The efficacy of ASSC-DNL method is analyzed with Markov Random Field (MRF) model [1] and Sparse and low-rank illustration with contextual regularization (SLRC) model [2] respectively.

IV. RESULTS

Here the simulation result of the projected ASSC-DNL Technique is presented and analyzes the existing two methods.

4.1 Detection Accuracy

In ASSC-DNL Technique, Detection Accuracy ‘DA’ is determined as the relative amount of moving objects that are precisely detected as,

$$DA = \frac{k_{CD}}{k} * 100 \quad (20)$$

With reference to the (20), the correctness of moving objects detection in a video sequence is estimated with respect to a varied number of video frames. Here, ‘ k_{CD} ’ denotes video frames, ‘ k ’ refers the overall video frames within the simulation work.

Sample Mathematical Calculation:

Proposed ASSC-DNL Technique: number of video frames correctly identified is 24 and total video frames are 25. The detection accuracy is as follows,

$$DA = \frac{24}{25} * 100 = 96 \%$$

❖ **MRF:** number of video frames accurately detected is 21 and the total video frames are 25. The detection accuracy is as follows,

$$DA = \frac{21}{25} * 100 = 84 \%$$

❖ **SLRC:** number of video frames exactly detected is 18 and the total video frames are 25. Then detection accuracy is determined as,

$$DA = \frac{18}{25} * 100 = 72 \%$$

The simulation of moving objects and the accuracy is illustrated in below Table 1.

Table 1 Tabulation Result of Detection Accuracy

Number of Video Frames (k)	Detection Accuracy (%) (%)		
	ASSC-DNL	MRF	SLRC
25	96	84	72
50	94	86	80
75	96	85	79
100	95	87	80
125	97	88	82
150	95	90	78
175	95	90	84
200	93	89	82
225	95	91	80
250	97	93	87

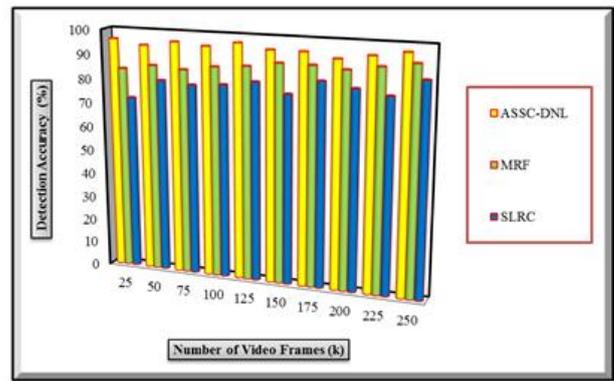


Figure 2 Comparative Result Analysis of identifying Accuracy with reference to Video Frames

Figure 2 presents the impact of moving objects detection accuracy based on a diverse in video frames in the range of 25-250 using three methods i.e. proposed ASSC-DNL Technique and existing MRF model [1] and SLRC model [2]. As shown in the above graphical diagram, proposed ASSC-DNL Technique provides higher accuracy to detect the moving objects with an increasing manner. The ASA, TS-BS and C-OSE and proposed ASSC-DNL Technique. on the contrary to conventional works.

By using the above TS-BS and C-OSE and deep neural learning concepts, proposed ASSC-DNL Technique automatically learns significant features from each video frame and subsequently applies Gaussian activation function for finds the moving objects on the contrary to state-of-the-art works. In the proposed ASSC-DNL Technique, Gaussian activation function compares the discovered features of each video frames with pre-stored templates to accurately generate detection result. This helps for proposed ASSC-DNL Technique to enhance the ratio of number of moving objects in video frames that are exactly detected as compared to state-of-the-art works MRF model [1] and SLRC model [2]. Hence, the proposed ASSC-DNL Technique increases the accuracy of moving objects discovery by 8 % when compared to MRF model [1] and 19 % when compared to SLRC model [2].

4.2 Detection Time

In ASSC-DNL Technique, Detection Time ‘DT’ measures the amount of time required for detecting the moving objects in video frames. The detection time is determined in terms of milliseconds (ms) and mathematically obtained using below,

$$DT = k * t \quad (DSF) \quad (21)$$

From (21), the time needed for detecting the moving objects in a compressed video sequence is evaluated. Here, ‘ k ’ point outs the number of input video frames whereas ‘ t (DSF)’ indicates the time utilized to identify the moving objects in single video frame.

Sample Mathematical Calculation:

❖ **Proposed ASSC-DNL Technique:** time taken for discovering moving objects in a single video frame is 0.95 ms and the total number of the video frame is 25. Then detection time is estimated as,



$$DT = 25 * 0.95 = 24 \text{ ms}$$

❖ **Existing MRF:** the time required for finding moving objects in a single video frame is 1.1 ms and the total number of the video frame is 25. Then detection time is calculated as,

$$DT = 25 * 1.1 = 28 \text{ ms}$$

❖ **Existing SLRC:** time used for detecting moving objects in a single video frame is 1.25 ms and the total number of the video frame is 25. Then detection time is obtained as,

$$DT = 25 * 1.25 = 31 \text{ ms}$$

The performance result analysis of moving objects detection time is demonstrated in below Table 2.

Table 2 Tabulation Result of Detection Time

Number of Video Frames (%)	Detection Time (ms) (ms)		
	ASSC-DNL	MRF	SLRC
25	24	28	31
50	30	35	37
75	41	44	47
100	50	53	58
125	56	60	67
150	60	65	72
175	67	72	81
200	70	76	86
225	74	81	92
250	78	85	98

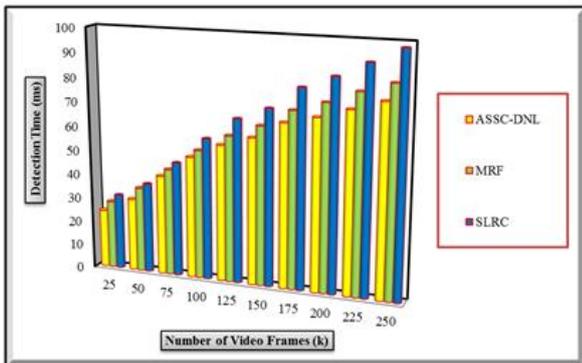


Figure 3 Comparative Result Analysis of Detection Time versus Number of Video Frames

Figure 3 shows the impact of moving objects detection time with respect to a diverse number of video frames in the range of 25-250 using three methods i.e. proposed ASSC-DNL Technique and conventional MRF model [1] and SLRC model [2]. As demonstrated in the above graphical figure, proposed ASSC-DNL Technique provides minimal time complexity for effective moving objects detection while an increasing number of video frames as input when compared to existing MRF model [1] and SLRC model [2]. This is due to the application of ASA, TS-BS and C-OSE and deep neural learning in proposed ASSC-DNL Technique on the contrary to state-of-the-art works.

With the application of TS-BS and C-OSE and deep neural learning concepts, proposed ASSC-DNL Technique automatically extracts key features such as shape, color, texture and size of each video frames with a minimal amount of time utilization. By considering the identified significant features, then proposed ASSC-DNL Technique correctly recognize the moving objects with a lower amount of time

consumption. This supports for proposed ASSC-DNL Technique to minimize the amount of time required for detecting the moving objects in video frames as compared to state-of-the-art works MRF model [1] and SLRC model [2]. Therefore, the proposed ASSC-DNL Technique reduces the detection time of moving objects by 9 % when compared to the MRF model [1] and 18 % when compared to the SLRC model [2].

4.3 False Positive Rate

In ASSC-DNL Technique, False Positive Rate ‘FPR’ evaluated as the ratio of number of moving objects in video frames that are incorrectly detected to the total number of video frames as input. The false-positive rate is determined in terms of percentage (%) and mathematically estimated as,

$$FPR = \frac{k_{ID}}{k} * 100 \text{ (22)}$$

From (22), the false-positive rate of moving objects detection is measured. Here, ‘ k_{ID} ’ represents number of video frames that are inaccurately detected and ‘ k ’ refers a total number of video frames.

Sample Mathematical Calculation:

❖ **Proposed ASSC-DNL Technique:** number of video frames wrongly detected is 1 and the total number of video frames is 25. Then false positive rate is calculated as,

$$FPR = \frac{1}{25} * 100 = 4 \%$$

❖ **Existing MRF:** number of video frames inaccurately detected is 4 and the total number of video frames is 25. Then false positive rate is determined as,

$$FPR = \frac{4}{25} * 100 = 16 \%$$

❖ **Existing SLRC:** number of video frames inexactly detected is 7 and the total number of video frames is 25. Then false positive rate is measured as,

$$FPR = \frac{7}{25} * 100 = 28 \%$$

The tabulation result analysis of false-positive rate for moving objects discovery is presented in below Table 3.

Table 3 Tabulation Result of False-positive Rate

Number of Video Frames (%)	False-positive Rate (%) (%)		
	ASSC-DNL	MRF	SLRC
25	4	16	28
50	6	14	20
75	4	15	21
100	5	13	20
125	3	12	18
150	5	10	22
175	5	10	16
200	8	11	18
225	5	9	20
250	3	7	13

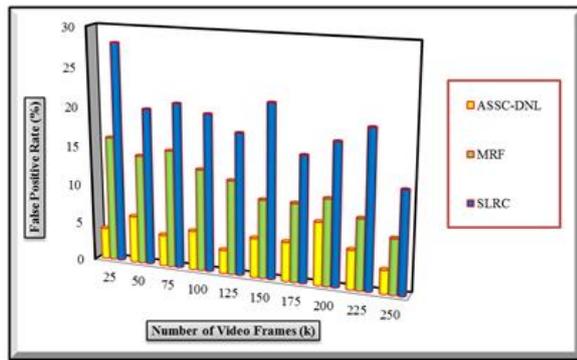


Figure 4 Comparative Result Analysis of False Positive Rate versus Number of Video Frames

Figure 4 depicts the impact of a false positive rate of moving objects detection along with a different number of video frames in the range of 25-250 using three methods i.e. proposed ASSC-DNL Technique and existing MRF model [1] and SLRC model [2]. As presented in above graphical representation, proposed ASSC-DNL Technique provides a lower false-positive rate for accurately recognize moving objects with an increasing number of video frames as input when compared to conventional MRF model [1] and SLRC model [2]. This is because of the application of ASA, TS-BS and C-OSE and deep neural learning in proposed ASSC-DNL Technique on the contrary to existing works.

With the help of TS-BS and C-OSE and deep neural learning concepts, proposed ASSC-DNL Technique increases the moving objects discovery performance as compared to existing works. This assists for proposed ASSC-DNL Technique to reduce the ratio of number of moving objects in video frames that are incorrectly detected as compared to state-of-the-art works MRF model [1] and SLRC model [2]. As a result, the proposed ASSC-DNL Technique minimize the false positive rate of moving objects discovery by 57 % when compared to MRF model [1] and 75 % when compared to SLRC model [2].

V. LITERATURE SURVEY

An Algorithmic Lateral Inhibition-Template Matching (ALI-TM) algorithm was presented in [11] for predicting moving objects in the dynamic background with higher accuracy. A novel method was introduced in [12] with aiming at discovering the highly articulated moving objects based on co-segmentation using athletic video sequences.

A robust person tracking method was presented in [13] with the help of particle swarm optimization (PSO) algorithm for multiple persons in real-time video. A distributed tracking algorithm was proposed in [14] to decrease the number of computational resources owing to detecting people in a sparse set of frames.

Size-aware visual object tracking was accomplished in [15] by means of a dynamic fusion of correlation filter-based part regressors to diminish the computational complexity. A sophisticated and powerful convolutional neural network (CNN) based faster-RCNN was designed in [16] to solve the challenges in discovering extremely low-resolution targets.

A high-speed video salient object discovery method was introduced in [17] by application of temporal propagation and correlation filter. A compressed video saliency detection

algorithm was applied in [18] using discrete cosine transformation (DCT) coefficients.

A compressed sensing (CS)-based algorithm was employed in [19] for the recognition of moving object in video sequences with minimal time. A novel video coding method was designed in [20] with the aid of a reference frame created by dynamic background modeling with lower time.

VI. CONCLUSION

An efficient ASSC-DNL Technique is designed with the goal of improving moving object detection performance by resolving dynamic changing background, and degradation of video frames and object size variations problem. The goal of ASSC-DNL Technique is achieved with the help of ASA, TS-BS and C-OSE and a deep neural learning concept on the contrary to conventional algorithms. The proposed ASSC-DNL Technique increases the ratio of a number of moving objects in video frames that are precisely detected as compared to state-of-the-art works. Also, the proposed ASSC-DNL Technique diminishes the amount of time needed for detecting the moving objects in video frames as compared to state-of-the-art algorithms. Further, the proposed ASSC-DNL Technique decrease ratio of a number of moving objects in video frames that are wrongly detected as compared to existing works. The ASSC-DNL Technique performs simulation work using parameters such as detection accuracy, detection time and false positive rate with respect to a diverse number of video frames and compared with two existing works. The simulation result depicts that the proposed ASSC-DNL Technique provides better performance with an enhancement of moving object detection accuracy and minimization of moving object detection time when compared to state-of-the-art works.

REFERENCES

- Ming Ma and Houbing Song, "Effective moving object detection in H.264/AVC compressed domain for video surveillance", *Multimedia Tools and Applications*, Springer, Pages 1-15, September 2019
- Bo-Hao Chen, Ling-Feng Shi, Xiao Ke, "A Robust Moving Object Detection in Multi-Scenario Big Data for Video Surveillance", *IEEE Transactions on Circuits and Systems for Video Technology*, Volume 29, Issue 4, Pages 982 - 995, April 2019
- Xin Sun, Junyu Shi, Lipeng Liu, Junyu Dong, Claudia Plant, Xinhua Wang, Huiyu Zhou, "Transferring deep knowledge for object recognition in Low-quality underwater videos", Elsevier, Pages 1-12, 2017
- Lingling Tong, FengDai, DongmingZhang, DongWang, YongdongZhang, "Encoder combined video moving object detection", *Neurocomputing*, Elsevier, Pages 1-13, 2014
- Chulyeon Kim, Jiyoung Lee, Taekjin Han and Young-Min Kim, "A hybrid framework combining background subtraction and deep neural networks for rapid person detection", *Journal of Big Data*, Springer, Volume 5, Issue 22, Pages 1-24, 2018
- Marcus Laumer, Peter Amon, Andreas Hutter, André Kaup, "Moving object detection in the H.264/AVC compressed domain", *APSIPA Transactions on Signal and Information Processing*, Volume 5, 2016
- K. L. Chan, "Segmentation of moving objects in image sequence based on the perceptual similarity of local texture and photometric features", *EURASIP Journal on Image and Video Processing*, Springer, Volume 62, Pages 1-16, 2018
- Natal Henrique Cordeiro, Emerson Carlos Pedrino, "A new methodology applied to dynamic object detection and tracking systems for visually impaired people", *Computers and Electrical Engineering*, Elsevier, Volume 77, Pages 61-71, 2019

9. Ong Kok Meng, Ong Pauline, Sia Chee Kiong, Low Ee Soong, "Effective moving object tracking using modified flower pollination algorithm for visible image sequences under complicated background", *Applied Soft Computing Journal*, Elsevier, Volume 83, Pages 1-30, 2019
10. Midhula Vijayan, Mohan Ramasundaram, "Moving object detection using vector image model", *Optik - International Journal for Light and Electron Optics*, Elsevier, Volume 168, Pages 963–973, 2018
11. Yufei Zhao, Yong Songa, Shangnan Zhao, Yun Li, Xu Li, Qun Hao Zhengkun Guo, "ALI-TM: A moving objects detection algorithm for infrared images with dynamic background", *Infrared Physics & Technology*, Pages 1-18, 2018
12. Walid Barhoumi, "Detection of highly articulated moving objects by using co-segmentation with application to athletic video sequences", *Signal, Image and Video Processing*, Springer, Volume 9, Issue 7, Pages 1705–1715, October 2015
13. Ching-Han Chen, Chien-Chun Wang, Miao-Chun Yan, "Robust tracking of multiple persons in the real-time video", *Multimedia Tools and Applications*, Springer, Volume 75, Issue 23, Pages 16683–16697, December 2016
14. D. A. Kuplyakov, E. V. Shalnov, V. S. Konushin, and A. S. Konushina, "A Distributed Tracking Algorithm for Counting People in Video", *Programming and Computer Software*, Volume 45, Issue 4, Pages 163–170, 2019
15. Alireza Memarmoghadam, Payman Moallem, "Size-aware visual object tracking via dynamic fusion of correlation filter-based part regressors", *Signal Processing*, Volume 164 Pages 84–98, 2019
16. Huaizhong Zhang, Chunbo Luo, Qi Wang, Matthew Kitchin, Andrew Parmley, Jesus Monge-Alvarez, Pablo Casaseca-de-la-Higuera, "A novel infrared video surveillance system using deep learning-based techniques", *Multimedia Tools and Applications*, Springer, Volume 77, Issue 20, Pages 26657–26676, October 2018
17. Qi Qi, Sanyuan Zhao, Wenjun Zhao, Zhengchao Lei, Jianbing Shen, Libin Zhang, Yuanyuan Pang, "High-speed video salient object detection with temporal propagation using correlation filter", *Neurocomputing*, Elsevier, Volume 356, Pages 107-118, September 2019
18. Qin Tu, Aidong Men, Zhuqing Jiang, Feng Ye, Jun Xu, "Video saliency detection incorporating temporal information in the compressed domain", *Signal Processing: Image Communication*, Elsevier, Volume 38, Pages 32-44, October 2015
19. Bin Kang, Wei-Ping Zhu, Jun Yan, "Object detection oriented video reconstruction using compressed sensing", *EURASIP Journal on Advances in Signal Processing*, Springer, Pages 1-15, and December 2015
20. Manoranjan Paul, Weisi Lin, Chiew Tong Lau & Bu-Sung Lee, "Video coding with dynamic background", *EURASIP Journal on Advances in Signal Processing*, Volume 11, Pages 1-17, 2013
21. 168VJ Clips video dataset repository:
https://archive.org/details/Architects_of_Tomorrow.