International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Towards Enhanced Anomaly Object Detection and Face Recognition (EAODFR) in Surveillance Videos using Recurrent Neural Networks

P.Ragupathy, P.Vivekanandan

Abstract*: In the present decade, anomaly object detection and face recognition from surveillance videos from diverse environments have become interesting and challenging research areas in computer vision. This paper works on developing an Enhanced Anomaly Object Detection and Face Recognition (EAODFR) model using Recurrent Neural Networks (RNN). Moreover, fractional derivative based background separation has been incorporated for framing efficient background subtraction model and foreground segmentation with appropriate pixel definitions on each frame of the surveillance videos. The Region of Interest detection has been done using optimal thresholding and for detecting anomaly objects. Further, efficient face recognition has been accomplished by designing the Recurrent Neural Networks (RNN), which is implemented with Long Short-Term Memory (LSTM). The recurrent NN are trained in terms of determining anomalous objects using the extracted features in the each frame of the video. The obtained results are analyzed in terms of precision, recall and f-measure and compared with some existing face recognition models. The comparative analysis provides better results and outperforms others.

Keywords*: Anomaly Object Detection, Face Recognition, Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Background Subtraction Model.

I. INTRODUCTION

The increasing needs of security measures in current scenario, automated video surveillance plays a significant role in people, objects, vehicles, etc, concurrently involves in identification of events, interactions and actions among all [1]. In that process, some issues may happen in moving object detection with high accuracy rate, but which is more important in face identification and tracking.

Manuscript published on November 30, 2019.
* Correspondence Author
P.Ragupathy*, Research Scholar, Department of Computer Science and Engineering, Park College of Engineering and Technology, Tamil Nadu, India.
P.Vivekanandan Professor and Head, Department of Computer Science and Engineering, Park College of Engineering and Technology, Tamil Nadu, India.

A standard explanation aids to variant the regions of concern in the surveillance videos for particular instances, based on the background region of moving objects [2]. The basic model
for face recognition from surveillance video is presented in Fig.1. In countries having dense population, it is very complicated to identify the face of a person from the surveillance video clippings. Hence, an automated system is required for processing. Moreover, while designing a face recognition model, some of the important criteria like lighting conditions, dynamic environment, shadows, illumination changes, etc, are to be considered. Previously, face recognition model has been limited to the recorded images and videos, but in present cases, it has been improved in so many ways by enforcing for surveillance video, multi-view videos for achieving better accuracy in results.
In general, the face images obtained from the real-time videos are low in resolution and off-front. Therefore, the images may not match exactly to the frontal face images from the database. Here, the videos are converted into frames, which are to be processed with feature extraction process. The image oriented face recognition has more advancement in video based recognition. The improved results can be obtained through [3], [4],

- Utilization of temporal face information
- Obtaining three dimensional image view from the video input

Moreover, the proposed Enhanced Anomaly Object Detection and Face Recognition (EAODFR) model utilizes the Recurrent Neural Network with LSTM for examining the changes occurred in each frame in a video sequence. The Recurrent Neural Networks are comprised with the number of connected neurons with three units called input, hidden and output units, with the time establishment at ‘t’, which is used for selective data processing. When a single element is processed at an instant, the output can be modeled on the basis of the sequential elements that are dependent [5]. The RNN framework gives power to process and find the hidden patterns from the spatio-temporal data such as text, audio and video contents. In RNN, the data have been processed in a sequential manner at the time instant ‘t’, further, the input has been obtained from the previous hidden state F_{t-1} and a new input x(t). The data has been multiplied with their corresponding weights and given for further activations. Since there are huge numbers of computations, after few layers, the impact of the initial data becomes insignificant for the sequence that is processing in the future.

This may cause vanishing gradient problems. This can be solved by incorporating the Long Short-Term Memory (LSTM) concept, which comprises memory unit, and three gates called input, output and forget gate. By this, state of an element can be maintained for certain time [6]. In order to process with a sequential data, different categories of LSTM such as bi-directional, multi-layer has been used. The proposed model can handle the complex video patterns, which are not able to be handled effectively by the simple of multi-layer LSTM [7]. In this proposed model, the input video features are analyzed and the facial images are efficiently recognized by using the two-layer LSTM, which is presented in the Fig.2.

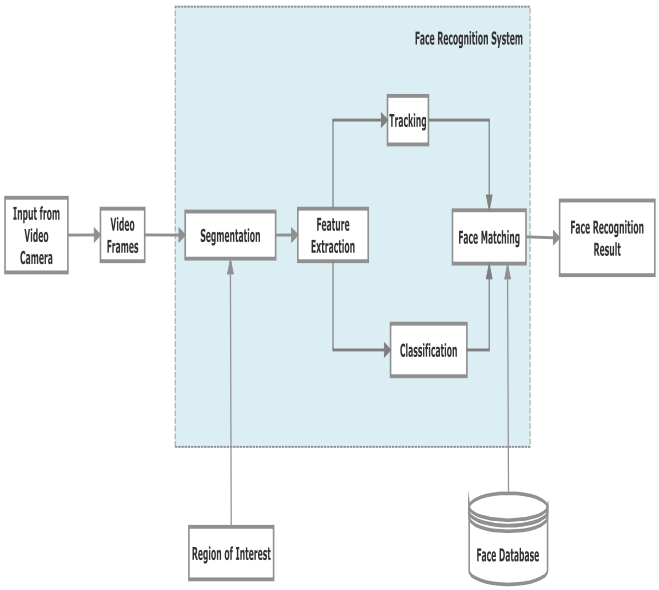

Fig.1. Basic Model for Face Recognition in Video Processing

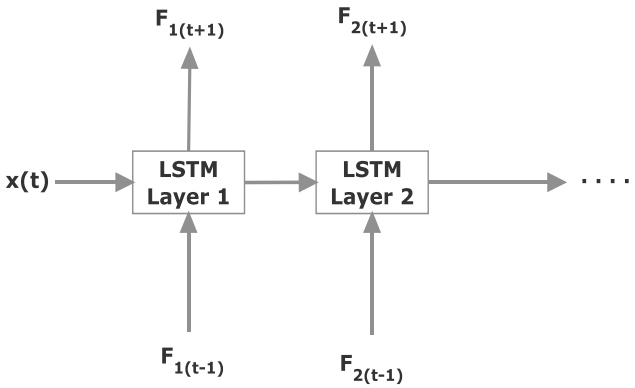

Fig.2. Double Layered LSTM

The proposed model investigates a methodology for background separation based on fractional derivative and foreground segmentation with appropriate pixel definitions on each frame of the surveillance videos. The Region of Interest detection from the video clips has been done using optimal thresholding and thereby effectively detection the faces.

The remainder of the paper is organized as follows: Section 2 discussed about the related works on anomaly object detection and face recognition in video processing. Section 3 presents working procedure of the proposed Enhanced Anomaly Object Detection and Face Recognition (EAODFR) model. The results and the performance evaluation of the proposed model are presented in the Section 4. Section 5 concludes the paper by highlighting the efficiency of the proposed methodology and also some pointers for future enhancement.

II. RELATED WORKS

In [8], the authors have discussed about two kids of features such as MHI- Motion History Imaging and HOG- Histogram of Oriented Gradients. The MHI is the method of foreground image has been eliminated from the background region, whereas the HOG handles with the direction and extends of the image edges. Based on the fusion, the further classification has been made by the model called Simulated Annealing Multiple Instance Learning SVM (SMILE-SVM). The features of the object motion and their static features are derived from the real time videos and processed in [9]. The static features have been obtained from the application of motion statistics to the noisy motion. Additionally, the paper has also utilized the page ranking model for building the visual terminologies. But, the process has some limitations in handling with multiple actions at an instant. Hence, combined models including motion detection [10], background separation [11], HOG, etc.

SIFT (Scale Invariant Feature Transform) was developed and explained in [12]. It involved in invariant based feature detection. The process of key point extraction is based on the following steps: Scale-Space extrema detection, Localization of Key points, Assignment orientation and Description of key points. In order to speed up the interest point detection process, Speeded-Up Robust Features (SURF) model has been developed in [13]. For the efficient detection of feature point, SURF model used the conventional method of Hessian matrix approximation.

Specifically in image classification, object detection, bio-informatics and person identification, deep learning concepts are applied widely in significant manner [13]. In [14], an action recognition model has been implemented based on deep learning methods of neural networks, specifically, three-dimensional CNN. In that technique, each video frame has been applied with the 3-D kernels based on the spatio-temporal time axis. Moreover, the model has also been concentrating on detecting the optical flow of the moving objects, since the frames are under fully-connected layers. Further, in [15], a multi-resolution based CNN architecture has been given for detecting the local space and time related data of the video sequences.

The authors of [16] utilized two CNN methodologies for action recognition from the obtained video sequences, processing each frame of the video. The intermediate layer between the two CNN frameworks has been operated by 1×1 kernels. Finally, 30-frames unrolled LSTM has been utilized for training the data. In [17], for object tracking and detection, a CNN based model has been framed. The model comprised individual stages for proposal generation, object classification and rescoring. This might reduce the efficient of detection results from the video clips. In [18], the model adapted RNN and LSTM for sequence tracking from the obtained video.

A data association methodology using the learning based approach has been given in [19]. Moreover, in [20], in each LSTM layer, social pooling layer has been added to mere the hidden states of the neighbour objects inside the spatial region. Nevertheless, the model used the deep learning models; object detection has not been performed in problem detection.

In [21], discussions have been made on the process of background subtraction involves in identifying the moving objects by subtracting the current frame pixel-by-pixel from a referral background image that is framed by averaging the images according to time in the starting period of time. While doing, when the pixel value becomes above the defined threshold is categorized into foreground. Formation of foreground pixel mapping has been done with that.

The process of foreground detection matches up the input video frame with the background model and finds out the candidate pixels of foreground frame from the input image [22]. For attaining this classification, the map framed based on the differences are binarized by the method called thresholding. The appropriate threshold value is dependent on the scene, the noise level and the illumination occurrences.

Entropy based robust window thresholding is given in [23], that provides a framework for video object detection system. In the paper, the model involved in the threshold computation of the blocks of an image robustly based on the RoC (Regions of Change). The computations are made with the local adaptive thresholding and averaging the threshold values for determining the global entropy values.

Video based face recognition model has been given in [24] based on joint sparsity technique. Based on the illumination conditions and poses, the faces are segmented for training. Using sequential sampling and update model, the face recognition technique has been given in [25]. The pose variations have been typically identified using robust and resilient object features. Region based modelling has been used for solving the multiple sample problems. The authors of [26], Hidden Markov Model has been used for facial identification, but the model could handle with the limited dataset having minimal pose variations.

III. METHODOLOGY

In the proposed Enhanced Anomaly Object Detection and Face Recognition (EAODFR) model, the significant features are extracted from the input surveillance video efficiently, and then matched up with the pre-defined training dataset for recognizing the face in the video. For better recognition of faces and anomaly objects, the proposed model combines the efficiencies of Recurrent Neural Networks and LSTM, since it produces efficient results on visual and sequential data processing. Moreover, EAODFR model comprises three phases of work, as follows:

1. Static Background Separation using Fractional Derivative
2. Region of Interest (RoI) Identification
3. Optimal Thresholding based Face Recognition

Initially, the input surveillance video is obtained from the repository. The obtained video is divided into several frames. For effective segmentation and précised object detection and face recognition, optimal thresholding is incorporated in the work. The pre- processing has been performed for every frame for removing the unwanted noise using the Gaussian filter. At that instant, fractional derivative is figured out for each pre-processed frame. Following that, absolute difference is calculated based on the past procedure. The segmentation has been performed based on the method of optimal thresholding.

A. *Static Background Separation using Fractional Derivative:*

The pre-processed frames of the obtained video sequence are further given for static background separation using fractional derivative. In that, the derivations of Fractional Calculus (FC) are used for framing optimal background pixels for background replication.

The Fractional Calculus oriented to the simplification process of differentiation infused with integration can be termed as fractal sets. According to the differential hypothesis initialization combined with integral calculus, there are number of arithmetic functions have been established for the identification of non-integer derivative orders based on the integrals. Nevertheless, the pertinence of the FC stands underprovided still; the modern improvement in science provoked a modification over interest in this region of interest. Based on these progressions, the utilization of Fractional Derivatives in Background separation for the construction of optimal background pixels is determined.

For each pixel of the pre-processed video frames (a', b'), $BM_t(a', b')$ is assumed to be the corresponding background model derived from the following derivations.

$$BM_t(a', b') = BM_{t-1}(a', b') + \frac{1}{K}(CV_t(a', b') - BM_{t-1}(a', b')) \quad (1)$$

From the equation (1), $BM_{t-1}(a', b')$ is the earlier background model, $CV_t(a', b')$ is the denotation of the current video series, 'K' represents the equalizing constant, which is assumed as 8 here. Moreover, the equation renovated corresponding to the transformation motive for the background with fractional derivative. The equation is derived from (1) as follows,

$$BM_t(a', b') - BM_{t-1}(a', b') = \frac{1}{K}(CV_t(a', b') - BM_{t-1}(a', b')) \quad (2)$$

As given earlier, the Fractional Calculus is a component of the exponential analysis that frames to real or complex integers the differential order based integral functions. Because of the simplification feature of the fractional derivative and also the integral operator to a non-integer N^{th} order based on variant models. With respect to time 't', the Laplace definition of the fractional derivative of the functional order '$N \in CV$' of the mentioned signal *a(t)*, $BM_k[a'(t)]$ is the direct generalization model of the basic integer-order method. $BM_t(a', b') - BM_{t-1}(a', b')$ is the discrete variant of the functional derivative order N=1, with that consideration t=1, the following equation (3) is framed,

$$BM^N[BM_{t-1}(a', b')] = \frac{1}{K}(CV_t(a', b') - BM_{t-1}(a', b')) \quad (3)$$

From the above derivations, the background separation is performed from the obtained input video frame $F''(x', y')$.

With respect to time 't', the resultant background model is further provided for moving object detection and face recognition using the optimal thresholding methodology.

B. *Region of Interest (ROI) Identification*

The Region of Interest is identified specifically for,

- Combining the unstructured data into formal data
- Separating order regions based on the spatio-temporal information of a particular video frame.

In the proposed model, the ROI (Region of Interest) states the object that is on motion in each frame of a video clip and is mentioned by the red marking, which is depicted in the Fig.3.

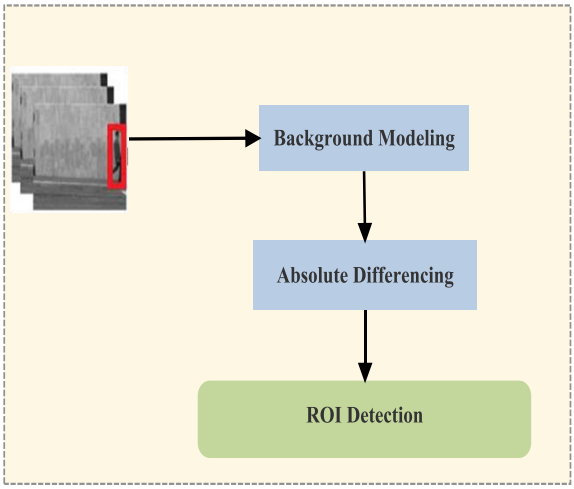

Fig.3. Phase of ROI Identification

The phase of interesting region identification makes the process of motion detection for evaluating each block that comprises the moving objects, which can be the human faces. After that, when a background separation is done based on the fractional derivative computations in static background separation in each frame, the absolute variation, (which is stated as $\Delta_t(a,b)$) is obtained using the absolute differential computation between the past data about the updated background model and the present obtained video frame ($F_t(a,b)$). For improving the execution speed of the motion detection process in an efficient manner, the ROI identification phase that combines present block based motion identification based on morphological operations. Based on the given block based motion estimation, the identification of each motion in each video frame has been made efficient. The overall block for the operations involved in the proposed model is presented in Fig.4.

C. *Optimal Thresholding based Face Recognition*

Face recognition is the process of deriving specific objects called faces, which involves in the foreground object extraction from frame sequences. Here, optimal thresholding are the efficient method for processing face recognition. The process is performed using by distributing the intensity rates called thresholds. For each pixel, that points every other pixel is either clustered as the object point or the background point.

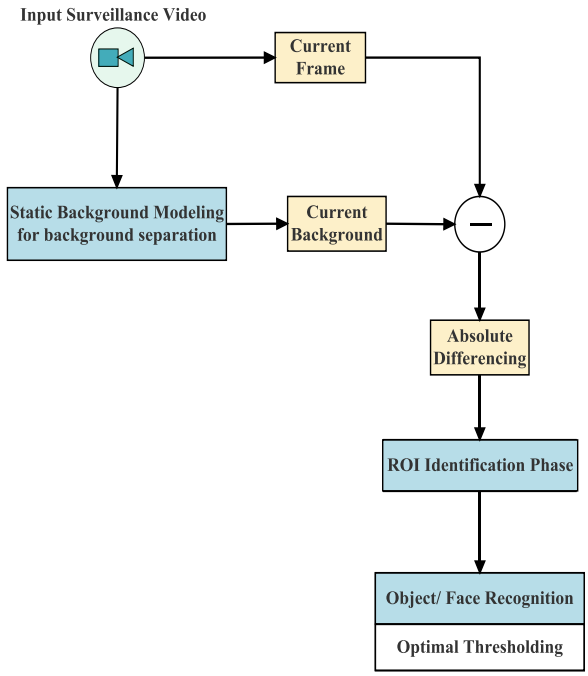

Fig.4. Operations Involved in Proposed Model

Further, the process involves in separating the image pixels into two classes such as CL_0 and CL_1 at the gray level called 'gl', where $CL_0 = \{1, 2, \ldots, gl\}$ and the next class, $CL = \{t+1, t+2, \ldots, l-1\}$. Based on the consideration that p_0 and p_1 as the estimated of class probabilities, which are computed in following equations.

$$p_0 = \sum_{l=0}^{gl} PR_r(l) \quad (4)$$

$$p_1 = \sum_{l=gl+1}^{l-1} PR_r(l) \quad (5)$$

Following, the individual class variances (σ) are derived as follows,

$$\sigma_0 = \left(\sum_{l=0}^{gl}\left[i - \sum_{l=0}^{gl} l.PR_r(l)/p_0(gL)\right]^2 \frac{PR_r(l)}{p_0}\right)^{\frac{1}{2}} \quad (6)$$

$$\sigma_0 = \left(\sum_{l=gl+1}^{l-1}\left[i - \sum_{l=gl+1}^{l-1} l.PR_r(l)/p_0(gl)\right]^2 \frac{PR_r(l)}{p_1}\right)^{\frac{1}{2}} \quad (7)$$

where 'PR_r' represents the frame histograms. Moreover, the difficulty of lessening within the class variance in represented by the extension of class variance commection. It is connected with respect to the difference between the collective variance and the inside class variance.

$$\sigma_n^2(gl) = p_0(gl)[1 - p_0(gl)]\left[\sum_{l=gl+1}^{l-1} l.PR_r(l)/p_0(gl) - \sum_{l=0}^{gl} l.PR_r(l)/p_0(gl)\right] \quad (8)$$

Moreover, the steps involved in Object Detection and Face Recognition using optimal thresholding is as follows,

1: Histogram computation and deriving probabilities of intensity level of each frames computation
2: Initialization of $p_0(0)$ and $l.PR_r(l)/p_0(gl)$ at gl=0
3: Calculation of intensity levels for each threshold
4: Adjust p_0 and $\sum_{l=0}^{gl} l.PR_r(l)/p_0(gl)$
5: Calculate $\sigma_n^2(gl)$

6: Threshold is selected based on their highest rates

D. *Incorporation of LSTM in RNN in the EAODFR Model*

The Recurrent Neural Networks (RNN) is an efficient source for the image classification. It strengthens the process and aids in identifying the hidden patterns of the input source efficiently. In the obtained video sequence, each converted frames are denoted by the features of RNN, followed by determining the sequential data between those frames using the Deep Bidirectional-LSTM. In general, a video is defined as the collection of frames that are running at 30 to N number of frames per second. From that, an efficient object detection and face recognition process needs efficient deep learning technique for image identification. This may require higher number of images to be processed. In RNN, the process can be managed by the layer based framework. In the proposed model, the RNN utilizes a double layered LSTM structure. By this, the RNN obtains more level of sequential data about the surveillance video. The diagrammatical representation of the incorporation of the LSTM in the proposed work is given in Fig.5. The typical structure can manage the long video processing in object detection. Here, according to LSTM, x(t) is the input given at an instant 't', f(t) represents the forget data at time 't', involves in deleting the data from the memory unit of LSTM, when there is a need. And, the data log about the previous frame is kept. The output unit o(t) handles the data about the future frames, where 'r' is the recurrent factor for function activation of 'tanh'. It can be calculated using the present frame input and the condition of the previous frame F_{t-1}. Moreover, the hidden layer conditions of RNN are computed using the value of 'tanh' and memory unit M(t). The immediate output of the LSTM is considered to be the final stated of the RNN, which is taken for face recognition.

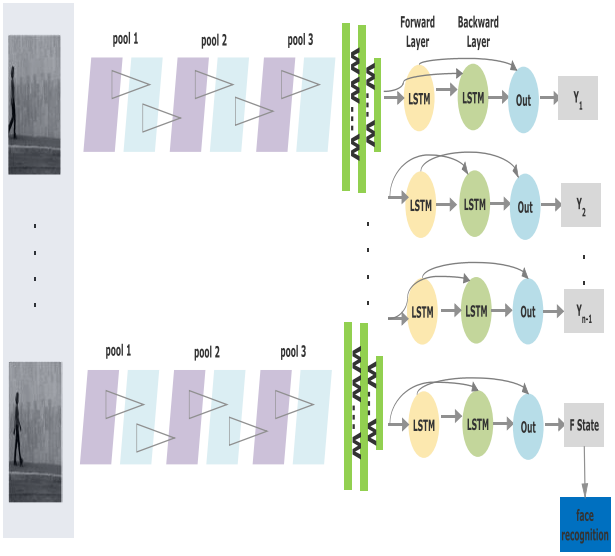

Fig.5. Incorporation of LSTM in the Proposed Work

The computations involved in this section are as follows:

$$i_t = \sigma((x(t) + F_{t-1})Wt^i + bs^i) \quad (9)$$

$$f_t = \sigma((x(t) + F_{t-1})Wt^f + bs^f) \quad (10)$$

$$o_t = \sigma((x(t) + F_{t-1})Wt^o + bs^o) \quad (11)$$

$$r = \tanh((x(t) + F_{t-1})Wt^r + bs^r) \quad (12)$$

$$M(t) = M(t-1).f_t + r.i_t \quad (13)$$

From the above equations, i_t, f_t, o_t are the input, forget and output gates of LSTM framework, respectively. Wt^i, Wt^f, Wt^o, Wt^r denotes the weights of the corresponding gates and bs^i, bs^f, bs^o, bs^r are the respective biases values of LSTM unit, which are adaptive based on back propagation. Training the complicated videos are not be effectively handled in single layer LSTM, whereas, stacking number of LSTM layers can process the video input precisely.

IV. RESULT AND DISCUSSION

In this section, the proposed EAODFR Model is experimentally analyzed and the obtained results are provided with respect to the different benchmark face recognition datasets like SCface - Surveillance Cameras Face Database [27] and PIE Database, CMU [28]. A few sample images from the benchmark datasets used are given in Fig.6. Further, the obtained images are separated for machine learning process for the categories, training, testing and validation in the ratio of (6:2:2) respectively. The images of training data sets are given for deep learning at the rate of 0.001 in sets of 512 sizes. Moreover, the proposed model is evaluated based on the parameters such as accuracy of the recognition rate and the processing time. The results are compared with the obtained accuracy and time of Entropy based model [23] and SMILE-SVM [9]. The experimentation of the model is done using MATLAB tool.

A. Surveillance Cameras Face Database

The Surveillance Camera Face dataset is a database having a collection of static human facial images. Those images were captured using five surveillance cameras of different qualities in closed environments. Moreover, the database comprises 4160 static facial images of 130 people. The cameras of different qualities are used to get variant recorded images, thereby making the efficient testing and accentuating of variant positions to enable the robustness of the face recognition mechanisms.

B. PIE Database, CMU

PIE Database dataset contains 41, 368 facial images of 68 persons. Each person images are taken with 43 variances of illumination states and 13 different poses and also with 4 varieties of facial expressions. Moreover, the dataset comprises >750 thousand pictures of 337 persons taken about five months of duration with 4 categories. The face expressions are captured under 15 oriented views and 19 illumination variations. In total, the database of the dataset contains 305 GB of facial images for assisting various researches. The Fig.7 portrays the processing of frontal facial image in the face recognition model, which is obtained from the SCface dataset.

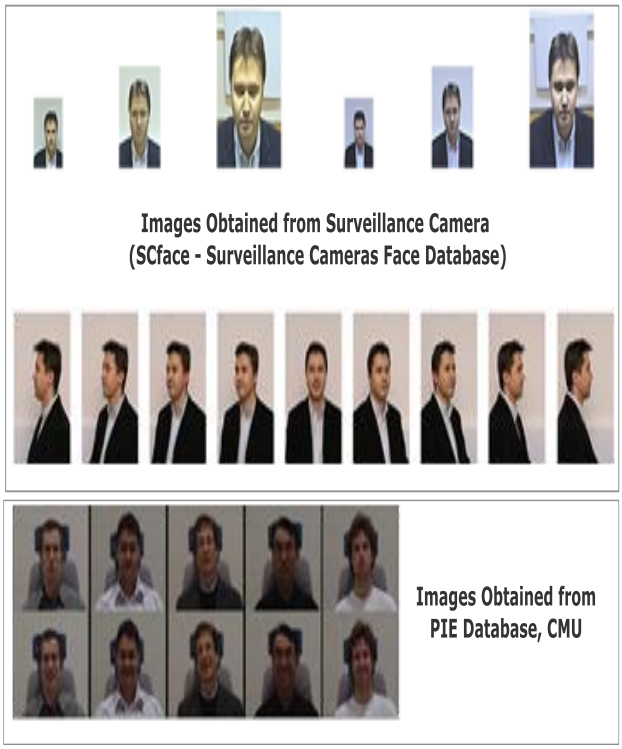

Fig.6. Sample Faces Obtained from SCface and PIE Database Datasets

Fig.7. Processing with the Frontal Facial Image Obtained from the Dataset

The confusion matrix is which the true positive intensity rates are high in the proposed model, analysing with the images obtained from the datasets SCface dataset and PIE Dataset. This provides the evidence for the efficiency of the proposed model. The Fig.8 presents the comparison chart for the rate of accuracy in face recognition obtained by the proposed model and the compared works. The analysis has been made between the accuracy rates with respect to the image categories. From the chart, it is explicit that the proposed model produces higher rate of accuracy than others with the datasets utilized. Further, the Fig.9 presents the comparison on processing time for face recognition. It is shown from the graph that the proposed model produces lesser amount of processing time than others.

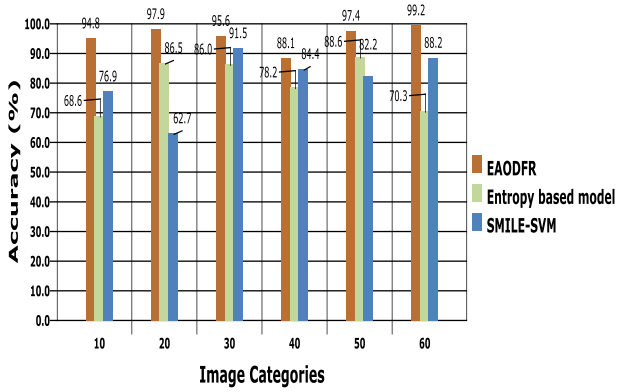

Fig.8. Accuracy Rate Comparison

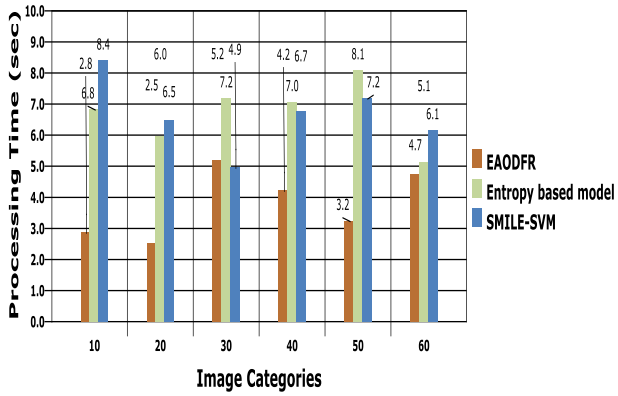

Fig.9. Analysis on Processing Time

V. CONCLUSION

In this paper presents an Enhanced Anomaly Object Detection and Face Recognition (EAODFR) model, using the features of RNN with the combined efficiency of DB-LSTM. Initially, the video frames are processed for background separation and Identification of ROI. Further, the RNN characteristics are derived from the video frames, which are given to the DB-LSTM, comprises the double layer. This aids in identification of complex frames from the hidden patterns. Moreover, the long-term complex sequences are processed efficiently using the proposed model. The evaluation has been made with the benchmark datasets namely, SCface - Surveillance Cameras Face Database and PIE Database, CMU. The experimental results show that proposed model produces higher accuracy rate of face recognition than others. The work can be further enhanced by utilizing intelligent techniques for processing in crowded environments and dense video patterns. Action recognition based methodologies can also been included in the part of future work.

REFERENCES

1. Robert T. Collins, Alan J. Lipton and Takeo Kanade, “Introduction to the Special Edition on Video Surveillance”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, Pp. 745-757, 2000.
2. P.Chiranjeevi and S. Sengupta, “Robust Detection of Moving Objects in Video Sequences through Rough Set Theory Framework”, Image and Vision Computing, 2012.
3. J. Liu, J. Luo, and M. Shah, ``Recognizing realistic actions from videos `in the wild,'" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2009, pp. 1996_2003.

4. S. K. Choudhury, P. K. Sa, S. Bakshi, and B. Majhi, ``An evaluation of background subtraction for object detection vis-a-vis mitigating challenging scenarios," IEEE Access, vol. 4, pp. 6133_6150, 2016.
5. C. Lipton, J. Berkowitz, and C. Elkan. (2015). ``A critical review of recurrent neural networks for sequence learning." [Online]. Available: https://arxiv.org/abs/1506.00019
6. K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, ``LSTM: A search space odyssey," IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 10, pp. 2222_2232, Oct. 2017.
7. A. Graves, S. Fernández, and J. Schmidhuber, ``Bidirectional LSTM networks for improved phoneme classi_cation and recognition," in Proc. 5th Int. Conf., Warsaw, Poland, Sep. 2005, p. 753.
8. Y. Hu, L. Cao, F. Lv, S. Yan, Y. Gong, and T. S. Huang, ``Action detection in complex scenes with spatial and temporal ambiguities," in Proc. IEEE 12th Int. Conf. Comput. Vis., Sep./Oct. 2009, pp. 128_135.
9. K. Soomro, A. R. Zamir, and M. Shah. (2012). ``UCF101: A dataset of man actions classes from videos in the wild." [Online]. Available: https://arxiv.org/abs/1212.0402
10. S. K. Choudhury, P. K. Sa, S. Bakshi, and B. Majhi, ``An evaluation of background subtraction for object detection vis-a-vis mitigating challenging scenarios," IEEE Access, vol. 4, pp. 6133_6150, 2016.
11. S. K. Choudhury, P. K. Sa, K.-K. R. Choo, and S. Bakshi, ``Segmenting foreground objects in a multi-modal background using modi_ed Z-score," J. Ambient Intell. Hum. Comput., pp. 1_15, Apr. 2017, doi: https://doi.org/10.1007/s12652-017-0480-x
12. Lowe, D., 2004, 'Distinctive image features from scale-invariant keypoints', International journal of computer vision, vol. 60, no. 2, pp. 91-110.
13. S. K. Choudhury, P. K. Sa, R. P. Padhy, S. Sharma, and S. Bakshi, ``Improved pedestrian detection using motion segmentation and silhouette orientation," Multimedia Tools Appl., pp. 1_40, Jun. 2017, doi: https://doi.org/10.1007/s11042-017-4933-1
14. S. Ji, W. Xu, M. Yang, and K. Yu, ``3D convolutional neural network for human action recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 1, pp. 221_231, Jan. 2013.
15. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, ``Large-scale video classi_cation with convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 1725_1732.
16. J. Y.-H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, ``Beyond short snippets: Deep networks for video classification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 4694_4702.
17. K. Kang, W. Ouyang, H. Li, and X. Wang. Object detection from video tubelets with convolutional neural networks. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.
18. J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. In CVPR, 2015.
19. Y. Xiang, A. Alahi, and S. Savarese. Learning to track: Online multi-object tracking by decision making. In The IEEE International Conference on Computer Vision (ICCV), December 2015.
20. A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.
21. Javed O., Shafique K., Shah M, 2002, 'A hierarchical approach to robust background subtraction using color and gradient information', Motion Workshop on Motion and Video Computing (MOTION' 02), pp. 22-27.
22. L. Li, W. Huang, I.Y.H. Gu and Q. Tian, 2004, 'Statistical modeling of complex backgrounds for foreground object detection', vol. 13, no. 11, pp. 1459–1472.
23. Anuradha.S.G, K.Karibasappa, B.Eswar Reddy, 2013, 'Video Segmentation For Moving Object Detection Using Local Change & Entropy Based Adaptive Window Thresholding', Computer Science & Information Technology (CS & IT), pp 155-165.
24. Chen, Yi-Chen, "Video-based face recognition via joint sparse representation." Automatic Face and Gesture Recognition (FG), 2013, 10th IEEE International Conference and Workshops on. IEEE.
25. G.Aggarwal, S.Biswas, P.J.Flynn, K.W.Bowyer, "A.Sparse Representation approach to face matching across plastic surgery, in:Proceedings of IEEE Workshop on Applications of Computer Vision, Colorado Springs, CO, pp.113–119, January 2012, IEEE.
26. X. Liu, T. Chen, "Video-based Face Recognition using Adaptive Hidden Markov Model", CVPR 03, proceeding of 2003 IEEE computer society conference eon computer vision and pattern recognition, pp 340-345, 2003, IEEE.
27. Mislav Grgic, Kresimir Delac, Sonja Grgic, SCface - surveillance cameras face database, Multimedia Tools and Applications Journal, Vol. 51, No. 3, February 2011, pp. 863-879.
28. http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html

AUTHORS PROFILE

P.Ragupathy received his B.Tech (Information Technology) from Anna University, Chennai and M.E (Computer Science and Engineering) from Karpagam University, Coimbatore. He is currently Research Scholar in the Department of Computer Science and Engineering at Park College of Engineering and Technology, Coimbatore, Tamil Nadu, India. His research interest is mainly focused on computer vision, object recognition, emotion classification with machine learning techniques and deep learning.

Dr.P.Vivekanandan is currently working as a Professor, Department of Computer Science and Engineering, Park College of Engineering and Technology, Coimbatore, Tamilnadu, India. He has more than twelve years of teaching experience. He obtained his B.E (Computer Science and Engineering) from Bharathiar University, Coimbatore, India and his M.Tech (Distributed Computing Systems) from Pondicherry University, Pondicherry, India.and his Ph.D from Anna University Chennai. At present, He is guiding 10 research scholars of Anna University, India. His research interests include Knowledge Discovery and Data Mining, Soft Computing and Distributed Computing. He has published many research papers in National/International Conferences and Journals. He has attended several seminars and workshops in the past ten years. He has also organized several symposiums and workshops. He has guided more than 20 UG and 15 PG projects. He is a life member of ISTE and also a member of Computer Society of India.