Abnormal Event Detection using Additive Summarization Model for Intelligent Transportation Systems

G. Balamurugan[1]
Research Scholar
Department of Computer Science and Engineering
Puducherry Technological University, Pondicherry, India

Dr. J. Jayabharathy[2]
Associate Professor
Department of Computer Science and Engineering
Puducherry Technological University, Pondicherry, India

***Abstract*—Video surveillance is used for capturing the abnormal events on roadsides that are caused due to improper driving, accidents, and hindrances resulting in transportation lags and life-critical issues. It is essential to highlight the accident keyframes in videos to achieve intelligent video surveillance. Video summarization plays a vital role in summarizing the keyframe for an abnormal event from the stacked video surveillance input. The observed video is converted into frames and analyzed for providing an accurate summarization for accident analysis forecast and guiding the users in avoiding such events. The main issues in summarization arise from the inconsistency between the spatiotemporal redundancies and the classification of sequence verification in video surveillance. This article introduces an Additive Event Summarization Method (AESM) for projecting classified events through a gated recurrent unit learning paradigm. In this process, the gates are assigned for unclassified and active frames for sequence verification. Based on the sequence, the abnormality is classified and summarized with higher accuracy than the state of art techniques. This proposed method relies on heterogeneous features for classifying events with better structural indices. The proposed method's performance is analyzed using the metrics accuracy, false rate, analysis time, SSIM, and F1-Score.**

***Keywords*—Event detection; gated recurrent unit; summarization; intelligent transportation system**

I. INTRODUCTION

Road transportation is one of the cheapest and easiest among the other types of the transportation system. Many people around the world are traveling via road to travel from one place to another. Abnormal event detection is one of the critical tasks to perform in transportation systems [1]. Closer circuit television (CCTV) plays a major role in detecting events which are occurred on the roadside. A transport monitoring system plays a vital role in analyzing every detail which is occurred on the roadside and helps to protect people. Accurate abnormal event detection helps to reduce the crime rate and death rate on roadsides [2]. Abnormal events on roadsides are classified based on the physical attributes and behavior, postures, and gestures of the vehicle's position. The abnormal Event detection process is done based on two stages namely classification and video summarization process [3].

The video summarization process is done by analyzing the events which are occurred on the roadside based on certain keyframes or parameters from the given video clips. Keyframe plays a major role in the summarization process which helps to identify the exact features of the video which is done by comparing it with an important set of features [4]. The classification process is processed by combining both normal and abnormal events which are occurred on the roadside and then it produced a dataset that contains the cause of abnormal events in a detailed manner [5]. The Video Summarization process is used in every monitoring system to enhance the network by understanding the exact cause of events by analyzing the given set of videos [6]. The video summarization process helps to control accidents and crime on roadsides. A keyframe is generated to identify the exact actual cause of the events and it also helps to find out the upcoming events based on people's activities [7]. The machine learning algorithm is mostly used in the summarization process which helps to increase the accuracy rate in the detection process and also helps to reduce the time consumption rate in processing data [8]. A dynamic hierarchical clustering algorithm is used in the summarization process which is done by training the data which are captured by CCTV and producing trained data for further uses. It is done by combining the current clips or data with the previously collected data and generating detailed information which helps to prevent the upcoming accident [3, 9]. A reinforcement algorithm is also used here to identify the keyframes based on features such as gestures, signs, and postures of people and produce sequenced keyframes which help to reduce the crime rate on roadsides [10]. The main disadvantages of the video summarization cause contradiction between the spatiotemporal redundancies and sequence prediction. The proposed Additive Event Summarization Method (AESM) is used for projecting classified events through a gated recurrent unit. The proposed system is used to increase the chances of early classification due to limited state-based classification and summarization. The main advantages of the proposed system are to decrease the computational complexity caused by recurrent replication-based classification and reduce the sensitivity of the output. The experimental analysis of the proposed work is conducted using the dataset of UCSD to find the predominance compared to state of art techniques. The remaining sections of the paper are organized as follows. Section 2 presents the analysis of the related work with the merits and limitations. Section 3 presents the proposed work with a detailed mathematical

analysis of the summarization and classification process. Section 4 describes the experimental analysis of the proposed work. Section 5 concludes the paper with its contributions and the scope of the research.

II. Related Work

Yang et al. [11] proposed an algorithm for the learning model in the real-time event summarization process HRES. The learning model is proposed to capture the information which is stored in the knowledge base (KB) and implicitly the information based on the queries which are given to the users. The proposed HRES method improves the robustness and effectiveness and reduces the time consumption rate.

Wan et al. [12] proposed a long video retrieval algorithm based on a superframe segmentation process for ITS event detection. A long video stream is used to identify the unwanted frames which are present in the database and helps to reduce the unnecessary frame. The segment of Interest (SOI) is generated by using the superframe segmentation process. The proposed method increases the effectiveness by reducing the retrieval time.

Thomas et al. [13] proposed video summarization based on a perceptual model for the roadside event detection process. This method is used to find out the optimal solutions by analyzing the vast number of videos that are captured during accidents time. The surveillance camera is used here to capture video on the roadside. The proposed method increases the accuracy rate in the detection process.

Ji et al. [14,15] proposed a summarization method based on a multi-video by using archetypal analysis on a multi-modal weighted method. To create WAA weight, the multi-modal graph is used which is done based on the query. A multi-modal graph is used to fuse the information which is generated such as the tags, frames, and video clips for the prediction process. The proposed method outperformed the traditional summarization method by increasing the accuracy rate. The proposed method introduced a sparse coding framework for video summarization using query-aware. The proposed framework uses web images for identifying the exact information of the events. Unsupervised multi-graph fusion is used here to find out the keyframes which are available in the database based on the priority of the queries.

Elharrouss et al. [16] proposed multiple human action detection methods for the recognition and summarization process. Human activities are analyzed and generated into sequences to form a dataset. Then the sequence is divided into shots for the detection process. The histogram of oriented gradient (HOG) is framed based on the frames which are generated by based on the given video clips. The proposed method increases the efficiency and accuracy ofthe recognition and summarization process.

Zhang et al. [17] proposed a method that uses the key contents of the frames from the given video. A discriminator is used to find out the keyframes for the summarization process. The proposed approach increases the efficiency and accuracy rate.

Yang et al. [18] proposed a new framework using a deep neural network to leverage the benefits generated by the systems. It uses LSTM to represent the priority of the queries which are captured by the network. The proposed method increases the accuracy rate.

Gao et al. [19] proposed a key framework for the video summarization process of surveillance videos. Videos are sequenced based on the overlapping maps features. The clustering approach is used to finalize the key frames and generate an accurate set of frames for further use. The proposed method increases the performance and effectiveness of the system.

Lei et al. [20] have introduced a video summarization model using action parsing driven by a reinforcement algorithm. Action parsing is used to divide the videos into a sequenced part which is used in the final stage. The proposed system deals with recurrent neural networks used in the summarization process which selects the frames based on the actions and activities. The proposed method increases the accuracy rate and classification rate of key frames.

Ji et al. [21] have proposed a new video summarization method by combining a deep attentive ad semantic preserving approach. The Huber loss approach is used to replace the error loss which is occurred during summarization. A deep learning approach is used to ensure the security and safety of the keyframes. The proposed framework increases the performance and robustness of the system.

The proposed method is designed for mitigating the inconsistencies in the frame series detection process. In the MWAA process, the graph alignments are based on weights that imbalance the detection due to frame segregation. Contrarily, the sparse representations in the proposed ERA-SS increase the complexity due to multiple superframes. In this process, the computing time is hiked due to frequent switches over. Therefore, these drawbacks increase the difference in pixel representations, resulting in errors.

III. Proposed Method

The proposed method intakes video inputs for analyzing its sequence and event detection. The input videos are segregated as frames from which distinct features are extracted for analysis and classification. The data from external dataset is used for validating the proposed method. The input is split into different parts for individual processing as presented in the below Fig. 1. Based on the gate assignments for the observed variations are presented for analysis. In Fig. 1, the proposed method's process is illustrated.

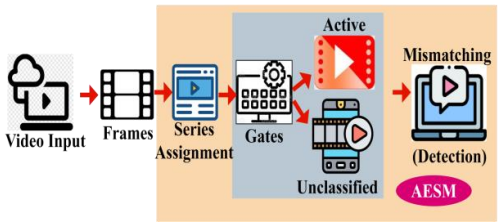


Fig. 1. Proposed Method.

In the series assignment (as in Fig. 1), the mixed heterogeneous features namely contrast (∇) and entropy ($t°$) are analyzed. First, these two features are extracted from the frames as defined in (1).

$$\left.\begin{array}{c}\Delta = \sum_{i=1,j=1}^{n} d_{i,j}, diff\ (i,j)^2 \\ and \\ E° = \sum_{i=1,j=1}^{n} -\ln(d_{i,j}).d_{i,j} \\ such \\ i,j \in \underbrace{n \times m}_{even}(or)\underbrace{m \times n}_{uneven}\end{array}\right\} \quad (1)$$

In equation (1) the variables $d_{i,j}$ and $(m \times n)$ represent the pixel density for a frame of size $(m \times n)$. Here i and j balance to m and n for uneven pixel frames or n and n for even pixel frames. The computation $diff(i,j)$ illustrates the variations between two consecutive pixels. Let T denote the time frame sequence for observing $d_{i,j}$ such that the series assignment is mapped as denoted in (2).

$$\left.\begin{array}{c}\forall\ d_{i,j} \in n\ or\ m, \Delta = d_{i,j} dT \\ provided\ i * j\ and\ \nabla = 1, if\ i = j \\ and \\ E° = \begin{cases} -\ln(d_{i,j}) \forall\ i * j \\ 0\ \forall i * j \end{cases} \\ Therefore \\ \nabla :: (i,j) \forall T = \begin{cases} 1 \\ o, otherwise \end{cases} \\ E° :: \begin{matrix}(i,j) \\ or \\ (j,i)\end{matrix}\ \forall\ T = \begin{cases} 1 \\ 0, otherwise \end{cases}\end{array}\right\} \quad (2)$$

This series assignment as in (3) is used for assigning gates in the learning process. This assignment requires $d_{i,j}$ based assignment in improving the fidelity of summarization. The contrary process of active (sequence) and unclassified is performed. For this purpose, we define current and update gates for state updates. This process is explained in the following subsection.

A. Event Classification

The events are classified by t_o variations in the observed sequences, for which the mapping in equation (2) is used. First, the gates are defined for sequence mapping as in equation (3).

$$\left.\begin{array}{c}\complement = \tan h\left[\frac{n(i,i)+n(j,j)}{n(i,j)} + \gamma_T \odot \frac{d(i,j)}{n(i,j)}\right] \\ and \\ \gamma_T = \frac{d(i,j)}{n(i,j)}.\frac{n(i,i)+n(j,j)}{n(i,j)} + \int d_{ij}\ \forall \begin{matrix} i \in n \\ j \in n \end{matrix} or\ m\end{array}\right\} \quad (3)$$

In equation (3), the variables $\complement$ and γ_T represents the current and update states at time T. Based on the further requirement, the gate states are changed and hence the classifications are performed. In the classification process, the mapping discreteness is observed for detecting active and unclassified series. This detection is performed until the end of the frame. If N is the end of the frame $\forall\ n, m \in N$, then:

$$\left.\begin{array}{c}\Delta_1 = 1 \\ \Delta_2 = 1 - \frac{\complement_1}{n(i,j)} - \frac{\complement_1}{\gamma_2} \\ \vdots \\ \underbrace{\Delta_N = 1 - \frac{\complement_{N-1}}{n(i,j)} - \frac{\complement_{N-1}}{\gamma_N}}_{for\ active\ N\ classification}\end{array}\right| \left.\begin{array}{c}\Delta_1 = 0 \\ \Delta_2 = \frac{\gamma_1}{N} + \frac{\gamma_1}{n(i,j)} \times \frac{1}{N} \\ \vdots \\ \underbrace{\Delta_N = \frac{\gamma_{N-1}}{N} + \frac{\gamma_{N-1}}{n(i,j)} \times \frac{1}{N}}_{for\ unclassified\ N}\end{array}\right\} \quad (4)$$

Based on the above classification, the mismatching and detection processes are differentiated. In this process, the $E°$ and ∇ based mismatching for mapped instances as in (2) is performed. The two conditions for ∇ and $E°$ based on T requires multi-feature analysis for a gate assignment. The similarity feature is verified for the stored and acquired features from the mapping as in (3).

$$\left.\begin{array}{c}S(x|i,y|j) = \frac{(2 \times \mu_n \mu_m + k_1)(2 \times \sigma_{nm} + k_2)}{(\mu_n^2 + \mu_m^2 + k_1)(\sigma_n^2 + \sigma_m^2 + k_2)} \\ where \\ \mu = \frac{1}{n-1}\sum_{i=1}^{n}(n-i)(m-j) \\ and \\ \sigma = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(i-j)^2} = \frac{1}{\sqrt{n-1}}\sum_{i=1}^{n}(i-j)^2 \forall\ i \neq j\end{array}\right\} \quad (5)$$

This similarity verification is performed for different $(i,j) \in (n,m) \in N$ wherein the mismatch for the above requires an alternate mapping such that $\complement$ is assigned with a new $n(i,j)$ and γ_T is updated for $(n-1)^{th}\ d(i,j)$. This means the variations in consecutive pixels are violated in detecting an event. The detection is performed in multiple intervals from 1 to N such that the mismatching $d(i,j)$ are segregated. A contrary part of the abnormal event detection is the synchronization of the $\left(1 - \frac{S}{N}\right)$ and mapping as in (2) for the different features. In the proposed method, the abnormal events at different T are considered non-cumulative (due to different occurrences). Therefore, the occurrences are synchronized based on $S(x|i,y|j) \forall n$ and m until N is achieved. This is validated for $d(i,j)$ such that the alternating sequences are varied until the end of classification. In Fig.2, the gate assignment and classification processes are illustrated.

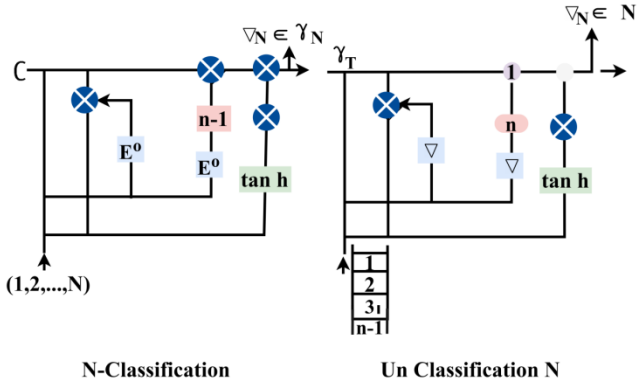


Fig. 2. Classification Process.

In the classification process, as presented in Fig. 2, the " × " and " + " symbols represent the product and sum of the mapping presented in equation (2). First, the product represents the $\nabla \otimes E° \; \forall T$ in 1 to n (or) 1 to m; the sum is the joint set of ∇ and $E°$. For an abnormal event summarization, the $T \notin (\nabla \odot E°)$ is segregated for classifying it as a whole interval. In contrast to the augmenting and mapping processes, the variations are detected for event detection and categorization. Therefore,

$$\left.\begin{array}{c} d_{i,j} \in T \notin (\nabla(\odot E°) \; is \; expected \; for \; T = 1 \\ rather, \\ (i,j) \in n \notin N \forall \nabla \; is \; achieved \; for \; T = o \\ such \; that \\ \Delta_N \in n(i,j) \forall \complement and \\ \Delta_N \in N \; \forall \; \gamma_T \; \mu \; is \; high \end{array}\right\} \quad (6)$$

In equation (6) the abnormal (Post the matching) T is identified for summarization. In the summarization process, the distinct event occurrences are augmented cumulatively. The differences are mitigated without augmenting the Δ_N as classified for $\gamma_T \in N$ and $S(x|i, y|i)$. The summarization process is described below.

B. Summarization Process

In the summarization process, the γ_T that encloses both $\Delta_N \in \gamma_T$ and $s(x|i, y|j)$ (failing) conditions are augmented based on T. This is either discrete/ sequential depending on multiple updates as in $E°$ and ∇ . The process requires unidentified $d(i,j)$ post the gate allocation for maximizing event aggregation. If the event is observed in $T \; \forall \; \Delta_N \in \gamma_T$ interrupts, then.

$$\left.\begin{array}{c} t_{i=1}^{T} = |\Delta_N^2 - \max\{\complement, E°\}|_{i=1 \; to \; T} \; and \\ (or) \\ t_{i=T-t}^{T} = \left|\frac{\Delta_N}{N} - \min\{E°, \gamma_T\}\right| \forall \; i = T - t \; to \; t \end{array}\right\} \quad (7)$$

In equation (7), the augmenting events classified for $i = t$ and $i - (T - t)$ are identified. Depending on the $\max\{\complement, E°\}$ and $\min\{E°, \gamma_T\}$ the abnormal classifications is grouped. This is required for projecting $d(i,j) \in T$ and hence the deviations are identified. The proposed method performs a cumulative augmentation of the above observation post $(\nabla \odot E°)$ assessment and hence the summary is an allocation of $d(i,j) \in$ distinct T and $(T - t)$. This is non-recurrent and hence new frames in T and $(T - t)$ (intermediate) are identified without false rates.

IV. Experimental Results and Discussion

This section discusses the proposed method's performance assessment using MATLAB simulations. The dataset from UCSD [22] is used for validating the proposed method's performance for accuracy, false rate, analysis time, SSIM, and F1-Score. The inputs are classified based on the available objects; the objects are used as in the dataset labeling. Depending on 4 textural features, the classification is performed; the pixel un-matching inputs are alone mitigated. In this comparative analysis, the identified objects and state updates are varied for the proposed and existing MVS-MWAA [14] and ERA-SS [12] methods. The data set provides multiple video frames observed at 30fps in 800x480 pixel resolution. A total of 9390 frames are observed in this dataset.

A. Accuracy

In Fig. 3, the accuracy for different objects and state updates is analyzed. The proposed method maximized accuracy by improving the γ_T and $E°$ detection in T and $\complement$ and Δ_N detection $\forall \; (T - t)$. This is non-recurrent based on the available $n(i,j)$ in multiple T such that accuracy is maximized. Another detection is the $s(x|i, y|j) \in N$ where multiple pixels with the observed features are validated. This is consistent for different objects identified in the frames wherein accuracy is high.

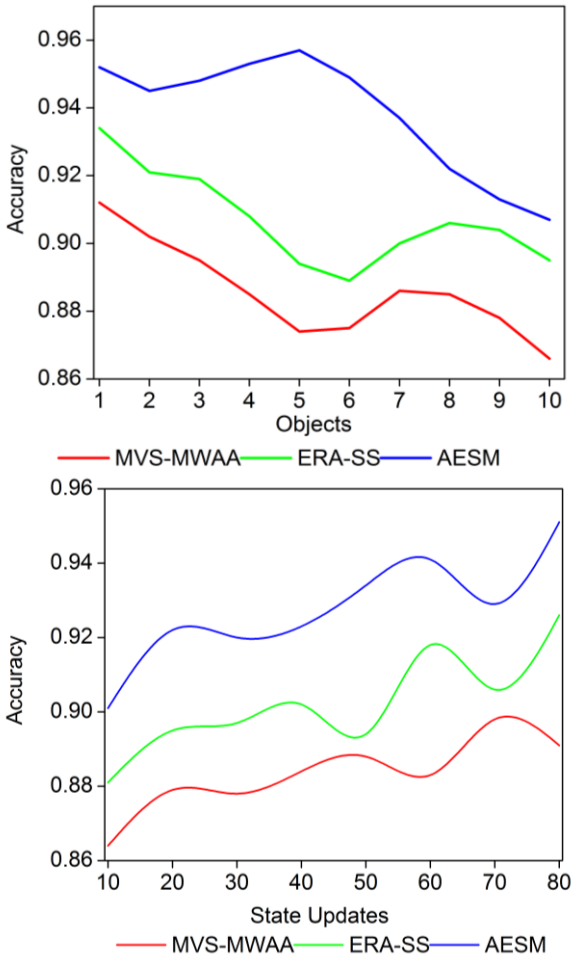


Fig. 3. Accuracy Analysis.

B. False Rate

The augmentation of $(\Delta \odot E°)$ and $(\nabla \cup E°)$ relies on multiple factors of $\complement$ and $E°$ such that no error arises. The proposed requires Δ_N classification based on $\complement$ and γ_T such that in t interference is avoided. In distinct instances, interferences are modeled independently.

The gate updates are non-linear $\forall \; S(x|i, y|j)$ for unclassified N such that C is high. In this process, the unclassified instances are reduced which achieves less False rate (refer to Fig. 4).

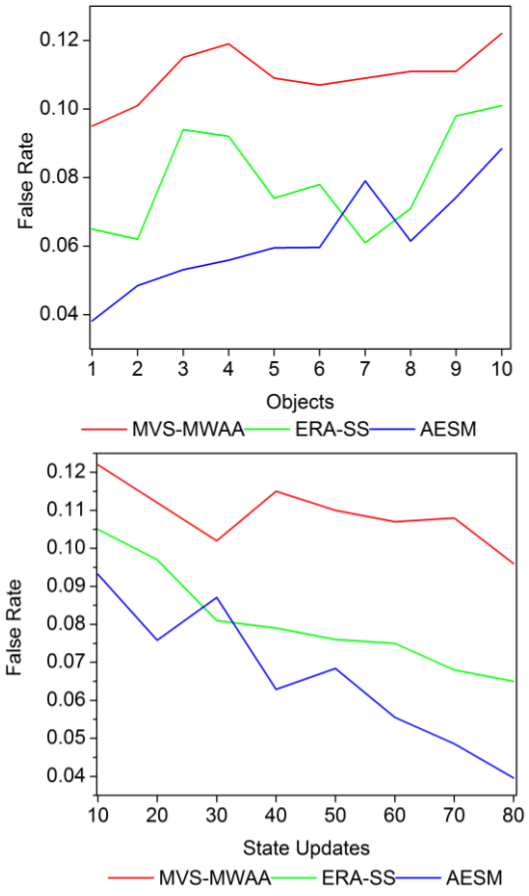
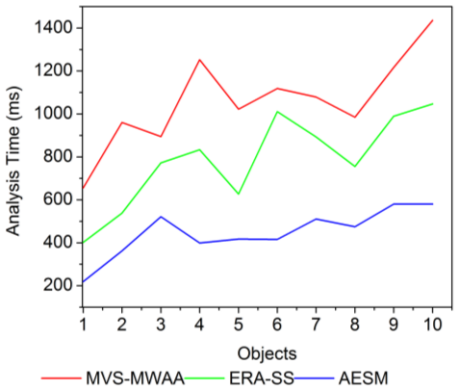


Fig. 4. False Rate Analysis.

C. Analysis Time

The proposed method achieves less analysis time as the proposed method classifies $\complement$ and $\gamma_T \forall\, d(i,j)$. In the active classification and (μ, σ) estimation, independent assessments are performed.

These are validated based on the mapping and hence $(T - t)$ and T as independent rather than cumulative and joint analysis. Therefore, the proposed method achieves less analysis time for identified objects and state updates (Fig. 5).



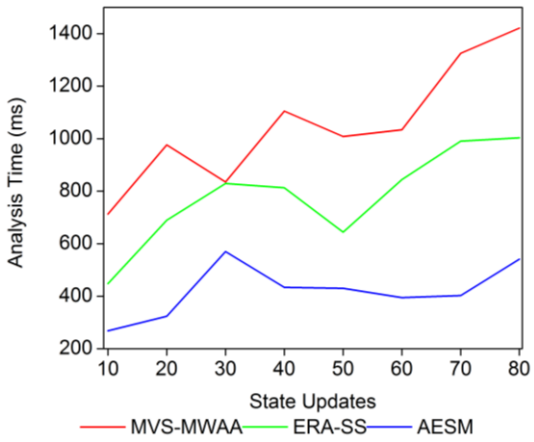


Fig. 5. Analysis Time.

D. SSIM

In Fig. 6, the SSIM for the proposed method is compared for different objects and state updates.

In multiple state updates, the classification is performed under different N. These classifications are performed for $\Delta_N \in \gamma_N$ and $\Delta_N \in N$ for detecting multiple SSIM for $d(i,j)$ such that ∇ is achieved in different $(T - t)$.

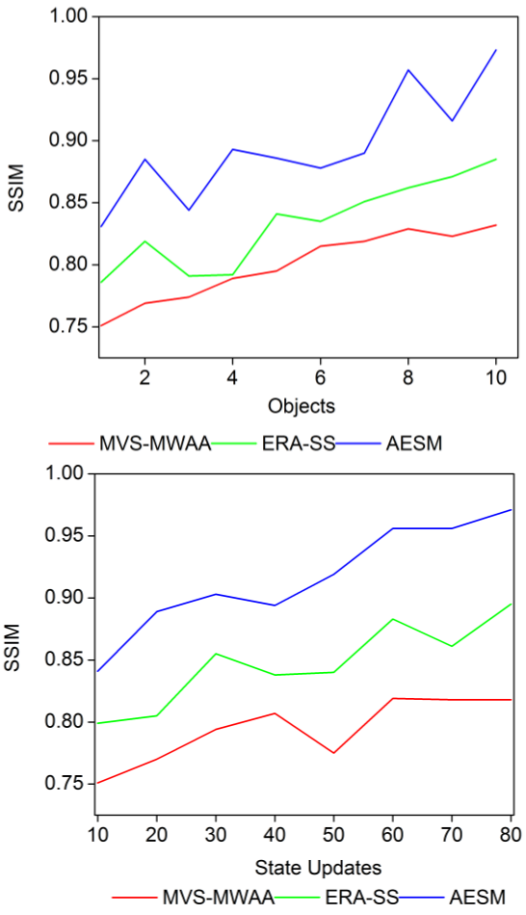


Fig. 6. SSIM Analysis.

This is unanimously observed for distinct intervals and objects whereγ_T is less, maximizing SSIM.

E. F1-Score

For any density of objects and state updates, the F1-score is high for the proposed method (Fig. 7). The proposed method achieves a high F1score by mitigating $(\nabla \cup E^{\circ})$ instances. This is performed based on the classification and mapping of distinct $d(i,j)$. In the classification process, $(T-t)$ and T instances are distinguished for maximizing the F1-score, preventing false rate. The comparative analysis results are tabulated in Tables I and II for different objects and state updates.

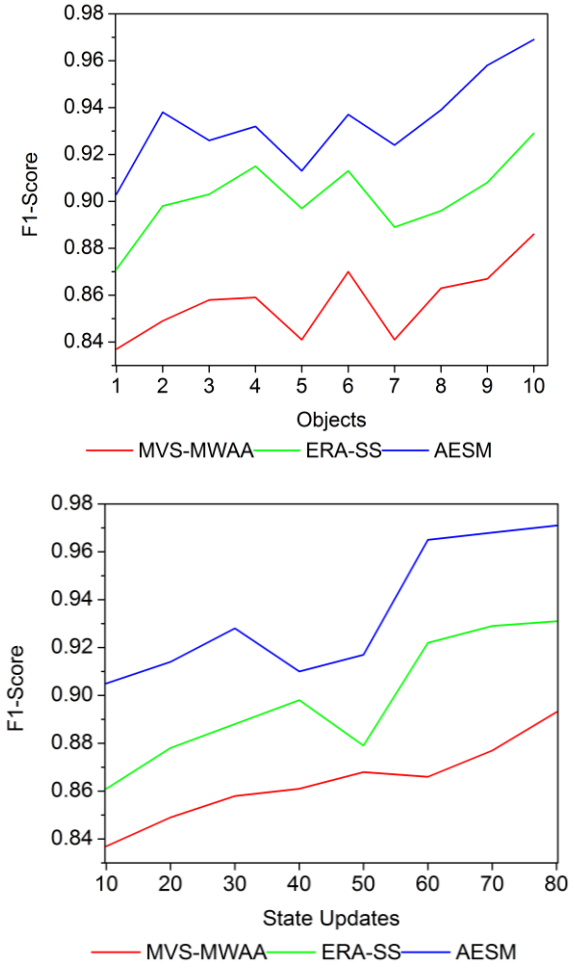


Fig. 7. F1-Score Analysis.

TABLE I. COMPARATIVE ANALYSIS RESULTS FOR OBJECTS

Metrics	MVS-MWAA	ERA-SS	AESM	Findings
Accuracy	0.866	0.895	0.907	8.83% High
False Rate	0.122	0.101	0.0884	7.7% Less
Analysis Time (ms)	1436.06	1046.51	580.226	8.88% Less
SSIM	0.832	0.885	0.973	11.45% High
F1-Score	0.886	0.929	0.969	12.3% High

TABLE II. COMPARATIVE ANALYSIS RESULTS FOR STATE UPDATES

Metrics	MVS-MWAA	ERA-SS	AESM	Findings
Accuracy	0.891	0.926	0.951	7.08% High
False Rate	0.096	0.065	0.0396	6.82% Less
Analysis Time (ms)	1420.73	1003.2	541.227	9.22% Less
SSIM	0.818	0.895	0.971	11.55% High
F1-Score	0.893	0.931	0.971	11.8% High

The significance of the proposed method is adaptable for varying objects and computations (state updates). In a video processing, the variations due to objects and computations are practically addressed using this proposed method. The variations are suppressed using the gate assignment and hence the accuracy, SSIM, and F1-Score are improved.

V. CONCLUSION

This article discussed an additive event summarization method for reducing the inconsistencies in video event summarization. The classification events are identified using different state assignments through gated recurrent units. The recurrent unit identifies unclassified and active frames for preventing false rates in event extraction. This classification is performed based on the heterogeneous features over the varying pixel densities over different sequences. From the analyzed sequences, the abnormal feature exhibiting pixels are segregated for providing a summarized output. For the different objects classified, the proposed method achieves 8.83% high accuracy, 11.45% high SSIM, 12.3% high F1-score, 7.7% less false rate, and 8.88% less analysis time. Though the proposed method is reliable in summarizing event related to abnormal occurrences the varying textural features result in pixel errors. Therefore, a spatiotemporal feature classification pre-processing is planned to be integrated in the future work.

REFERENCES

[1] Wang, H., Feng, J., Sun, L., An, K., Liu, G., Wen, X., ...& Chai, H. (2020). Abnormal Trajectory Detection Based on Geospatial Consistent Modeling.IEEE Access, 8, 184633-184643.

[2] Wang, X., Song, H., & Cui, H. (2018). Pedestrian abnormal event detection based on multi-feature fusion in traffic video. Optik, 154, 22-32.

[3] Wu, J., Xu, H., Zheng, Y., &Tian, Z. (2018). A novel method of vehicle-pedestrian near-crash identification with roadside LiDARdata. Accident Analysis & Prevention, 121, 238-249.

[4] Alhussain, T. (2021). Density-scaling traffic management for autonomous vehicle environment—predictive learning-based technique.Soft Computing, 1-15.

[5] Jiang, Z., Liu, Y., Fan, X., Wang, C., Li, J., & Chen, L. (2020). Understanding urban structures and crowd dynamics leveraging large-scale vehicle mobility data.Frontiers of Computer Science, 14(5), 1-12.

[6] Yang, L., & Yang, N. (2019). An integrated event summarization approach for complex system management.IEEE Transactions on Network and Service Management, 16(2), 550-562.

[7] Muhammad, K., Hussain, T., Del Ser, J., Palade, V., & De Albuquerque, V. H. C. (2019). DeepReS: A deep learning-based video summarization strategy for resource-constrained industrial surveillance scenarios. IEEE Transactions on Industrial Informatics, 16(9), 5938-5947.

[8] Fei, M., Jiang, W., & Mao, W. (2021). Learning user interest with improved triplet deep ranking and web-image priors for topic-related video summarization.Expert Systems with Applications, 166, 114036.

[9] Lan, L., & Ye, C. (2021). Recurrent generative adversarial networks for unsupervised WCE video summarization.Knowledge-Based Systems, 222, 106971.

[10] Zhang, J., Shi, Y., Jing, P., Liu, J., & Su, Y. (2019). A structure-transfer-driven temporal subspace clustering for video summarization.Multimedia Tools and Applications, 78(17), 24123-24145.

[11] Yang, M., Qu, Q., Shen, Y., Zhao, Z., Chen, X., & Li, C. (2020). An Effective Hybrid Learning Model for Real-Time Event Summarization.IEEE Transactions on Neural Networks and Learning Systems.

[12] Wan, S., Xu, X., Wang, T., &Gu, Z. (2020). An intelligent video analysis method for abnormal event detection in intelligent transportation systems.IEEE Transactions on Intelligent Transportation Systems.

[13] Thomas, S. S., Gupta, S., & Subramanian, V. K. (2017). Event detection on roads using perceptual video summarization.IEEE Transactions on Intelligent Transportation Systems, 19(9), 2944-2954.

[14] Ji, Z., Zhang, Y., Pang, Y., Li, X., & Pan, J. (2019). Multi-video summarization with query-dependent weighted archetypal analysis.Neurocomputing, 332, 406-416.

[15] Ji, Z., Ma, Y., Pang, Y., & Li, X. (2019). Query-aware sparse coding for web multi-video summarization.Information Sciences, 478, 152-166.

[16] Elharrouss, O., Almaadeed, N., Al-Maadeed, S., Bouridane, A., &Beghdadi, A. (2021). A combined multiple action recognition and summarization for surveillance video sequences. Applied Intelligence, 51(2), 690-712.

[17] Zhang, Y., Kampffmeyer, M., Liang, X., Zhang, D., Tan, M., & Xing, E. P. (2019). Dilated temporal relational adversarial network for generic video summarization.Multimedia Tools and Applications, 78(24), 35237-35261.

[18] Yang, M., Tu, W., Qu, Q., Lei, K., Chen, X., Zhu, J., &Shen, Y. (2019). MARES: multitask learning algorithm for Web-scale real-time event summarization. World Wide Web, 22(2), 499-515.

[19] Gao, Z., Lu, G., Lyu, C., & Yan, P. (2018). Key-frame selection for automatic summarization of surveillance videos: a method of multiple change-point detection. Machine Vision and Applications, 29(7), 1101-1117.

[20] Lei, J., Luan, Q., Song, X., Liu, X., Tao, D., & Song, M. (2018). Action parsing-driven video summarization based on reinforcement learning.IEEE Transactions on Circuits and Systems for Video Technology, 29(7), 2126-2137.

[21] Ji, Z., Jiao, F., Pang, Y., & Shao, L. (2020). Deep attentive and semantic preserving video summarization.Neurocomputing, 405, 200-207.

[22] https://gram.web.uah.es/data/datasets/rtm/index.html