International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

Video Summarization Based on Gaussian Mixture Model and Kernel Support Vector Machine for Forest Fire Detection

B. Pushpa, M. Kamarasan

Abstract— Exponential growth in the generation of multimedia data especially videos resulted to the development of video summarization concept. The summary of the videos offers a collection of frames which precisely define the video content in a considerably compacted form. Video summarization models find its applicability in various domains especially surveillance. This paper intends to develop a video summarization technique for the application of forest fire detection. The proposed method involves a set of processes namely convert frames, key frame extraction, feature extraction and classification. Here, a Merged Gaussian Mixture Model (MGMM) is applied for the process of extracting key frames and kernel support vector machine (KSVM) is employed for classifying a frame into normal frame and forest fire frame. The simulation analysis is performed on the forest fire video files from FIRESENSE database and the results are assessed under several dimensions. The final outcome proves the efficiency of the presented MGMM-KSVM model in a considerable way.

Keywords— KSVM Classification; Forest fire; Keyframe extraction; MGMM.

I. INTRODUCTION

Recent developments in the storage space and advanced imaging devices leads to the exponential increase in the massive quantity of videos. These huge quantities of videos are uploaded to various websites like Netflix, Hot star and so on at every second. Since numerous recommendations will be given to every searching content, identifying the particular videos from the wide-ranging consumes more time. It is also difficult to retrieve the videos at a faster rate. For addressing this issue, several works has been done on video summarization for provides a summary of the whole video in a small duration. Video summarization offers the advantage of fast browsing of massive amount of videos and highly efficient of indexing as well as accessing the videos. The summary will be created by the selection of key frames which optimally define the video. The extraction of key frames from the detection of change point, low level features or clustering based on objects. The key frames are advantageous to index videos; however, they are cancelled of motion information.

It limits the usage of particular retrieval processes. In case of video abstraction, video is divided into segments, and more important and interesting segments are selected for a shorter form. Small portions of the videos will be chosen for summarization. On the selection of portions, more attention will be given which defines the whole video and also fascinating to the user. However, the key frames are adaptable to the devices with restricted bandwidth and it offers the entire substance of the video in less number of frames.

The summary created from the video are generally in the form of static key frame sets [1], or dynamic video summaries are the concatenation of audio-visual segments extracted from the original videos. Additionally, some models assume the time stamp of the frames or solely content. Time aware models might comprise a frame which is identical to a present key frame when it defines a shot distinct in time. Time-oblivious models eliminate many upcoming frames which are identical to the available frame present in the past frames of the video. The suitable option for the form of the summary is based on the application. Video summarization is to process video sequence with more interesting, valuable and useful to the user. The major task in video summarization is to segment the original video into shots and extract those video frames from the original video that would be most informative and concise representation of the whole video. Several video summarization models have been developed for various domains.

Fig. 1. Forest Fire Images

The way of acquiring the attention by the user is employed for generating a summary. N. Ejaz [1] the visual attention modeling schemes have proved to be effective in video summarization.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
B. Pushpa*, Department of Computer and Information Science Annamalai University
M. Kamarasan, Department of computer and Information Science Annamalai University

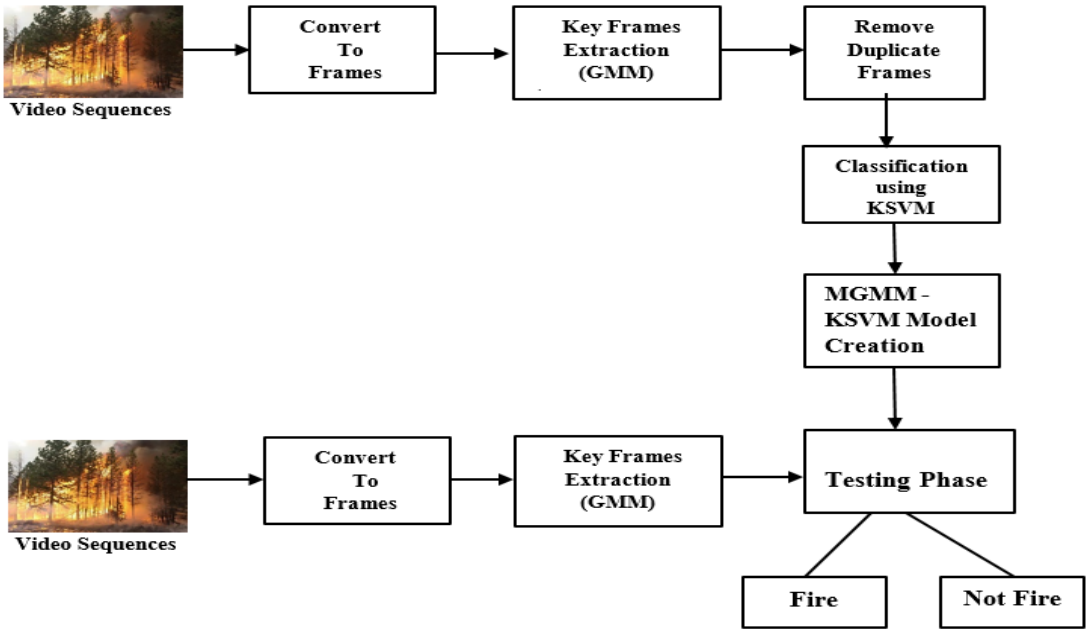

Fig. 2. Overall Process of Proposed Method

Then, Y.F.Ma [2] designed an attention curve based visual saliency detection to select key frames in a video. Through the acquiring process of the physiological responses of viewer, a temporal position of the most important sub-portions of the video is recognized in an automatic way [3, 4], Cong, J. Yuan [5] considered the problem of summarizing videos as a dictionary selection problem by the use of sparsity consistency. G.Evangelopoulos[6]the extension of the concept of audio-visual curve depending upon the way of summarizing movies by incorporating the textual cue for recognizing important actions that can be utilized to generate a summary of video. S.D.Thepade [7] make use of discrete cosine transform coefficients of every frame and is applied to the frame retrieval process with high video information. H.Liu [8] presented a method to make use of the low level features of image for key frame retrieval. The limitation is that it eliminates the high-level semantic information about the video. In S. D. Thepade and A. A.Tonge [9], a content based video retrieval process is applied for retrieving the key frames. For improving the scalability and response time, temporal sparse method comprising the detection of key frames is applied. But, this approach is not robust to spatial editing and therefore degrades the results. Clustering is also an alternate way of choosing the key frames. J.Peng [10] determines the color histogram initially which is applied for frame clustering. Then, the important frame from every cluster is chosen as a key frame. The weakness of the approach is the temporal order of the key frames will be vanished as K-clustering technique is applied.

Several applications of video summarization exist in the literature. In this study, we have focused on the issue of forest fire detection through a video summarization method. Forest fires define a continuous danger to environmental systems, infrastructure and lives of human being. The major intention to reduce damage lies in the earlier identification and proper reaction at a faster rate. More works has been carried out to recognize the fire at the earlier stage. Some kinds of forest fire are shown in Fig. 1.

The present paper intends to develop a video summarization technique for the application of forest fire detection. The proposed method involves a set of processes namely frame splitting, key frame extraction, feature extraction and classification. Here, a Merged Gaussian Mixture Model (MGMM) is applied for the process of extracting key frames and kernel support vector machine (KSVM) is employed for classifying an image into normal image and forest fire image. The simulation analysis is performed on the forest fire dataset and the results are assessed under several dimensions. The final outcome proves the efficiency of the presented MGMM-KSVM model in a considerable way.

II. PROPOSED METHOD

The overall working principle of the presented MGMM-KSVM model is shown in Fig. 2. As shown, several sub-processes exist. Among various processes, key frame extraction using MGMM and classification using KSVM plays a vital role in the overall identification process. Initially, the video sequence will be provided as input to the presented model. The conversion of videos to a collection of frames takes place. Then, key frames are extracted from the total number of frames which will be useful to generate a video summary. After the key frames are extracted, some of the duplicate frames will be discarded. By using of features each key frame will be selected and the set of features are color - color autocorrelgram, spatial analysis, temporal analysis, shape, motion, intensity, dynamic texture analysis. Then, KSVM model will be invoked to classify the images and it builds a model. Once the training model gets completed, testing of input images can takes place. To test an image, MGMM based key frame extraction takes place. Finally, the input images undergo classification whether the image includes the presence of fire or not.

The principle for merging Gaussian components works as follows:

1. Start with all components of the initially estimated Gaussian mixture as current clusters.
2. Find the pair of current clusters most promising to merge.
3. Apply a stopping criterion to decide whether to merge them to form a new current cluster, or to use the current clustering as the final one.
4. If merged, go to 2.

A. *MGMM for feature extraction*

The provided raw video in the feature space is considered as a high dimensional "feature process." At time t, GMM treats every feature x_t as an integration of n-dimensional data points, can figure out probability of it belonging to each of the clusters K Gaussian components, i.e.

$$x_t = \sum_{k=1}^{K} \omega_t \eta(X_t, \mu_k, \Sigma_k) \qquad (1)$$

where ω_k, μ_t and Σ_k indicates the weight, mean and covariance matrix of kth component respectively. Compute probability of each point belonging to each of the k clusters. The probability density function (pdf) is represented as follows

$$\eta(x, \mu, \Sigma) = \frac{1}{(2\pi)^n |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \qquad (2)$$

The covariance matrix Σ is an $n \times n$ matrix that needs more training for estimation. For simplifying the learning procedure, it is assumed as follows.

$$\Sigma = \sigma^2 I \qquad (3)$$

It considers that every dimension of the feature holds similar variance. Though it is not practical, it is employed as an approximation in the high dimensional space and outcome is acceptable. For the estimation of the above mentioned parameters are quite satisfactory.

At every input feature x_t at time t, the distance from x_t to mean μ_k for every component is determined to identify a matched component. The matching takes place when the distance from x_t and μ_k is lesser than thrice the variance σ_k, i.e.

$$(x_t - \mu_k)^T (x_t - \mu_k) < 3\sigma_k \qquad (4)$$

The weights gets updated as

$$\omega_k, t = (1-\alpha)\omega_{k,t-1} + \propto M_{k,t} \qquad (5)$$

Where α represents fixed learning rate, $M_{k,t}$ equals to 1 for matched component or 0 otherwise. The other variables are also updated as follow.

$$\mu_{k,t} = (1-\rho)\mu_{k,t-1} + \rho x_t \qquad (6)$$

$$\sigma_{k,t}^2 = (1-\rho)\sigma_{k,t-1}^2 + \rho(x_t - \mu_{k,t})^T (x_t - \mu_{k,t}) \qquad (7)$$

Where ρ is the second learning rate and

$$\rho = (\alpha \eta x_t, \mu_k, \Sigma_k) \qquad (8)$$

When no matching components are identified, then the one with least possibility gets updated as $\mu_k = x_t$ with a lower weight and higher variance.

The MGMM can be update to explain the distribution of a data stream. But, instead of using a fixed number of clusters, this method enables fresh ones to be included upon needed and also offers a way to combine identical clusters in an equivalent way. The MGMM model is applied to cluster the generic data. In case of applying MGMM model to the video summarization models, an extra process of selecting a representative from every cluster as a key frame. In every level of processing, the frame located nearest to every cluster mean is saved as the present key frame. The frames might undergo replacement when a succeeding frame is nearer to the mean. Since the cluster mean is not static, the last set of key frames might not be the optimal ones which will be selected when the entire dataset is placed in a storage area.

B. *KSVM*

Linear SVM holds the drawback on linear hyperplane that could not split the difficult distributed realistic data. For generalizing it to non-linear hyperplane, the kernel concept is employed to SVM [18]. The resultant model is technically identical instead that every dot product undergo replacement with the nonlinear kernel function. At the same time, the KSVM enables to fix the maximum-margin hyperplane in a converted feature space. The conversion might be nonlinear, and the converted space could be high dimensional. Hence, even though the classifier is a hyperplane in the high-dimensional feature space, it could be non-linear in the imaginative input space. In case of every individual kernel, a minimum of one modifiable variable is needed to make the kernel flexible and tailor itself to realistic data. This paper makes use of Radial Basis Function (RBF) kernel is because of its outstanding results. The kernel can be defined as given in Eq. (9):

$$k(x_a, x_b) = exp\left(-\frac{\|x_a - x_b\|}{2\sigma^2}\right) \qquad (9)$$

Then, Eq. (9) will be applied to Eq. (10) to derive the final training unction of SVM in Eq. (11):

$$\max_{\alpha} \sum_{b=1}^{N} \alpha_b - \frac{1}{2}\sum_{b=1}^{N}\sum_{a=1}^{N} \alpha_a \alpha_b y_a y_b k(x_a, x_b) \qquad (10)$$

$$\max_{\alpha} \sum_{b=1}^{N} \alpha_b - \frac{1}{2}\sum_{b=1}^{N}\sum_{a=1}^{N} \alpha_a \alpha_b y_a \alpha_b \, exp\left(-\frac{\|x_a - x_b\|}{2\sigma^2}\right) \qquad (11)$$

It is still a quadratic programming issue and interior point approach is used to resolve it. The major benefit of the dual form function given in Eq. (10) is that the slack variables vanish from the dual problem, with the constant C appearing only as an additional constraint on the Lagrange multipliers.

III. EXPERIMENTAL RESULTS ANALYSIS

A. *Dataset*

For ensuring the optimal performances of the applied MGMM-KSVM model for the identification of forest fire, a detailed validation takes place on a FIRESENSE database [11] and videos are gathered from the Internet sources.

Eleven videos containing actual fire and ten videos containing flame-colored moving objects were used for the evaluation.

In addition, a set of sample frames under normal and forest fire classes are shown in Fig. 3. As shown in Fig. 3, it can be seen that forest fire frames are provided in Fig. 3a and the non-fire frames are given in Fig. 3b.

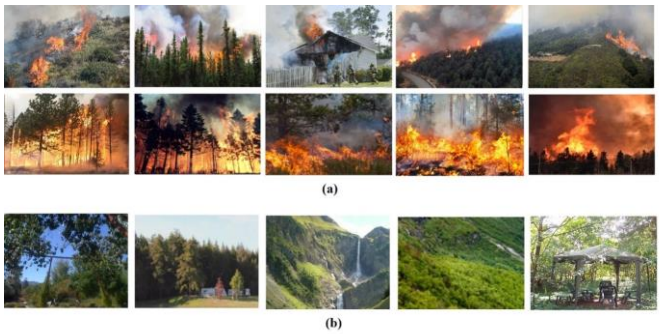

Fig. 3. (a) Forest Fire Frames (b) Non-fire Frame

B. Results analysis

Fig. 4 displays the number of frames extracted regarding the Forest Fire Images from Video Sequences. Fig. 5 illustrates an example to understand the results of the key extraction process. In the Fig. 5a, the key frames are provided. Then, the redundant key frames are avoided as shown in Fig. 5b.

Fig. 4. Frames Extracted from Video Sequences of Forest Fire Images

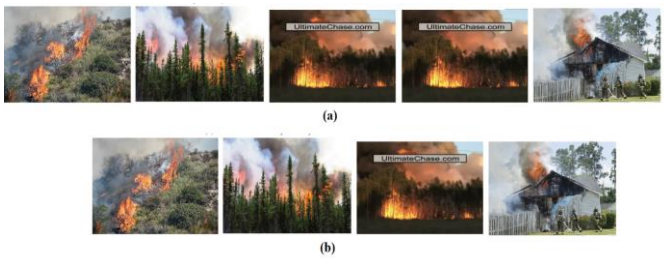

Fig. 5. a) Key Frame Extraction b) Eliminate of Redundant Frames

Table 1 provides the results attained by the presented MGMM-KSVM based fire detection model under three measures. Fig. 6 shows the investigation of the precision values attained by various models for forest fire detection. The figure clearly exhibited that the MGMM-KSVM shows effective detection with the maximum precision value of 89.98. At the same time, the rule based model shows almost closer detection performance to MGMM-KSVM by achieving a higher precision value of 86.76. In line with, it is also shown that the NN model tries to handle the detection process, but ended with the moderate precision value of 85.40. Simultaneously, NB shows poor forest fire detection with the lower precision value of 84.89. However, the existing DT model shows inappropriate detection with the lowest precision value of 83.90. As a whole, better detection rate is achieved by the MGMM-KSVM model with the highest precision value.

TABLE I

COMPARISON OF PROPOSED WITH EXISTING METHODS

Methods	Precision	Recall	Accuracy
Proposed	89.98	90.82	90.56
Rule Based	86.76	88.53	87.23
Neural Network [NN]	85.40	86.18	85.78
Naïve Bayes [NB]	84.89	85.39	85.14
Decision Tree [DT]	83.90	81.23	82.35

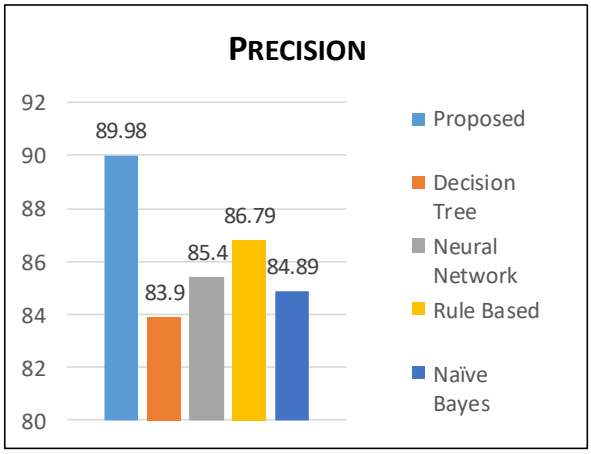

Fig. 6. Analysis of various models interms of precision

Fig. 7 displays the comparison results of diverse models attained interms of recall for forest fire detection. The figure clearly exhibited that the MGMM-KSVM shows effective detection with the maximum recall value of 90.82. At the same time, the rule based model shows almost closer detection performance to MGMM-KSVM by achieving a higher recall value of 88.53. In line with, it is also shown that the NN model tries to handle the detection process, but ended with the moderate recall value of 86.18. Simultaneously, NB shows poor forest fire detection with the lower recall value of 85.39. However, the existing DT model shows inappropriate detection with the lowest recall value of 81.23. In short, better detection rate is achieved by the MGMM-KSVM model with the highest recall value over other methods.

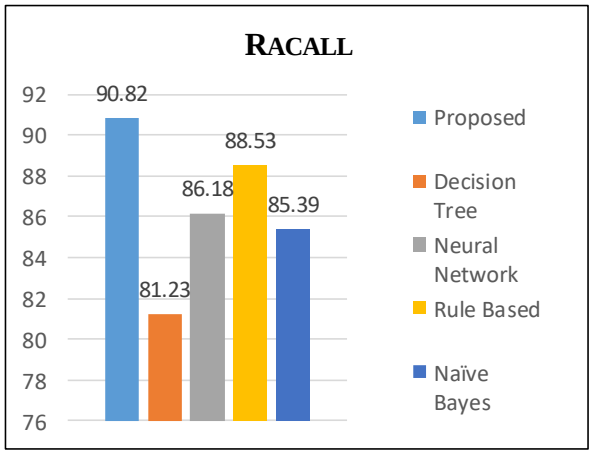

Fig. 7. Analysis of various models interms of recall

An important metric which properly assess the detection rate of the forest fire is accuracy. The analysis of the performance interms of accuracy is depicted in Fig. 8. The figure clearly exhibited that the MGMM-KSVM shows effective detection with the maximum accuracy value of 90.56.

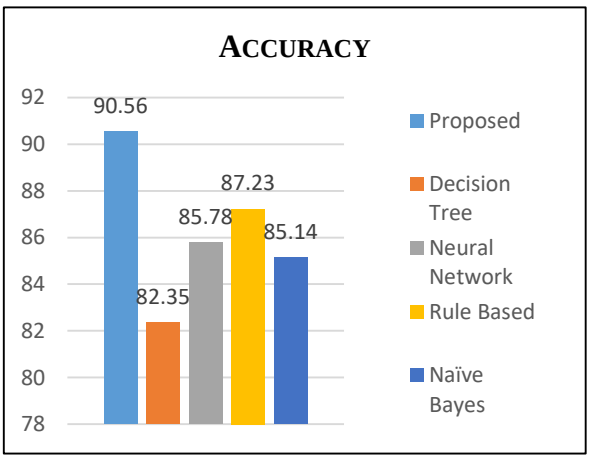

Fig. 8. Analysis of various models interms of accuracy

At the same time, the rule based model shows almost closer detection performance to MGMM-KSVM by achieving a higher accuracy value of 87.23. In line with, it is also shown that the NN model tries to handle the detection process, but ended with the moderate accuracy value of 85.78. Simultaneously, NB shows poor forest fire detection with the lower accuracy value of 85.14. However, the existing DT model shows inappropriate detection with the lowest accuracy value of 82.35. In short, better detection rate is achieved by the MGMM-KSVM model with the highest accuracy value over other methods. As a whole, the entire results analysis section strongly pointed out that the presented MGMM-KSVM model is an efficient choice for forest fire detection.

IV. CONCLUSION

Video summarization offers the advantage of fast browsing of massive amount of videos and highly efficient of indexing as well as accessing the videos. In this paper, we have focused on the problem of forest fire detection through a video summarization method. The proposed method involves a set of processes namely convert frames, key frame extraction, feature extraction and classification. Here, MGMM is applied for the process of extracting key frames and KSVM is employed for classifying an image into non-fire frame and forest fire frame. The MGMM-KSVM model shows its superiority over other methods with the maximum precision of 89.98, recall of 90.82 and accuracy if 90.56 respectively. In addition, future work is focused to analyze more parameters to get better performance of the proposed method.

REFERENCES

1. N. Ejaz, I. Mehmood and S. W. Baik, Efficient Visual Attention Based Framework for Extracting Key Frames from Videos, Signal Processing: Image Communication, vol. 28(1), pp. 34–44, (2013).
2. Y. F. Ma, X. S. Hua, L. Lu and H. J. Zhan, A Generic Framework of User Attention Model and its Application in Video Summarization, IEEE Transactions on Multimedia, vol. 7(5), pp. 907–919, (2005).
3. C. Chˆenes, G. Chanel, M. Soleymani and T. Pun, Highlight Detection in Movie Scenes Through Inter-Users, Physiological Linkage, In Social Media Retrieval, Springer, pp. 217–237, (2013).
4. A. G. Money and H. Agius, Elvis: Entertainment-Led Video Summaries, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 6(3), pp. 17, (2010). [8] Y.
5. Cong, J. Yuan and J. Luo, Towards Scalable Summarization of Consumer Videos Via Sparse Dictionary Selection, IEEE Transactions on Multimedia, vol. 14(1), pp. 66–75, (2012).
6. G. Evangelopoulos, A. Zlatintsi, G. Skoumas, K. Rapantzikos, A. Potamianos, P. Maragos, Video Event Detection and Summarization using Audio, Visual and Text Saliency, In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3553–3556, (2009).
7. S. D. Thepade and A. A. Tonge, Extraction of Key Frames from Video using Discrete Cosine Transform, In International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp. 1294–1297, (2014).
8. H. Liu and T. Li, Key Frame Extraction based on Improved Frame Blocks Features and Second Extraction, In IEEE 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 1950–1955, (2015).
9. S. D. Thepade and A. A. Tonge, An Optimized Key Frame Extraction for Detection of Near Duplicates in Content based Video Retrieval, In IEEE International Conference on Communications and Signal Processing (ICCSP), pp. 1087–1091, (2014).
10. J. Peng and Q. Xiao-Lin, Keyframe-Based Video Summary using Visual Attention Clues, IEEE MultiMedia, (2), pp. 64–73, (2009)
11. FIRESENSEdatabase: https://zenodo.org/record/836749.

AUTHORS PROFILE

Mrs.B.Pushpa, received her M.C.A., Degree in Computer Application from Bharathidhasan University, India, in 1998 and M.phil in Computer Science from Manonmaniam Sundaranar University, India in 2006. She is currently doing research in Computer Science and Faculty Member in the Department of Computer and information Science, Faculty in science, Annamalai University, India. Her research area Digital Image Processing and Remote Sensing.

Dr. M.Kamarasan, received the MCA Degree from Annamalai University. Chidambaram, India, in 1999 and M.phil in computer science from Annamalai University in 2005 and PhD in Computer Science from Annamalai University in 2015. He is a Assistant Professor of Department of Computer and Information Science, Annamalai University, Tamilnadu, India. His research interest includes Multimedia. Data Mining and Image Analysis, Scene analysis, and Computer Vision and Machine learning. He also life Member of Computer Society of India.