Design Prognostic Framework for Scene Classification in Video Processing

Mr. Ajay B Gadicha¹, Dr. M.V.Sarode², Dr. V.M.Thakare³

¹Research Scholar, Department of Computer Science & Engg, P.R.Pote (Patil) College of Engineering & Management
²Professor and Head, Department of Computer Engg, Government Polytechnic, Yavatmal.
³Professor and Head, P.G. Department of Computer Science, Sant Gadge Baba Amravati University, Amravati (MH)

ajjugadicha@gmail.com¹

Abstract-- This paper provides the technique over selecting phenomenal frames extraction structure using SBD, KFE, and Semantic mechanism which quite easy to extract the valuable scene from the entire video. The proposed system will concentrates on findings the accuracy, precision and recall measure of inputted video from TRACEVID dataset. Here the proposed system signifies the use of thresholding technique and TBD of individual frame and calculate the various parameters like standard deviation, Mean etc. In order to legalize this claim, content based video reclamation systems were furnished using color histogram, features extraction and different approaches are applied for the supervision of the semantic temperament of each frame in the video. Also it gives major idea about the classification of the scene and algorithm which generates significant result.

Keywords: VSBD, SBD, KFE, TBD

1. INTRODUCTION

In the present days everything is shrinking, whole contents are available online or on social media hence there is some need of reducing or shorten the content so that user get benefited and saves actual time. Compression concerning the statistics is positive over the essential steps into in recent times specifically records era. In any video, lots of digital data impact more on social platform and generate significant information which might be concerning no uses. Figure 1 depicted the working of video summarization, the video is divided into shots and scene further it divided into reference frame i.e. first frame of video and another called general frame. If these contents are extracted from the entire video then we can engrave the video or lessen the video. This makes the video smaller but in the end requires an awful about statistics i.e. the viewer won’t be able to spent much of their time to browse the entire content of video and process the digital content fast. Reading positive kind regarding video with lesser quantity concerning frames can also improve the general overall performance of the guideline assorted. Each and Every frame stuck between the video as include important facts are called key frames. This important frame is the summary of the video characteristics. The criteria for scene technique are certain at the algorithm for extracting the solution meaningful content from the video. Many researcher works on video processing to reduce the work of summarization. The video comprises of various frames, shots and scene, but user don’t have time to view each and every aspect of video hence we need to execute some mechanism to balance the time and generate the efficient histogram to get actual content and result. A video summarization is basically divided into two broad categories. First Video abstraction and second is video synopsis. These both version of VSUMM uses in the detection of frontiers in shot and scene and rectify the actual content from the video.
A. VSU MM SHOT BOUNDARY DETECTION (VSBD):
In this method a frame is depicted into blocks with ‘x’ rows and ‘y’ columns, and then the divergence of the corresponding blocks among two chronological frames is computed. In the end, the very last difference of frames is acquired via manner of merging remaining frames of the variations. Below stated algorithm will give an explanation for the shot frontiers boundary as follows:

Algorithm 1: VSU MM Shot Frontiers Detection

1: Let $M(f_i)$ be the $f_i$th frame in video sequence, where

   $F_1, 2, \ldots, F_g$ ($F_g$ refer to the whole variety of Video sequence)

2: Segregation of respective block as of frame with its proper co-ordinates i.e. Let us consider here x rows and y column and $L(m, n, f_i)$ means the block by (m, n) in the $f_i$th frame.

3: Evaluate the $x^2$ between the consecutive blocks. $R(m, n, f_i)$ and $R(m, n, f_i+1)$ stand for the histogram of blocks at (m, n) inside the $f_i$th and $(f_i+1)$th frame respectively. Block’s calculation is computed with the following equation:

   $$M'(f_i, f_i+1, m, n) = \frac{\sum_{i=1}^{f_g-1} [R(m, n, f_i) - R(m, n, f_i+1)]^2}{R(m, n, f_i)}$$

4: Computing $x^2$ histogram difference between two consecutive frames:

   $$M'(f_i, f_i+1) = \sum_1^x \sum_1^y W_{mn} D'(f_i, f_i+1, m, n)$$

where $w_{ij}$ refers to the influence of block at (m, n);

5: Computing threshold mechanically: computing the mean and standard variance of $x^2$ histogram differentiation over the complete video sequence [7]. The subsequent formulas are used for calculating MD and STD given follows:

   $$MD = \sum_{f_i=1}^{f_v-1} \frac{D(f_i, f_i+1)}{f_v - 1}$$

   $$STD = \sqrt{\sum_{f_i=1}^{f_v-1} \frac{(D(f_i, f_i+1) - MD)^2}{f_v - 1}}$$

6: Shot boundary detection

Let the threshold $T=Mean\ Deviation +b* STD$
B. VSUMM Using KeyFrame Extraction (VKFE):

Algorithm 2: Key frame extraction

1: Calculate the differentiation between RF frames and GF frame using algorithm 1:

\[ D_k(1,k) = \sum_{i=1}^{x} \sum_{j=1}^{y} wij \ D_{fib}(1,k,m,n), k = 1,2,3,4 \ldots \ Fc \]

Where ‘Dfib’ is the difference between the frame with their block difference measures.

2. Trace the highest distinction:

\[ \max(i) = \{Dc (1,k)\}_{\max}, k=2,3,4……Fc. \]

3. Calculate “ShotType” according to the relationship between max(i) and MD: DynamicShot:

\[ \text{SHOTu} = \begin{cases} 
1 & \text{if } \max(i) > MD \\
0 & \text{otherwise} 
\end{cases} \]

4. Calculate the exact position of key frame;

if \( \text{SHOTu}=0 \), with regards to odd numbers of shot, then the middle frames is chosen as key frame otherwise i.e. in even case the frame in between the two frame in the center of shot is treated as key frame. If \( \text{SHOTu}=1 \), the frame having MD is declared as key frame.

The figure 2 explains the foundation analysis of structuring the video contents along with mechanism of key frames generation. The proposed system concentrates on shot frontier detection and key frame using prognostic framework i.e. A video document is separated into set of shots then these shots are further process to club all key frames using the formula:

Key frame = \{ if the MD > Threshold it is key frame \\
Otherwise skip the duplicate frame \}

Here another important aspect called semantic approach of trained scene in datasets. Once the trained frame is identified these frames are taken into consideration for finding the category using prognostic framework.
C. Video Segmentation:
Segmentation of Video is the first step towards content based video search analysis hence the entire video is broken up into particular scenes, afterwards those scene is converted in shots again shots are further divided in frames at the end. [10]

Fig. 2: Structuring of Video

Fig. 3: Video Segmentation

2. LITERATURE SURVEY
Masoud Mazloom et al. [1] this review work is dedicated to shot boundary detection and some gleam of video generalization & summarization. The various fundamental issues are provided and its countermeasures identified and calculated thoroughly.

Richard M. Jiang et al. [2] proposed the detection of shot using Gabor transformed first order method. This technique is applied on the datasets of TRECVID 2001 and compared the result.

Manasa Srinivas et al. [3] gives significant way of getting the key frame and calculate the precision and recall for the important key frame to get rid of group values on different dataset.

G’erard Medioni et al. [4] proposed approach for key frames have extracted the use of the dominant set of clustering for calculating similarity matrix

Krishna Choudhari et al. [5] gives the idea of calculating scene from the video. A function is used by the author in which all the values of matrix gets stores and check the status of edge and color histogram.

Zhong Zhou et al. [6] as the video is available in digital form hence it is tedious to detect SBD automatically therefore this paper gives the idea over the research based on seven years of TRECVID activities and outperforms the 57 different research groups to generates various best merits of SBD.

Naveed Ejaz et al. [7] proposed the adaptive key frame extraction for video summarization using an aggregation mechanism, this approach is to get the important reference frame from video using video processing also calculate the visual features and uses aggregate formula to reduce the redundancy
between the key frames. These methods applied to standard dataset to get actual rid of information and produce the meaningful result

Xiaomu Song et al. [8] Proposed access is in balmy of a appointment window of accelerating casings that slides over the accomplished case amalgamation (shot). The amalgamation of edges blanketed into every window is attempted for agreeable actual accord utilizing a acceptable unimodality yield a attending at. in this way, every window is authentic as unimodal or no best and the beginning amalgamation of every non-unimodal window is splitted into sections. Each video attempt is sectioned into unimodal bits and the key-outlines are ample due to the actuality the agent outlines of every element.

Yingying Zhu et al. [9] proposed the unique ideology of detection of shot boundaries which depends upon the concept of FC, then classify these shot into different classes, via SVM Classifier.

Huayong Liu [10] augmented transferring item detection set of guidelines primarily based via frame distinction and aspect detection. This method no longer best keeps the small postscript from frame distinction method and the impregnability of outrage from place detection entre.

Changsheng Xu et al. [11] accord acknowledged video abridgment access that is accomplished aural the arranged place. Our 4-enhance proposed address depends on a basal but boundless descriptor and a scene-recognition method, to accept bit-by-bit and abrupt advances with outstanding accuracy

A.B.Gadicha, Dr. M.V.Sarode, Dr. V.M.Thakare[12] proposed the basics methods to generate summaries are static and dynamic.it present extraordinary techniques for every mode inside the literature and describe a few capabilities used for producing video summaries

Azra Nasreen et al.[13] proposing new algorithm for detection of shot automatically specially in flashlight events.here it uses long formula to compress and enhance the frame diffrence.

3. PROPOSED WORK

From aloft analysis it's far empiric that, video is basically addle into shots, afresh shots are categorized into frames added frames are taken access as a continued way as video processing or video statistics retrieval is involved. These all strategies works on concepts of pixel via pixel apparatus the proposed plan is adherent to accomplish attempt borderland apprehension the acceptance of saliency map and affection extraction, bend map apparatus and key anatomy abstraction from assorted categories of video application KNN type

A. Key frame extraction using total block difference:

For arty the “key anatomy abstraction the acceptance of accepted block aberration mechanism”, our access is to abstract frames from a video. First a fall, frames are extracted from the ambition video and are stored in a called directory. An action is created in which every arena or anatomy is again adapted to its agnate blah calibration frame. For anniversary generation, two after gray calibration photographs are taken and their Histogram acumen is calculated. The sum of the factors of that histogram is again affected and returned.
B. **Object Segmentation:**

This methodology plays a very vital role to generate the shot by doing object segmentation approach. In this approach, the proposed system is capable of applying the saliency map techniques to find the object in the video. Then in the second part, the proposed system is concentrating on SVM for creating different frames from the entire video and generating the most valuable key frames out of it for mining purpose.
C. Semantic Mechanism:

From figure 5, for enhancing semantic intelligence in video summarization, first a fall the proposed mechanism accepts input as video sequence then apply adaptive threshold or shot frontiers algorithm [Algorithm 1 stated above] to get the absolute amount of shots in the unconditional video. Afterwards total number of shots is utilized for extraction of important key frames.

D. Frame Extraction

This research section extract the entire frames from an original video, the intact frames are bound together get the desired output.
E. Key Frame Algorithm

For computing the key frame from entire video first the proposed system calculates the total block difference between the $f_i$ and $f_{i+1}$ frame respectively. If TBD (total block differentiation) is more than the average value then it would be considered as key frame. For key frame identification motive enter is taken into consideration as all the factors of the difference matrix for every successive frame. After this suggest and standard deviation of the whole matrix for solving the brink cost. Threshold value is required for perceive the key frames. If the threshold value is more than output pair then first image of that pair is taken into consideration as key frame otherwise second image of that pair is considered as a key frame. Process is forestalled whilst the frame listing is empty.

The figure shown below gives the apparent idea about the key frames extracted from above frame extraction phase. As far as hypothetical analysis is concerned total number of frames extracted 1794, in it total key frames are 276 having values of threshold and different block difference and standard deviation as well.

F. Scene Classifier Algorithm:

Total Number of Iterations (NUM_ITERATIONS) = 10;
NUM_SOLUTIONS = 20;
solution_nns = {};}
solution_accuracy = []; 
sols_to_change = 1:NUM_SOLUTIONS; 
for iter=1:NUM_ITERATIONS 
    fprintf(‘Iteration %d
’, iter); 
    random_data = rand([1 NUM_SOLUTIONS]); 
    for sol=sols_to_change 
        x = [edges;ehists]; 
        t = cats; 
        t = convertToNeural(t,size(x,2)); 
        t = t'; 
        NumNeurons = 3*size(x,2); 
        net = patternnet(NumNeurons); 
        net = train(net,x,t); 
        y = net(x); 
        out_classes = vec2ind(y); 
        difference = logical(abs(cats - out_classes)); 
        difference = double(difference); 
        difference = mean(difference); 
        difference = double(difference) / length(cats); 
        accuracy = 1-difference; 
        solution_nns{sol} = net; 
        solution_accuracy(sol) = accuracy; 
        fprintf(‘Solution %d has accuracy:%0.04f %%
’, sol, accuracy*100); 
    end 
mean_acc = mean(solution_accuracy); 
sols_to_change = find(solution_accuracy < mean_acc); 
end.

![Graph1: TPVs FP](image.png)
Fig.9: True positive and False positive rate of trained frames

The above figure gives True positive and False positive rate of trained frames. Here in this research the first 50 frames are taken into consideration for generating fruitful result and adaptive throughput.

4. RESULTS & DISCUSSION

The proposed mechanism is able to find the meaningful frames, shots and scene from the video and gives proper summary of the video so that user gets minimum time frame to view or browse it. In this section the proposed system uses saliency map technique and edge histogram methodology to get gray level images, furthermore these images are useful to calculate the time taken by each frame and total key frames available in video.

A. Mechanism of Frame

The video containing different ‘Climate’ Category frames in which the different scene of climate along with some person frames is identified some of the key frames are utilized for the processing of video summarization. To generate the efficient and smooth generation of various values of precision and recall the proposed system takes some of the keyframes given below:

Fig.10: Mechanism of Frame


**B. Calculate Performance on Different Epoch**

This Evaluation will give complete idea about the cross entropy of frames which trained during training phase vs Epoch. The first graph gives hypothetical study of various trained (blue), validation (green) and test (red) result at epoch 68 and second graph gives the different result at epoch 8.

![Fig.11: Calculate Performance on Different Epoch](image)

**C. Perfomarce Evalauion**

In this current research work, We have calculate the Error Histogram between the Instances against the Error. For the calculation of Error here we used the following formula.

Error = Target – Output.

The above formula identify the best match solution from the various parameters like validation, test and zero error also after evaluate best fit in the event or scene classification the significant result produces shown in figure below.

![Fig.12: Performace Evalauion](image)

Precission and Recall ration on different set of video datasets. The result on various categories of video vs Proposed Algoritm shown below:

**Table 1: categories of video vs Proposed Algoritm shown below**

| Sr. No | Video Type | Precision (%) | Recall (%) | Accuracy (%) | Proposed Approach (%) |
|-------|------------|---------------|------------|--------------|-----------------------|
| 1     | Movie      | 80            | 93         | 81           | 98.22                 |
5. Conclusion
The above all significant methods and algorithm is dedicated to detection of proper scene in video processing (VSUMM SFT) using support vector mechanism so the user get the summary of video in minimum time and complexity to get the idyllic frames which are desirable for supplementary processing also Scene Classifier algorithm will detect the scene available in the video and process it, generates the summary of entire scene as whole. Also it gives cumulative study on various parameters like Performance, accuracy, precision and recall ratio.

6. References
[1]. Masoud Mazloom, Efstratios Gavves, and Cees G. M. Snoek "Conceptlets: Selective Semantics for Classifying Video Events" IEEE TRANSACTIONS ON MULTIMEDIA, VOLUME. 16, NO. 8, pp 2214-2228, 2014.

[2]. Richard M. Jiang, Abdul H. Sadka, Danny Crookes “Hierarchical Video Summarization in Reference Subspace” IEEE Transactions on Consumer Electronics, Volume. 55, No. 3, pp 1551-1557, 2009.

[3]. Manasa Srinivas, M. M. Manohara Pai and Radhika M. Pai “An Improved Algorithm for Video Summarization – A Rank Based Approach “Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016) Procedia Computer Science 89, pp 812 – 819, ©2016

[4]. G’erard Medioni Ram Nevatia Isaac Cohen “Event Detection and Analysis from Video Streams”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 23 , Issue 8 , pp 873-889, 2001.

[5]. Krishna Choudhari,Vinod K Bhalla “Video Search Engine Optimization Using Keyword and Feature Analysis” Procedia Computer Science 58 ( 2015 ) Second International Symposium on Computer Vision and the Internet (VisionNet’15) Elsevier B.V, pp 691 – 697 © 2015.

[6]. Zhong Zhou, Feng Shi, and Wei Wu “Learning Spatial and Temporal Extents of Human Actions for Action Detection” IEEE TRANSACTIONS ON MULTIMEDIA, VOLUME. 17, NO. 4, pp 512-525, 2015.

[7]. Naveed Ejaz , Tayyab Bin Tariq , Sung Wook Baik “Adaptive key frame extraction for video summarization using an aggregation mechanism” ELSEVIER Journal of Visual Communication and Image Representation Volume 23(2012) pp 1035-1040 © 2012.
[8]. Xiaomu Song, Guoliang Fan “Joint Key-frame Extraction and Object Segmentation for Content-based Video Analysis” IEEE Transactions on Circuits and Systems for Video Technology, Volume 16, Issue 7, pp 904-914, 2006.

[9]. Yingying Zhu, Zhong Ming “SVM-based Video Scene Classification and Segmentation” IEEE International Conference on Ubiquitous Engineering 978-0-7695-3134-2/08 DOI 10.1109/MUE.2008.92 © 2008.

[10]. Huayong Liu, Tao Li “Key Frame Extraction Based on Improved Frame Blocks Features and Second Extraction” IEEE 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) 978-1-4771-0313-4/12 ©2015.

[11]. Changsheng Xu, Alan Hanjalic, Shuicheng Yan ”Special Section on Object and Event Classification in Large-Scale Video Collections” IEEE Transactions on Multimedia, Volume 14, Issue 1, pp 1-2, Feb. 2012.

[12]. A.B. Gadicha, Dr. M.V. Sarode, Dr. V.M. Thakare, “Mechanism of Event Summarization in Video” International Journal of Scientific Research and Review Volume 7, Issue 9, pp 372-379, 2018, ISSN NO: 2279-543X.

[13]. Azra Nasreen, Dr. Shobha G "Key Frame Extraction from Videos - A Survey” International Journal of Computer Science & Communication Networks, Volume 3 Issue 3, pp 194-198. JUNE 2013.

[14]. Huaping Liu & Yulong Liu & Fuchun Sun "Video key-frame extraction for smart phones” Springer Science Business Multimedia Tools Application, Volume 75, Issue 4, pp.2031-2049, doi.10.1007/s11042-014-2390-7, November 2014.

[15]. Houari Sabirin and Munchurl Kim “Moving Object Detection and Tracking Using a Spatio-Temporal Graph in H.264/AVC Bitstreams for Video Surveillance” IEEE Transactions On Multimedia, Volume. 14, NO. 3, pp.657-668, June 2012.

[16]. Mei Huang, Ling Xia, Jin Zhang, Hui Dong "An Integrated Scheme for Video Key Frame Extraction" 2nd International Symposium on Computer, Communication, Control and Automation pp 258-261 Atlanta Press © 2013.

[17]. Pb Agus Ristono*,”Design Of Reliable And Efficient Manchester Carry Chain Adder Based 8-Bit Alu For High Speed Applications”, Journal Of VLSI Circuits And Systems, 1 (01), 1-4,2019

[18]. NHK K. ISMAIL*,”Estimation Of Reliability Of D Flip-Flops Using Mc Analysis”, Journal of VLSI Circuits And Systems 1 (01), 10-12,2019.