A Research on Multi-View Video Summarization Techniques

M.Raja Suguna, A.Kalaivani

Abstract— Video Surveillance System uses video cameras to capture images and videos that can be compressed, stored and send to place with the limited set of monitors .Now a Days all the public places such as bank, educational institutions, Offices, Hospitals are equipped with multiple surveillance cameras having overlapping field of view for security and environment monitoring purposes. A Video Summarization is a technique to generate the summary of entire Video Content either by still images or through video skim. The summarized video length should be less than the original video length and it should covers maximum information from the original video. Video summarization studies concentrating on monocular videos cannot be applied directly to multiple-view videos due to redundancy in multiple views. Generating Summary for Surveillance videos is more challenging because, videos Captured by surveillance cameras is long, contains uninteresting events, same scene recorded in different views leading to inter-view dependencies and variation in illuminations. In this paper, we present a survey on the research work carried on video summarization techniques for videos captured through multiple views. The summarized video generated can be used for the analysis of post-accident scenarios, identifying suspicious events, theft in public which supports Crime department for the investigation purposes.

Index Terms— Video Summarization, Multi-View Videos, Still Images, Video Skimming

I. INTRODUCTION

The fundamental unit of a video is a Frames.The frames when recorded by the uninterupted camera at regular interval of time constitutes a shot. A scene is a collection of shots. A Summarized video contains only important video frame called key frames. Generally, Video Summarization is a process of finding these keyframes from original lengthy video Content. Structural Hierarchy of video is given by Fig 1.

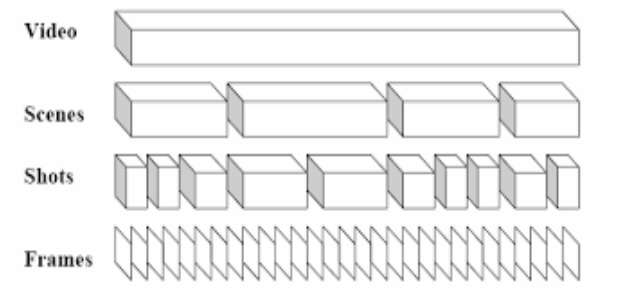

Figure 1: Structural Video Hierarchy

Now a Days, Users create a large repository of video content in the internet. The Summarized video content can be used for video indexing and quick retrieval process. Similarly accessing videos captured by Surveillance camera such as CCTV's requires large time to view all the content to get user interest of information from it. Hence the summarized video content reduces the task overhead in Post analysis of video in application such as Remote video Monitoring, Facility Protection, Car Parking Lots, Event Video Surveillances, Public Safety, Traffic Monitoring, Outdoor Perimeter Security and Internet Security Systems.

The rest of the paper is organized as follows: Section II gives the brief introduction about Video Summarization. In Section

III the research work of multi-views video Summarization techniques is discussed. Section IV gives the Comparison of Multiview video Summarization Techniques discussed in previous section. Section V describes the inference of the survey done. Section VI discuss about the Quantitative Evaluation metrics for Video Summarization Techniques. Section VII Concludes the Paper..

II. VIDEO SUMMERIZATION

Automatic Video Summarization generates the abstract of the lengthy video content holding all useful information within the short duration. It is divided into two types Static Video Summarization and Dynamic Video Summarization. Static Video Summarization generates only the keyframes (still images) irrespective of the time and sequence. Dynamic Video Summarization arranges the keyframe in temporal order (moving story Board) also called Video Skim.

Video Summarization techniques holds the following common steps such as Video Pre-processing, Feature Extraction ,Key Frame Extraction and summarization. The general architecture of Video Summarization is in Figure 2.

Revised Version Manuscript Received on October 15, 2019.

M.Raja Suguna, Assistant Professor, Department of CSE, Saveetha School Of Engineering, Tamil Nadu, India

A.Kalaivani, Associate Professor, Department of CSE, Saveetha School Of Engineering, Tamil Nadu, India

The input video is converted into frames. Different techniques are found in the literature to extract the frames from the input video. After frame extraction the next phase is to extract the visual features from the frames. Visual Features constitutes Color Features, Edge detection, Block Correlation. Color is the important aspect of an image. A image can be represented by different color Spaces such as grey Scale, RGB, HSV, YUV etc., Color Histogram is constructed for all input frames. Difference in Color Histogram of consecutive frames will be small for the similar Frame.

These differences can be used for key frame extraction.

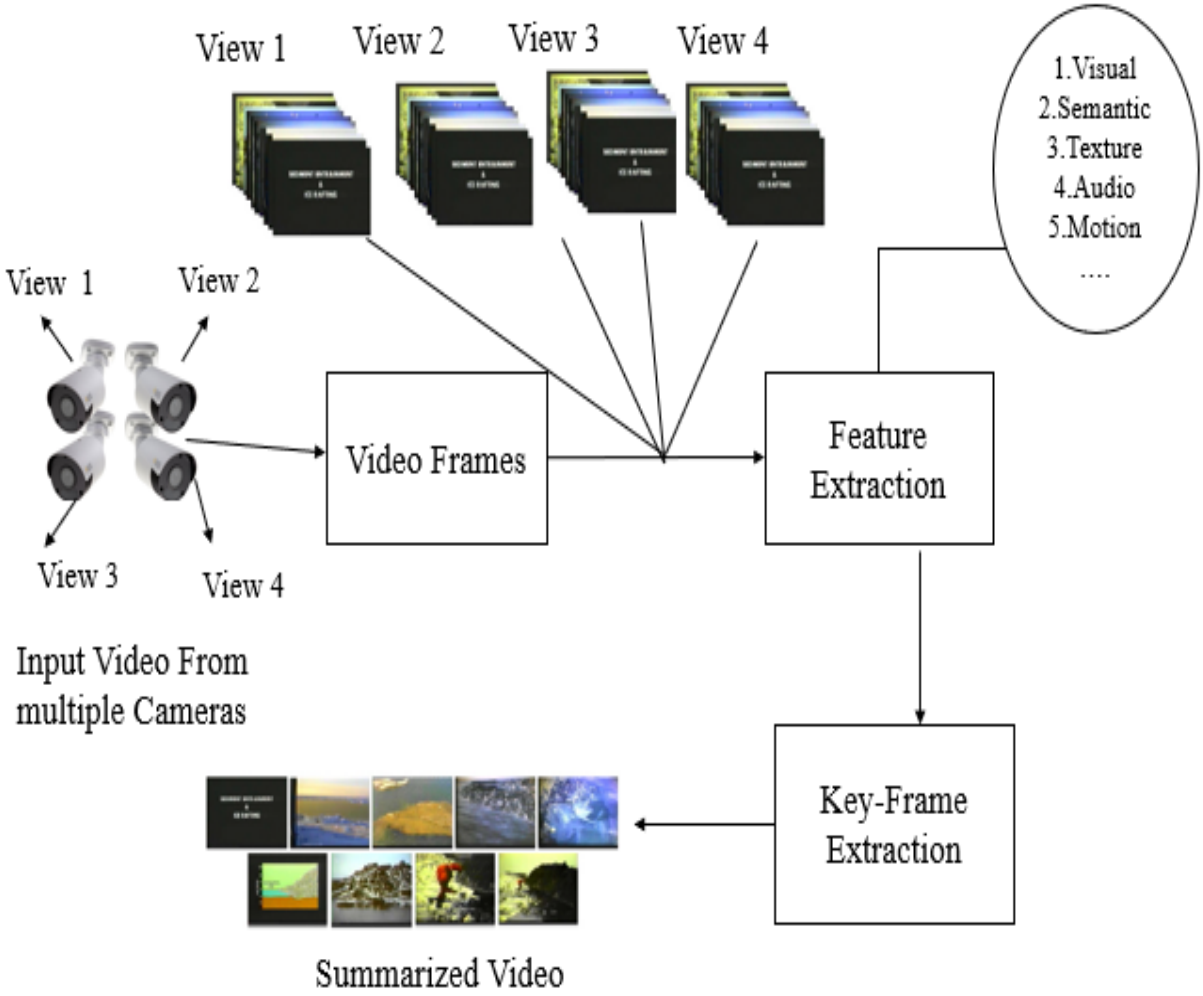

Figure 2: General Architecture of Multi-View Video Summarization

Edge Detection is important part of image analysis. Edge is the boundary between the object and background in the frame and boundary between overlapping objects. Edge Detection in the frame used for image segmentation and object detection. Redundant frames can be eliminated by edge detection rate. In Block Co-relation, the frames are partitioned in blocks of pixels (e.g. macro blocks of 16×16 pixels in MPEG) by block motion compensation (BMC)method. Each block is predicted from a reference frame which contains block of pixel of equal size. The blocks are shifted to the position of the predicted block given by a motion vector.. The Key Frames are identified based on the type of visual Features extracted. Summarization constitutes arranging the Key Frame in Temporal order which gives meaningful abstract video

III. RELATED WORKS

Content Based Video Analysis is the important domain of Video Summarization Technique. It has a wide range of application such as in Consumer Video Analysis, Video Database Management System and Video Surveillances System. Video database Management includes application areas such as Digital Video Library, Video Search Engine, Object tracking, Automatic Object labelling and Object classification, Video Indexing and retrieval.

Video Surveillances System has a wide range of application such as Remote video Monitoring [10], Mobile Video Surveillances, Facility Protection, Car Parking Lots, Event Video Surveillances, Public Safety, Traffic Monitoring, Outdoor Perimeter Security and Internet Security Systems. For the above-mentioned Video Surveillance applications, Multi-View Video Summarization Technique reduces the very long recorded video into small, removes low activity (frames without any events) frames, resolves the inter-view dependencies if the same scene recorded in different views. A Robust Solution is provided for the Post Analysis of videos Captured by the surveillance cameras without loss of any single useful information

In this section we study the previous work carried out in multi-view Video Summarization

A. Spatio-Temporal Shot Graph Based Video Summarization:

Yanwei Fu et al [11] proposed a method for summarizing multi-view videos. Initially a spatio-temporal shot graph is constructed for the input video .Graph Labelling is done to generate the summarized video. A hypergraph is initially created in which edges contains the Correlation of different attributes of multi-view video shots. The spatio-temporal shot graph is derived from a hypergraph, the shot graph are then partitioned and event-centered shots clusters are identified via random walks

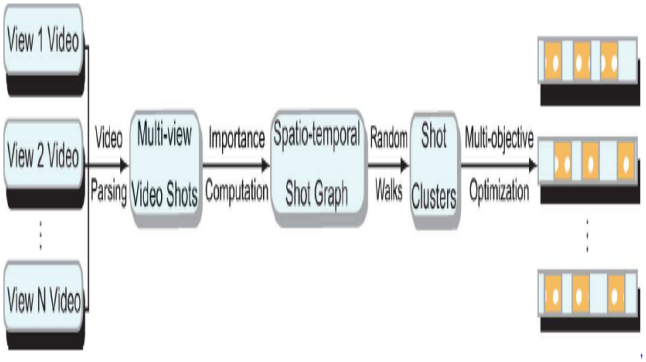

Figure 3: Proposed Multi-View Video Summarization [11]

The summarization result is generated through solving a multi-objective optimization problem based on shot importance evaluated using a Gaussian entropy fusion scheme. The different summarization objectives, such as minimum summary length and maximum information coverage are obtained in this framework. Multi-View summaries are proposed by multi-view storyboard and event board shown in Fig 4.

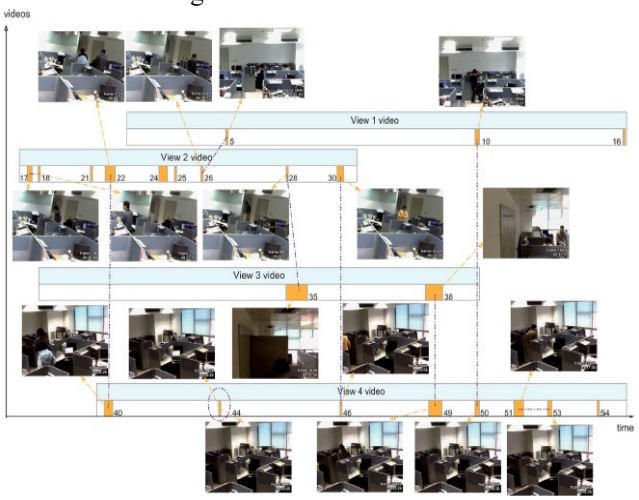

Figure 4: Multi-View Video Story Board Summarization [11]

The storyboard naturally reflects correlations among multi-view summarized shots which describes the important events. The event-board serially assembles event-centered multi-view shots in temporal order. The above method is implemented using office1, campus, office lobby, and road multi-view video surveillance dataset. The proposed method outperforms well with User Attention Model proposed by Yu-Fei Ma et al [16] and Graph methods [17],[18]

Video summarization can be improved by extracting visual features and semantic features from the video frames.

B. Bi-partite Graph Based Video Summarization:

Sanjay K. Kuanar et al [8] proposed graph-based approach for summarizing Multi-View Videos. Shot Detection method is used to select the input frames for video processing. Both the Visual and semantic features are extracted from the input video frame. Visual features such as HSV Color Space, texture features using Edge Histogram Descriptor and three-pixel level Tamura are extracted from the frames. Semantic features are obtained by Visual Bag of words Model (BoVW).Visual Bag of words are obtained by extracting SIFT features from the video frames and K-means clustering algorithm is applied to obtain the clusters.. Each Cluster forms the Visual Word. After combining all the features, a video frame is represented by a 439-dimensional feature vector (256 for color + 80 for texture + 3 for Tamura + 100 for visual bag of words). The Proposed Architecture is given by the Fig 9.

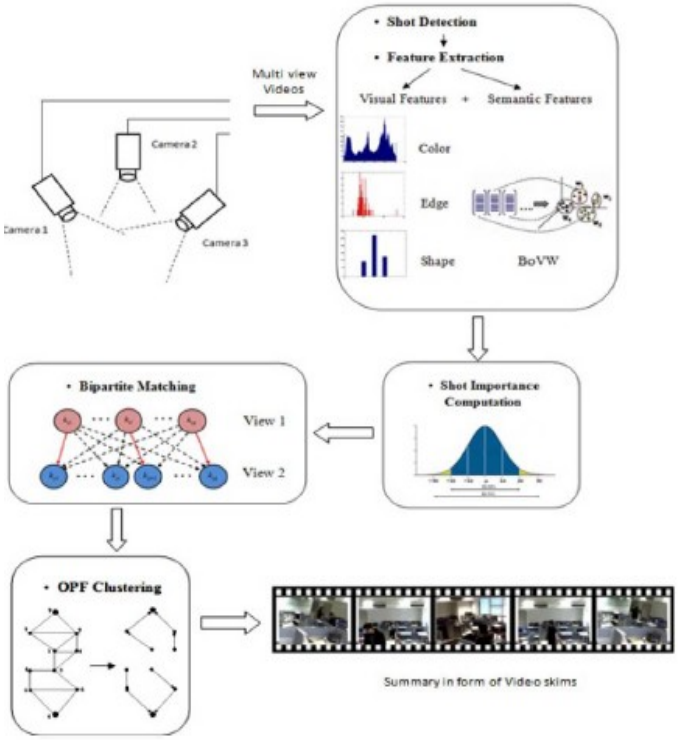

Fig:5 Proposed architecture of Bi-partite path Matching [8]

Gaussian Entrophy fusion model is used to remove the shots which has low or no activity. Correlation among the multiple view is matched to Maximum Cardinality Minimum Weight (MCMW) of Bi-partite Graph. A Bi-partite graph G= (N, V) is constructed for the video for

each view. MCMW algorithm, is applied to graph which obtains actual key frame correspondences between any pair of views. As input video is of high dimensional vector OPF Algorithm is applied for clustering the shots in bi-partite graph to obtain the summarized video.

The above method is applied to four Multiview data sets Office1, Office lobby, Campus, BL-7F and compared with the corresponding ground truth. Experimental results show this method outperforms well with the existing mono-view summarization technique [16],[17],[18] and multi-View Summarization technique learning based multi view method [6].[9],[11].

C. Multi-View Metric Learning Framework:

Linbo Wang et al [6] method uses Multiple kernel learning to solve the multiple view problem and optimal distance metric is used to obtain the consistent clusters. It proposes Unified Metric learning framework by integrating both the Disagreement Minimizing Criterion (DMC)and Maximum Margin Criterion (DMC). The Input Video is converted in sequence of frames. Each view video is represented in its own Feature dimensional vector. These frames is fed into Metric Learning framework which constructs the Common Metric Space i.e. each view's high-dimensional low level features are embedded in the same common low dimensional space . After K-mean Clustering Algorithm is applied on the frames to extract the key frames. Key Frames are arranged in temporal order to get the summarized video.

The above method is implemented in Office 1 and road dataset and compare with ground truth for measuring objectiveness performance. It outperforms well with Multi-view video Summarization [11]. Efficient Optimization Algorithm needed to improve the Objectiveness and to address several Multi-View Summarization issues such as summarization for large duration videos.

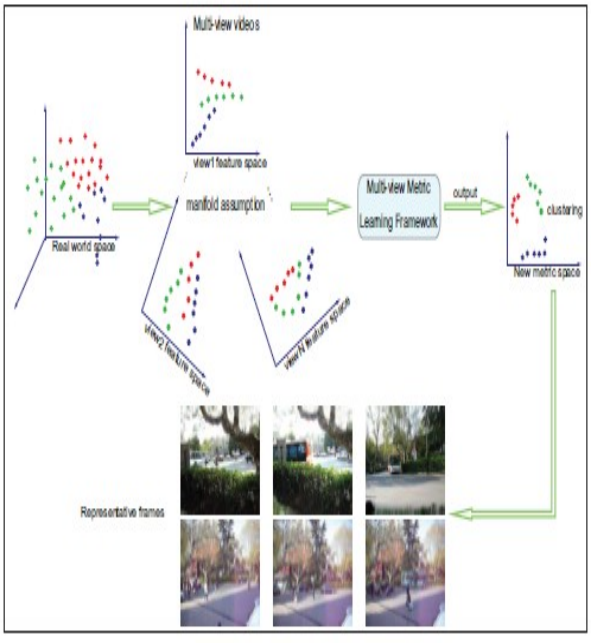

Fig:6 Multi-View Metric Learning Framework[6]

D. Multi-View Video Synopsis Framework:

Ansuman Mahapatra et al [4] proposed a framework for creating a synopsis of multi-view videos captured by surveillance cameras (indoor and outdoor) having overlapping field of views. In video synopsis ,the spatial locations of objects are unchanged but the objects are shifted along the temporal axis and represented simultaneously in a common ground plane.

A Common Ground Plane is created for videos captured by multiple cameras. For outdoor videos, PETS 2009 dataset, Top view of the site is found by Google Map and for indoor videos common ground plan is identified. The proposed work is restricted to human actions identified in the video. The synopsis creation is obtained by three techniques: contradictory binary graph coloring (CBGC), table-driven approach and simulated annealing (SA) based approach. Action Recognition Module is used to identify the important actions of humans in the video. These important actions reduces the synopsis length. Fuzzy Inference System calculates the visibility score of each object track in the video which further reduces the synopsis length. In CBGC approach, the maximum reduction in synopsis length is achieved. The stochastic approach using SA, on the other hand, achieves a better trade-off among the multiple optimization criteria.. The overview of the proposed work is given by Fig 6.

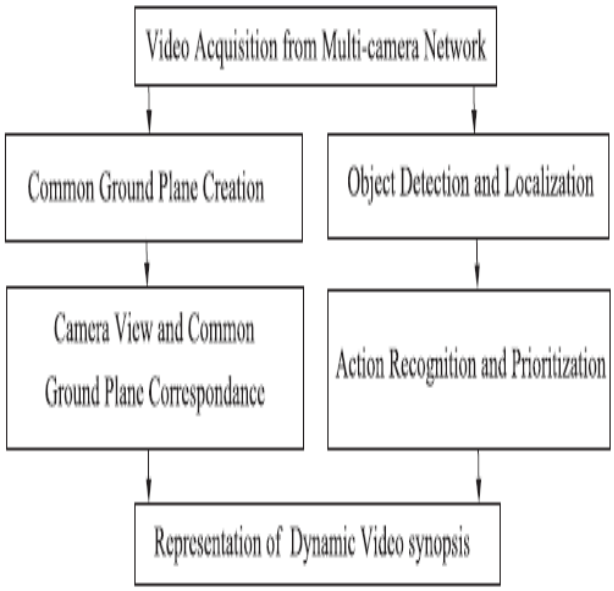

Fig 7 . Proposed framework for Multi-View Video Synopsis Framework [4]

The proposed methods are validated using four different datasets; KTH, WEIZMANN, PETS 2009 and LABV. The latter two datasets containing multi-view videos are used for synopsis generation. PETS 2009 dataset has four multi-view videos each having 794 frames. The LABV is an indoor sequence captured in the cameras, each having 34,438 number of frames. Sample frames of PETS 2009 datasets for each camera view are shown in Fig 7 and corresponding Video synopsis generated is shown in Fig 8 .

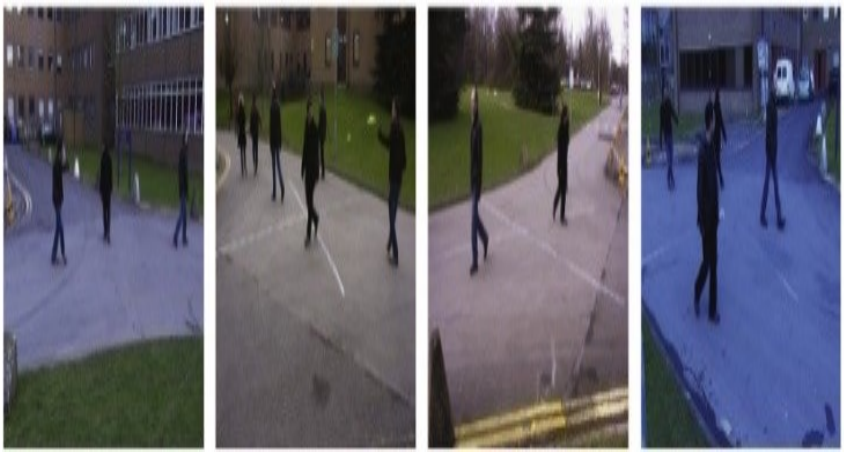

Fig:8 Sample frames for PETS 2009 dataset.

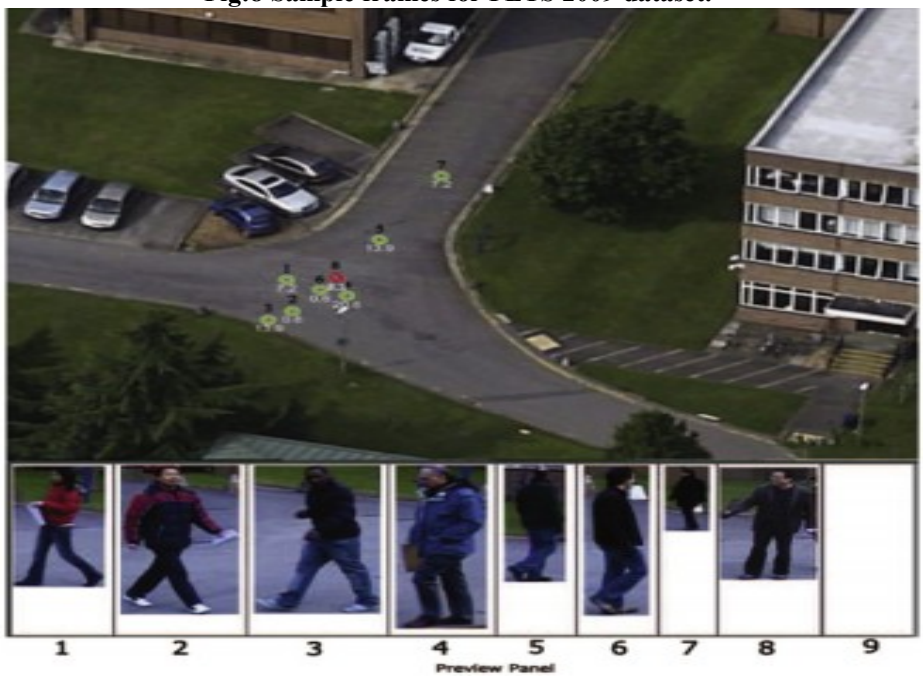

Fig 9. Video Synopsis

E. Video Summarization using MMR Algorithm:

Yingbo Li and Bernard Merialdo [12] proposed the Key frame extraction technique based on Video Maximal Marginal Relevance Algorithm analogy to classical algorithm of text summarization, Maximal Marginal Relevance for multi-video Summarization. Video-MMR retains relevant keyframes and removes redundant keyframes.

Histogram of Visual Words are the features extracted from the video frames. SIFT descriptor is computed by detecting the Local interest points (LIPs) in the image, by taking the difference of Gaussian and Laplacian of Gaussian. K-means is applied to the SIFT descriptors to compose a visual vocabulary with 500 words. Cosine of similarity between the successive frame is calculated and Video-MMR algorithm is applied to select the representative key frames. It also proposes two methods Global Summarization and Individual video Summarization. Individual Summarization generates a summary for each video in the set and concatenate those summaries. Global Summarization considers both inter- and intra- relations of individual videos simultaneously and avoid the redundancy of individual Summarization.

The above method is implemented in videos collected from the internet news aggregator web site (http://www.wikio.fr/). It outperforms well with K-means clustering algorithm [15] and also resolves the intra- and inter- relations of summary and produces static video summarization.

F. Multi-View Video Summarization on Many core GPU:

Pandurang Matkar et al [5] proposed a Framework for Multi-View Video Summarization on Many core GPU. A Graphics Processing Unit, a single-chip processor primarily used to manage and boost the performance of video and graphics. The input video is split into adjacent data cubes by temporal segmentation algorithm. Two consecutive video frames are DWT transformed and then the differences of statistical features of the two frames is calculated. If the difference value of a pair is greater than threshold, the last frame of the pair is considered as a key frame. A video Summarization is created by extracted Key-frames. The output is static video summarization. This method doesn't extract any semantic features from the input key frame.

G. Event Bagging, Ensemble Video Summarization:

Krishan Kumar et al [3] proposed Machine learning ensemble method to summarize the video content. Bootstrap Aggregation method is used. The input video is converted into Frames N. In Training phase, bootstrap samples of different scenes from the individual view of video is taken. The sample size is m which should be less than N (m<N). For P views, P bootstrap samples are taken and given as input to P Classifiers which gives decision tree as output. A Node in the decision tree is the Frame and the tree is trained by variance(σ) between the frames. The decision tree is not pruned and hence they have high variance. The proposed framework is given by Figure 10.

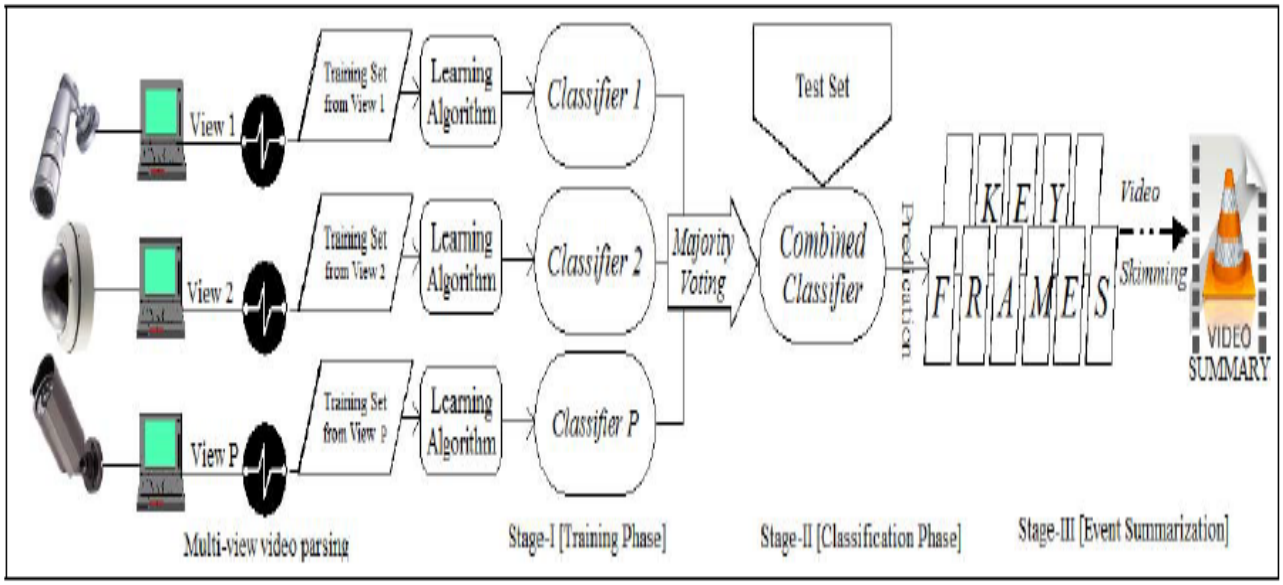

Fig .10 Event Bagging [3]

After training, based on the output of the previous classifier ,the current classifier is weighed and added to the ensemble set.

In Testing Phase, video frames from the individual view of the camera is given as a Input to the Combined Classifier. If a frame of any view is appeared in more than 70% of classified trees it is declared as the keyframe otherwisw discarded. This is also called majority voting policy. Duplicate frames are also removed in this stage. Next stage is Event summarization. If Euclidean distance between a frame and the key frame of the event is equal to or greater than Event Boundary Threshold value, then the current frame counted in the current event otherwise discarded.

The above method is applied to two multi-view dataset namely Lobby: three views of Lobby dataset [9] and Office: four views of Office dataset [9] and compared with the ground truth. The information of number of events in the video is not required

H. Multi-View Videos Summarization Via Joint Embedding And Sparse Optimization:

Rameswar Panda et al [13] proposed a novel method of unsupervised framework for summarizing multi-view videos via joint embedding and sparse optimization. The embedding is used to capturing content correlations in multi-view dataset. The sparse representative selection is used to generate Multiview summaries based on user length request without additional computational cost.

The video is segmented into multiple shots by measuring the amount of difference of RGB and HSV color spaces of two consecutive frames in the video.Visual features are extracted by applying 3D convolutional filters to a set of 16 input video frames and the responses are recorded at the layer FC6. The local ordering structure within a shot is maintained by temporal mean pooling scheme. The pooling scheme gives the final feature vector of a shot (4096 dimensional)which is used for the sparse optimization. All the shots are embedded a joint latent space by considering the feature similarities between two shots in an individual video (Inter-View) and in two different videos.(Intra-View) .The summary of the multi-view videos is the optimal subset of all the embedded shots

The above method is implemented on the 6 multi-view datasets namely Office, Lobby, campus, Road, Badminton, BL-7F, produces Multiview summaries as per user length request. It can be improved by including semantic features including attention model and user defined semantic preferences.

I. Online Multi-view Video Summarization:

Shun-Hsing Ouet et al [9] proposed an Online Video Summarization technique for Wireless Video Sensor Network. Wireless Video Sensors are deployed at numerous places with overlapping field of views. Battery lifetime is very important for these networks. streaming lengthy video content to the server need more power which drains the energy of the sensor. In order to reduce the power consumption, online multi view summarization technique is proposed. The architecture of the proposed system is shown in Fig:11

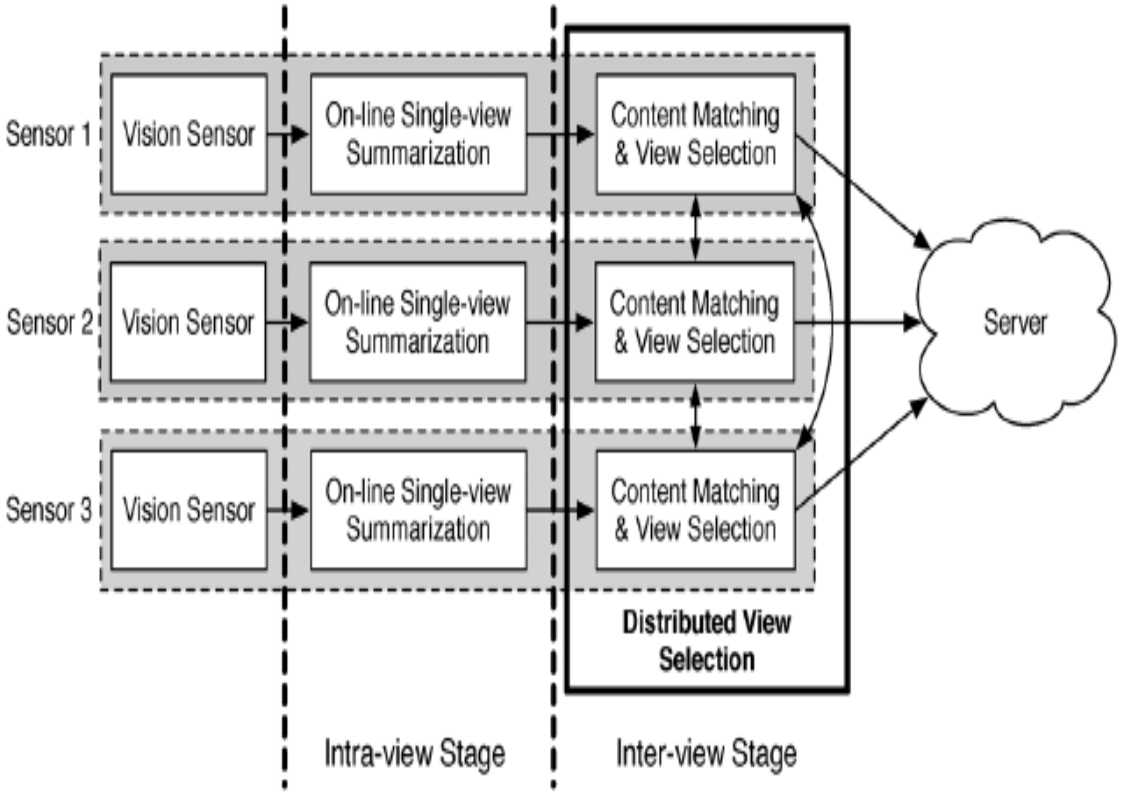

Fig 11: Architecture of Online Multi-View Video Summarization [9]

In the Intra-View Stage, the Video recorded by each sensor is first processed through On-line Single-View Summarization. Color feature is extracted from the each video frame by using MPEG-7 color layout descriptor. GMM, a On-line clustering Algorithm is used for clustering the input video frames. Key frames are selected from the clusters to generate Single-view Summary.

In Inter-View Stage, Distributed View Selection Algorithm removes the redundant frames from multiple views. Besides color layout descriptor ,the algorithm also extract the position of the foreground object in the frame for cross view matching. Then the important video frames and non-redundant video frames are streamed to the server.

The above method is implemented using Office dataset provided by [11] and lobby dataset provided by [11] and on BL-7F generated by the authors by installing 17 cameras ,all synchronized perfectly. Both the on-line single view summarization and Inter-view stage is implemented separately and compared with Ground truth (User created Summaries). This method outperforms well than the multi-view video summarization technique using Spatio-temporal Shot Graph[11].

J. FASTA :

Krishan Kumar et al [1] proposed FASTA approach which is local alignment based method to summarize the events in Multiview videos. Convolutional Neural Network (CNN) is trained with RGB input images with multiple multi-channel filters .Initially N frames of equal length of a single view is fed into these CNN to extract the visual features and object detection . CNNs features extracted is used for further video processing. A frame can be categorized to fall on any of the following type based on shreds of evidence(number of moving objects).1. NE: No Evidence. 2) SH: Some Hints 3) SE: Significant Evidence . 4) SV:,the frame has more than two moving objects.

Nucleotide sequence is formed by assigning a labels “A”,”C”,”G”,”T” to the frames which has maximum cosine similarity between the current frame and previous frame. FASTA ,a rapid local alignment algorithm is used to remove the inter-view redundancy .and to captures the correlations among multiple views using an optimized alignment approach. Further redundant frames are removed by using Object tracking method. The architecture of proposed method is given by the Fig 12.

The extracted key frame are then arranged in temporal order to get the summarized video.

The above method is executed in three multi-view datasets Office, Lobby and BL -7F.It outperforms well with all the previous multi-view summarization techniques [3],[8],[9] ,[11] in Precision, Recall , F-Measure and Summary length.

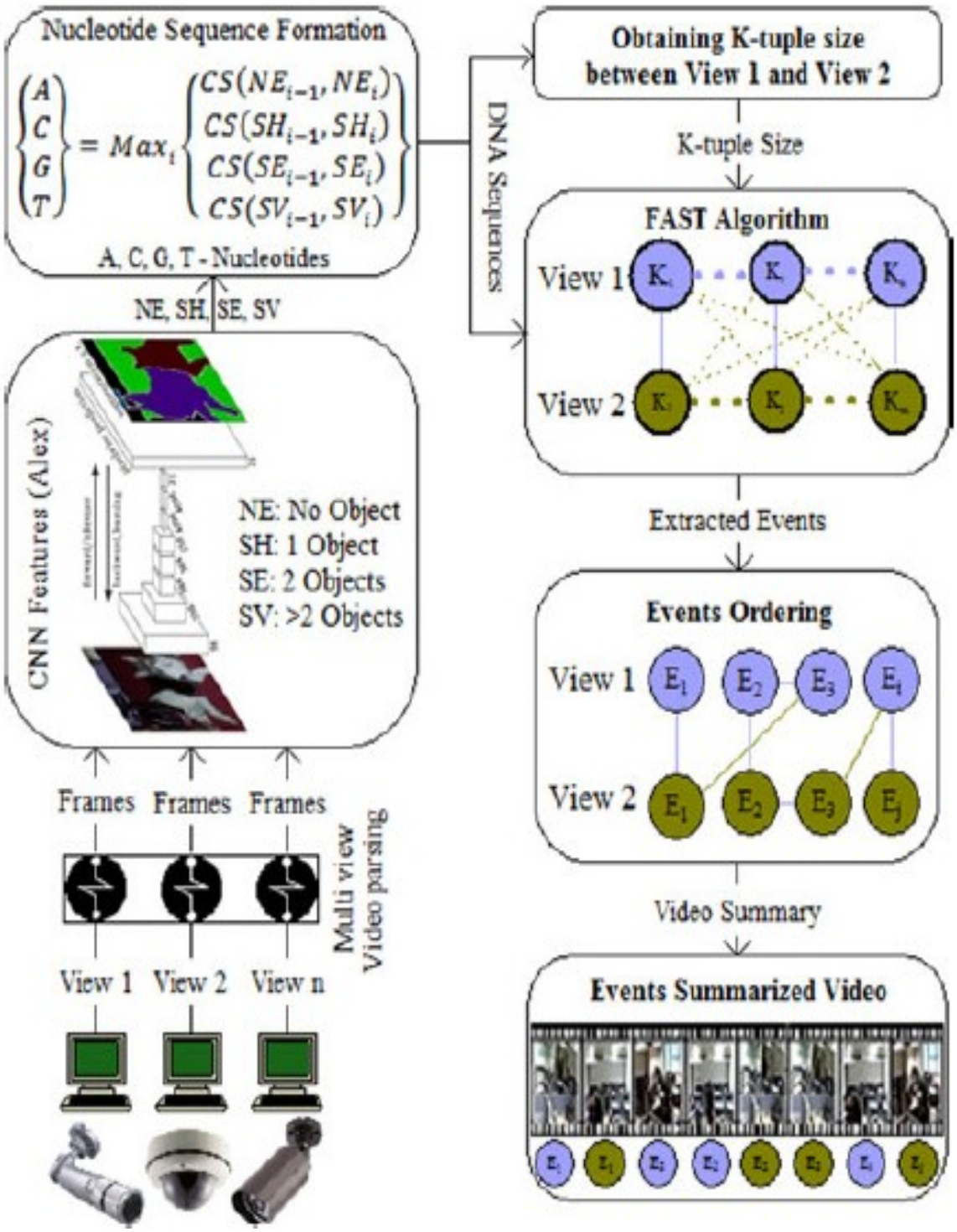

Figure :12 Proposed Architecture of F-DES [1]

Table 1:Benchmark dataset description

Dataset	No of views	Durations (mins)
Office	4	46:19
Lobby	3	24:42
BL-7F	19	136:10

IV. PERFORMANCE ANALYSIS

All the proposed method are implemented in the Three benchmark multi-view datasets namely Office provided by [11], Lobby provided by [11], BL-7F provided by [9] .Detailed Description of three dataset is given the Table 1. Comparison of Multi-View video Summarization Techniques of State of Art Methods is given by Table 2

Table 2: Comparison of Multi-View Video Summarization Techniques of State of Art Methods

Dataset	Algorithm	Total video length	Summary length(s)	# Events Detected
Office	Online Multi-view[9]	3016	402	N/A
	Bi-partite Graph[8]	3016	59	18/32
	Event Bagging [3]	3016	80	17/32
	F-DES[1]	3016	69	24/32
Lobby	Online Multi-view[9]	3016	484	N/A
	Bi-partite Graph[8]	1482	176	33/37
	Spatio -temporal Shot Graph[11]	1482	158	34/37
	Event Bagging [3]	1482	152	36/35
	F-DES[1]	1482	153	34/37
BL-7F	Online Multi-view[9]	3016	516	N/A
	Bi-partite Graph[8]	8170	633	N/A
	F-DES[1]	8170	315	51/59

V. INFERENCE FROM THE SURVEY

The inference on the survey of multi-view video summarization is as follows

¤ Yingbo Li [12] and Pandurang Matkar [5] proposed Key-frame extraction strategies to summarize the multi-view video Video-MMR method and DWT differences are used to extract key frame respectively. They outperforms with the single view video summarization technique [16].

¤ Linbo Wang et al [6] and Ansuman Mahapatra et al [4] uses Common multi-view metric learning framework and Common plane generation technique to resolve the inter-view dependencies for multi-view video frames respectively. After computing common platform for multi-view co-ordinates Linbo Wang et al [6] uses the k-means clustering algorithm to extract key frames and Ansuman Mahapatra et al [4] uses contradictory binary graph coloring (CBGC) , simulated annealing (SA) for prioritize and human object detection methods to generate the video synopsis.

¤ Yanwei Fu et al [11] and Sanjay K. Kuanaret al [8] proposed graph-based approach for resolving inter-view dependencies for the videos captured by multiple cameras of surveillance system. Yanwei Fu et al [11] uses Spatio-temporal graph and keyframe is extracted by using random walks. Sanjay K. Kuanaret al [8] generates the Bi-partite Graph by extracting both visual and semantic features from input video frames and MCMW algorithm is used to resolve the inter-view dependency and OPF algorithm is applied to cluster the shots in Bi-partite graph and extract the key frames to generate the video skim.[8] outperforms well with [11]

¤ Rameswar Panda et al [3] proposed unsupervised framework via joint embedding and sparse representative selection to resolve inter-view dependencies among multi-view videos. They extract CNN visual features from the video frames .It outperforms well with the [11] and [8]

¤ Krishan Kumar et al [1] method extract the CNN features from the input video frame and nucleotide sequence is created based on the cosine similarity of CNN feature between the frames. FASTA algorithm is used to remove the inter-view redundancy and the correlations among multiple views is cptured using an optimized alignment approach. Further redundant frames are removed by using Object tracking method. It outperforms well with method proposed in [3],[8],[9] ,[11]. The method can be further improved by adding user attention model and for long duration multi-view video

In the literature more concentration is given on removing the inter-view dependencies. Visual Features extracted using CNNs, shows the considerable improvement by including all semantic information. The work can be further improvised by extracting audio features,Optimization algorithm for removing inter-view dependencies[1] and object Detection methods.

VI. EVALUATION METRICS FOR VIDEO SUMMARIZATION

Sandra E. et al [16] proposed the technique in which the Automatically generated video is compared with User Generated video Summary (ground truth). Each key frame is given a score .This score represents the user selection of key frame to identify the video content. The mean score for each video summary is gives the quality level of video summaries produced. This can be calculated as:

SCOREM = Sum of Keyframe Scores/Number of Keyframes

Mei Huang, et al. [19] defined video summarization can be evaluated by using recall, precision, redundancy.

Recall is the ratio of number of frames in the ground truth matched to the frames in the summarized video. Recall is computed as:

$$\text{Recall} = \text{Nref\_m} / \text{N ref} \qquad (1)$$

where Nref is the total number of frames in a reference summary, and Nref m is the number of frames being matched with the reference frame.

Precision is the ratio of number of frames in the candidate summary to matching frames in the reference summary, and is defined as

$$\text{Precision} = N_{can\_m} / N_{can} \quad (2)$$

where N can is the total number of frames in a candidate summary; and Ncan_m is the number of candidate frames being matched.

Redundancy in video refers when one frame is repeated more than one time in the summary. Redundancy for one reference frame is the ratio of how many frames in the candidate's summary can match this reference frame. For the whole candidate summary, Average Redundancy Rate is:

$$AR = N_{can\_v} / N_{ref\_m} \quad (3)$$

Ncan _ v is the number of frames in the candidate summary that can match at least one frame in the reference summary.

CONCLUSION

Automatic Multi-View Video Summarization technique generates the summarized video content by removing the unimportant frames and inter-view dependencies. A detailed survey of the research work carried out in the area of multi-view video summarization techniques are discussed. Future research direction can be focused on extracting more semantic features from the video frames to produce most meaningful video summaries ,large duration multi–view surveillance videos and including object detection methods.

REFERENCES

1. K. Kumar and D. D. Shrimankar. F-des: Fast and deep event summarization. IEEE Transactions on Multimedia, 20(2):323–334, Feb 2018.
2. Behrooz Mahasseni, Michael Lam, and Sinisa Todorovic. Unsupervised video summarization with adversarial lstm networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2982– 2991, 2017.
3. K. Kumar, D. D. Shrimankar, and N. Singh. Event bagging: A novel event summarization approach in multiview surveillance videos. In 2017 International Conference on Innovations in Electronics, Signal Processing and Communication (IESC), pages 106–111, April 2017.
4. Pandurang Matkar, Aditya Tajne, Sushil Bomane, Piyush Bansal, Prof. S. A. Saoji, 2016, Framework for Multi-View Video Summarization on Many core GPU, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 05, Issue 01 (January 2016),
5. Ansuman Mahapatra, Pankaj K. Sa, Banshidhar Majhi, and Sudarshan Padhy. Mvs: A multi-view video synopsis framework. Signal Processing: Image Communication, 42:31 – 44, 2016.
6. L. Wang, X. Fang, Y. Guo, and Y. Fu. Multi-view metric learning for multi-view video summarization. In 2016 International Conference on Cyberworlds (CW), pages 179–182, Sep. 2016.
7. J. Zhu, S. Liao, and S. Z. Li. Multicamera joint video synopsis. IEEE Transactions on Circuits and Systems for Video Technology, 26(6):1058– 1069, June 2016.
8. S. K. Kuanar, K. B. Ranga, and A. S. Chowdhury. Multi-view video summarization using bipartite matching constrained optimum-path forest clustering. IEEE Transactions on Multimedia, 17(8):1166–1173, Aug 2015.
9. S. Ou, C. Lee, V. S. Somayazulu, Y. Chen, and S. Chien. On-line multiview video summarization for wireless video sensor network. IEEE Journal of Selected Topics in Signal Processing, 9(1):165–179, Feb 2015.
10. J. Yin et al., "Remote Video Surveillance Applications in Landslide Monitoring", Advanced Materials Research, Vols. 594-597, pp. 1086-1092, 2012
11. Y. Fu, Y. Guo, Y. Zhu, F. Liu, C. Song, and Z. Zhou. Multi-view video summarization. IEEE Transactions on Multimedia, 12(7):717–729, Nov 2010.
12. Merialdo. Multi-video summarization based on video-mmr. In 11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10, pages 1–4, April 2010.
13. R. Panda and A. K. Roy-Chowdhury. Multi-view surveillance video summarization via joint embedding and sparse optimization. IEEE Transactions on Multimedia, 19(9):2010–2021, Sep. 2017.
14. Tao Xiang and Shaogang Gong. Activity based video content trajectory representation and segmentation. In BMVC, 2004.
15. S. E. F. de Avila, A. d. Jr., A. de A. Arau´jo, and M. Cord. Vsumm: An approach for automatic video summarization and quantitative evaluation. In 2008 XXI Brazilian Symposium on Computer Graphics and Image Processing, pages 103–110, Oct 2008.
16. Y. F. Ma, X. S. Hua, L. Lu, and H. J. Zhang, "A generic framework of user attention model and its application in video summarization," IEEE Trans. Multimedia, vol. 7, no. 5, pp. 907–919, Oct. 2005
17. Chong-Wah Ngo ; Yu-Fei Ma ; Hong-Jiang Zhang. Video summarization and scene detection by graph modeling. IEEE Transactions on Circuits and Systems for Video Technology, 15(2):296–305, Feb 2005.
18. S. Lu, I. King, and M. R. Lyu. Video summarization by video structure analysis and graph optimization. In 2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763), volume 3, pages 1959–1962 Vol.3, June 2004.
19. A. Mahajan and Daniel DeMenthon. Automatic performance evaluation for video summarization. 2004