

Multi-view Video Summarization: A Review

Vishal Parikh¹ and Rutvi Nathwani²

¹⁻²Computer Science and Engineering Department, Nirma University, S G Highway, Ahmedabad, 382481, Gujarat, India

Abstract—Video Camera use cameras to capture videos and photos that may then be compressed, saved and transmitted to a location with a limited number of monitors. For security and environmental monitoring, all public buildings such as banks, educational institutions, offices, and hospitals now have several cameras installed with the field of view overlapping. Video summarization is a method of generating a summary of entire video content using still photos or video skimming. The length of the summary video should be smaller than that of the original video and it should include all of the highest information frames from the original video. Video summaries are more difficult to create because the video recorded by cameras is longer, also, include non-informative events, and the same scene is recorded multiple times with different views because of that inter-view dependencies and illumination variations.

Index Terms— Keyframes selection, Multi-view Videos, Surveillance Videos, Video Summarization.

I. INTRODUCTION

The advancement of technology and resources has resulted in a huge increase in the number of surveillance cameras deployed for real-time monitoring purposes in various areas such as highways, offices, and other public locations. These cameras are capable of producing hours of video content with overlapping fields of view. Reviewing such big recordings for the purpose of collecting key events (for example, in the case of a theft or burglary) can take more time. As a result, it becomes critical to create an effective and comprehensive summary of these movies in order for operators to quickly identify the key and intriguing events and take appropriate action. A-Frame is the most basic unit of a video. A shot is made up of the frames recorded by a continuous camera at regular intervals of time. The grouping of photographs is called a scene. There are a few strategies that can be used to shorten long videos like Fast forward [1], keyframe extraction [2], video skimming [3], video digest [4], video condensation [5], video carving [6], video summarization [7], video synopsis [8].

II. VIDEO SUMMARIZATION

Video Summarization creates a summary of a long video that contains all relevant and also non-relevant information in a short amount of time. Video Summarization is a method to extract the frames from the original video. The structure of the video is given in figure 1. There are two types of video summarization: static video summarization and dynamic video summarization. Regardless of the time or sequence, Static Video Summarization generates simply the keyframes (pictures). Dynamic Video Summarization, also known as Video Skim, places the keyframes in the given sequence. Video Summarization creates a summary of a long video that contains all relevant and also non-relevant information in a short amount of time. There are two types of video summarization techniques: single view video summarization or monocular video summarization and multi-view

video summarization. Single view capture event from the fixed angle with the only single view where multi-view video capture same event but on different angles using different cameras [9]. In [10, 11], from single camera captured, there are numerous methods have been proposed for summarizing videos. Multi-camera networks, which record the same event from various perspectives for better coverage, have been increasingly popular in recent years. Innovative methods for summarising videos have been proposed as the multicamera network has developed and multi-view videos with overlapping field of views acquired from multiple cameras have highly correlated contents [12].

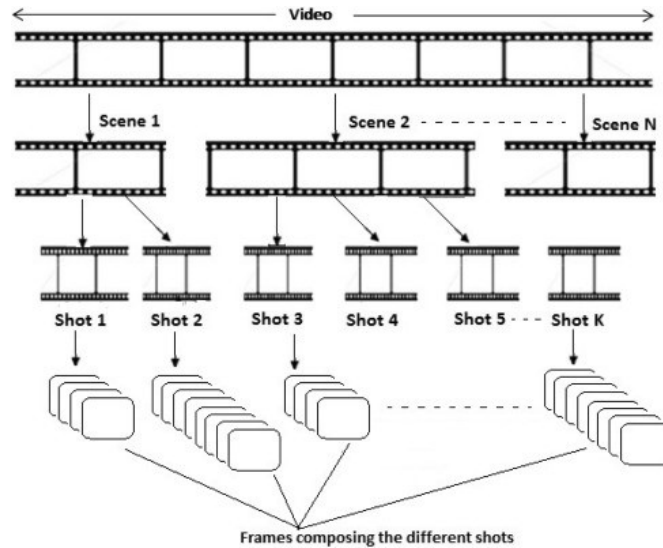


Fig 1: Structure of a Video

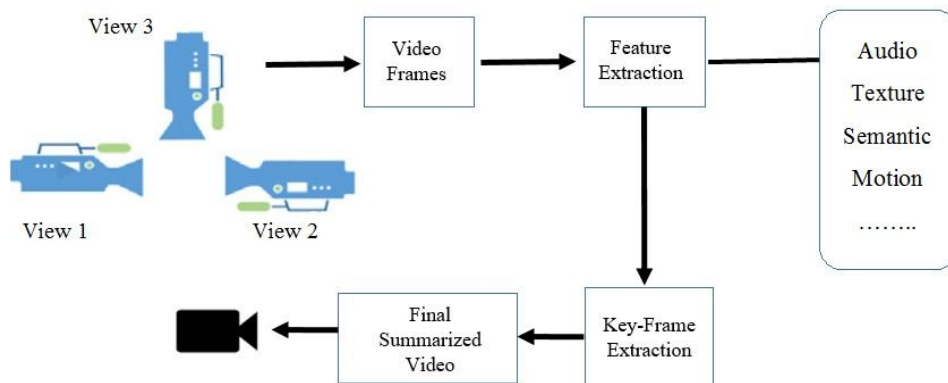


Fig. 2: Video Summarization Process

Standard mono-view video summary generating methods cannot be easily used to multi-view videos since each view would be evaluated separately, ignoring inter-video relationships, resulting in a summarised video with duplicated contents [12].

Video pre-processing, extraction of features, keyframe retrieval, and summary are some of the most typical phases in video summarization methodologies. Figure 2 shows the overall structure of Video Summarization. Frames are created from the given videos. Various ways for extracting frames from the given video can be identified in the literature such that Sufficient Content Change, Maximum Frame Coverage, Minimum Correlation, Curve Simplification, Clustering, etc. [13]. Video pre-processing, extraction of features, keyframe retrieval, and summary are some of the most typical phases in video summarization methodologies. Visual features such as colour, texture, and shape are combined with semantic features in the form of a visual bag of words [14]. After feature extraction, compute intra and inter-interview correlations and generate final summary [14]. Figure 2 shows the overall structure of Video Summarization. Frames are created from the given videos. Various ways for extracting frames from the given video can be identified in the literature such that Sufficient Content Change, Maximum Frame Coverage, Minimum Correlation, Curve Simplification, Clustering, etc

The distinction between keyframe based summarization and video skimming is represented in table 3 also keyframe based summarization is divided further into sampling, scene segmentation, and shot segmentation given in figure 3 [17]. There has been some work on Multiview video synopsis described, and they all output n number of synopsis, where n is the number of cameras, each view has a different background, multiple synopses are generated [12]. The fundamental benefit of video skimming is that it may include both motion and audio data. Users will find the summary more interesting and insightful as a result of this. In any case, these two methods of video summary can be switched back and forth.

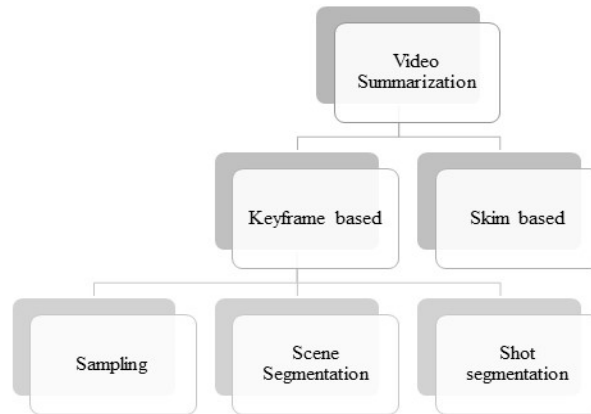


Fig. 3: Types of Video Summarization

TABLE I. PAPER REFERENCE WITH USED DATASET

Paper	Dataset
[15]	Personalized videos
[16]	Personalized videos
[17]	Office, Lobby
[17]	Office, Lobby
[18]	BL-7F, Office, Lobby
[19]	Ballroom Multi-view videos
[20]	Office, Campus, Lobby, Open Video Project, YouTube
[21]	Office, Campus, Lobby, Road, Badminton, BL-7F
[14]	Office 1, Office lobby, Campus, BL-7F
[22]	Office, Campus, Lobby
[23]	PETS 2001, Two Cameras, Eleven Cameras
[24]	Office
[25]	Road, Office, YouTube

TABLE II. DATASETS

Dataset	No. of views	Duration (Mins.)
Office	4	46:19
Lobby	3	24:12
BL-7F	19	136:10
Campus	4	56:43

TABLE III. KEYFRAME BASED SUMMARIZATION VS VIDEO SKIMMING

Keyframe based summarization	Video skimming
Using the keyframe of the video, it will generate a summary	Using video, it will generate a summary but the length of that video is very small as compared to the input video.
There is no moving information	There is moving information
Contain frame of video	Contain audio, text, and video
Used for searching from video	Used for searching from the video but less efficient compared to keyframe based approach

III. RELATED WORKS

T. Hussain et al. in [24], Using multi-cameras, track objects and capture the maximum area. In[35] this work focuses on the multi-camera system to obtain a single event and addresses the problem of summarising event sequences collected in an office setting. Fu et al. in [31] first, give a method for multi-view summarization. To generate a summary, a Spatio-temporal shot graph used and using random walks, identify similar cluster of event, Gaussian entropy fusion scheme is used for solving optimization and generate final summary [31]. In [23] detects activity motifs and compresses the depiction of individual activities to shorten the input films and by eliminating from the summary activity events that may be predicted by previous occurrences, redundancy is removed across camera perspectives. In [36], observed local motion from original input video and using spatio-temporal, final summary generated.

Li et al. in[37] proposed correlation map that effectively describes correlations with all multi-keyframes, using probabilistic theory to compute correlations, using the center of keyframes, identifying event class. Li et al. in [38] proposed method that focuses on motion, speed of motion, and all motion objects from input video included in the final summary. Video summary generate through semantic-face, video caption as well as visual feature clustering in [39]. Y. Li al. in [34] builds algorithm on the Maximal Marginal Relevance algorithm for text summarising. Gamhewage C et al. in [40] proposed method For each person in the house, data from floor sensors was used to extract video sequences and key frames without missing any event.

R. Panda et al. in [22] proposed a framework for summarization of multi-view videos that considers both intra- and inter-view correlations and generates a summary using sparse coding., the proposed method focuses on motion, speed of motion, and all motion objects from input video included in the final summary. Kawashima et al. in [41] proposed a method for a baseball game that extracts frame from the telecast. Parikh et al. in [42] give a camera placement algorithm that covered the maximum surveillance area. Shen et al. in [43] proposed a technique that selects the camera automatically and based on camera view performance, the Quality Of View is calculated. Parikh et al. in [44] give a comparative analysis of keyframe extraction techniques for closed room scenarios.

In Table IV, the output format with the proposed method, and in Table 5, which type of feature extract with paper reference is given.

TABLE IV. SUMMARY TYPE, METHOD USED FOR GENERATING SUMMARY

Paper	Year	Type	Method	Remarks
[26]	2021	Key frame based	Camera placement strategy as well as a keyframe-based technique by preserving intra-view and inter-view correlation for creating video summaries [26]	Placing multiple cameras such that maximum area can be covered. Keyframes are selected from inter-view and intra-view and final summaries are generated.
[24]	2019	Video skimming	CNN and Bidirectional LSTM	Segmented video into shots, stored in the lookup table, shots send to the cloud, using CNN extract the feature, using DB-LSTM generate video summary.
[25]	2019	Key frame based	IIoT based framework	Capture data using a camera that connects with client RPi, frame send to master RPi. It decodes the frame and extracts information and the highest information keyframes use during the summary.
[22]	2016	Key frame based	Sparse coding method over the embedding	For video summary, Joint embedded sparse coding was used that considered inter-view and intra-view correlations.
[20]	2016	Key frame based	summarizing video using content correlations via joint embedding [20]	Introduced intra and inter-view correlations and for generating the summaries using sparse coding [20]
[14]	2015	Key frame based	Optimum-path forest algorithm	Using bipartite graph matching, removing inter-view dependencies, and using the optimum path forest algorithm, the final summary is generated.
[27]	2015	Key frame based	On-line multi-view video summarization [27]	The proposed method has a low complexity and a limited memory resource need also Experiments indicate that when compared to existing offline summarising approaches, the suggested scheme produces good results [27].

[28]	2014	Video skimming	View Tendency by Support Vector Machine (SVM) [28]	Select viewpoint based on user's interest.
[29]	2014	Key frame based	On-line, distributed multi-view summarization system [29]	It saves bandwidth and server storage, and the proposed approach could be a good first step toward online, distributed summarization [29].
[30]	2014	Key frame based	Multiview metric learning Framework [30]	In each view, this method achieves a good balance between cluster specificity and similarity to the original metric [30].
[31]	2010	Video skimming	Spatio-temporal shot graph is use for representation of multi-view videos, cluster the event-centered using random walks and multi-objective optimization to get the final	In this method requires the establishment of various parameters, the values of which are now determined experimentally through tests, but it would be even better if the parameters could be set automatically [31].
[26]	2021	Key frame based	Camera placement strategy as well as a keyframe-based technique by preserving intra-view and inter-view correlation for creating video summaries [26]	Placing multiple cameras such that maximum area can be covered. Keyframes are selected from inter-view and intra-view and final summaries are generated.

TABLE V. FEATURES AND SUMMARY GENERATION

Paper	Year	Features	Summary generation
[24]	2020	Deep Features	The frames with the highest probability are combined in this stage to create a single summary video [24].
[25]	2019	Both high level and low-level features	To compute the information included in frames using two methods: entropy and complexity, these approaches produce a real value between 0 and 1, and the keyframes with the highest sum of complexity and entropy are chosen [25].
[21]	2017	C3D feature	Summarizing multi-view videos via joint embedding learning and 2,1 sparse optimizations
[12]	2016	Low level features	Using fuzzy inference system, calculate the visibility score and if the visibility score is higher, the synopsis will show the related object track [12].
[22]	2016	Low level features	Sparse representative selection method [22]
[32]	2015	Its extraction includes two tasks: 1) moving object segmentation and 2) object tracking	Sticky tracking algorithm may include multiple moving objects, of which the chronological orders are naturally preserved within the tube
[14]	2015	Both visual (e.g. color, texture and shape) and semantic features (visual bag of word)	Visual bag of words are obtained by applying K-means clustering. K-means clustering algorithm is applied to visual features (color, edge and shape) OPF clustering is finally used for summarizing the multi-view videos.
[33]	2015	Moving Object Detection and Localization [33]	machine learning based classifiers
[30]	2014	Low level feature	K-means
[29]	2014	Low level feature (256-bin color histogram in HSV color space)	K-means
[23]	2014	Extracting features from the local motion	Spectral clustering
[27]	2014	Color histogram or edge histogram	Clustered by an on-line Gaussian mixture model (GMM)
[18]	2014	Object detection	Graph matching algorithm
[31]	2010	Low-level features (color histogram and wavelet coefficients, and compute an importance score) for shot information	Graph clustering algorithm
[34]	2010	Low-level features	K-means clustering algorithm

IV. DATASET

The popular and most used datasets of multi-view summarization are office dataset [17], Lobby dataset [17], Campus dataset [21], BL-7F [27]. In table 1, the dataset name with paper reference is given and table 2 give information of datasets like number of cameras used and total duration.

- Office dataset: This dataset was made up of four cameras set in an office room where the field of view was overlapped [29]. Here, all the cameras were not synced.
- Lobby dataset: This dataset was made up of three cameras set in the lobby where the field of view was overlapped [29]. This dataset has a lot of complex scenes, which makes summarising more difficult.
- BL-7F dataset [27]: Among all multi-view summarization datasets, BL-7F is the largest. This dataset was made up with nineteen cameras set in Taiwan University's Berrylam building and the recorded videos are exceedingly diverse and complex.
- Campus dataset: The campus dataset was shot outdoors on a university campus with a lot of random activities. This dataset was made up of four camera sets. The campus dataset is difficult to work with since they are not synced and contain camera motion.

V. CONCLUSION

Using the multi-view video summarization method, less informative frames as well as redundant frames can be removed. Because of that final video of the summary includes only informative scenes and events. Many approaches and algorithms have been presented in the recent development of multi-view video summarization. We present a detailed review of the multi-view video summarization papers published to date. We include dataset descriptions used in video summarization.

REFERENCES

- [1] O'callaghan, D., Lew, E.L.: Method and apparatus for video on demand with fast forward, reverse and channel pause. Google Patents. US Patent 5,477,263 (1995)
- [2] Liu, T., Zhang, H.-J., Qi, F.: A novel video key-frame-extraction algorithm based on perceived motion energy model. *IEEE transactions on circuits and systems for video technology* 13(10), 1006–1013 (2003)
- [3] Ma, Y.-F., Zhang, H.-J.: A model of motion attention for video skimming. In: *Proceedings. International Conference on Image Processing*, vol. 1, p. (2002). IEEE
- [4] Pavel, A., Reed, C., Hartmann, B., Agrawala, M.: Video digests: a browsable, skimmable format for informational lecture videos. In: *UIST*, vol. 10, pp. 2642918–2647400 (2014). Citeseer
- [5] Feng, S., Lei, Z., Yi, D., Li, S.Z.: Online content-aware video condensation. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2082–2087 (2012). IEEE
- [6] Chen, B., Sen, P.: Video carving. In: *Eurographics (Short Papers)*, pp. 63–66 (2008)
- [7] Ajmal, M., Ashraf, M.H., Shaker, M., Abbas, Y., Shah, F.A.: Video summarization: techniques and classification. In: *International Conference on Computer Vision and Graphics*, pp. 1–13 (2012). Springer
- [8] Pritch, Y., Rav-Acha, A., Peleg, S.: Nonchronological video synopsis and indexing. *IEEE transactions on pattern analysis and machine intelligence* 30(11), 1971–1984 (2008)
- [9] Parihar, A.S., Pal, J., Sharma, I.: A study on multi-view video summarization techniques. In: *2nd International Conference on Communication & Information Processing (ICCIP)* (2020)
- [10] Money, A.G., Agius, H.: Video summarisation: A conceptual framework and survey of the state of the art. *Journal of visual communication and image representation* 19(2), 121–143 (2008)
- [11] Truong, B.T., Venkatesh, S.: Video abstraction: A systematic review and classification. *ACM transactions on multimedia computing, communications, and applications (TOMM)* 3(1), 3 (2007)
- [12] Mahapatra, A., Sa, P.K., Majhi, B., Padhy, S.: Mvs: A multi-view video synopsis framework. *Signal Processing: Image Communication* 42, 31–44 (2016)
- [13] Parikh, V., Mehta, J., Shah, S., & Sharma, P. (2021). Comparative analysis of keyframe extraction techniques for video summarization. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, 14(9), 2761–2771.
- [14] Kuanar, S.K., Ranga, K.B., Chowdhury, A.S.: Multi-view video summarization using bipartite matching constrained optimum-path forest clustering. *IEEE Transactions on Multimedia* 17(8), 1166–1173 (2015)
- [15] Schuldts, C., Laptev, I., Caputo, B.: Recognizing human actions: a local svm approach. In: *Proceedings of the 17th International Conference on Pattern Recognition*, 2004. *ICPR 2004*, vol. 3, pp. 32–36 (2004). IEEE
- [16] Blank, M., Gorelick, L., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. In: *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, pp. 1395–1402 (2005). IEEE
- [17] Parihar, A.S., Pal, J., Sharma, I.: Multiview video summarization using video partitioning and clustering. *Journal of Visual Communication and Image Representation* 74, 102991 (2021)

- [18] Zhu, X., Liu, J., Wang, J., Lu, H.: Key observation selection-based effective video synopsis for camera network. *Machine vision and applications* 25(1), 145–157 (2014)
- [19] Zhang, M.-Y., Cai, W.-Y.: Multi-view video summarization algorithm for wmsn. In: 2014 International Conference on Wireless Communication and Sensor Network, pp. 213–216 (2014). IEEE
- [20] Panda, R., Dasy, A., Roy-Chowdhury, A.K.: Video summarization in a multi-view camera network. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 2971–2976 (2016). IEEE
- [21] Panda, R., Roy-Chowdhury, A.K.: Multi-view surveillance video summarization via joint embedding and sparse optimization. *IEEE Transactions on Multimedia* 19(9), 2010–2021 (2017)
- [22] Panda, R., Das, A., Roy-Chowdhury, A.K.: Embedded sparse coding for summarizing multi-view videos. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 191–195 (2016). IEEE
- [23] Leo, C.d., Manjunath, B.S.: Multicamera video summarization and anomaly detection from activity motifs. *ACM Transactions on Sensor Networks (TOSN)* 10(2), 1–30 (2014)
- [24] Hussain, T., Muhammad, K., Ullah, A., Cao, Z., Baik, S.W., Albuquerque, V.H.C.: Cloud-assisted multiview video summarization using cnn and bidirectional lstm. *IEEE Transactions on Industrial Informatics* 16(1), 77–86 (2019)
- [25] Hussain, T., Muhammad, K., Del Ser, J., Baik, S.W., Albuquerque, V.H.C.: Intelligent embedded vision for summarization of multiview videos in iiot. *IEEE Transactions on Industrial Informatics* 16(4), 2592–2602 (2019)
- [26] Parikh, V., & Sharma, P. (2022). Intelligent Video Analytic Based Framework for Multi-view Video Summarization. *International Journal of Computing and Digital System*, 619-628.
- [27] Ou, S.-H., Lee, C.-H., Somayazulu, V.S., Chen, Y.-K., Chien, S.-Y.: On-line multiview video summarization for wireless video sensor network. *IEEE Journal of Selected Topics in Signal Processing* 9(1), 165–179 (2014)
- [28] Muramatsu, Y., Hirayama, T., Mase, K.: Video generation method based on user's tendency of viewpoint selection for multi-view video contents. In: *Proceedings of the 5th Augmented Human International Conference*, pp. 1–4 (2014)
- [29] Ou, S.-H., Lu, Y.-C., Wang, J.-P., Chien, S.-Y., Lin, S.-D., Yeti, M.-Y., Lee, C.-H., Gibbons, P.B., Somayazulu, V.S., Chen, Y.-K.: Communication-efficient multi-view keyframe extraction in distributed video sensors. In: 2014 IEEE Visual Communications and Image Processing Conference, pp. 13–16 (2014). IEEE
- [30] Fu, Y., Wang, L., Guo, Y.: Multi-view metric learning for multi-view video summarization. *arXiv preprint arXiv:1405.6434* (2014)
- [31] Fu, Y., Guo, Y., Zhu, Y., Liu, F., Song, C., Zhou, Z.-H.: Multi-view video summarization. *IEEE Transactions on Multimedia* 12(7), 717–729 (2010)
- [32] Zhu, J., Liao, S., Li, S.Z.: Multicamera joint video synopsis. *IEEE Transactions on Circuits and Systems for Video Technology* 26(6), 1058–1069 (2015)
- [33] Mahapatra, A., Sa, P.K., Majhi, B.: A multi-view video synopsis framework. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 1260–1264 (2015). IEEE
- [34] Li, Y., Merialdo, B.: Multi-video summarization based on video-mm. In: 11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10, pp. 1–4 (2010). IEEE
- [35] Park, H.-S., Cho, S.-B.: A fuzzy rule-based system with ontology for summarization of multi-camera event sequences. In: *International Conference on Artificial Intelligence and Soft Computing*, pp. 850–860 (2008). Springer
- [36] Leo, C.D., Manjunath, B.: Multicamera video summarization from optimal reconstruction. In: *Asian Conference on Computer Vision*, pp. 94–103 (2010). Springer
- [37] Li, P., Guo, Y., Sun, H.: Multi-keyframe abstraction from videos. In: 2011 18th IEEE International Conference on Image Processing, pp. 2473–2476 (2011). IEEE
- [38] Li, C., Wu, Y.-T., Yu, S.-S., Chen, T.: Motion-focusing key frame extraction and video summarization for lane surveillance system. In: 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 4329–4332 (2009). IEEE
- [39] Zhuang, Y., Xiao, R., Wu, F.: Key issues in video summarization and its application. In: *Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint*, vol. 1, pp. 448–452 (2003). IEEE
- [40] De Silva, G.C., Yamasaki, T., Aizawa, K.: Evaluation of video summarization for a large number of cameras in ubiquitous home. In: *Proceedings of the 13th Annual ACM International Conference on Multimedia*, pp. 820–828 (2005)
- [41] Kawashima, T., Tateyama, K., Iijima, T., Aoki, Y.: Indexing of baseball telecast for content-based video retrieval. In: *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269)*, vol. 1, pp. 871–874 (1998). IEEE
- [42] Parikh, V., Sharma, P., Shah, V., Ukani, V.: Optimal camera placement for multimodal video summarization. In: *International Conference on Futuristic Trends in Network and Communication Technologies*, pp. 123–134 (2018). Springer
- [43] Shen, C., Zhang, C., Fels, S.: A multi-camera surveillance system that estimates quality-of-view measurement. In: 2007 IEEE International Conference on Image Processing, vol. 3, p. 193 (2007). IEEE
- [44] Parikh, V., Sharma, P., Patel, C., & Patel, H. (2021). Multi-view Keyframe Extraction Techniques: A Comparative Analysis for Closed Room Scenario. *Turkish Online Journal of Qualitative Inquiry*, 12(6).