International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:9, No:4, 2015

Video Summarization: Techniques and Applications

Zaynab El khattabi, Youness Tabii, Abdelhamid Benkaddour

***Abstract*—**Nowadays, huge amount of multimedia repositories make the browsing, retrieval and delivery of video contents very slow and even difficult tasks. Video summarization has been proposed to improve faster browsing of large video collections and more efficient content indexing and access. In this paper, we focus on approaches to video summarization. The video summaries can be generated in many different forms. However, two fundamentals ways to generate summaries are static and dynamic. We present different techniques for each mode in the literature and describe some features used for generating video summaries. We conclude with perspective for further research.

***Keywords*—**Semantic features, static summarization, video skimming, Video summarization.

I. Introduction

IN recent years, because of the rapid growth in multimedia information, the advance in internet communication and digital video technologies, multimedia information indexing and retrieval has become more and more important and lots of research efforts have been devoted to the video retrieval and video analysis based on audio or visual features. This analysis shows that, when developing retrieval applications and video indexing, we first have to consider the issue of structuring the huge and rich amount of heterogeneous information related to video content. In addition, to retrieve information from the audio or visual content is a very challenging since it requires the extraction of high-level semantic information from low-level audio or visual data. Video summarization is an important process that facilitates faster browsing of large video collections and also more efficient content indexing and access. There are two main video summarization techniques in the literature: static video summarization (video summary) and dynamic video summarization (video skimming). In order to summarize a video, most of the methods have consists on visual features computed from video frames. Also, there are methods that consider the semantic meaning implied in the video to produce a more informative summary.

This paper is organized as follows. In Section II, we describe some concepts that are generally true regardless of the modality of features used for video summarization. In Section *A*, we describe static video summarization approaches presenting the different features used. Section *B* includes dynamic video summarization methods. Finally, some application areas are presented. We provide conclusions in Section IV and also suggest some ideas for future research in this area.

Z. El Khattabi and A. Benkaddour are with LIROSA Laboratory, Faculty of Sciences, University of Abdelmalek Essaadi, Tetouan, Morocco (e-mail: zaynabelkhattabi@gmail.com, ham.benkaddour@yahoo.fr).

Y. Tabii is with LIROSA Laboratory, National School of Applied Sciences, University of Abdelmalek Essaadi, Tetouan, Morocco (e-mail: youness.tabii@gmail.com).

II. Video Summarization

Due to the increasing volume of video content on the Web, and the human effort taken to process it, new technologies need to be researched in order to develop efficient indexing and search techniques to manage effectively and efficiently the huge amount of video data. One of the most evolving research areas is Video summarization. As the name implies, video summarization is a mechanism to produce a short summary of a video to give to the user a synthetic and useful visual abstract of video sequence, it can either be a images (keyframes) or moving images (video skims). In terms of browsing and navigation, a good video abstract will enable the user to gain maximum information about the target video sequence in a specified time constraint or sufficient information in the minimum time [1]. Automatically generated summaries can support users in navigating large video archives and in taking decisions more efficiently regarding selecting, consuming, sharing, or deleting content [2]. Video abstracts can also be used as an end product to be shared, digested, and enjoyed by the user. Retaining only the essential information of a video sequence improves storage, bandwidth, and viewing time [1]. We found that the developed techniques in video summarization touch various domains, such as movies, sports, news, home videos, e-learning, etc.,

The video abstraction process usually has three phases: Video information analysis, meaningful clip selection and output synthesis. To analyze video information, it is necessary to detect salient features, structures or patterns in the visual component, the audio component and the textual component like closed captions. Fig. 1 presents these steps.

Video summarization can be represented into two modes: A static video summary (storyboard) and a dynamic video skimming. In one hand, static video summary represents a video sequence in a static imagery form (one or more selected representative frames from the original video, or a synthesized image generated from the selected keyframes). According to different sampling mechanisms, a set of keyframes are extracted from shots of the original video. Then, the selected keyframes are arranged or blended in a two-dimensional space [4]. On the other hand, dynamic summarization consists in selecting the most relevant small dynamic portions (video skims) of audio and video in order to generate the video summary [5].

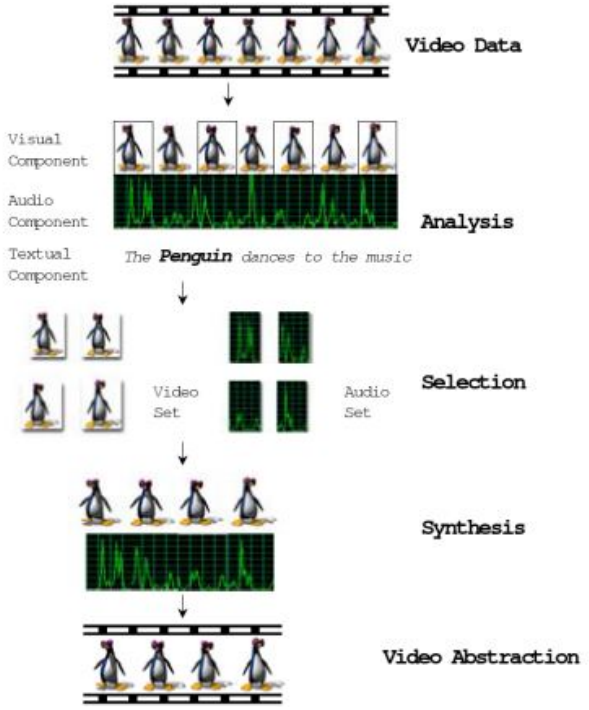

Fig. 1 Video abstraction scheme [3]

A. Static Video Summarization

In this section, we present some of existing methods for static video summarization. In [6], authors consist of extracting the keyframes by pre-sampling uniformly or randomly the original video sequence. Keyframe extraction is fundamental process in video content management. It involves selecting one or multiple frames that will represent the content of the video and used for generating video summaries. Fig. 2 shows hierarchical structure in a video sequence in the extraction of such keyframes.

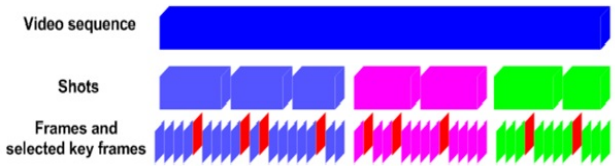

Fig. 2 Hierarchical structure of a video sequence [7]

B. Methods Based On Clustering Techniques

The basic idea is clustering together similar frames/shots and then extraction some frames (generally one frame) per cluster as key frames. These methods are different in features (e.g., color histogram, luminance, and motion vector) and clustering algorithms (e.g., k-means, hierarchical) [2].

In [8], an approach based on shot detection using color histograms is proposed. First, a RGB histogram is used to provide distribution information of colors for a given video frame. Then, the PCA (principal component analysis) is executed on the one dimensional vector representing the frequency of the color histogram in order to reduce the dimension of feature vector. In order to detect the different shots, two algorithms are performed, Fuzzy-ART and Fuzzy C-Means algorithms to automatically detect the number of clusters in the video and consequently extract the shots from the original video. Once all the keyframes are detected from the video, 10 frame neighborhood surrounding are extracted for each keyframe to produce the storyboard. The process is entirely automatic and no a priori human interaction is needed. The storyboards produced by this model are evaluated and the results show that the model is effective for finding keyframes and is not computationally expensive.

In VGRAPH approach, presented in [9], the keyframes extraction process is also a shot-based method, which requires video segmentation by detecting the shot boundaries. First, the original video is pre-sampled to reduce the number of frames to be processed. Second, the pre-sampled video is segmented into shots using the color features which are extracted using the color histogram computed from the HSV. Third, noise frames are eliminated and the second frame is selected as a shot representative. Finally, the keyframes are extracted using nearest neighbor graph which is built from the texture features extracted from the shots representative frames using Discrete Haar Wavelet Transforms.

The proposed method in [2] uses fuzzy c-means clustering algorithm. Initially, the original video is split into a set of meaningful and manageable basic elements (e.g., shots, frames). Secondly, color features are extracted to form a color histogram in HSV color space. Not all the video frames are considered, but takes a sample instead. After that, the frames are grouped by fuzzy c-means clustering algorithm. Then, one frame per cluster is selected. Finally, the remaining keyframes are arranged in the original temporal order to facilitate the visual comprehension of the result. One advantage of using FCM clustering from other techniques is the type of clustering output that is a membership matrix. This matrix specifies the most representative frames of each cluster directly.

In [10], content complexity and the content change ratio of video are defined by the SIFT (Scale-invariant feature transform) feature. proposed approach detects video shots boundaries using both the features of the content complexity and the content change ratio feature. In the meantime, it merges similar shots according to their similarities, and furthermore, estimates the number of key frame based on the segmented shots. At last, it extracts key frames from video based on the above results. The SIFT descriptor, described in [11], has been generally used in computer vision for its ability to handle intensity, rotation and scale variations, this makes it a good descriptor but it is high computational cost.

All these approaches rely on detecting shot changes, and therefore, dependent on having the correct shot detection.

Detecting shot changes automatically is still a difficult problem, due to the variety of transitions that can be used between shots. A cut is an abrupt transition between two shots that occurs between two adjacent frames. A fade is a gradual change in brightness, either starting or ending with a black frame. A dissolve is similar to a fade except that it occurs between two shots. The images of the first shot get dimmer and those of the second shot get brighter until the second shot

replaces the first one. Other types of shot transitions include wipes and computer generated effects such as morphing [12].

Shot boundary detection is not needed in [13]; an algorithm called content-based adaptive clustering (CBAC) is used. All the frames of the video are analyzed together for content changes without segmenting video into shots. The changes of the content in iteratively increasing units are compared globally to determine which part of the video is important for description. The proposed approach has been applied to histogram analysis and can extract as small as 5 percent length of representative frames from the video and it is suitable for applications on video content-browsing and content-summary.

One disadvantage of the most of the methods that relay on complicated clustering algorithms is making them too computationally complex for real time applications.

1) Methods Based On Semantic Features

Video summarization based on visual and temporal features cannot guarantee the preservation of all informative content. In other words, such a summary is a kind of lossy representation without keeping the semantic fidelity, especially when compression rate is very high [4]. Some techniques use the extraction of interesting events and objects in an attempt to find the semantically relevant key frames. A more informative summary can be obtained if the method considers the semantic meaning implied in the video. However, to summarize videos taking into account the semantic information, the methods have relied on object detection.

Bag-of-visual-words is a popular image representation that produces high matching accuracy and efficiency. Research on representative local descriptors shows that with similarity based clustering, the intra-cluster similarity extent of descriptors plays the same role in straightforward matching as vocabulary size in visual words matching [14].

The BoW(Bag of Words) is used in [5] with temporal segmentation. Initially, the temporal segmentation procedure detects the shot boundaries using color histograms. Then, each shot is clustered to detect frame samples from each shot using X-means algorithm. Afterward, a Bag-of-Visual-Words approach is adopted. The detected local features are clustered to generate the Visual Word Vocabulary. Next, the histograms of occurrences of visual words are computed from each frame of the detected frame samples. The histograms of occurrence are clustered; the method finds the frames that are closer to each clusters centroid. The frames that represent the centroids are considered as keyframes. The method filters the results to eliminate possible redundant keyframes. Finally, the keyframes are ordered in chronological order.

In [15], a method based on concept preservation is proposed. BoW model is used. First the video is segmented into shots, then for each shot the SIFT descriptor (Scale-invariant feature transform) is used to extract the local features from detected keypoints. Later these features are clustered to produce a visual word dictionary. In addition, for each shot a histogram of occurrences of visual words is generated using a visual word dictionary. Then, the histograms are grouped, meaning that similar visual entities will be grouped together. Finally, the video summary is produced by extracting the frames that contain the important visual entities.

A novel approach is presented in [4], called Near-Lossless Semantic Summarization (NLSS), to summarize a video stream with the least high-level semantic information loss. The basic idea is to segment the visual and audio streams into basic units, extract representative information from these units, and then employ efficient compression techniques to make this information compact. For that, a subsegment of a shot (subshot) is selected as the basic unit for metadata extraction. The visual track is decomposed into a series of subshots by a motion-based method, where each subshot is further classified into one of four categories on the basis of dominant motion. Regarding the aural track, an Audio Content Analysis Compression (ACAC) scheme is proposed to first segment the audio track into units, and then classify each unit into five categories. The NLSS summary consists of three components, an XML file describing the temporal structure and motion information, as well as the compressed image stream and the audio stream compressed by ACAC. The method is evaluated on TRECVID and other video collections, and demonstrates that it is a powerful tool for significantly reducing storage consumption, while keeping high-level semantic fidelity.

From another side, Bayesian methods can be used to resolve the problem of video summarization and browsing on a semantic basis. Bayesian architecture is presented in [16] for content characterization. A computational framework based on Bayesian principles is performed. This framework consists of a set of extremely simple visual sensors trained to detect relevant visual features and a probabilistic (Bayesian) network that infers the state of a set of semantic content descriptors from all the sensory information.

2) Methods Based On Visual Descriptors

In [17], an approach does not rely on complex clustering algorithms is developed. It combines MPEG-7 Color Layout Descriptors with adaptive threshold technique to detect shot boundaries. At first, video frames are sampled in order to reduce further computational burden. Then, efficient MPEG-7 Color Layout Descriptor (CLD) is extracted on pre-sampled video frames. The method sequentially computes local threshold of frame differences inside the sliding window. Shot changes are detected at places where the frame difference is maximal within the window and larger than the local threshold. For each shot, a representative keyframe is extracted and similar keyframes are eliminated in a simple manner. As a final result the most informative keyframes are selected as a video summary.

Another main approach to keyframe extraction is motion analysis. The automatic video summarization technique based on motion analysis in [6] uses optical flow computations to identify global extrema and local minimum between two maximums in the motion and using two key frame selection criteria. The advantage of proposed approach is capturing motion information which is crucial for videos containing dynamic content, particularly sports videos. Therefore, it can

provide more meaningful summary by capturing the high action content through motion analysis. A new technique for keyframe extraction based on comparison of frames; called VSUKFE (video summarization using key frame extraction) is presented in [18]. The idea is to use inter-frame differences calculated based on the correlation of RGB color channels, color histogram and moments of inertia. The three frame difference measures are then combined using an aggregation mechanism to extract key frames. An adaptive formula is used to combine frame difference measures of the previous and current iterations, which generates a smooth function and helps in handling gradual changes in lighting conditions. This technique is evaluated to demonstrate the benefits of the proposed aggregation mechanism, to show some tradeoffs of varying system parameters and to compare it with other techniques. The approach presented in [19] of static video summarization consists on a visual attention model. It gives saliency maps which highlight area of frames containing more information and which attract human gaze. These saliency maps are used to detect changes on frames during the video which make it possible to select keyframes. Finally redundant frames are eliminated.

Storyboards or video summaries are not restricted by any timing or synchronization issues and therefore, they offer much more flexibility in terms of organization for browsing and navigation purposes. But it still usually more impressive and interesting to watch a skim than a slide show of keyframes.

C. Dynamic video Summarization

The fundamental idea of video skim which is a short video composed of informative scenes from the original video presented to the user to be able to receive an abstract of the video story, but in video format [20]. For dynamic summarization (skimming), most mechanisms extract and segment video clips from the original video. Compared with static storyboard summary, there are relatively few works being addressed for dynamic video skimming. Techniques for dynamic video skimming include applying SVD (Singular Value Decomposition), motion model [21] and semantic analysis in [22] and [12]. Most techniques are based mainly on visual information and some others approaches where audio and linguistic information are also incorporated in order to derive semantic meaning.

A method is presented in [12] for the automatic extraction of summaries in soccer video based on shot detection, shot classification and FSM (Finite State Machine). It consists of four stages: playfield segmentation as a preprocessing step, shot detection using the Discrete Cosine Transform Multi-Resolution (DCT-MR) to extract the different types of shot transition, shot classification in three major classes (long shot, medium shot and close-up shot) using statistical method and finally, soccer video word extraction and finding out the appropriate sub-words which present summaries using the FSM and domain knowledge, where a set of rules are defined to present the semantic states in soccer game and explore the interesting relations between syntactic structure and the semantic of the video.

In [22], a dynamic video abstraction scheme for movie videos is presented. It is based on the progress of stories. The proposed approach attempts to comprehend video contents from the progress of the overall story and human semantic understanding. Firstly, the property of two-dimensional histogram entropy of image pixels is adopted to segment a video into shots. Then, semantically meaning scenarios are obtained according to the spatio-temporal correlation among detected shots. Finally, general rules of special scenario and common techniques of movie production are exploited to grasp the progress of a story in terms of the degree of progress between scenarios to the overall story.

More recently, spatial interest point operators have been extended to the third temporal dimension and the concept of spatio-temporal feature detectors has been introduced. In [23], a technique based on the study of spatio-temporal activity within the video is presented. First the spatio-temporal features are extracted using the spatio-temporal Hessian matrix and provides a measure of the activity within each frame. Then, this information is processed to retrieve the keyframes and thereafter the video segments (clips) where activity level is high which form the candidate set used to build the summary. As the video rushes contain a lot of redundancy, two steps are designed to first detect redundant clips and second to eliminate clapperboard images. The final step consists in fusing together of all these pieces of information to achieve the final summary taking into account the time constraint. New algorithm is proposed in [24] with a two-level redundancy detection procedure. After video segmentation step into shots using color histogram and optical-flow motion features, the cast indexing procedure is employed to generate the storyboards of cast in the video. Then similar key frames are removed using HAC in each scene. An original video summary is constructed by extending the selected key frames with impact factors of scenes and key frames. A further repetitive frame segment detection step is designed to remove redundant information left in the initial video summary.

A successful skimming approach involves using information from multiple sources, including sound, speech, transcript, and video image analysis. In [25] an example of this approach is presented, which automatically skims news videos with textual transcriptions by first abstracting the text using classical text skimming techniques and then looking for the corresponding parts in the video. This method creates a skim video, which represents a short synopsis of the original. The goal was to integrate language and image understanding techniques for video skimming by extracting significant information, such as specific objects, audio keywords, and relevant video structure.

Compared to static storyboards, dynamic videos skimming also support the recognition of objects in the content, and their representativeness is enough even for replacing the original video content [26].

III. Applications

Many professional and educational applications that involve generating or using large volumes of video and multimedia data are prime candidates for taking advantage of video content analysis techniques [27]. The developed techniques in video summarization touch various domains; we find in [3], three categories presented: Consumer video applications, Image-Video databases management and surveillance. For each category, some of the exemplar applications are listed. With the increasing in the storage and computational capacity of consumer electronic devices such as personal video recorders (PVR), consumer video applications enables the end user of browsing the recorded content in efficient ways and view the interesting parts quickly. On the other hand, Image and video databases management includes different application areas like video search engine, digital video library, object indexing and retrieval, automatic object labeling and object classification.

Consequently, Media organizations and TV broadcasting companies have shown considerable interest in these applications, especially in organizing and indexing large volumes of video data to facilitate efficient and effective use of these resources for internal use. These large video libraries create a unique opportunity for using intelligent media analysis techniques to create advanced searching and browsing techniques to find relevant information quickly and inexpensively. Intelligent video segmentation and sampling techniques can reduce the visual contents of the video program to a small number of static images. We can browse these images to spot information and use image similarity searches to find shots with similar content and motion analysis to categorize the video segments. Higher level analysis can extract information relevant to the presence of humans or objects in the video. Audio event detection and speech detection can extract additional information to help the user find segments of interest [27].

IV. Conclusion

Recently, video summarization has attracted considerable interest from researchers and as a result, various algorithms and techniques have been proposed. In this work, we have carried out a review of the research in two dominant forms of video summarization: static summary and the skim. We identified important elements and described how they are addressed in specific works. Regardless of whether static or dynamic forms are employed, the evaluation process showed that the techniques proposed produces video summaries of high visual quality and some approaches are suitable for real time video processing of diverse types of compressed videos. However, a valid evaluation method can support the field to advance and the best technique for abstracting a video sequence to be identified.

Although, video abstraction is still largely in the research phase; practical applications are still limited in both complexity of method and scale of deployment. For example, video search services such as Yahoo and Alta Vista use a single keyframe to represent the video, while Google provides a context-sensitive keyframe list of the video. In addition, the amount of research carried out in the domain of video summarization using machine learning is quite less even the importance of learning algorithms which allow potentially the computer to understand the media collection on a semantic level. Video summaries created with semantic analysis were the most similar to the users summaries, for that, a more informative summary can be obtained if the method considers the semantic meaning implied in the video and combine it with other visual descriptors.

References

[1] B. T. Truong and S. Venkatesh, “Video Abstraction: A Systematic Review and Classification,” ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 3, no. 3, February 2007.

[2] E. Asadi and N. M. Charkari, “Video Summarization Using Fuzzy C-Means Clustering,” in 20th Iranian Conference on Electrical Engineering, (ICEE2012). IEEE, 15-17 May 2012, pp. 690 – 694.

[3] Z. Xiong, R. Radhakrishnan, A. Divakaran, Y. Rui, and T. S. Huang, A Unified Framework for Video Summarization, Browsing Retrieval: with Applications to Consumer and Surveillance Video. Academic Press.2006.

[4] T. Mei, L.-X. Tang, J. Tang, and X.-S. Hua, “Near-Lossless Semantic Video Summarization and Its Applications to Video Analysis,” ACM Trans. Multimedia Computing Communications and Applications, June 2013.

[5] E. J. Y. C. Cahuina and G. C. Chavez, “A New Method for Static Video Summarization Using Local Descriptors and Video Temporal Segmentation,” in 26th Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, 5-8 Aug 2013, pp. 226–233.

[6] E. Mendi, H. B. Clemente, and C. Bayrak, “Sports video summarization based on motion analysis,” Computers and Electrical Engineering, vol. 39, pp. 790–796, April 2013.

[7] E. Mendi and C. Bayrak, “Shot Boundary Detection and Key Frame Extraction using Salient Region Detection and Structural Similarity,” in ACM SE ’10 Proceedings of the 48th Annual Southeast Regional Conference, no. 66, 2010.

[8] E. J. Y. Cayllahua-Cahuina, G. Camara-Chavez, and D. Menotti, “A static video summarization approach with automatic shot detection using color histograms,” 2012.

[9] M. K.M., G. N.M., and I. M.A., “VGRAPH: An Effective Approach for Generating Static Video Summaries,” in IEEE International Conference on Computer Vision Workshops (ICCVW). IEEE, 2-8 Dec 2013, pp. 811 – 818.

[10] J. Li, “Video Shot Segmentation and Key Frame Extraction Based on SIFT Feature,” in 2012 International Conference on Image Analysis and Signal Processing (IASP). IEEE, 9-11 Nov 2012, pp. 1–8.

[11] L. D.G., “Object Recognition from Local Scale-Invariant Features,” in The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, vol. 2. IEEE, 20-27 Sep 1999, pp. 1150–1157.

[12] Y. Tabii and R. O. H. Thami, “A new method for soccer video summarization based on shot detection, classification and finite state machine,” in SETIT 2009 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, 22-26 March 2009.

[13] X. Sun and M. S. Kankanhalli, “Video Summarization Using R-Sequences,” Real-Time Imaging, vol. 6, pp. 449–459, December 2000.

[14] J. Hou, J. Kang, and N. Qi, On Vocabulary Size in Bag-of-Visual-Words Representation, ser. Lecture Notes in Computer Science, 21-24 September 2010, vol. 6297, pp. 414–424.

[15] Z. Yuan, T. Lu, D. Wu, and H. Y. Yu Huang, “Video Summarization with Semantic Concept Preservation,” in MUM ’11 Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia, 2011, pp. 109–112.

[16] V. N. and L. A., “Bayesian Modeling of Video Editing and Structure: Semantic features for Video Summarization and Browsing,” in Proceedings of 1998 International Conference on Image Processing. ICIP 98, vol. 3. IEEE, 4-7 Oct 1998, pp. 153–157.

[17] Cvetkovic, S. Jelenkovic, and S. V. M. Nikolic, “Video summarization using color features and efficient adaptive threshold technique,” Przeglad Elektrotechniczny, vol. R. 89, nr 2a, pp. 247 – 250, 2013.
[18] N. Ejaza, T. B. Tariqb, and S. W. Baika, “Adaptive key frame extraction for video summarization using an aggregation mechanism,” Journal of Visual Communication and Image Representation, vol. 23, no. 7, pp.1031–1040, October 2012.
[19] M. Sophie, G. Mick¨ael, and P. Denis, “Résumé de vidéo à partir d’un modèle d’attention visuelle,” in GRETSI - Actes de Colloque, G. d’Etudes du Traitement du Signal et des Images, Ed., 2007.
[20] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, “Content-Based Multimedia Information Retrieval: State of the Art and Challenges,” ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 2, pp. 1–9, February 2006.
[21] Y.-F. Ma and H.-J. Zhang, “A model of motion attention for video skimming,” in Proceedings. International Conference on Image Processing, vol. 1. IEEE, 2002, pp. 129–132.
[22] S. Zhu, Z. Liang, and Y. Liu, Automatic Video Abstraction via the Progress of Story, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, vol. 6297, pp. 308–318.
[23] R. Laganière, R. Bacco, A. Hocevar, P. Lambert, G. Païs, and B. E. Ionescu, “Video summarization from spatio-temporal features,” in TVS’08 Proceedings of the 2nd ACM TRECVid Video Summarization Workshop, 2008, pp. 144–148.
[24] Y. Gao, W.-B. Wang, J.-H. Yong and H.-J. Gu, “Dynamic video summarization using two-level redundancy detection,” Multimedia Tools and Applications, vol. 42, pp. 233–250, April 2009.
[25] M. G. Brown, J. T. Foote, G. J. F. Jones, K. S. Jones, and S. J. Young, “Automatic content-based retrieval of broadcast news,” in MULTIMEDIA ’95 Proceedings of the third ACM international conference on Multimedia. ACM Digital Library, 1999, pp. 35–43.
[26] S. P. Rajendra and K. N, “A Survey of Automatic Video Summarization Techniques,” International Journal of Electronics, Electrical and Computational System (IJEECS), vol. 3, April 2014.
[27] N. Dimitrova, H.-J. Zhang, B.Shahraray, I. Sezan, T. Huang, and A. Zakhor, “Applications of video-content analysis and retrieval,” in IEEE Multimedia, IEEE, Ed., 2002.