JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Extending the Horizon: Key Frame Extraction from Video Sequences - A Comprehensive Comparison and In-Depth Analysis

[1]Swati J. Patel, [2]Mehul C. Parikh

[1]Assistant Professor, [2]Associate Professor,
[1,2]Information Technology Department,
[1,2]L.D.College of Engineering, Ahmedabad, India

Abstract: Media on the Internet has entirely taken over the world as a result of the digital revolution. Because of the expanding use of internet services, falling prices for digital storage devices, and the introduction of 4-G technology, digital video may now be made available to everyone. An enormous video collection is continually expanding, and analyzing such a large amount of data always takes time. The video sequence is made up of several still images known as frames. A movie has a lot of information, and as a result, the frames frequently contain unnecessary and identical material that is worthless if the film's substance is important. A practical and instructive presentation is required for the proper processing of video content. It is critical to automatically select relevant and instructive content from videos. Keyframe extraction is deemed appropriate for complete video analysis since it removes replications and extracts important frames from the movie. A key frame is a representative frame that includes the video collection's facts, representing important information from the video. It is not only useful for recognizing the entire video, but it may also minimize the processing time, computational expenses, and storage needs of each video sequence in a variety of applications. One of the most essential tasks in video processing is the extraction of these frames. This document discusses various key-frame extraction strategies that have been developed in the past.

IndexTerms **- Key Frame Extraction; Video Processing; Video Key Frame.**

1. INTRODUCTION

The advent of video recording devices, including smartphones, portable cameras, and surveillance equipment, has streamlined the processes of capturing, sharing, and creating videos. This has led to an explosive growth in video data. With the widespread use of digital media on the Internet for purposes such as information dissemination, education, entertainment, business, and surveillance, video processing has emerged as a prominent area of research in the field of image processing. [1]. Processing an entire video sequence, composed of numerous frames at a frame rate of at least 24 frames per second (fps) for high-definition video, is not recommended. Instead, it is preferable to employ methods that can extract key frames from a video sequence. These key frames are adequate for representing the video and can be effectively utilized for recognizing the entire video sequence. [2]. Keyframes offer a rapid overview of video content and contribute to minimizing computational complexity in various video analysis and retrieval applications. The video can be reconstructed by utilizing the extracted keyframes.[3] Keyframes serve as the foundational elements for a variety of tasks, encompassing video browsing, summarization, searching, comprehension, and chapter titles in DVDs. Their application extends to various domains, such as surveillance, medical imaging, underwater exploration, web browsing videos, sports and news programs, as well as indoor and outdoor video scenarios.

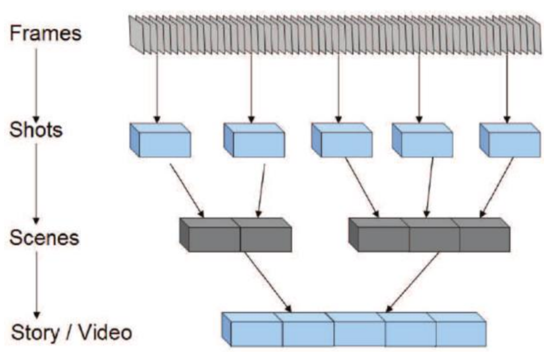

Figure 1: Structural-hierarchy-of-a-video [4]

Video content comprises an extensive collection of data objects, often characterized by a considerable amount of redundant and nonessential information. The intricate structure of a video, depicted in Figure 1, involves the arrangement of scenes, shots, and frames. [5]. A shot refers to a continuous, successive sequence of frames captured by a single camera during uninterrupted action. Meanwhile, the key frame serves as a segment of the video that encapsulates a visual summary, containing meaningful information about the sequence. [6]. The key Frame must contain the high-priority entities and events of the video and be free of repetition and redundancy [7]. Video processing plays a crucial role in numerous applications, encompassing tasks like watermarking, scene segmentation, detection of shot boundaries within those scenes, and extraction of key frames from those shots. A key frame, whether singular or a set of frames, functions as a representation of the entire content of a video clip. It denotes the image frame in the video sequence that is highly representative, reflecting a comprehensive summary of the video content by incorporating most salient features. [3]. The essence of key-frame extraction lies in pinpointing the most distinctive segments of a video sequence, chosen to enable the reproduction of the entire video. The number of key-frames extracted from a single shot depends on the complexity of the content within that shot. A shot, defined as an uninterrupted sequence of frames captured by a single camera, serves as the foundational unit of video. In scenarios where video data comprises multiple shots, it becomes imperative to identify and delineate individual shots for the purpose of key-frame extraction. [8]. Selecting key-frames from a video is a ranking process of unique frames regarding their representativeness to the video [3]. Utilizing keyframes allows for a concise representation of the primary content in video data, leading to a reduction in the required memory for video processing and simplifying overall complexity. The selection of key-frames is guided by three crucial properties: continuity, priority, and repetition. Continuity emphasizes the need for minimal interruption in the video sequence. Priority involves highlighting specific objects or events that hold greater significance in a given application, requiring key-frames to feature these high-priority elements—an aspect heavily dependent on the task at hand. Repetition underscores the importance of avoiding redundant representation of identical events. Successfully incorporating these semantic properties can pose a challenge in the keyframe selection process. [9].

2. CLASSIFICATION OF APPROACHES FOR KEY FRAME EXTRACTION:

2.1. Uniform Sampling Method:

The prevalent technique for key frame extraction is uniform sampling, where every kth frame is selected from the video sequence. The value of k is predetermined based on the video's length, with larger values for longer sequences and smaller values for shorter ones. Typically, the goal is to extract 5% to 15% of key frames from the original video. While this method is straightforward, it lacks semantic relevance as it doesn't consider the content's meaning.[10] [7]. As it is based on the predefined fixed value, these approaches are not content-based and do not consider the dynamics of the visual content, and selected frames are often unstable [11]. This technique is very easy to implement and computationally efficient, but lacks semantic relevance and may miss capturing important visual information.

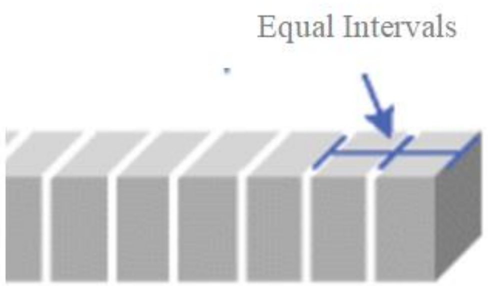

Figure 2: Uniform sampling

2.2. Pixel Compare Method:

In this approach, each successive frame is subjected to a pixel-wise comparison, and a frame is designated as a keyframe when the difference in comparison exceeds a predetermined threshold. However, this method is characterized by high time consumption and excessive sensitivity to the motion of objects within the frame. [12].

2.3. Image Histogram Method:

The image histogram provides information on the distribution of pixels across brightness values, ranging from 0 to 256. Leveraging this, keyframes can be extracted. In this approach, the histogram for each frame is computed, and the dissimilarity between two consecutive frames is assessed based on the histogram difference. If the histograms of two successive frames exhibit a dissimilarity of 50% or more, the system identifies and extracts that frame as a keyframe. [7].

2.4. Scale-Invariant Feature Transform:

The Scale-Invariant Feature Transform (SIFT) is a pivotal technique for feature detection, enabling the identification and description of local features within an image. Widely employed in computer vision applications, the SIFT feature descriptor exhibits invariance to uniform scaling, orientation, illumination changes, translation, and rotation, and partial invariance to affine distortion. Leveraging these properties, SIFT features can be effectively utilized for keyframe extraction.

The process begins by defining significant locations using a scale space created from smoothed and resized images. The application of Difference of Gaussian functions on these images reveals maximum and minimum responses. Non-maxima suppression is then implemented, and putative matches are discarded to ensure a collection of highly distinctive and interesting key points. Further, a Histogram of Oriented Gradients is conducted by dividing the image into patches, determining the dominant orientation of localized key points. In essence, these key points serve as extracted local features for subsequent keyframe identification. [7] [13]

Advantages of this approach are robust feature detection, invariant to scale, orientation, and illumination changes, and drawbacks are computational complexity and sensitivity to affine distortions.

2.5. Cluster-Based Method:

Clustering serves as a widely embraced method for keyframe extraction, leveraging algorithms capable of automatically categorizing video data based on their similarities. In this approach, keyframe clusters are formed by utilizing data points and diverse features from video sequences. The set of keyframes is then composed of frames with the shortest distance from the cluster's center. While this method effectively captures the global characteristics of the scene, it comes with the drawback of necessitating a substantial computational investment for both cluster generation and feature extraction from the scene. [12]. The main drawback of these methods is that, depending on the number of clusters, keyframes can be either redundant or fail to represent the content of the whole shot efficiently. Captures global characteristics of scenes andadaptable to various visual content are the advantages and computational cost for cluster generation and potential sensitivity to specific data characteristics is the most disadvantage of this method.

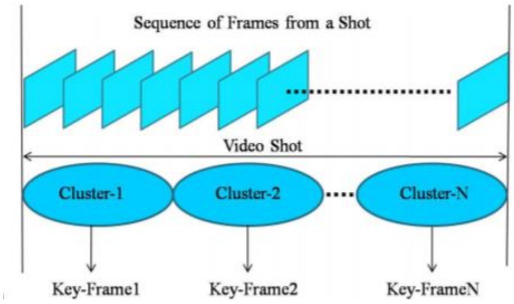

Figure 3: Cluster-based [14]

2.6. Shot-Based Method:

One method for identifying significant changes in a video's content involves shot boundary detection. In this process, keyframes are extracted, with one keyframe designated per shot. The determination of the number of keyframes used to encapsulate a shot aligns with the visual complexity present in the shot, and the positioning of these keyframes aims to represent the most noteworthy visual content.To achieve this, shots within the video are subdivided into sub-shots. For each sub-shot, the entropy is computed, and the extraction of keyframes within each shot is contingent on the maximum entropy value. However, it's important to note that this method has limitations, as it doesn't fully account for content complexity and may not be as suitable for accurately handling videos with large shots. [6].

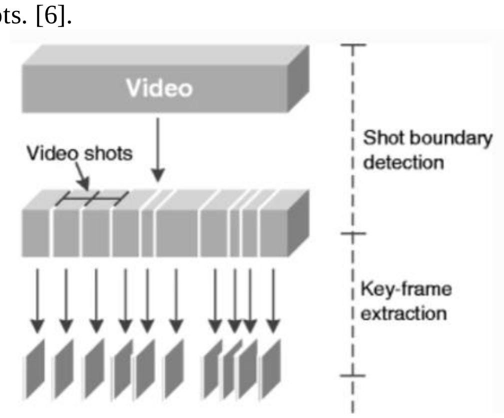

Figure 4: Shot based

2.7. Content- Analysis-Based Method:

In this approach, keyframes are chosen by evaluating color, texture, and other pertinent visual information in each frame. Frames exhibiting significant changes in this information are identified as keyframes. The process begins by selecting the initial frame as a reference, and subsequent frames are compared to this reference. If the distance between the kth frame and the reference frame surpasses a predefined threshold, the kth frame becomes the new reference. This method determines keyframes by assessing the extent of content change in each frame. However, it displays sensitivity to camera movement, leading to the selection of unstable and less efficient keyframes. [12]

2.8. Histogram-Based Method:

Histogram-based methods for key frame extraction rely on analyzing the distribution of pixel intensities in the video frames. This approach utilizing image histograms to capture key frames based on significant changes in brightness values. histogram-based key frame extraction is a straightforward and efficient technique, especially suited for scenarios where sudden changes in pixel intensity are indicative of key events. However, it may not be as effective in capturing nuanced semantic information or spatial relationships within the video content. Advantages of this methos is efficient in capturing sudden changes in content, but ot may not consider spatial relationships and semantic content.

2.9. Motion-Based Method:

This motion-centric approach initially partitions an input video clip into cohesive segments, categorized by major types of camera motion like pan, zoom, pause, and steady movement. Specific rules tailored to each segment are then applied to extract keyframes. Detection and analysis of movement in video shots are facilitated by examining the optical flow of the video sequence. In this method, a keyframe is identified as the local minimum in the movement. However, a limitation of this approach is its relatively low robustness, as it relies on local information without considering global factors for keyframe extraction. [15]. This method allow specific to detecting changes in camera motion, but has low robustness, dependency on local information as drawbacks.

2.10. Sparse Representation-Based Method:

This method, based on sparse representation, involves projecting video frames onto a low-dimensional feature space through a random projection matrix. The sparse representation is then leveraged within this random feature space to analyze the spatiotemporal information embedded in the video data, ultimately generating keyframes. [15]. This approach does not require shot(s) detection, segmentation, or semantic understanding and is computationally efficient. This methos has efficient representation of spatiotemporal information, But it require fine-tuning, and sensitivity to noise.

3. Comparative Analysis of Key Frame Extraction Methods:

Table 1 compares keyframe extraction methods based on their characteristics, merits, and demerits.

Method	Characteristics	Merits	Demerits
Uniform Sampling	Most common method	Straightforward method	Not content-based and Selected frames are often unstable
Pixel Compare	Pixel-wise comparison	Easy to evaluate	Time-consuming
Image Histogram	Similarity measure between keyframe	High-level segmentation	Don't consider the local similarities
Scale-invariant feature transform	Describe the local features in an image	Most prominent local feature	-
Cluster-based	Clustering similar frames/shots	Covers global characteristics of the scene	High computational cost (Takes 10 times the video length)
Shot based	keyframe in each shot is based on the maximum entropy value of each shot	-	Not appropriate for a big shot
Content-analysis based	Keyframes extraction based on the degree of change in the content of the Frame	Maintain good segmentation results	Insensitive to camera movement.
Motion-based	Adopts advantage of the digital capture device.	Reduce the spatiotemporal effects	High-quality video expected

4. Prior Art of Key frame extraction approaches:

This section provides a concise summary of earlier research conducted in the field of key frame extraction.

The first clustering-based keyframe extraction algorithm was published in 1998 by Zhuang et al [16]. The determination of keyframes is based on the number and size of clusters, each comprising visually similar frames with content involving color, texture, and shape. This approach is characterized by efficiency, rapid computation, and ease of application for online processing. Its effectiveness was tested on two films: a romantic comedy (Movie 1) and an action movie (Movie 2), with the latter exhibiting a higher number of keyframes compared to the former.

The key frame is selected from the collection based on its frequent dissimilarity from its consecutive neighbor. Utilizing the fuzzy C-means clustering algorithm, visually similar frames are grouped into clusters. Following the clustering process, frames exhibiting change ratios—indicating content variation—higher than the average value of the cluster are designated as keyframes.

The effectiveness of this technique was assessed using various video datasets, including football footage and sports videos sourced from YouTube. [17].

By fusing the key aspects of the video, Jiaxin Wu et almethod .'s [18] allows comparable frames to cluster together. In the initial phase, pre-sampling is employed to reduce redundancy in video frames and create a set of potential frames. These candidate frames are then characterized using the Bag of Words (BoW) model to capture their visual content. Subsequently, the Video Representation based High-Density Peaks Search (VRHDPS) clustering technique is applied to organize the candidate frame data into clusters. The central value of each cluster is then aggregated to form keyframes.

Keyframe extraction requires two phases, according to Besiris et al. [19]. The initial step involves constructing the MST (Minimum Spanning Tree) graph, linking each node to a distinct frame within the shot. Subsequently, keyframes are identified in the second stage through the application of the maximum speed approach. The adaptive threshold dynamically regulates the quantity of selected keyframes.

On the basis of spatial and temporal color distribution, Zhonghua et al [20].'s research focused on video keyframe extraction. First, a frame is built during the video shot that takes into account the spatial and temporal distribution of the pixels. The shot calculates the weighted separation between each Frame's color histogram. As keyframes, they choose the frames that are closest to the distance curve's peaks.

According to Spyrou et al. [21], Keyframes are selected from video clips by considering their semantic context. To extract color and texture features, keyframe regions are utilized. A hierarchical clustering method is employed to create a local region thesaurus for each frame. This thesaurus is then locally extracted from every photo, ensuring a comprehensive representation of visual elements.

Every video frame is endowed with a set of features, including semantic and frame-based characteristics. Semantic features gauge the presence of semantic concepts within a frame. Each frame in every segment of the video is associated with at least one semantic attribute. A score is then computed for each group of frames based on their semantic values. Ultimately, the representative frame is selected based on the corresponding score value. [22].

Keyframe extraction was created by Ling Shao et al. [23] based on intra-frame and inter-frame motion histogram analysis. Keyframes are derived from frames exhibiting intricate motion and greater significance compared to their adjacent frames, capturing a more comprehensive representation of the video's actions and activities. The initial step involves identifying peaks in the entropy curve, generated using motion histograms for each video frame. These peaked entropies are then weighted using inter-frame saliency, employing histogram intersection, resulting in the generation of final keyframes. This approach leverages the maxima of motion complexity in foreground objects and the variance in motion between successive frames to extract keyframes.

The keyframe extraction approach that was performed hierarchically to produce a keyframe with a tree-structured was discussed by Hyun Sung Chang et al. in their study [24]. There are a lot fewer frame comparisons as a result. It creates the video's multilevel abstract. By utilizing the depth-first search technique with pruning, it offers an effective content-based retrieval.

Keyframe extraction and object segmentation are concurrently built by a unified feature space, according to Xiaomu Song and his colleagues [25]. The process of selecting keyframes is framed as a feature selection within the context of the Gaussian Mixture Model (GMM) for object segmentation. Two divergence criteria are employed for keyframe extraction in this scenario. The first involves maximizing pairwise interclass divergence between GMM components. Following this, the focus shifts to maximizing marginal divergence, which evaluates how the mean density varies across frames. Through this method, representative keyframes are extracted for object segmentation. The integration of keyframes and objects enables the execution of content-based video analysis. This approach showcases an integrated content-based video analysis, providing a novel and adaptable functionalization of frames and objects.

The entropy difference approach was investigated by Markos Mentzelopoulos et al., [26] in an effort to segment spatial frames. The entropy that the dominating item possesses can be used to extract the keyframe. When the object can be distinguished from the backdrop, this work produces good results. Yet, when transient changes like flashes happen, performance suffers.

Keyframe attributes such as texture, edge, and motion were leveraged for content-based video indexing and retrieval. Keyframes were obtained through clustering techniques, specifically employing K-means. The effectiveness of this method was compared to the Volume Local Binary Pattern (VLBP).[27].

The Joint Kernel Sparse Representation technique was devised to alter essential attributes of human motion capture data, facilitating keyframe extraction. This method adeptly models the sparseness and Riemannian manifold structure inherent in human motion data, regardless of the manner in which motions are captured. Joint representation is employed to capture the internal structure of the motion capture data. Additionally, the imposition of the triangle restriction ensures the validity of locally extracting keyframes, especially for periodic motion sequences. Experimental results demonstrate the superiority of this approach compared to other state-of-the-art methods. [28].

In generating keyframes for consumer films, a process involves projecting video frames into a low-dimensional random feature space, followed by keyframe recovery through sparse representation. The utilization of sparse signal representation allows for the evaluation of both spatial and temporal information in the video, leading to the identification of keyframes. This technique eliminates the need for shot detection, segmentation, or semantic comprehension.[15].

The keyframe selection process employs a key point-based architecture, considering local features. The selection of keyframes is based on the discernible parameters of coverage and redundancy. This approach stands out as a promising technique for keyframe extraction.[29].

In order to extract keyframes, Badre et al. [30] described the Haar wavelet transform with different levels and thepade's sorted pentnary block truncation coding. The Alias Canberra distance, Sorensen distance, Wavehedge distance, Euclidean distance, and mean square error similarity measurements are used to measure variety among successive frames.

5. SUMMARY AND CONCLUSION:

The key frame extraction process serves as a crucial step in eliminating redundant frames from videos, forming a fundamental unit for structural video analysis. It provides an accurate representation of the entire shot, holding significance across various applications such as video summarization, content-based video indexing and retrieval, video searching, and video compression. This paper conducts a comprehensive analysis of methods employed for key frame extraction, exploring their advantages, disadvantages, and the challenges users face in the extraction process. Although standardized metrics for evaluating key frame extraction methods are lacking, the identified approaches should exhibit high effectiveness, reliability, and computational simplicity. Extracted key frames must be compact yet representative of the complete video sequence. The cluster-based approach emerges as an advanced strategy for key frame extraction, offering versatility depending on the intended use.

The key frame extraction process, as a fundamental unit in structural video analysis, plays a crucial role in removing unnecessary frames from videos and delivering an accurate representation of the complete shot to the user. Its significance spans across diverse applications, including video summarization, content-based video indexing and retrieval, video searching, and video compression.

This paper has thoroughly examined various methods for locating key frames, shedding light on their respective benefits and drawbacks, along with the challenges users face during the extraction process. Despite the absence of standard metrics for evaluating key frame extraction methods, it is imperative that these approaches exhibit high effectiveness, reliability, and computational simplicity. Extracted key frames should be compact while faithfully reflecting the entire video sequence.

In addition to the existing knowledge, it is worth noting that the success of key frame extraction methods also hinges on their adaptability to different video content and user preferences. A noteworthy advanced strategy explored in this context is the cluster-based approach, offering a versatile solution depending on the specific use case. As the field continues to evolve, future research could focus on refining these methods, potentially establishing standardized metrics and benchmarks to further enhance the evaluation and comparison of key frame extraction techniques.

REFERENCES

[1] M. Ma, S. Mei, S. Wan, J. Hou, Z. Wang, and D. D. Feng, "Video summarization via block sparse dictionary selection," *Neurocomputing*, vol. 378, pp. 197–209, 2020.

[2] G. M. E. M. Elahi and Y. H. Yang, "Online learnable keyframe extraction in videos and its application with semantic word vector in action recognition," *arXiv*, 2020.

[3] S. Mei, G. Guan, Z. Wang, S. Wan, M. He, and D. D. Feng, "Author ' s Accepted Manuscript reconstruction Video Summarization via Minimum Sparse Reconstruction," *Pattern Recognit.*, 2014.

[4] M. K. Asha Paul, J. Kavitha, and P. A. Jansi Rani, "Key-Frame Extraction Techniques: A Review," *Recent Patents Comput. Sci.*, vol. 11, no. 1, pp. 3–16, 2018.

[5] I. H. Ali and T. T. Al-Fatlawi, "Key Frame Extraction Methods," *Int. J. Pure Appl. Math.*, vol. 119, no. 10, pp. 485–490, 2018.

[6] U. Gawande, K. Hajari, and Y. Golhar, "Deep Learning Approach to Key Frame Detection in Human Action Videos," *Recent Trends Comput. Intell.*, pp. 1–16, 2020.

[7] S. Jadon and M. Jasim, "Unsupervised video summarization framework using keyframe extraction and video skimming," *2020 IEEE 5th Int. Conf. Comput. Commun. Autom. ICCCA 2020*, pp. 140–145, 2020.

[8] C. V. Sheena and N. K. Narayanan, "Key-frame Extraction by Analysis of Histograms of Video Frames Using Statistical Methods," *Procedia Comput. Sci.*, vol. 70, pp. 36–40, 2015.

[9] C. Gianluigi and S. Raimondo, "An innovative algorithm for key frame extraction in video summarization," *J. Real-Time Image Process.*, vol. 1, no. 1, pp. 69–88, 2006.

[10] D. ju Jeong, H. J. Yoo, and N. I. Cho, "A static video summarization method based on the sparse coding of features and representativeness of frames," *Eurasip J. Image Video Process.*, vol. 2017, no. 1, pp. 1–14, 2016.

[11] W. Barhoumi and E. Zagrouba, "On-the-fly Extraction of Key Frames for Efficient Video Summarization," *AASRI Procedia*, vol. 4, pp. 78–84, 2013.

[12] M. S. V. Kavita Sahu, "Key Frame Extraction From Video Sequence : A Survey," *Int. Res. J. Eng. Technol.*, vol. 04, no. 05, pp. 1346–1350, 2017.

[13] H. Jin, Y. Yu, Y. Li, and Z. Xiao, "Network video summarization based on key frame extraction via superpixel segmentation," *Trans. Emerg. Telecommun. Technol.*, no. February, pp. 1–11, 2020.

[14] N. J. Janwe, "Video Key-Frame Extraction using Unsupervised Clustering and Mutual Comparison," no. 10, pp. 73–84, 2016.

[15] M. Kumar and A. C. Loui, "Key frame extraction from consumer videos using sparse representation," *Proc. - Int. Conf. Image Process. ICIP*, no. 1, pp. 2437–2440, 2011.

[16] Y. Zhuangt, Y. Rui, T. S. Huang, and S. Mehrotra, "ADAPTIVE KEY FRAME EXTRACTION USING UNSUPERVISED CLUSTERING," no. 94, pp. 866–870, 1998.

[17] S. Angadi and V. Naik, "Entropy based fuzzy C means clustering and key frame extraction for sports video summarization," *Proc. - 2014 5th Int. Conf. Signal Image Process. ICSIP 2014*, pp. 271–279, 2014.

[18] J. Wu, S. Zhong, and J. Jiang, "A novel clustering method for static video summarization," *Multimed. Tools Appl.*, 2016.

[19] D. Besiris, "Key frame extraction in video sequences : a vantage points approach," pp. 0–3, 2007.

[20] H. Chen, "Video Key Frame Extraction Based on Spatial-temporal Color Distribution," pp. 196–199, 2008.

[21] E. Spyrou and Y. Avrithis, "Keyframe Extraction using Local Visual Semantics in the form of a Region Thesaurus," pp. 98–103, 2007.

[22] J. V. C. I. R, J. Lai, and Y. Yi, "Key frame extraction based on visual attention model," *J. Vis. Commun. Image Represent.*, vol. 23, no. 1, pp. 114–125, 2012.

[23] L. Shao and L. Ji, "Motion Histogram Analysis Based Key Frame Extraction for Human Action / Activity Representation," 2009.

[24] H. S. Chang, S. Sull, S. U. Lee, and S. Member, "for Content-Based Retrieval," vol. 9, no. 8, pp. 1269–1279, 1999.

[25] X. Song and G. Fan, "Joint Key-Frame Extraction and Object Segmentation for Content-Based Video Analysis," vol. 16, no. 7, pp. 904–914, 2006.

[26] M. Mentzelopoulos and A. Psarrou, "Key-Frame Extraction Algorithm using Entropy Difference," pp. 39–45, 2004.

[27] M. Ravinder and T. Venugopal, "Content-Based Video Indexing and Retrieval using Key frames Texture , Edge and Motion Features," vol. 6, no. 2, pp. 672–676, 2016.

[28] G. Xia, H. Sun, X. Niu, G. Zhang, and L. Feng, "Keyframe Extraction for Human Motion Capture Data Based on Joint Kernel Sparse," vol. 2, no. c, 2016.

[29] G. Guan, Z. Wang, S. Lu, J. Da Deng, and D. D. Feng, "Transactions Letters," vol. 23, no. 4, pp. 729–734, 2013.

[30] S. R. Badre and A. B. T. Coding, "Summarization with Key Frame Extraction using Thepade ' s Sorted n-ary Block Truncation Coding Applied on Haar Wavelet of Video Frame," pp. 332–336, 2016.