CONTENT BASED VIDEO RETRIEVAL USING AN EFFICIENT KEY-FRAME EXTRACTION TECHNIQUE

[1]Soumya Balan P, [2] Prof. Leya Elizabeth Sunny
[1]Student, [2]Assistent Professor
[1] Department of Computer Science and Engineering,
[1] Mar Athanasius College of Engineering, Kothamangalam, Kerala, India

Abstract : ***Video retrieval is a young field which has its genealogy rooted in artificial intelligence, digital signal processing, natural language understanding, databases, psychology, computer vision, and pattern recognition. The Video Retrieval system includes various steps: Video Segmentation, Key frames Selection, Feature Vector Creation and finally Video Retrieval. Length of these videos may be large. User may require only viewing the summary of video. So in such cases, video content summarization is used. A video comprise of number of frames. Key frame represent the main content of the video. For video content summarization key frame extraction is considered. Thus here used a method to extract key frames from video using Thepade's sorted n-ary block truncation coding. Using this key frames video is retrieved using an input video query.***

IndexTerms - key-frame, Feature Extraction, Local Binary Pattern, Random Sampling, Feature Vector.

I. INTRODUCTION

Content based Video Retrieval (CBVR), in the application of video retrieval is the issue of searching for digital videos in large databases with less input keywords. "Content-based" is the search which analyze the actual content of the video. CBVR system works more effectively as these deals with content of video rather than video metadata. The extensive appearance of video cameras together with the lot of social networking and video sharing website has resulted in the increase in the no of videos in the world which has never seen before. These video's are not in proper format and well structured. Length of these video's are not properly defined, they can be of any length. Video summarization helps to summarize such type of video. For video summarization, key frame extraction is done. Key frame is frame which is having major difference as compared to previous frame in the series of frames. Entire video can be converted into small no of frames which is having major content of video.

Video summarization can have application such as video skimming in which important audio and video are extracted which create skim video which represent the concise form of video. Like that video summarization is used in video indexing and filmmaking. Key frame extraction helps out to find the most important frames from video. While extracting key frame, if we found that the two consecutive frames are having major difference then that frames are termed as key frame. If we are working with whole video content then it will take large processing time but if key frame extraction is considered it reduces maximum time. Until now, large amount of work has been done in video processing such as video indexing [5], shot boundary detection, video classification and content based video retrieval. Block truncation coding was developed in 1979 and initially used for gray- scale images. Block truncation coding is widely used for video content summarization. Initially, block truncation coding was used for only gray scale images [1]. In block truncation coding, image is divided into no of blocks. Individual formulation is done on each block. Block truncation coding proves to be better in color feature extraction from video and threshold is considered here for formulation [10].

II. LITERATURE SURVEY

There are a number of techniques methods are there in the field of content based video retrieval systems. They are required to effectively search, index and retrieve videos from video databases. But the reliable and effective systems are still awaited for huge databases [1]. For this reason, text based searches are still in practice for the video retrieval systems [2]. A content based retrieval system was developed for commercial uses [3]. Face detection method was used for image and video searches in this systems. But this method also proved to be very poorly performing [4] by the automatic systems participated in video retrieval track [5]. Other useful information from videos can bring performance of video retrieval systems to a great level of success. Researchers still face a challenge to utilize some important information such as sequence of shots, temporal and motion information [2]. To compensate this problem and to get better retrieval performance, a video retrieval system [6] utilized all frames of a shot instead of only the key frames. Another system [7] integrated color and motion features for better utilization of spatiotemporal information. But a fact is still relevant that an efficient image retrieval technique results in an efficient video retrieval technique [4].

A detailed review of the existing techniques is done by Li et al. [8]. But this may not be proper choice, since the rest of the frames are not inspected to determine whether they are also represented by selected key- frame [9]. The plotted curve is then analyzed to estimate the sharp corners and the frame corresponding to the sharp corner are treated as key-frame of the shot [10].

Normally features are extracted on the entire frame, where as in A fuzzy based key-frame selection approach each frame is segmented using the multi-resolution recursive shortest spanning tree algorithm. For each segment a fuzzy color and motion histogram is extracted and stored as a representative. The frames which are not similar to each other are selected based on the cross-correlation criterion by the use of genetic algorithm [11]. Also to take a decision based on both spatial and wavelet features Dempster – Shafer theory of evidence is used [12]. Li Liu et al. developed a new method of key-frame extraction using correlated pyramidal motion-feature for human action recognition. To select key- frame for each action sequence he used Ada - Boost learning algorithm [13]. A method of key frame extraction based on unsupervised clustering was adopted by Zhuang et al. They used color histogram of each frames of video computed in HSV color space and a threshold to control clustering density. The key frame selection is employed only to the clusters which are big enough to consider as key

cluster. From each key cluster a representative is selected as a key-frame [14]. In Gong and Liu proposed a technique, calculated color histogram of video frames in RGB color space. To incorporate spatial information each frame is divided into 33 blocks and a 3D histogram is created for each block. The nine histogram are then concatenated together to form a feature vector. Then these feature vectors are used to form clusters and for each cluster the frame closest to the cluster center is selected as a key frame [15].

A video flow consists of images and audios [25]. Feature extraction is to convert the raw information in recorded videos into abstract information. Although caption [25] and audio [13] features are often selected in some news and film video summary systems. Unlike image pixels, image features are not affected by image transformation such as transition and hence are more reliable. Image feature extraction is an essential step in image/video content analysis [30]. The typical high-level features are shapes or objects [26]. Low-level features include global [32], local [11] and motion [33] features. For a global feature, the entire image is the subject of interest, whereas only a set of blobs in the image is interested for a local feature. Shots may be generated by camera cut, edited changeover [34] and camera work state change. Shot or scene boundary detection is usually achieved by measuring difference between several consecutive image frames [35]. According to number of compared frames, the commonly used boundary detection methods can be divided into 2-frame method [36] and 3-frame method [37].

Existing key frame extraction methods can be categorized into four classes: shot boundary method clustering method, and visual difference curve based method. Shot boundary method uses shot boundaries as key frames [38]. UTS has two drawbacks: (1) the key frames may not be most representative of the scene; (2) redundancy in the key frames is high especially in low progressing videos. Based on the cumulative difference from a reference frame, ATS extracts key frames by predefined frame difference interval [26]. Visual difference curve based method is to compare the feature difference between the current frame and the previous frame or a reference frame. Some researchers proposed using curvature of points on the cumulative VDC to locate the key frames [27]. Some of the studies also based on image feature extraction techniques. Examples include the detection and tacking of construction equipment [28] and content-based image query [29]. Some other studies directly use certain visual contents for the assessments of construction quality [30] or the measurement of work progress [31].

III. PROPOSED SYSTEM

3.1 Content Based Video Retrieval Using Key-frame Extraction

In todays electronic world huge amount of useful digital information like images, audio and video data apart from textual data exists online. It is available to government authorities, professionals and researchers easily and accessible at reasonably cheaper cost and it is due to the rapid growth in availability of user friendly and cheaper multimedia acquisition devices at a very large scale like high resolution camera in mobile phones, handy cams and other advanced digital devices large scale usage of internet by rapidly growing number of applications used by digital devices to upload huge amount of multimedia information and internet infrastructure. Video data possesses a lot of information for those using multimedia systems and applications like digital libraries, publications, broadcasting and entertainment. Such applications are useful only when video retrieval systems are efficient enough to retrieve videos and from large databases as quick as possible. Most of the web based video retrieval systems work by indexing and searching videos based on texts associated with them. Since video retrieval is not effective using conventional query-by-text retrieval technique. Content Based Video Retrieval is considered as one of the best practical solutions for better retrieval quality. Steps for retrieving the video is given below:

3.1.1 Segmentation of the Video

Segmentation of video is the first step in most of existing content based video analysis techniques. These shots contain a sequence of frames recorded one after another. These are organized and edited with cut transitions or gradual variation of visual effects forming a video scene or sequence during video sorting. Therefore the process of video segmentation is nothing but converting a video into various smaller video clips representing different scenes where each scene is decomposed again into different shots.

3.1. Comparison Between the Two Key-frame Extraction Methods

Here a comparison study between two key-frame extraction methods namely “Video content summarization using Thepade’s Sorted n-ary Block Truncation coding” and “Key-frame extraction by analysis of histograms of video frames using statistical methods” are performed.

3.1.1 Key-frame Extraction by the Analysis of Histograms of Video Frames Using Statistical Methods

One of the methods to summarize video data is the extraction of key- frame. This paper proposes a method of key-frame extraction using thresholding of absolute difference of histogram of consecutive frames of video. Key-frame extraction from video data is an active research problem in video retrieval.

Here computes thresholding point using mean and standard deviation of absolute difference of histogram, for comparative study of feature difference of consecutive video frames. Here explores the method of key-frame extraction algorithm based on absolute difference of histogram of consecutive image frames. It is a two phase method, in which first phase compute threshold using mean and standard deviation of histogram of absolute difference of consecutive image frames. Second phase extract key–frames comparing the threshold against absolute difference of consecutive frames. The algorithm starts with extracting video frames one by one. After pre-processing each video frames histogram difference between two consecutive frames were calculated. The mean an d standard deviation of absolute difference of histogram calculated to fix a threshold point. The threshold (T) is computed using following equation (1).

$$T = \mu adh + \sigma adh \qquad (1)$$

Where μadhmean is absolute difference and σadhis is the standard deviation of absolute difference. Once the threshold obtained next phase determine the key-frames by comparing the absolute difference of histogram against threshold. The proposed algorithm is given below;

Step.1 Extract frames
Step. 2 Calculate histogram difference between two consecutive frames
Step. 3 Find mean and standard deviation of absolute difference
Step. 4 Computing threshold
Step. 5 Compare the difference with T. If it is
>Tselects it as a key-frame else go to step 2
Step. 6 Continue until end of video

Figure 1: Key-frame extracted for a 13sec video

The above figure shows key-frames extracted for a 13 sec video. It yields 32 key-frames from 323 frames and the key-frames are extracted within 6 min.

3.1.2 Video Content Summarization using Thepade's Sorted n-ary Block Truncation coding

Here a novel method to extract key frames from video using Thepade's sorted n-ary block truncation coding is proposed. Video summarization helps to summarize such type of video data. Key frame is frame which is having major difference as compared to previous frame. If we are working with whole video content then, it will take large processing time but if key-frame extraction is considered it reduces the maximum time BTC is widely used for video content summarization. Initially, block truncation coding was used only for gray-scale images. In block truncation coding, image is divided into blocks and Individual formulation is done on each block. Here we get two values for each colour component. Such as two values for red, green and blue. If we are working with BTC less computational complexity can achieve. Features are extracted from image with the help of block truncation coding. These are properties of image. From that key frame can be extracted from video.

Thepade's Sorted Ternary Block Truncation Coding

Intensity values of frames are extracted from the frame in Thepade's sorted ternary block truncation coding with respect to that colour component. These values are arranged in one feature vector with respect to colour. Then these feature vectors are sorted in ascending order, divided into three parts and an average of each part is taken. The nine values that can be obtained from this are the feature vector of respective frame. Then these feature vectors are considered for key frame extraction purpose. Various similarity measures can be applied on theses feature vector which determine the key frames from video. In the proposed methodology, five variations of TSNBTC has tried. Feature vector that are getting from this on that various similarity measure has been applied on that to extract key frames from video data.

There are five variation that are introduced in this paper. These are such as:

- Thepade's sorted quaternary block truncation coding.(TSQBTC)
- Thepade's sorted pentnary block truncation coding.(TSPBTC)
- Thepade's sorted hexnary block truncation coding.(TSHBTC)
- Thepade's sorted septnary block truncation coding.(TSSBTC)
- Thepade's sorted octanary block truncation coding.(TSOBTC)

Total intensity of R component of mn can be presented in form of a single dimensional array (SDR), having elements with indices 1 to m∗n [1]. Red component four values can be formulated as given below;

$$lR = (4/m*n)* \sum_{i=1}^{(m*n)/4} sortedSDR(i) \quad (2)$$

$$muR = (4/m*n)* \sum_{i=(m*n)/4+1)}^{(m*n)/2} sortedSDR(i) \quad (3)$$

$$mlR = (4/m*n)* \sum_{i=(m*n)/2+1)}^{(m*n)/2} sortedSDR(i) \quad (4)$$

$$uR = (4/m*n)* \sum_{i=(m*n)/4+1)}^{(m*n*3)/4} sortedSDR(i) \quad (5)$$

$$lG = (4/m*n)* \sum_{i=1}^{(m*n)/4} sortedSDR(i) \quad (6)$$

$$muG = (4/m*n)* \sum_{i=(m*n)/42+1)}^{(m*n)/2} sortedSDR(i) \quad (7)$$

$$mlG = (4/m*n)* \sum_{i=(m*n)/2+1)}^{(m*n*3)/4} sortedSDR(i) \quad (8)$$

$$uG = (4/m*n)* \sum_{i=(m*n*3)/4+1)}^{(m*n)/4} sortedSDR(i) \quad (9)$$

$$lB = (4/m*n)* \sum_{i=1}^{(m*n)/4} sortedSDR(i) \quad (10)$$

$$muB = (4/m*n)* \sum_{i=(m*n)/(4+1)}^{(m*n)/4} sortedSDR(i) \quad (11)$$

$$mlB = (4/m*n)* \sum_{i=(m*n)/2+1)}^{(m*n)/4} sortedSDR(i) \quad (12)$$

$$uB = (4/m*n)* \sum_{i=(m*n*3)/4+1)}^{(m*n)/4} sortedSDR(i) \quad (13)$$

After solving this equations we are getting the values that, we are getting a feature vector of frame. Therefore feature vector will be be [lR, muR, mlR, uR, lG, muG, mlG, uG, lB, muB, mlB and uB]. Same equations can be made for TSPBTC, TSHBTC, TSSBTC, TSOBTC. Like that for TSHBTC, TSSBTC and TSOBTC six, seven, eight values can be generated for each color component. Here, similarity measure is used for key-frame extraction. Here, the similarity measures used here are: Euclidean distance, L1 or Manhattan distance, Canberra distance, Chi-square (Chisq) or 2 distance, Cosine distance, and D1 distance. With the help of six similarity measures mean is calculated of all frames. After that standard deviation and threshold is calculated. Threshold can be calculated by the addition of mean and standard deviation.

$$Mean(M) = (\sum_{n=1}^{N} diff(i))/N - 1 \quad (14)$$

$$Standard\ Deviation(S) = \frac{\sqrt{\Sigma(diff(i)-M)^2}}{N-1} \quad (15)$$

$$Threshold = M+a*S \quad (16)$$

"a" is constant

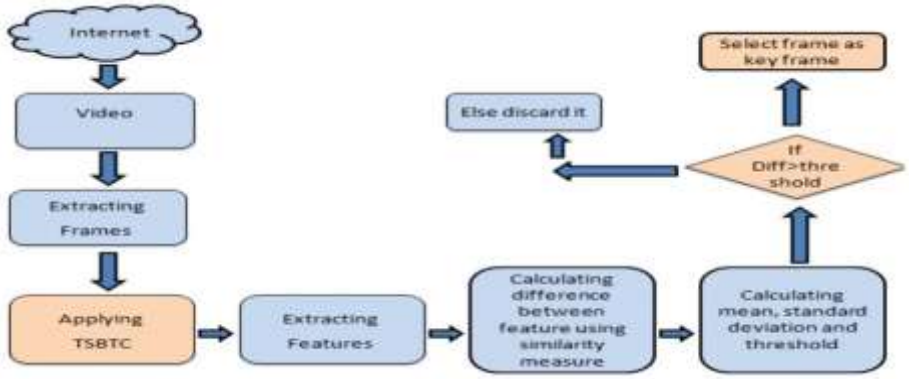

Figure 2: Block diagram of proposed algorithm

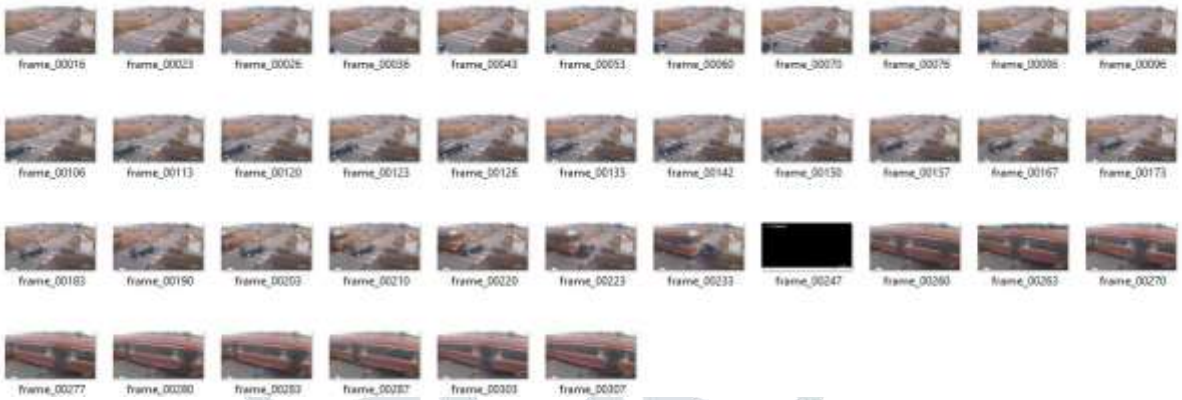

Figure 3: Key-frame extracted for a 13sec video

The above figure shows key-frames extracted for a 13 sec video. It yields 39 key-frames from 323 frames and the key-frames are extracted within 1 min.

Performance Evaluation Between Two Key-frame Extraction Methods

Performance is greater when using video content summarization using Thepade's Sorted n-ary Block Truncation coding than key-frame extraction by analysis of histograms of video frames using statistical methods. For example, a Key-frames from a 1min video can be extracted within 6min when using Thepade's method whereas it takes more than half hour to extract key-frames when using absolute histogram methods. The graph given below shows key-frame extraction from a 13sec video by using the above two methods.

3.1.3 Query by Selected Key-frame Using Random Sampling

From the Key-frames that is extracted by using Thepade's Sorted n-ary Block Truncation coding a key-frame is selected by random sampling. Here the similarity measures used are Euclidean, Chi-square and L distance. The most similar video is then retrieved by checking the feature vector of query and the feature vectors of the dataset videos.

IV. PERFORMANCE EVALUATION

The key frame extraction method based on the similarity between continuous frames. The given graphs compare the performance of the two key-frame extraction methods called Video Content Summarization using Thepade's Sorted n-ary Block Truncation coding and Key-frame Extraction by Analysis of Histograms Using Statistical Methods.

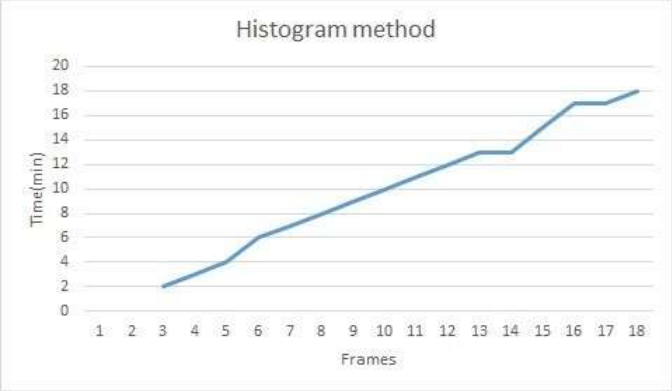

Figure 4: Graph for absolute histogram method

The above graph shows "Key-frame extraction by analysis of histograms of video frames using statistical methods. It extracts 39 frames from a 13 sec video in 1 min.

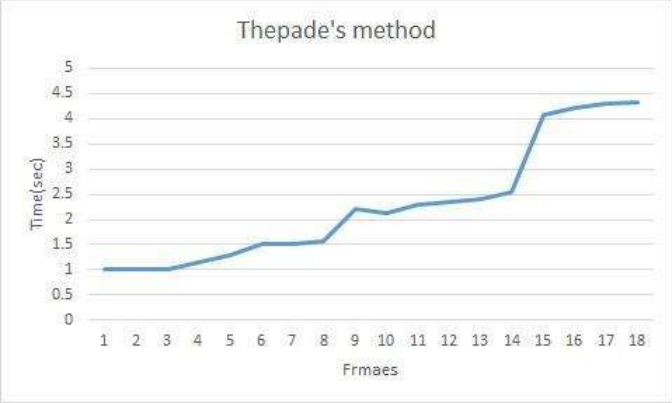

Figure 5: Graph for Thepade's method

The above graph shows "Video content summarization using Thepade's Sorted n-ary block truncation coding". It extract 32 frames from a 13 sec video in 6 min.

Video summarization aimed at reducing the amount of data that must be examined in order to retrieve the information desired from information in a video, is an essential task in video retrieval. So Thepade's method performs better than the absolute histogram method. And we choose that method for content based video retrieval. For performance comparison, average accuracy is used.

Percentage accuracy = Actual correct extracted frames/Total expected extraction of frames

Graph shows that the percentage accuracy of various TSBTC with sorensen similarity measure. In this Thepade's sorted pentnary block truncation coding is giving highest performance for sorensen similarity measure.

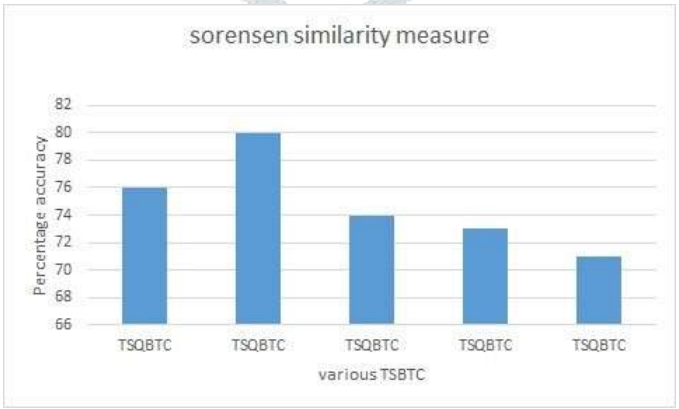

Figure 6: Comparison of various TSBTC over sorencen similarity measure

V. CONCLUSION

Content based video retrieval techniques are widely distributed among two types: one of them is comparison of frames and their corresponding features within two clips and a set of frames is obtained which are sequentially matching which helps in the retrieval of videos. But the computational cost depends upon the features size and is very high. These techniques have a drawbacks such as, synchronization between frames as different clips may have used different rate to encode them. To overcome the drawback, a key frame is used to represent an entire shot. With the quick growth of multimedia technology, a huge amount of videos are available across the world. Here video content summarization is used. Key-frame represent the main content of video data. For video content summarization the key frame extraction is mainly considered. Thus the method to extract key frames from video using Thepade's sorted n-ary block truncation coding gives the better

result. Here also used two multichannel decoded local binary patterns are introduced namely multichannel adder local binary pattern (maLBP) and multichannel decoder local binary pattern (mdLBP) for feature vector creation and is used for retrieving the video. Thus the proposed method has been envisioned for the purpose of video retrieval from the multimedia database by using an efficient algorithm. Which increase the performance of the system which is difficult in traditional video retrieving system.

VI. ACKNOWLEDGMENT

I owe my sense of gratitude and sincere thanks to Prof. Leya Elizabeth Sunny for her guidance, constant supervision, encouragement and support throughout the period of this thesis work.

REFERENCES

[1] Nicu Sebe, Michael S. Lew, Arnold W.M. Smeulders, "Video retrieval and summa- rization", Computer Vision and Image Understanding, vol. 92, no. 2-3, pg 141-146, 2003.
[2] Ja-Hwung Su, Yu-Ting Huang, Hsin-Ho Yeh, Vincent S. Tseng , "Expert Systems with Applications", 37, pg 5068-5085, 2010.
[3] Corporate Web Site, FaceIt Developer Kit Software, http://www.visionics.com, 2002.
[4] Alexander G. Hauptmann, Rong Jin, and Tobun D. Ng, "Video Retrieval using Speech and Image Information", Electronic Imaging Conference (EI'03), Storage Retrieval forMultimedia Databases, Santa Clara, CA, January 20-24, 2003.
[5] The TREC Video Retrieval Track Home Page, http://wwwnlpir.nist.gov/projects/trecvid/.
[6] Liang-Hua Chen, Kuo-Hao Chin, Hong-Yuan Liao, "An integrated approach to video retrieval", Proceedings of the nineteenth conference on Australasian database- Volume 75, 49–55, 2008.
[7] Mohd. Aasif Ansari, Hemlata Vasishtha," CBVR and Classification of Video Database–Latest Trends, Methods, Effective Techniques, Problems and Chal- lenges", International Journal of Computer Applications (ISSN : 0975 – 8887),Vol- ume 125 – No. 6, September 2015.
[8] Y. Li et al. Techniques for movie content analysis and skimming: tutorial and overview on video abstraction techniques, IEEE signal processing magazine 23(2)2006 p. 27-50.
[9] Suresh C Raikwar, Charul Bhatnagar and Anand Singh Jalal, "A frame work for key-frame extraction from surveillance Video", 5th International Conference on Computer and Communication Technology", IEEE, 2014, p. 297-300.
[10] Zhao et al. "Key-frame extraction and shot retrieval using nearest feature line",Proceedings of ACM Workshop on Multimedia, 2000,p. 217- 220.
[11] Doulamis et al. "A fuzzy video content representation for video summarization and content based retrieval", Journal of signal processing, 2000,p.1049-1060.
[12] Mukhargee et al., "Key-frame estimation in video using randomness measure of feature point pattern", IEEE transactions on circuits on systems for video technol- ogy, vol.7,no.5,May 2007, p. 612-620.
[13] Li Liu, Ling Shao, Peter Rockett, "Boosted key-frame and correlated pyramidal motion feature representation for human action recognition", Pattern Recognition 45 (2013), p. 1810-1818.
[14] Zhuang Y, Rui Y, Huang T.S and Mehvotra S, " Adaptive key-frame extraction using unsupervised clustering", Proceedings of International conference on Image Processing, 1998, p 866-870.
[15] Gang Y and Liu, "Video summarization using singular value decomposition", Pro- ceedings of Computer Vision and Pattern Recognition, 2000, p 347-358.
[16] I. Ide, H. Mo, N. Katayama, S.i. Satoh, Exploiting topic thread structures in a news video archive for the semi-automatic generation of video summaries, 2006 IEEE International Conference on Multimedia and Expo, Proceedings, IEEE, 2006, pp. 1473–1476.
[17] Z. Rasheed, M. Shah, Scene detection in Hollywood movies and TV shows, 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings, vol. 2, IEEE, 2003, pp. 343–348.
[18] L. Ott, P. Lambert, B. Ionescu, Coquin, Animation movie abstraction: Key frame adaptative selection based on color histogram filtering, 14th International Confer- ence on Image Analysis and Processing Workshops, Proceedings, 2007, pp. 206–211.
[19] C.-M. Tsai, L.-W. Kang, C.-W. Lin, W. Lin, Scene-based movie summariza- tion via role-community networks, IEEE Transactions on Circuits and Systems for Video Technology 23 (11) (2013) 1927–1940, http://dx.doi.org/10.1109/tcsvt.2013. 2269186.
[20] W. Chang, N. Yang, C. Kuo, C.H. Lin, Template-based scene classification for baseball videos using efficient playfield segmentation, Third International Confer- ence on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007), vol. 2, IEEE, 2007, pp. 543–548.
[21] C.-M. Kuo, W.-H. Chang, M.-Y. Fang, C.-H. Lin, A template-based baseball video scene classification using efficient playfield segmentation, Multimedia Tools and Applications 55 (3) (2011) 399–422, http://dx.doi.org/10.1007/s11042-010- 0555-6.
[22] W. Hua, M. Han, Y. Gong, Baseball scene classification using multimedia features, 2002 IEEE International Conference on Multimedia and Expo, Proceedings, vol. 2, IEEE, 2002, pp. 821–824.
[23] Q. Luo, T.M. Khoshgoftaar, E. An, Hierarchical indexing of ocean survey video by mean shift clustering and MDL principle, IEEE International Conference on Information Reuse and Integration, Proceedings, 2005, pp. 404–409.
[24] K. Schoeffmann, M. Del Fabro, T. Szkaliczki, L. B¨osz¨ormenyi, J. Keckstein, Keyframe extraction in endoscopic video, Multimedia Tools and Applications 74 (24) (2015) 11187–11206, http://dx.doi.org/10.1007/s11042-014-2224-7.
[25] B. Munzer, K. Schoeffmann, L. Boszormenyi, Relevance segmentation of laparo- scopic videos, 2013 IEEE International Symposium on Multimedia (ISM), IEEE, 2013, pp. 84–91.
[26] H. Mo, M. Yamagishi, I. Ide, N. Katayama, S.i. Satoh, M. Sakauchi, Key image extraction from a news video archive for visualizing its semantic structure, 5th Pacific Rim Conference on Multimedia (PCM 2004), vol. 3331, Springer, Berlin, Heidelberg, 2004, pp. 650–657.
[27] I.P. Tussyadiah, D.R. Fesenmaier, Mediating tourist experiences: access to places via shared videos, Ann. Tour. Res. 36 (1) (2009) 24–40.
[28] Y. Yuan, M.Q.H. Meng, Hierarchical key frames extraction for WCE Video, 2013 IEEE International Conference on Mechatronics and Automation, 2013, pp. 225–229.
[29] R. Brunelli, O. Mich, C.M. Modena, A survey on the automatic indexing of video data, J. Vis. Commun. Image Represent. 10 (2) (1999) 78–112, http://dx.doi.org/ 10.1006/jvci.1997.0404.
[30] W. Hu, N. Xie, L. Li, X. Zeng, S. Maybank, A survey on visual content-based video indexing and retrieval, IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 41 (6) (2011) 797–819, http://dx.doi.org/10.1109/TSMCC.2011.2109710.

[31] H. Zhang, A. Kankanhalli, S.W. Smoliar, Automatic partitioning of full-motion video, Multimedia Systems 1 (1) (1993) 10–28.
[32] I.K. Sethi, N.V. Patel, A statistical approach to scene change detection, IS T/SPIE's Symposium on Electronic Imaging: Science Technology, International Society for Optics and Photonics, 1995, pp. 329–338.
[33] D. Zhang, Y. Rao, J. Zhao, J. Zhao, A. Hu, B. Cai, Feature based segmentation and clustering on forest fire video, IEEE International Conference on Robotics and Biomimetics, 2007. ROBIO 2007, IEEE, 2007, pp. 1788–1792.
[34] S.H. Han, K.J. Yoon, I.S. Kweon, A New Technique for Shot Detection and Key Frames Selection in Histogram Space, Changes 2 (1) (2000) 1.
[35] C. Gianluigi, S. Raimondo, An innovative algorithm for key frame extraction in video summarization, J. Real-Time Image Proc. 1 (1) (2006) 69–88, http://dx.doi. org/10.1007/s11554-006-0001-1.
[36] J. Gong, C.H. Caldas, An object recognition, tracking, and contextual reasoning- based video interpretation method for rapid productivity analysis of construction operations, Autom. Constr. 20 (8) (2011) 1211–1226, http://dx.doi.org/10.1016/j. autcon.2011.05.005.
[37] I. Brilakis, L. Soibelman, Content-based search engines for construction image databases, Autom. Constr. 14 (4) (2005) 537–550.
[38] K.-L. Lin, J.-L. Fang, Applications of computer vision on tile alignment inspection, Autom. Constr. 35 (2013) 562–567, http://dx.doi.org/10.1016/j.autcon.2013.01. 009.
[39] X. Zhang, N. Bakis, T.C. Lukins, Y.M. Ibrahim, S. Wu, M. Kagioglou, G. Aouad, A.P. Kaka, E. Trucco, Automating progress measurement of construction projects, Autom. Constr. 18 (3) (2009) 294–301, http://dx.doi.org/10.1016/j.autcon.2008.