Video Summarization Using a Key Frame Selection Based on Shot Segmentation

Wisnu Widiarto
Informatics Department
Sebelas Maret University
Surakarta, Indonesia
wisnu.widiarto@staff.uns.ac.id

Eko Mulyanto Yuniarno
Electrical Engineering Department
Sepuluh Nopember Institute of Technology
Surabaya, Indonesia
ekomulyanto@ee.its.ac.id

Mochamad Hariadi
Electrical Engineering Department
Sepuluh Nopember Institute of Technology
Surabaya, Indonesia
mochar@te.its.ac.id

***Abstract*— Video length can reach a few hours so that the user requires the system to summarize and obtain video abstract that has a search process effectively and efficiently. Summary process aims to allow users to quickly understand the content of a video sequence so that a summary of the video should not contain too many key frames. Selection of key frames is done by selecting the priority of the collection of frames that have been formed from a video. Selection of key frames that have been chosen from a collection of frames will be collected and combined in accordance with the order. The combined key frames will be used by the user to know the summary and video abstraction and to make the process of browsing and retrieval of video.**

Keywords*— *key frame; selection; video summary; browsing; retrieval

I. INTRODUCTION

Advances in technology make people familiar with audio-visual recordings. Technology makes everyone knows the media with the name of the video. They are very familiar with the use of a tape recorder to produce a video. Video content can be accessed by using a top-down (to get the essence of the content / video browsing) or bottom-up (to get something when it is known exactly what to look for in a video content / video retrieval) [1].

There are three main things in video browsing and video retrieval [2]: understanding the structure of video content, video summaries, and video matching. A video consists of several frames, making it difficult to choose and set key frames to display video content, except by understanding the structure of the video. The structure of the video is the video description in the form of images, which provide an abstract of the content of the video. Key frame is defined as a part of a video sequence representing the video content.

Video summarization is difficult, because usually the summary will reduce the amount of original content. The difficulty is the decision about what part should be included and what is excluded. Another difficulty is the selection of a video section that will attract audiences and preserve the original story of the video. There are four main parts to create a video summary: segmentation (shot segmentation), key frames (key frame extraction), the similarities (classification) and the generation of summaries [3].

Video matching technology is a new technology that is used to compare two videos, whether these two videos have similarity value. Then the video matching is considered as matching the video features of the multi-dimensional space. One benefit of this technology is that it can be used to find some parts of the video. This can be done by matching the part of the original video.

The simplest technique for summarizing videos is to use a compression technique, which displays the contents of the video by playing full high-speed video, but the video content can still be understood. Another popular technique is to summarize the video using skimming techniques [4], using a browsing tool to obtain a summary of the video which is the incorporation of audio and video information. Other research related to video summaries done by [5], to produce a video summary that is applied to the video on-line, and considers each frame as a separate image. Other research conducted by [6], the video segmentation automation plays an important role in the real-time systems, presents the algorithm of video segmentation using three things: change detection, technical background and techniques of adaptive real-time threshold.

Of the several applications for video processing, summarizing videos is one of the important things in the era of multimedia technology advances, including archiving and providing access to a collection of videos. Summary video is usually divided into two types [7], [8]: 1. Video static (summarization) is a process to select a group of interesting frame of the video division, which will be used to represent the video content called a key frame. 2. Dynamic video (skimming) is the original video that is displayed in the form of a video clip by making a long video into a short video using the technique of high-speed video plays.

The smallest segment of a video can be defined as a sequence of continuous frames from the shot, which was recorded from the same camera. Each video shot can be represented by one or several frames called key frames. The number of key frames can not be determined in advance, due to differences content variations for each shot. Camera shot

978-1-4799-8386-5/15/$31.00 ©2015 IEEE

detection is the first stage for video segmentation. The method is often used for shot detection based on color histogram [9].

II. Related Work

The digital technology is growing, including the technology in the field of audio and video. This development affects the process of managing video, the process of classification and process of video retrieval. Development of video classification and retrieval process can be done easily using the summary video, make a long video into a video shorter without reducing the significance and originality of the story in the video.

A. Video Summarization

The concept of automation and video analysis is important in some applications, including video management application called video summaries. Management and video summaries will provide convenience to the user in terms of search and retrieval of video content.

Extraction on the object motion video is taking in the scene that can be achieved by performing segmentation based on motion (motion segmentation). In [10], the authors propose a method to choose the activities that stand out using the points of motion features as a key parameter (a Gaussian mixture model).

Reference [11] proposed a method to automatically detect events in the game of football using three stages: to detect shot boundaries and to classify shot using a hidden Markov model, the visual feature extraction, and to obtain semantic features (events and concepts) using Bayesian network. This method has been applied and has been able to detect seven different events in a video soccer.

This paper [12] proposed a summary of the video using the LE (LE = Laplacian Eigenmap). A small group of selected reference frame of the video sequence is used as a reference to measure the difference between frames. Paper [13] presents a Binary Genetic Algorithm (BGA) based video summarization, to evaluate segmentation (the system uses the similarity function) and to get meaningful summaries (the system uses the binary crossover and mutation operators).

B. Key Frame Selection

Paper [14] presents an algorithm of the main selection frame based on three iso-content (Distance, Errors and Distortion), so some selected key frames is the same distance in the video content in accordance with the principles used.

This research [15] developed a video summary of the system used for classification and indexing of video, to present video content by extracting key frame. Clustering algorithm is a method used in this research to be applied to all frames in the video.

In the paper [16] presented a framework for the extraction of key frames using a model that focuses on the human visual attention to represent scenes that are considered most important or the most meaningful. In [17], the authors propose another method for managing key frame which uses a panorama technology to get different sizes of key frames.

In [18] proposed a keypoint-based framework to address the problem of selecting a key frame so that local features can be used to select key frames. In general, the selected key frames should represent the video content as well and contains a minimum redundancy.

III. Proposed Method

Fig. 1, presented a general overview of the proposed method. The method is divided into three stages. First, the video is divided into frames, then every frame count histogram value (using a color histogram). Second, detect the similarities between two consecutive frames by calculating the distance using the Euclidean distance. If the distance between the two frames is less than the threshold, then the two frames is considered to be in the same segment, otherwise the two frames are considered different segments. In the same segment would have been two key frames (first and last frame of each segment are selected as key frames). Third, all key frames are collected to generate video summaries.

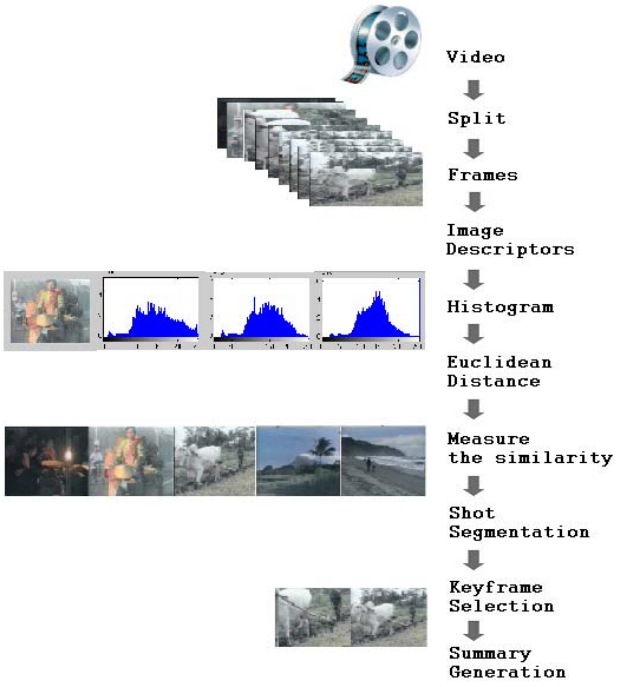

Fig. 1. A general overview of the proposed method

A. Pre Processing

A video is used to refer to a set of frame sequences, can be denoted as :

$$V = (I_{[x,y,1]}, I_{[x,y,2]},..., I_{[x,y,n]}).$$

$I_{[x,y,n]}$ represents $I_{[x,y]}$ frame with pixel (x, y) and a sequence number to n. The original video V will be divided into several frames in accordance with the length of the video, starting from the initial frame (f1) until the end of the frame (fn). The number of frames in the video V = N.

Feature selection is used to reduce the dimensionality of data (collection of frames f1 ... fn), by selecting a few frames were measured to create the model. The selection was based on criteria of a certain size, and based on the minimization of the prediction error. Feature selection is the process for selecting a relevant subset to build a model, which is also known as a selection of variables or attributes. Feature selection technique used for two reasons: 1. Data contains many redundant features, but these features do not provide more information than the currently selected features. 2. The feature is not relevant because it does not provide useful information. Feature selection is also useful as part of a process for data analysis.

B. Color Histogram

In the paper [7], there are two forms of video abstraction: a collection of key frames and video skim. In this research will be used first approach, a collection of key frames. Set of key frames in fig.2 is built from the selection frame, using a histogram, because a histogram is a good tool for video processing [19]. Histogram value is calculated for each frame in the original video, then the value histogram of each frame is compared. If two frames are compared have similarity, it is considered that the frames are redundant but did not provide more information than the currently selected frame, so the frame is removed. The histogram is used to compare two frame i and frame j. That is the color histogram Euclidean distance metric with the equation:

$$S = \sqrt{\sum_{i=1}^{I}(x_i - y_i)^2}$$

x_i, y_i are the histogram of the first and the second frame

C. Key Frame Selection

Split the input video into frames and shot segmentation. Take a first frame of each segment F(a) and compute the histogram from Fa then assign it Ha. Take a next frame from the same segment Fb and compute the histogram from Fb then assign it Hb. Comparing histograms Ha and Hb to measure similarity. If (Fa and Fb) have similarity then remove frame Fb and take a next frame as Fc. If (Fa and Fb) have dissimilar then take the frame Fa as first_keyframe (i) and the frame Fb as last_keyframe (i). The key frame selection algorithm is shown in Fig. 3.

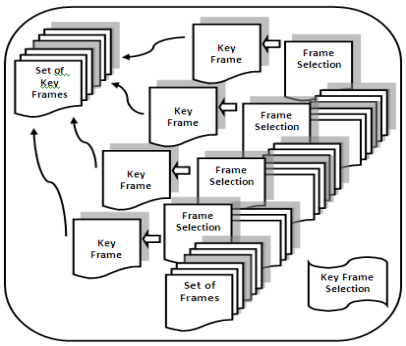

Fig. 2. Key Frame Selection

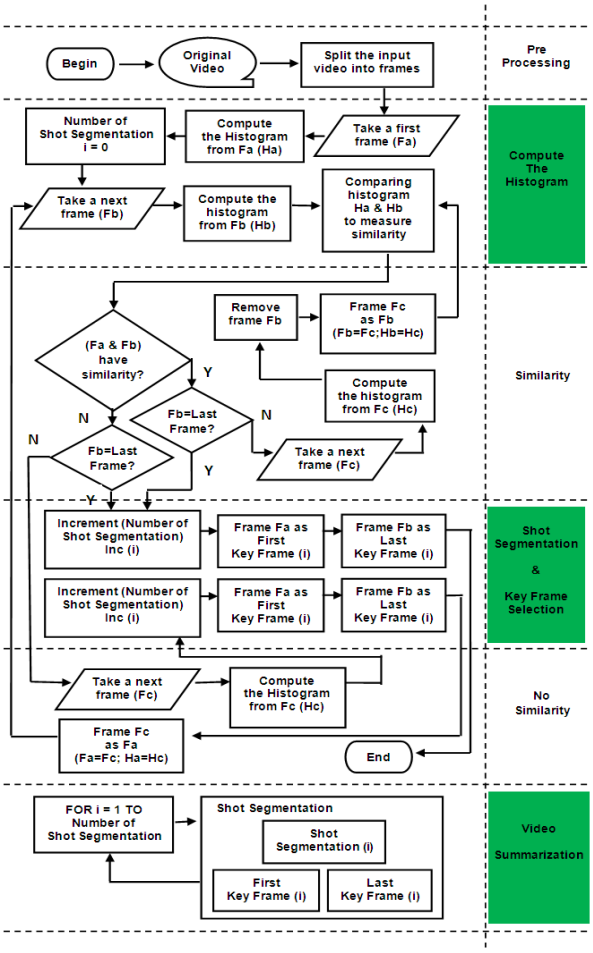

Fig. 3. The Key Frame Selection Algorithm

The process of formation of key frames is done as follows (Fig. 3):

1) Pre processing.

- Video is divided into several frames.

2) Compute the histogram of first frame.

- Each frame is given an index, all pixels in the frame is read, then the RGB color information obtained from each pixel. After all the color information is obtained, then the appearance of each color of RGB is summed to the color scale of 256 colors for each frame. After the appearance of the color scale is calculated, created histogram table. Table histogram is made for the three primary colors (Red / Green / Blue) with a corresponding array. Red color histogram calculated his own, as well as to the color histogram Green and Blue. Table histogram is also made to a gray color with a corresponding array.

- Frame #0091 was taken the color information, and then calculated the appearance of RGB basic colors, so the color histogram values obtained in accordance with the relevant array.

- Red color: R0091[1-256] = 1720, 430, 82, 45, 51, 48, 30, ..., 2671, 5610, 241898. Green color: G0091[1-256] = 0, 0, 2, 42, 126, 138, 197, ..., 290, 994, 249921. Blue color: B0091[1-256] = 34, 36, 19, 46, 67, 74, 85, ..., 1875, 731, 249578. Gray color: Y0091[1-256] = 0, 0, 0, 68, 101, 257, 376, ..., 1121, 4127, 245398

3) Compute the histogram of second frame.

- Frame #0092 are also taken color information, calculated appearance of RGB basic colors, so the color histogram values obtained in accordance with the relevant array.

- Red color: R0092[1-256] = 1151, 489, 134, 75, 61, 100, 106, ..., 2284, 6340, 240964. Green color: G0092[1-256] = 0, 0, 0, 15, 56, 142, 125, ..., 331, 981, 249902. Blue color: B0092[1-256] = 165, 92, 30, 45, 32, 46, 79, ..., 1212, 352, 249498. Gray color: Y0092[1-256] = 0, 0, 11, 70, 177, 223, 243, ..., 1133, 4124, 245445

4) Process the similarity.

- The histogram for each color of two consecutive frames are compared, according arrays are concerned, using the formula S = sqrt (sum of sqr ($X_i - Y_i$)) . If the result obtained is less than the threshold, then the result is called Similar. Apply dissimilar conditions to the contrary.

- Red Color: Sum of SQR9291 = SQR9291[1] + SQR9291[2] + ... + SQR9291[256] = 869822820; The square root of SQR9291= 29492.75877

- Green Color: Sum of SQG9291 = SQG9291[1] + SQG9291[2] + ... + SQG9291[256] = 1002638770; The square root of SQG9291= 31664.47173

- Blue Color: Sum of SQB9291 = SQB9291[1] + SQB9291[2] + ... + SQB9291[256] = 1137342556; The square root of SQB9291= 33724.50972

- Average of RGB color = The square root of SQR9291 + The square root of SQG9291 + The square root of SQB9291 = (29492.75877 + 31664.47173 + 33724.50972)/3 = 94881.74022 / 3 = 31627.24674

- Average of RGB color (31627.24674) < Threshold (60000) => similar; Frame #0091 and frame #0092 => similar

5) Shot segmentation.

- If the first frame and the second frame in a state similar, the two frames that are located on the same segment. If two frames in dissimilar conditions, it is considered that the two frames are located on different segments.

6) Key frame selection.

- One segment consists of several frames, the first frame and the last frame is taken to be used as key frames.

IV. EXPERIMENT

This research used three video. Video_01.avi, size 320x240 pixels, memory 5951.5 kb, length 34.00 seconds. Video_02.avi, size 624x256 pixels, memory 5484.0 kb, length 94.89 seconds. Video_03.avi, size 176x144 pixels, memory 94393.7 kb, length 121.13 seconds.

Color histogram represents the distribution of colors in a frame. Each video is divided into several frames, and then calculated the histogram to find the shot segmentation. Each shot consists of two or more frames. The first frame of each shot is used as baseline and reference distance for the next frame. If the distance exceeds a certain threshold, then the selected frame and serve as a new baseline. The last frame of one shot (frame 0092) and the first frame of the next shot (frame 0093) has a color histogram values are much different (dissimilar) are shown in Fig. 4. Another comparison between two frames in one shot has a histogram value in the threshold (similar). Shot boundary detection results can be improved, if the threshold is calculated for the entire video.

Each video is divided into several frames. Video1 is divided into 408 frames, video2 1139 is divided into frames, and Video3 is divided into 1453 frames. All the frames of each video are calculated color histogram and gray histogram to be compared and the distance is calculated using the Euclidean distance. The histogram of each frame is compared using key frames selection algorithm, as shown in Figure 3. If the difference in distance is at the threshold then called Similar and if the difference exceeds a threshold distance then called dissimilar. If there is dissimilar, then the shot will be replaced on the next shot. If a change of shot, then the first frame used in the final frame of the shot and the next frame is called the first frame of the next shot. And so on until the end of the frame on the last shot. Every shot is taken two frames as key frames, the first frame is used as the first key frame and the last frame is used as a second key frame, as shown in Figure 5.

Comparison within two frames is calculated using the Euclidean Distance by using the histogram color (Red/Green/ Blue) and gray histogram. An example is given in Figure 4 for the frame number (#0088: #0097). A comparison result is obtained:

Comparison of the distance frame #0091 and #0092 resulted in: red = 29492.75877; green = 31664.47173; blue = 33724.50972; average = 31627.24674; gray = 31023.87281. The first distance (average RGB) is 31627.24674 (below the limit) and the second distance (gray) is 31023.87281 (below the limit), so that the two frames are considered Similar conditioned.

Comparison of the distance frame #0092 and #0093 resulted in: red = 112855.44565; green = 108536.80897; blue = 94548.29386; average = 105313.51616; gray = 119837.16926. The first distance (average RGB) is 105313.51616 (over the limit) and the second distance (gray) is 119837.16926 (over the limit), so that the two frames are considered conditional dissimilar.

Comparison of the distance frame #0093 and #0094 resulted in: red = 15781.94380; green = 18186.41977; blue =

15873.61188; average = 16613.99181; gray = 19453.72057. The first distance (average RGB) is 16613.99181 (below the limit) and the second distance (gray) is 19453.72057 (below the limit), so that the two frames are considered Similar conditioned.

COLOR HISTOGRAM (Red-Green-Blue) - Euclidean Distance (VIDEO_03)

Frame1	vs	Frame2	Red(R)	Green(G)	Blue(B)
0088	vs	0089	44079.70547	45459.26907	73827.80725
0089	vs	0090	50814.38269	51491.60124	57867.87347
0090	vs	0091	25027.60656	24515.20320	26255.45924
0091	vs	0092	29492.75877	31664.47173	33724.50972
0092	vs	0093	112855.44565	108536.80897	94548.29386
0093	vs	0094	15781.94380	18186.41977	15873.61188
0094	vs	0095	21651.37649	24888.80270	28331.89203
0095	vs	0096	21104.18575	19758.71438	18891.38465
0096	vs	0097	23892.73731	26285.40397	24961.44860

Frame1	vs	Frame2	Average(RGB)	Gray	Similarity
0088	vs	0089	54455.59393	49719.70957	similar
0089	vs	0090	53391.28580	54883.04503	similar
0090	vs	0091	25266.08967	24078.54763	similar
0091	vs	0092	31627.24674	31023.87281	similar
0092	vs	0093	105313.51616	119837.16926	dissimilar
0093	vs	0094	16613.99181	19453.72057	similar
0094	vs	0095	24957.35708	25783.43096	similar
0095	vs	0096	19918.09493	20711.17404	similar
0096	vs	0097	25046.52996	25182.10682	similar

Fig. 4. Similarity Process (Video_03 Frame 0088:0097)

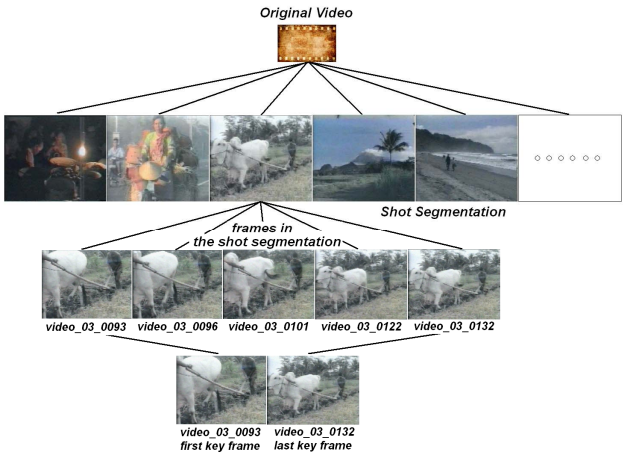

Fig. 5. The Structure Video of Key Frame Selection

From these results indicate that the frame #0091 and #0092 in conditions similar, the two frames that are in the same segment. Frame #0092 and #0093 in dissimilar conditions, so that the frame #0092 and #0093 is located on a different segment. The same thing for the frame #0093 and #0094 are in a condition similar, the meaning two frames are in the same segment. The same calculation is done for the other frames. After all frames calculated histogram and the distance between the frames has been calculated, then determined segmentation shot. Each segment is taken two frames as key frames. The original video has been divided into a number of frames will be grouped to form the appropriate segmentation of similarity.

After the video is divided into frames, calculate the histogram of each frame, calculate the distance between the frames using the Euclidean distance, and determining the similarity.

In this research, used 3 video with the following experiments:

Video #1

The original data:

Video_01.avi, size 320x240 pixels, memory 5951.5 kb, length 34.00 seconds.

The results:

Number of frame: 408, the number of shot: 5. Shot 1: the number of frames: 149 (0001: 0149), shot 2: the number of frames: 76 (0150: 0225), shot 3: the number of frames: 62 (0226: 0287), shot 4: number of frames: 68 (0288: 0355) shot 5: the number of frames: 53 (0356: 0408), and the detailed results are shown in Table 1a.

Video #2

The original data:

Video_02.avi, size 624x256 pixels, memory 5484.0 kb, length 94.89 seconds.

The results:

Number of frame: 1139, the number of shot: 16. Shot 1: the number of frames: 93 (0001: 0093), shot 2: the number of frames: 53 (0094: 0146), shot 3: the number of frames: 24 (0147: 0170), shot 4: number of frames: 114 (0171: 0284), shot 5: the number of frames: 25 (0285: 0309) and the detailed results are shown in Table 1b.

Video #3

The original data:

Video_03.avi, size 176x144 pixels, memory 94393.7 kb, length 121.13 seconds.

The results:

Number of frame: 1453, the number of shot: 41. Shot 1: the number of frames: 47 (0001: 0047), shot 2: the number of frames: 45 (0048: 0092), shot 3: the number of frames: 40 (0093: 0132), shot 4: number of frames: 30 (0133: 0162), shot 5: the number of frames: 35 (0163: 0197) and the detailed results are shown in Table 1c.

TABLE I. VIDEO IS DEVIDED INTO SHOT SEGMENTATION WHICH CONSIST OF A NUMBER OF FRAMES

a. video_01

Shot Number	Number of Frames	Frames Number
1	149	1-149
2	76	150-225
3	62	226-287
4	68	288-355
5	53	356-408
	408	

b. video_02

Shot Number	Number of Frames	Frames Number
1	93	1-93
2	53	94-146
3	24	147-170
4	114	171-284
5	25	285-309
6	92	310-391
7	49	392-440
8	248	441-688
9	74	689-762
10	133	763-895
11	25	896-920
12	18	921-938
13	32	939-970
14	75	971-1045
15	58	1046-1103
16	36	1104-1139
	1139	

c. video_03

Shot Number	Number of Frames	Frames Number	Shot Number	Number of Frames	Frames Number
1	47	1-47	21	35	779-813
2	45	48-92	22	35	814-848
3	40	93-132	23	27	849-875
4	30	133-162	24	34	876-909
5	35	163-197	25	31	910-940
6	63	198-260	26	25	941-965
7	48	261-308	27	26	966-991
8	28	309-336	28	31	992-1022
9	63	337-399	29	26	1023-1048
10	39	400-438	30	28	1049-1076
11	34	439-472	31	22	1077-1098
12	24	473-496	32	19	1099-1117
13	27	497-523	33	28	1118-1145
14	23	524-546	34	27	1146-1172
15	34	547-580	35	86	1173-1258
16	35	581-615	36	34	1259-1292
17	26	616-641	37	27	1293-1319
18	82	642-723	38	23	1320-1342
19	16	724-739	39	38	1343-1380
20	39	740-778	40	28	1381-1408
			41	45	1409-1453
	778			675	

V. CONCLUSION

Key frame selection is a tool to implement video summaries by identifying a set of key frames to represent the video sequence. Key frame selection is usually modeled as a grouping process to split one video into several groups. A video will be converted into a set of frames. Then a set of frames are compared from the beginning to the end of the frame. If there is a similar frame it will take a double frame (first frame and last frame) as a key frame (key frame selection). Key frame selection is used as a summary of the video that will allow users to understand the contents of the video sequences. The summary video aims to create facilities browsing, searching, retrieval and management of digital multimedia content. The resulting summary can support users in exploring large video files and supports a more efficient decision making in selecting, consume, share, or delete content.

REFERENCES

[1] Y. Rui, T.S. Huang, and S. Mehrotra, "Exploring video structures beyond the shots," in Proceedings of IEEE Conference on Multimedia Computing and Systems, 1998.

[2] M. Yeung and B.L.Yeo, "Video visualization for compact presentation and fast browsing of pictorial content," IEEE Transactions on Circuits System Video Technology, vol. 7, no. 5, pp. 771–785, May 1997.

[3] J. Ren and J. Jiang, "Hierarchical Modeling and Adaptive Clustering for Real-Time Summarization of Rush Videos," IEEE Transactions on Multimedia, Vol. 11, No. 5, pp. 906 – 917, August 2009.

[4] M.G. Christel, M.A. Smith, M.A., C.R. Taylor, and D.B. Winkler, "Evolving video skims into useful multimedia abstractions," In Proceedings of ACM Conference on Human Factors in Computing Systems, April 1998.

[5] D.P. Papadopoulos, S.A. Chatzichristofis, and N. Papamarkos, "Video summarization using a Self-Growing and Self-Organized Neural Gas Network," Department of Electrical and Computer Engineering, Democritus University of Thrace Xanthi 67100, Greece, 2011

[6] Y. Beevi and S. Natarajan, "An efficient video segmentation algorithm with Real time Adaptive Threshold Technique," International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 2, No.4, December 2009

[7] B.T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification," ACM Transactions on Multimedia Computing, Communications and Applications (TOMCCAP), Vol.3, No.1, Article3, pp.1-37, Feb 2007

[8] I. Yahiaoui, B. Merialdo and B. Huet, "Automatic Video Summarization," Multimedia Communications Department, Institut EURECOM, BP 193, 06904 Sophia-Antipolis, FRANCE, 2004

[9] H. J. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video," Multimedia Syst., vol. 1, no. 1, pp. 10–28, 1993.

[10] Sujatha C, A.R. Chivate, S.A. Ganihar, and U. Mudenagudi, "Time Driven Video Summarization using GMM," IEEE Fourth Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, ISBN: 978-1-4799-1586-6, pp. 1-4, Des 2013

[11] M. Tavassolipour, M. Karimian, and S. Kasaei, "Event Detection and Summarization in Soccer Videos Using Bayesian Network and Copula," IEEE Transactions on circuits and systems for video technology, Vol. 24, No. 2, February 2014

[12] R.M. Jiang, A.H. Sadka, and D. Crookes, "Hierarchical Video Summarization in Reference Subspace," IEEE Transactions on Consumer Electronics, ISSN:0098-3063, Vol.55, No.3, pp. 1551-1557, August 2009

[13] Y. Xue and W. Zhicheng, "Video Segmentation and Summarization Based on Genetic Algorithm," IEEE 2011 4th International Congress on Image and Signal Processing, pp. 460-464, 2011

[14] C. Panagiotakis, A. Doulamis, and G. Tziritas, "Equivalent KeyFrames Selection Based on Iso-Content Principles," IEEE Transactions on circuits and systems for video technology, Vol.19, No.3, pp.447-451, March 2009

[15] W. Sabbar, A. Chergui, and A. Bekkhoucha, "Video Summarization Using Shot Segmentation And Local Motion Estimation," IEEE Second International Conference on Innovative Computing Technology (INTECH), pp. 190-193, ISBN: 978-1-4673-2678-0, Sept 2012

[16] A.V. Potnurwar and M. Atique, "Visual Attention Key Frame Extraction for Video Annotations," International Journal of Computer Science Engineering (IJCSE), Vol. 3 No. 01, Jan 2014, ISSN: 2319-7323, pp. 39-42, 2014

[17] P. Tanapichet, N. Cooharojananone, and R. Lipikor, "Automatic Comic Strip Generation Using Extracted Keyframes from Caroton Animation," 2011 Interational Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), December 7-9,2011

[18] G. Guan, Z. Wang, S. Lu, J. D. Deng, and D. D. Feng, "Keypoint-Based Key frame Selection," IEEE Transactions on circuits and systemsfor video technology, Vol.23, No.4, pp. 729-734, April 2013

[19] J.S. Boreczky and L.A. Rowe, "Comparison of Video Shot Boundary Detection Techniques," Journal of Electronic Imaging Vol.5 No.2 pp. 122-128, April 1996