The Social Sciences 12 (2): 221-226, 2017
ISSN: 1818-5800

Key Frame Selection of Video CCTV Segmentation Based on Statistical Model

[1,2]Wisnu Widiarto, [2]Mochamad Hariadi and [2]Eko Mulyanto Yuniarno
[1]Department of Informatics, Sebelas Maret University, Surakarta, Indonesia
[2]Department of Electrical Engineering, Sepuluh Nopember
Institut of Technology, Surabaya, Indonesia

Abstract: Basically, Video CCTV is a collection of frames that are executed in sequence. The frame contains information of color values that will generate a color histogram values for determining the distance of two frames. The distance of two frames used to determine the position of the frame in the segment. This research is done to find the similarity between frames. Research activities are divided into five levels: frame generation, similarity calculation, shot segmentation, key frame selection and the final generation. Segmentation method is done by using a statistical model (Histogram difference and sum of absolute difference). Similarities between the two frames are calculated based on the difference within two frames (Euclidean distance). The similarities of the two frames will cause excessive frame. Similarities will also bring the same information with the selected frame (key frame) so it is recommended to be removed.

Key words: Retrieval, key frame, similarity calculation, segmentation, removed

INTRODUCTION

Video retrieval is important in the organization of multimedia data, including CCTV video management. Video retrieval included in the video summarization is divided into two types: static video summary and dynamic video skimming (Truong andVenkatesh, 2007). Analysis of the video mining can be carried out through four kinds: frame, shot, scene and video sequences (Zhao *et al*., 2007; Chergui *et al*., 2012).

CCTV video contains multiple image frames that have a specific pixel size. Each pixel has a different color value information. The video document has two principal layers: shot and scene (Zhu and Ming, 2008). Shot formed by multiple frame sequences. Shot formed based on a camera recording from the beginning to the end (Kang, 2001). Scene is a part of the video that presented the relationship between the content of which has some similarities shot (Chergui *et al*., 2012).

CCTV video is divided into several frames in accordance with the amount of recording time. Similar frames will be grouped in one group. Each group selected a single frame (or multiple frames) to be a key frame. The process to divide the video into a collection of frames a very meaningful can be done through a process called video segmentation. Usually, segmentation method performed by using the structure of the video sequences or statistical models. In this research used statistical methods using the Histogram Difference (HD) and Sum of Absolute Difference (SAD). HD and SAD are used to determine the similarity between the initial frame to the next frame based on the difference in distance (Euclidean distance).

Literature review: Some research has been done to determine the key frames selected from a collection of frames. In the first paper describes the use of key frames to perform video summary (Widiarto *et al*., 2015ab). In the second paper developed a classification system and video indexing using key frame extraction (Sabbar *et al*., 2012). In another paper utilizes key frames to build a comic strip (Widiarto *et al*., 2015a, b). Several methods have been performed to establish the key frames: based on the human visual attention (the most important or the most meaningful) (Potnurwar and Atique, 2014), based on visual attention clues (saliency driven, task independent and volition controled) (Peng and Q. Xiaolin, 2010), based on shot segmentation (segmentation, keyframe, the similarities and generation) (Widiarto *et al*., 2015ab), based on three iso-content (distance, errors and distortion) (Panagiotakis *et al*., 2009), based on the panorama technology (Tanapichet *et al*., 2011), based on two metrics (coverage and redundancy) (Ventura *et al*., 2013).

There are three approaches to determine key frame (Angadi and Naik, 2014) sampling-based approach (key frame is selected based on the sampling of the video frame), shot-based technique (key frames selected from

Corresponding Author: Wisnu Widiarto, Department of Informatics, Sebelas Maret University, Surakarta, Indonesia

each shot segmentation) and object-based technique (key frame selected by the object specified). Shot-based technique is done by making a shot classification. Classification and segments formed based on similarity between frames. Similarity is determined by measuring the color histogram for each frame. Several techniques have been developed for the segmentation process: based on graph theory, clustering algorithm, the process of merging (Dal *et al*., 2012). Development of classification and segmentation scene has done: create segmentation using video and audio features (Sundaram and Chang, 2000) Chang, segmentation scene using the Normalized graph cuts (Ncuts) (Zhao *et al*., 2007), segmentation scene using Dominant sets, classification and segmentation of the scene based on support vector machine (Zhu and Ming, 2008; Song *et al*., 2010), using a Markov chain Monte Carlo technique (Zhai and Shah, 2006), using similarity measures (Burget *et al*., 2013).

MATERIALS AND METHODS

This research is done in five levels frame generation, similarity calculation, shot segmentation, key frame selection, the final generation. Five levels are done with the steps shown by the block diagram in Fig. 1. Frame generation used to generate frame images from a CCTV video. Similarity calculation will determine the position of the frame on a particular segment. From the similarity calculation process will be obtained shot segmentation to determine the key frame. Set of key frames that will be built as a means for video retrieval.

The original video is divided into multiple frames. Each frame has information RGB color values (red, green and blue) which will be read starting from the beginning to the end. RGB color values of each pixel of each frame will be calculated based on the number of color values to generate a color histogram value. The color histogram value is used as a guideline to determine the distance of the frame.

The distance between the frames is calculated using euclidean distance to determine the similarity of each frame. The similarity calculation process will determine the position of a segment of a frame. Segmentation is a fundamental step in accessing, retrieving and browsing which was formed based on the similarity. Similar frames that will be located on the same segment. Each segment taken some frame is called a key frame. Key frames are formed in each segment are collected and reassembled in sequence so that generate new video that represents the contents of the original video.

Definition 1: Sum of absolut difference:

$$SAD(i, j) = \sum_{(i,\ j)\in W} \left|F_1(i, j)\right| - F_2(x+i, y+j)$$

F1 is the first frame and F2 is the second frame. SAD is the sum of absolut difference.

Definition 2: The Euclidean distance of frames (the first frame and the next frame) is Sim (Fb, Fa):

$$Sim(Fb, Fa) = \sqrt{\sum_{ci}^{CI=3} \left(Hist(Fb)_{ci} - Hist(Fa)_{cl}\right)^2}$$

Hist (Fa) is the histogram value of frame Fa and Hist(Fa) is the histogram value of frame Fa. Sim (Fb, Fa) is the similarity calculation of frame Fa and frame Fb.

```
1   Fa←First_frame
2   Fb← Next_frame
3   Ha←Hist_value_of_Fa
4   Hb← Hist_value_of_Fb
5   Comparing Hb(Fb)← Ha(Fa)
6   BEGIN
7      IF Fb,Fa similar
8      BEGIN
9         IF Fb← lastframe
10           Begin
11              Increment              (number_of_shot_seg)
12              FirstKeyFrame(i)← Fa
13              LastKeyFrame(i) ← Fb
14           end
15        ELSE
16           Begin
17              Fc← Next_Frame
18              Hc← Hist_value_of_Fc
19              remove(Fb)
20              Fc← Fb
21           end
22     END
23  ELSE
24  BEGIN
25     IF Fb← lastframe
26        Begin
27           Increment           (number_of_shot_seg)
28           First Key Frame(i)← Fa
29           LastKeyFrame(i) ← Fb
30        end
31     ELSE
32        Begin
33           Fc← Next_ Frame
34           Hc← Hist_value_of_Fc
35           Increment           (number_of_shot_seg)
36           First Key Frame(i)← Fa
37           LastKeyFrame(i) ← Fb
38           Fc← Fa
39        end
40     END
41  END
```

The process of browsing and retrieval is done with the main focus of the video segmentation process based on key frames. Video segmentation is formed of video frames in sequence, each frame is analyzed. Frame analysis was performed using the color histogram.

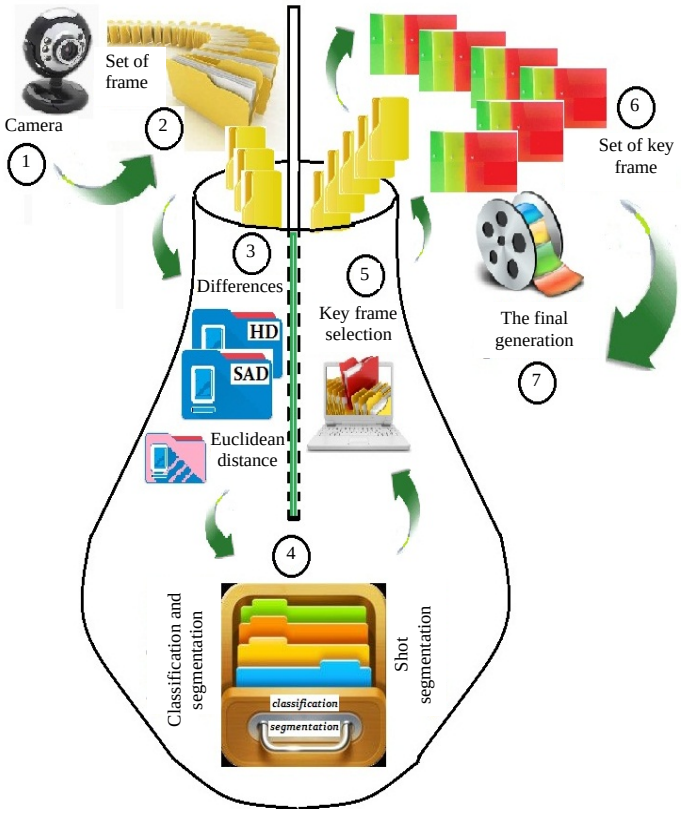

Fig. 1: Block diagram of the proposed method

Calculation of the color histogram is done to red, green and blue. Calculation of the histogram is also done for the histogram of gray color. Histogram value is calculated for each frame in the original video, then the value histogram of each frame in comparison. The color histogram of a frame compared to the next frame using the calculation of distances between frames (Euclidean distance). If the distance between the two frames is smaller than the specified threshold, then the two frames are Similar. If the distance of two frames exceeds a threshold, then the two frames are called dissimilar. If two frames are similar, the frames are said to be redundant and does not provide more information than the currently selected frame, so one of the frame is removed.

RESULTS AND DISCUSSION

This research calculates similarity using the Euclidean distance based on the Histogram Difference (HD) and the Sum of Absolute Difference (SAD). The similarity is calculated to determine the classification and segmentation shot. Histogram of the first frame (frame#0179); red color [1-256] = 120, 302, 285..., 944358; green color [1-256] = 234, 436, 1221 ..., 934334; blue color [1-256] = 2433, 2349, 4108 ..., 938755; gray color [1-256] = 10, 15, 21 ..., 947645. Histogram of the second frame (frame#0180), red color [1-256] = 123, 284, 430 ..., 944397, green color [1-256] = 145, 521, 434 ..., 934432, blue color [1-256] = 1323, 1421, 2131 ..., 937979, gray color [1-256] = 27, 14, 40 ..., 947942.

Euclidean distance of two frame (frame#0179 and frame #0180): red color [1-256] = 9, 324, 21025 …, 1521; green color [1-256] = 7921, 7225, 619369 …, 9604; blue color [1-256] = 1232100, 861184, 3908529 …, 602176; gray color [1-256] = 289, 1, 361 …, 88209. Sum of red color [1-256] = 2185754365, average of red color = 46752.05199. Sum of green color [1-256] = 2695769687, average of green color = 51920.80206. Sum of blue color [1-256] = 1912629656, average of blue color = 43733.62157. Sum of gray color [1-256] = 2893243562, average of gray color = 53788.87954.

The calculation of the distance (Euclidean distance) based on the HD's frame#0179 and #0180, data showed: red color = 46752.05199, green color = 51920.80206, blue color = 43733.62157, gray color = 53788.87954. The results of distance calculation is obtained that the value is below the threshold (60000) as shown in Fig. 2 so it is said that the frame #0179 and #0180 are similar.

Color Histogram of Frame1

Value	[1]	[2]	[3]	.....	[256]
Red	120	302	285	.....	944358
Green	234	436	1221	.....	934334
Blue	2433	2349	4108	.....	938755
Gray	10	15	21	.....	947645

Color Histogram of Frame2

Value	[1]	[2]	[3]	.....	[256]
Red	123	284	430	.....	944397
Green	145	521	434	.....	934432
Blue	1323	1421	2131	.....	937979
Gray	27	14	40	.....	947942

Euclidean distance (Frame2 - Frame1)

Value	[1]	[2]	[3]	.....	[256]	Sum	Sqrt
Red	9	324	21025	.....	1521	2185754365	46752,05199
Green	7921	7225	619369	.....	9604	2695769687	51920,80206
Blue	1232100	861184	3908529	.....	602176	1912629656	43733,62157
Gray	289	1	361	.....	88209	2893243562	53788,87954

Fig. 2: Histogram value of frame #0179 and #0180, similar

Sum of Absolut Difference (1667x2292)

	Frame1			Frame2			Absolut Difference			Sum (Komulatif)		
Cell	R	G	B	R	G	B				Red	Green	Blue
(1,1)	255	255	255	255	255	255	0	0	0	0	0	0
(1,2)	255	255	255	255	255	255	0	0	0	0	0	0
(1,3)	255	255	255	255	255	255	0	0	0	0	0	0
.....												
(2,1)	255	255	255	255	255	255	0	0	0	0	0	0
(2,2)	255	255	255	255	255	255	0	0	0	0	0	0
.....												
(677,221)	59	92	135	50	88	127	9	4	8	21978234	18979866	21240502
(677,222)	58	92	137	50	89	130	8	3	7	21978242	18979869	21240509
(677,223)	56	92	142	50	88	137	6	4	5	21978248	18979873	21240514
.....												
(1667,2290)	255	255	255	255	255	255	0	0	0	51724312	48972314	52139131
(1667,2291)	255	255	255	255	255	255	0	0	0	51724312	48972314	52139131
(1667,2292)	255	255	255	255	255	255	0	0	0	51724312	48972314	52139131
Sum of Absolut Difference (SAD)										51724312	48972314	52139131
Cell (1667x2292) = 3820764										3820764	3820764	3820764
Average of SAD (SAD / Cell)										13,53769	12,81741	13,64626

Fig. 3: Sum of absolut difference value of frame#0179 and frame#0180, similar

The distance calculation using the SAD applied to all frames. The calculation is performed on all pixels of the frame. Each frame is compared by calculating the absolute difference in color value for each cell. For example in the frame #0179 and #0180, the value of the red color in the cell (677×222) to frame #0179 is 58 and the value of the red color for the frame #0180 is 50, so the absolute value of the difference is |58-50| = 8. The previous cumulative value (at 677×221) is 21978234, so the cumulative value of the red color for the cells (677×222) is 21978234+8 = 21978242.

Implementation of the green color of the cell (677×223), the value of the frame #0179 cells (677×223) is 92, the value of the frame #0180 on the same cell is 88, so that the absolute value of the difference is | 92-88 | = 4. The cumulative value of green in the cell (677×223) is 18979869+4 = 18979873.

All values of the absolute differences are calculated, so that the SAD values ??obtained in the red, green, blue is 51724312; 48972314; 52139131. The size of 1667×2292 pixels (3820764 cells), the average value of SAD for red, green, blue is 13.53769; 12.81741; 13.64626.

The results of distance calculation by SAD (frame #0179 and #0180) obtained values as shown in Fig. 3. The distance by SAD (frame #0180 and #0181) is shown in Fig. 4. The distance between the frame #0179 and #0180

Sum of Absolut Difference (1667x2292)

Cell	Frame1 R	Frame1 G	Frame1 B	Frame2 R	Frame2 G	Frame2 B	Absolut Difference			Sum (Komulatif) Red	Green	Blue
(1,1)	255	255	255	255	255	255	0	0	0	0	0	0
(1,2)	255	255	255	255	255	255	0	0	0	0	0	0
(1,3)	255	255	255	255	255	255	0	0	0	0	0	0
.....												
(2,1)	255	255	255	255	255	255	0	0	0	0	0	0
(2,2)	255	255	255	255	255	255	0	0	0	0	0	0
.....												
(677,221)	50	88	127	24	20	19	26	68	108	59899682	79998874	121657389
(677,221)	50	89	130	24	20	19	26	69	111	59899708	79998943	121657500
(677,221)	50	88	137	25	21	20	25	67	117	59899733	79999010	121657617
.....												
(1667,2290)	255	255	255	255	255	255	0	0	0	229742812	241529637	237412028
(1667,2291)	255	255	255	255	255	255	0	0	0	229742812	241529637	237412028
(1667,2292)	255	255	255	255	255	255	0	0	0	229742812	241529637	237412028
Sum of Absolut Difference (SAD)										229742812	241529637	237412028
Cell (1667x2292) = 3820764										3820764	3820764	3820764
Average of SAD (SAD / Cell)										60,13007	63,21501	62,13732

Fig. 4: The first frame (frame #0180) and the second frame (frame #0181) are dissimilar

Table 1: Experimental results of shot segmentation using hd and SAD video 1)

HD		SAD	
Frames number	No. of frames	Frames number	No. of frames
1-36	36	1-36	36
37-180	144	37-180	144
181-288	108	181-288	108
289-576	288	289-576	288
577-780	204	577-780	204
781-876	96	781-876	96
877-1044	168	877-1044	168
1045-1164	120	1045-1164	120
1165-1212	48	1165-1212	48
1213-1248	36	1213-1248	36
1249-1692	444	1249-1692	444
1693-1884	192	1693-1884	192
1885-1980	96	1885-1980	96
1981-2064	84	1981-2064	84

Table 2: Experimental results of frames number (video 2)

HD		SAD	
Frames number	No. of frames	Frames number	No. of frames
1-72	72	1-72	72
73-204	132	73-204	132
205-300	96	205-300	96
301-564	264	301-564	264
565-792	228	565-792	228
793-912	120	793-912	120
913-984	72	913-984	72
985-1236	252	985-1236	252
1237-1344	108	1237-1344	108
1345-1608	264	1345-1608	264
1609-1704	96	1609-1704	96
1705-1956	252	1705-1956	252

did not exceed the threshold so the frame #0179 and #0180 are similar. The distance between the frame #0180 and #0181 exceeds the threshold, so the frame #0180 and #0181 are dissimilar. Overall, the process of determining similarity is applied to the two videos (video 1 and 2), obtained the number of segments and the number of frames shown in Table 1 (video 1) and Table 2 (video 2).

CONCLUSION

Video analysis is done on a frame that is part of the video. Each frame is analyzed on the information value of the color (red, green and blue) which raises the value of the color histogram. The value of the color histogram is processed to produce the distance between the frames that are used to determine the similarity of each frame. Similarities frame determine the position of the frame on a particular segment, the frame similar grouped in one segment. Each segment is taken a few frames to be used as key frames. Histogram difference and sum of absolute difference is used in the process and the Euclidean distance is used for the measurement of similarity. The similarities of the two frames causes: excessive frame and the duplicate information. Frame which has some similarities recommended for removal. Set of key frames are arranged in sequence so that generate a new video that is ready for use in the video retrieval process.

ACKNOWLEDGEMENT

This research was supported by Informatics Department and LPPM Sebelas Maret University (UNS) Surakarta; Electrical Engineering Department Sepuluh Nopember Institut of Technology (ITS) Surabaya.

REFERENCES

Angadi, S. and V. Naik, 2014. Entropy based fuzzy c means clustering and key frame extraction for sports video summarization. Proceedings of the 2014 5th International Conference on Signal and Image Processing (ICSIP), January 8-10, 2014, IEEE, Bagalkot, India, ISBN:978-1-4799-1394-7, pp: 271-279.

Burget, R., J.K. Rai, V. Uher, J. Masek and M.K. Dutta, 2013. Supervised video scene segmentation using similarity measures. Proceddings of the 2013 36th International Conference on Telecommunications and Signal Processing (TSP), July 2-4, 2013, IEEE, Brno, Czech Republic, ISBN:978-1-4799-0402-0, pp: 793-797.

Chergui, A., A. Bekkhoucha and W. Sabbar, 2012. Video scene segmentation using the shot transition detection by local characterization of the points of interest. Proceedings of the 2012 6th International Conference on Sciences of Electronics Technologies of Information and Telecommunications (SETIT), March 21-24, 2012, IEEE, Morocco, North Africa, ISBN:978-1-4673-1657-6, pp: 404-411.

Dal, M.C., P. Zanuttigh and G.M. Cortelazzo, 2012. Fusion of geometry and color information for scene segmentation. IEEE. J. Sel. Top. Signal Process., 6: 505-521.

Kang, H.B., 2001. A hierarchical approach to scene segmentation. Proceedings of the 2001 IEEE Workshop on Content Based Access of Image and Video Libraries (CBAIVL 2001), December 14, 2001, IEEE, South Korea, East Asia, ISBN:0-7695-1354-9, pp: 65-71.

Panagiotakis, C., A. Doulamis and G. Tziritas, 2009. Equivalent key frames selection based on ISO-content principles. IEEE. Trans. Circuits Syst. Video Technol., 19: 447-451.

Peng, J. and Q. Xiaolin, 2010. Keyframe-based video summary using visual attention clues. IEEE. MultiMedia, 17: 64-73.

Potnurwar, A.V. and M. Atique, 2014. Visual attention key frame extraction for video annotations. Int. J. Comput. Sci. Eng. (IJCSE.), 3: 39-42.

Sabbar, W., A. Chergui and A. Bekkhoucha, 2012. Video summarization using shot segmentation and local motion estimation. Proceedings of the 2012 2nd International Conference on Innovative Computing Technology (INTECH), September 18-20, 2012, IEEE, Morocco, North Africa, ISBN:978-1-4673-2678-0, pp: 190-193.

Song, Y., T. Ogawa and M. Haseyama, 2010. MCMC-based scene segmentation method using structure of video. Proceedings of the 2010 International Symposium on Communications and Information Technologies (ISCIT), October 26-29, 2010, IEEE, Sapporo, Japan, ISBN:978-1-4244-7007-5, pp: 862-866.

Sundaram, H. and S.F. Chang, 2000. Video scene segmentation using video and audio features. Proceedings of the IEEE International Conference on Multimedia and Expo, (ICME'00), Istanbul, Turkey, pp: 1145-1148.

Tanapichet, P., N. Cooharojananone and R. Lipikorn, 2011. Automatic comic strip generation using extracted keyframes from cartoon animation. Proceedings of the 2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), December 7-9, 2011, IEEE, Bangkok Thailand, ISBN:978-1-4577-2165-6, pp: 1-6.

Truong, B.T. and S. Venkatesh, 2007. Video abstraction: A systematic review and classification. ACM. Trans. Multimedia Comput. Commun. Appl., 3: 1-37.

Ventura, C., I.N.X. Giro, V. Vilaplana, D. Giribet and E. Carasusan, 2013. Automatic keyframe selection based on mutual reinforcement algorithm. Proceedings of the 2013 11th International Workshop on Content-Based Multimedia Indexing (CBMI), June 17-19, 2013, IEEE, Barcelona, Spain, ISBN:978-1-4799-0955-1, pp: 29-34.

Widiarto, W., E.M. Yuniarno and M. Hariadi, 2015a. Video summarization using a key frame selection based on shot segmentation. Proceedings of the 2015 International Conference on Science in Information Technology (ICSITech), October 27-28, 2015, IEEE, Surakarta, Indonesia, ISBN:978-1-4799-8384-1, pp: 207-212.

Widiarto, W., M. Hariadi and E.M. Yuniarno, 2015b. Shot segmentation of video animation to generate comic strip based on key frame selection. Proceedings of the 2015 IEEE International Conference on Control System Computing and Engineering (ICCSCE), November 27-29, 2015, IEEE, Surakarta, Indonesia, ISBN:978-1-4799-8251-6, pp: 303-308.

Zhai, Y. and M. Shah, 2006. Video scene segmentation using markov chain monte carlo. IEEE. Trans. Multimedia, 8: 686-697.

Zhao, Y.J., T. Wang, P. Wang, W. Hu, Y.Z. Du, Y.M. Zhang and G.Y. Xu, 2007. Scene segmentation and categorization using NCuts. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 17-22, Minneapolis, MN., pp: 1-7.

Zhu, Y. and Z. Ming, 2008. SVM-based video scene classification and segmentation. Proceedings of the 2008 IEEE International Conference on Multimedia and Ubiquitous Engineering 2008 MUE, April 24-26, 2008, IEEE, Shenzhen, China, ISBN:978-0-7695-3134-2, pp: 407-412.