International Journal of Grid and Distributed Computing
Vol. 13, No. 2, (2020), pp. 214 –219

Video Summarization: Keyframe Extraction based on Absolute Difference Method

R. P. Ram Kumar,
Professor, Department of Computer Science and Engineering,
Malla Reddy Engineering College (Autonomous), Secunderabad, Telangana State, India

Sanjeeva Polepaka,
Associate Professor, Department of Computer Science and Engineering,
Malla Reddy Engineering College (Autonomous), Secunderabad, Telangana State, India

Abstract

Due to the technological provoke in the digital era, there is an inevitable exponential growth in the video content generation and utilization for various applications. Browsing the complete video contents is tedious and time-consuming work. Hence, the summary of the video enables the user to browse the video content and proceed further (into the original video content), if interested. Henceforth, this paper aims to propose a video summarization method to extract keyframes from the input video based on the absolute frame difference method. When evaluated on the publicly available video, the examined results stated that the proposed method has improved performance regarding the keyframe extraction.

Keywords *– Video Summarization, Frames, Key Frames, Absolute Difference, RGB Channels*

1. INTRODUCTION

Segmentation, a contemporary technique, plays a significant role in processing the data acquired and making it simpler, meaningful and understandable for the user. In terms of research study, data may include digital media or patterns that aid in the analysis and understanding of video or interpretation of images or video summarization for detailed study. Irrespective of the field of study, interpretation or analysis of information needs to be in place [9]. As a matter of interest, Video Segmentation in digital media processing is practiced in medical imaging, image interpretation, surveillance, computer-guided surgery, motion object recognition, video summarization and many more.

Video segmentation is the method of segregating a video into multiple sets in subsequent frames that are homogeneous to specific criteria. The homogeneity in video segmentation is mostly in terms of pixels. Instances include segmentation of moving objects, object and background identification and so on. The following depicts the applications of video segmentation summarized from [10].

- Video monitoring to improve and track moving objects
- Object recognition to extract the features of the region of interest.
- Data compression to improve quality and reduce visual artifacts
- Computer vision
- Videophone applications to improve the foreground visually
- Digital entertainment to add effects and use for virtual scenes.
-

ISSN: 2005-4262 IJGDC

2. EXISING METHODS

The following section deals with few of the existing video segmentation methods.

Gopal, Kalpana and Bhose [1] proposed the detection and segmentation of objects in motion from an input video based on frames through Differencing and Summing Technique (DST). Initially, 8 successive frames from a video sequence are read and converted into greyscale for further processing. Consequently, each of the frames is subtracted from the background image and averaged. Further, averaged image is converted into a binary image and the noises in the resultant image are removed by morphological operation. Finally, the object with maximum displacement identified is covered with a bounding box for thorough labeling and scanning of pixels. As the samples for the study included indoor and outdoor videos, DST is robust and efficient in segmenting the object with displacement.

Onur, Ugur, Ozgur [2] claim content-based segmentation from videos by the Fuzzy Color Histogram method. The features of the input frame are computed by linear regression in the opening frame dropping detector step. The noises in the frame are pulled out through the median filter. Further, the mask is generated for the pattern or logos and the boundary is clipped out for the frame. Finally, the Fuzzy color histogram is applied where precision and recall values of other trending algorithms such as Color Histogram, Probabilistic Block intensity, Edge tracking, and Local Keypoint Matching are computed and compared. The technique revealed a 93.67% superior precision rate for color frames but reduced to 75.8% when the frames were transformed to greyscale.

Long and Wu [3] used the Canny Edge Detection technique for frame segmentation and pattern recognition-based technique to extricate the background from the segmented frame. The edge mask obtained in the Canny operator act as input for the pattern recognition technique. The edges are distinguished in either way. The pixel motion is decided based on the presence of an edge in the background or to the moving object. The other way is based on the pixel movement the edges belonging to background and object are decided. The edges that belong only to the object remain to discard the background. Techniques like Histogram, Threshold value help reduce the complexity.

Shigang, Xuegen and Hexin [4] put forth an algorithm fusing Disparity map with frame difference to fragment the object from the input video. The foremost step is to segment based on Disparity map to acquire the initial segmentation object. Then the objects are extracted by the Stereo Vision object Segmentation algorithm where the acquisition of an accurate motion picture is possible. As the disparity maps are of size, 160 X 120, the vision is clear. Less complexity is one of the vital features of this technique.

Chun Yu and Ming Yue [5] segmented the video by combining Kirsch edge operator and three frame difference. Three consecutive frames from a video and its greyscale conversion act as input. The moving object's border is identified obviously by the Three Frame Difference algorithm without noise or light or a shadow or background reappearance. Further, an adaptive Kirsch edge operator detects the edges with discontinuity which is eliminated using mathematical morphology.

Zhongkun, Yonping and Dechun [6] proved the LBP descriptor coding method segmented the video in ease without background construction and uncomplicated computation. The LBP method put forth by the authors is based on 3D spatial-temporal neighborhood value. Further, iterative threshold and morphological filter in binarization of frame ensure noise-free frame. The supremacy of the 3D LBP technique is proved by comparing the result with the difference algorithm.

Kavitha and Shanmugam [7] demonstrate a robust and proficient Edge Detection Technique by comparing Model matching and change detection of edges in Video object Extraction techniques. The comparison proved VOP extraction based on Model matching segmented even overlapped complex motion objects with partial occlusion. When combined with the motion vector and Haunsdroff measure improved segmentation results. The major pitfall, boundary inaccuracy in Change Detection Method was partially overcome by the Model Matching method and segmentation was amazing even in muddle background.

ISSN: 2005-4262 IJGDC

3. PROPOSED METHOD

The proposed video segmentation and key frame extraction method has two phases, namely, Phase-I and Phase-II. The Phase-I comprises the decomposition of video, i.e., the segmentation of input video into frames and write the corresponding frames into a specific directory, whereas the phase-II deals with the extraction of key frames from the generated frames using absolute difference parameter between the consecutive frames. Figures 1 and 2 depict the pseudo code for phase-I and II of the proposed method, respectively.

Input : *Video with ".avi" or ".mp4" format*
Output : *Frames*
Step 1: Read the video and store it in a variable 'v'.
Step 2: Create a destination directory 'd'.
Step 3: Extract the frames 'f' (i.e. the images) from the video 'v' and store it in the directory 'd'

Figure 1 Pseudo code for Phase-I

Input : *Frames from video*
Output : *Keyframes*
Step 1: Start reading the frames 'f' from the directory 'd'
Step 2: Number of available frames is stored in 'n'
Step3: Repeat the following for 'n' frames
Step 3 (a): Read the first frame and store it in 'f'
Step 3 (b): Euclidean norm or Euclidean length is determined for successive frames i.e. 'f1' and 'f2' regarding red, green and blue components and stored in '*eudiff*'.
Step 4: Threshold 't' is determined by calculating the mean value for '*eudiff*' parameter
Step 5: Repeat the following from the second frame to (n-2)
Step 5 (a): if the 'eudiff' of the current frame (j) is greater than the previous frame (j-1), next frame (j+1) and 't', then mark as "Keyframe"

Figure 2 Pseudo code for Phase-II

Initially, loading of input video followed by extracting the frames from it is carried out in the Phase-I. Later, Phase-II deals with the keyframes extraction from the resultant frames of Phase-I. Here, the Euclidean norm or Euclidean length of the successive frames for Red(R), Green (G), and Blue (B) components are determined regarding all the frames. Then, the threshold 't' is determined by the mean value of Euclidean norm values. Whenever the Euclidean norm values of RGB components of the current frame is greater than the previous frame, subsequent frame and determined threshold value 't', then the corresponding prevailing frame is identified as a "Keyframe". The procedure is repeated iteratively until all the frames are processed.

Later, the phase-II procedure is repeated with a frame difference of three; that is, the k^{th} frame is compared with $(k+3)^{th}$ frame.

4. EXPERIMENTAL RESULTS

The proposed work is implemented on MATLAB version 2018(b) installed on the Intel Core I3 2100 CPU @ 3.10 GHz machine with 4GB RAM. A 'Tom and Jerry' video is downloaded from the World Wide Web (WWW) [8], which plays for two minutes and fifty-nine seconds. In phase-I, the input video is divided into 5378 number of frames. The phase-II is made to execute twice with the frame difference one and three and extracted the keyframes from the input video. Table 1 shows the number of keyframes extracted for one and three frame difference methods.

ISSN: 2005-4262 IJGDC

Table 1 Keyframes extracted in the proposed method

Sl. No	Name of the input video	Duration of the Video	Extracted keyframes with a frame difference of	
			One	Three
1	Tom and Jerry Show – Beach Bully (1975).mp4	2 minutes and 59 seconds	143	47

Further, the efficiency of the proposed keyframe extraction method is assessed by Compression Ratio (CR) [11]. CR is defined as the ration of total frames in the input video to the extracted keyframes as represented in Equation 1.

$$CR = \frac{\text{total frames in the input video}}{\text{extrated keyframes}} \quad (1)$$

CR of keyframes extraction method based on single frame difference= $\frac{5378}{143}$ =37.608 and CR of keyframes extraction method based on three frame difference = $\frac{5378}{47}$ =114.425

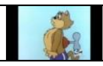
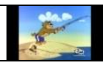
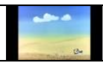
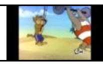
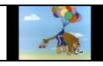
Observing Table 1 and extracted frames (manually), extraction based on the absolute difference of three subsequent frames method resulted in better performance than the single frame difference method. Table 2 shows the comparison of extracted keyframes with one and three frame differences. With respect to frame difference of one and three, the number of extracted frames is 143 and 47, respectively. Table 3 depicts the resultant 22 extracted common frames between the frame difference of one and three with their frame numbers.

Table 2 Comparison of Extracted Keyframes with Frame difference of one and three

Sl. No.	Description	Number of Frames	Extracted Frame Numbers
1	Input Video	-	5378
2	Extracted keyframes with a frame difference of One	143	41, 94, 173, 194, 208, 210, 228, 281, 285, 287, 289, 292, 294, 297, 299, 317, 383, 487, 543, 545, 547, 638, 757, 826, 935, 1009, 1064, 1276, 1288, 1290, 1316, 1526, 1592, 1662, 1746, 1776, 1814, 1994, 2090, 2096, 2101, 2198, 2230, 2276, 2344, 2404, 2459, 2579, 2650, 2652, 2654, 2659, 2661, 2687, 2733, 2773, 2842, 2847, 2874, 2880, 2882, 2887, 2889, 2891, 2894, 2896, 2899, 2904, 2914, 2921, 2938, 2953, 2959, 2961, 2971, 3010, 3035, 3037,3084, 3102, 3178, 3188, 3191, 3193, 3196, 3199, 3201, 3203, 3214, 3216, 3218, 3220, 3235, 3238, 3240, 3243, 3262, 3264, 3266, 3269, 3271, 3274, 3276, 3280, 3303, 3353, 3380, 3387, 3390, 3392, 3394, 3396, 3422, 3462, 3590, 3709, 3902, 3966, 3969, 3971, 3974, 3976, 4036, 4038, 4127, 4134, 4273, 4387, 4515, 4634, 4703, 4739, 4755, 4799, 4844, 4847, 4862, 4891, 4899, 4901, 4921, 5060, 5371
3	Extracted keyframes with a frame difference of Three	47	543, 545, 547, 755, 757, 824, 1007, 1316, 1662, 1744, 1774, 1776, 1814, 1993, 2196, 2198, 2276, 2403, 2458, 2577, 2771, 2889, 2969, 3082, 3084, 3176, 3392, 3460, 3462, 3708, 4127, 4132, 4273, 4386, 4513, 4515, 4703, 4753, 4755, 4797, 4799, 4860, 4862, 4919, 5058, 5060, 5370

ISSN: 2005-4262 IJGDC

Table 3 Resultant Extracted Common Frames between Frame Difference of One and Three

Frame No	543	545	547	757	1316	1662	**1776**
Frames							
Frame No	**1814**	**2198**	**2276**	**2889**	**3084**	**3392**	**3462**
Frames							
Frame No	**4127**	**4273**	**4515**	**4703**	**4755**	**4799**	**4862**
Frames							
Frame No	**5060**						
Frames							

5. CONCLUSION

The proposed method aimed to extract the keyframes from the input video using an absolute keyframe difference method. There are two phases, namely, (1) the first phase extracted the frames from the input video and (2) the second phase extracted the keyframes from the generated frames. The keyframes were determined using the absolute keyframe difference method with the frame difference of one and three frames. Experimental results concluded that the three frame difference method outperformed the one frame difference method regarding the extracted frames. Since the major pitfall of this approach is the execution time (the execution time grows exponentially with respect to the number of frames) Clustering-based technique and dimensionality reduction technique would be the choice for future enhancements in the keyframe extraction methods.

REFERENCES

[1] Gopal Thapa, Kalpana Sharma, M. K. Ghose, "Moving Object Detection and Segmentation using Frame Differencing and Summing Technique," IJCA , vol. 102, no. 7, pp. 20–25, 2014, doi: 10.5120/17828-8647.

[2] Onur Küçüktunç, Uğur Güdükbay, Özgür Ulusoy, "Fuzzy color histogram-based video segmentation," Computer Vision and Image Understanding, vol. 114, no. 1, pp. 125–134, 2010, doi: 10.1016/j.cviu.2009.09.008.

[3] Long Xiang, Wu Xiaoqing, "Motion Segmentation Based on Edge Detection," Procedia Engineering, vol. 29, pp. 74–78, 2012, doi: 10.1016/j.proeng.2011.12.671.

[4] Shigang Wang, Xuejun Wang, Hexin Chen, "A Stereo Video Segmentation Algorithm combining Disparity Map and Frame Difference," Proceedings of 3rd International Conference on Intelligent System and Knowledge Engineering, pp. 1121–1124, 2008, doi: 10.1109/iske.2008.4731098.

[5] Chun Yu Chen and Ming Yue Zhao, "Video Segmentation Algorithm Based on Improved Kirsch Edge Operator and Three-Frame Difference," Advanced Materials Research (AMR), vol. 981, pp. 335–339, 2014, doi: 10.4028/www.scientific.net/amr.981.335.

[6] Zhongkun He, Yongping Zhang, Dechun Zheng, "Video Segmentation using a novel LBP Descriptor," ICTACT Journal on Image and Video Processing: Special Issue on Video Processing for Multimedia Systems (IJIVP), vol. 5, no. 1, pp. 883–886, 2014, doi: 10.21917/ijivp.2014.0128.

[7] Kavitha Ganesan and Shanmugam Jalla, "Video Object Extraction Based on a Comparative Study of Efficient Edge Detection Techniques," The International Arab Journal of Information Technology, vol. 6, no. 2, pp. 107–116, 2009.

[8] Jonni Valentayn, “Tom And Jerry Show - Beach Bully (1975),” YouTube, 11-Nov-2019. [Online]. Available: https://www.youtube.com/watch?v=DuwRR2PkGHs. [Accessed: 22-Jan-2020].
[9] King Ngi Ngan, Hongliang Li, Video Segmentation and Its Applications. Springer Science & Business Media, 2011.
[10] Ling Liu and M. Tamer Özsu, Encyclopedia of database systems. New York: Springer, 2009.
[11] Sheena C V and N. K. Narayanan, “Key-frame Extraction by Analysis of Histograms of Video Frames Using Statistical Methods,” Procedia Computer Science, vol. 70, pp. 36–40, 2015, doi: 10.1016/j.procs.2015.10.021.

ISSN: 2005-4262 IJGDC