KEY FRAME EXTRACTION FOR VIDEO SUMMARIZATION USING MOTION ACTIVITY DESCRIPTORS

Supriya Kamoji[1], Rohan Mankame[2], Aditya Masekar[3], Abhishek Naik[4]

[1]*Assistant Professor, Computer Engineering, Fr. Conceicao Rodrigues College of Engineering, Maharashtra, India*
[2]*B.E. student, Computer Engineering, Fr. Conceicao Rodrigues College of Engineering, Maharashtra, India*
[3]*B.E. student, Computer Engineering, Fr. Conceicao Rodrigues College of Engineering, Maharashtra, India*
[4]*B.E. student, Computer Engineering, Fr. Conceicao Rodrigues College of Engineering, Maharashtra, India*

Abstract

Summarization of a video involves providing a gist of the entire video without affecting the semantics of the video. This has been implemented by the use of motion activity descriptors which generate relative motion between consecutive frames. Correctly capturing the motion in a video leads to the identification of the key frames in the video. This motion in the video can be obtained by using block matching techniques which is an important part of this process. It is implemented using two techniques, Diamond Search and Three Step Search, which have been studied and compared. The comparison process is tried across various videos differing in category, content, and objects. It is found that there is a trade-off between summarization factor and precision during the summarization process.

Keywords: *Video Summarization, Motion Descriptors, Block Matching*

----------***----------

1. INTRODUCTION

Video summary is the abstract of an entire video. It is the essence of the entire video provided in a shorter period of time. Video summarization can be defined as a non-linear content-based sampling algorithm, which provides a compact representation of a given video sequence [2].

The main purpose of video summary is due to viewing time constraints [2]. It helps us assess the value of information within a shorter period of time, while we make decisions. Its aim is to provide a compact video sketch, while it preserves the high priority entities of the original video. Video summarization can be deemed necessary in order to reduce large amount of data involved in video retrieval.

Video summarization plays a major role where the resources like storage, communication bandwidth and power are limited. It has several applications in security, military, data hiding and even in entertainment domains [7].

Consider the situation, of a military base which is situated in a remote location. The location is such that it causes bandwidth constraints. Videos which are high definition or are very large cannot be sent in and around this base easily. In scenarios like this, Video summarization can be used which creates an abstract of the whole video without losing on any important data. Thus, a shorter video of shorter length and of a shorter size is obtained which can be easily transmitted in and around the base even with the bandwidth constraints.

Another scenario where this would be applicable is of a surveillance video camera of an automated banking machine (ABM or ATM). The video tapes are generally checked by the respective security forces after a very long duration like 24 hours or 48 hours. It is humanly impossible to scrutinize a 24 hour video. In addition to that, the parts of video wherein there is some motion present in the ABM is highly important than the other parts of the video sequence. We can use video summarization in such a scenario which will provide us with the relevant video. The output video will contain the parts of the sequence which has motion in them thereby reducing our effort and making it possible for the security service to keep a proper surveillance.

2. RELATED WORK

Video summarization can be carried out in different methods. Each method is suitable in its own domain and can thus give variable results based on a number of parameters.

Liu et al. in [5] define a key as the key image of a video shot. Some key frame extraction methods are described in brief as follows:
1) Video Shot Method - It has frame average method and histogram average method. The key frames are extracted after computing maximum distance of the feature space.
2) Content Analysis Method - In this method we extract key frames based on color, texture and other visual information of each frame, whenever this information changes significantly, the current frame is considered as the key frame.

3) Cluster-based Method - This method uses cluster efficiency analysis; the frame which is most close to the cluster center is selected as the key frame.
4) Motion-based Analysis - This method searches for the local minimum in the movement of key frames.

In [5] a method based on improved optimization of frame difference is implemented. It concentrates on the following main points in a video:
1) When the directors shoot the videos, most of the times they put the most important part at the center of the shots and
2) The bodyline and the four corners of the shot don't seem so interesting comparatively.

In this method more importance is given to the center of the image rather than the other parts. Furthermore, the inter-frame distance is calculated using a weightage matrix which stresses out on the central block in the images. The key frames are selected after this part.

Zeinalpour et al. in [2] take the help of genetic algorithm to summarize a video. It is a search technique which is used in computing to find approximate solutions to optimization and search problems. The procedure is discussed as follows:
1) Sampling - A video may have many frames, and a large part of these frames which are adjacent are likely to be similar. Reduce this set of images by removing the images which look similar.
2) Encoding - To make chromosome, take a string of 0's and 1's. The value of 0 indicates those frames which are not selected while 1 denotes that the frame is selected.
3) Fitness Function - It is used to calculate the fitness of the chromosomes.
4) Crossover and Mutation - Genetic algorithm then works by selecting pairs of individual chromosomes, depending on their fitness function values. Later, any two chromosome strings will swap their gene's values from a random split point. The termination condition computes average mean of whole chromosome's fitness function values. If the mean value is more than the specified threshold, the generation loop will be broken. The winner would be the chromosome that has the maximum fitness value.

Sony et al. in [3] use Euclidean distance after clustering to obratin summarized frames. This method is based on the removal of redundant frames from a video and maintaining the user defined number of unique frames. Visually similar looking frames are clustered into one group using the Euclidean distance. After the clusters are formed, the frames that have larger distance metric are retrieved from each group to form a sequence. This makes up the desired output.

The algorithm is discussed as follows:
1) Video Acquisition - This is the process where an analogy video signal is converted to digital form.
2) Video Framing - This is used to convert the video into frames.
3) Euclidean Distance - In this the root of square differences are measured. The portions of video where motion changes considerably are detected. Two frames will be considered similar when the Euclidean distance between two frames is very less.
4) Iterative boundary scene change detection - After finding the approximate average Euclidean distance. Using iterations and depth the nodes are split as per the algorithm.
5) Frame Reduction - To preserve maximum continuity and less redundancy the number of frames to be taken from each node is to be properly selected.
6) Video Composition - The selected frames which are obtained from each node are combined to form the summarized video and it is saved as a new '.avi' file.

Doulamis et al. in [10] have discussed key frame extraction using cross correlation criterion which is implemented by forming a multidimensional fuzzy histogram

3. PROPOSED ALGORITHM

The aim of the algorithm is to provide a summarized video which produces a gist of the original video without losing semantics of the video. Fig-1 provides the blueprint for our process.

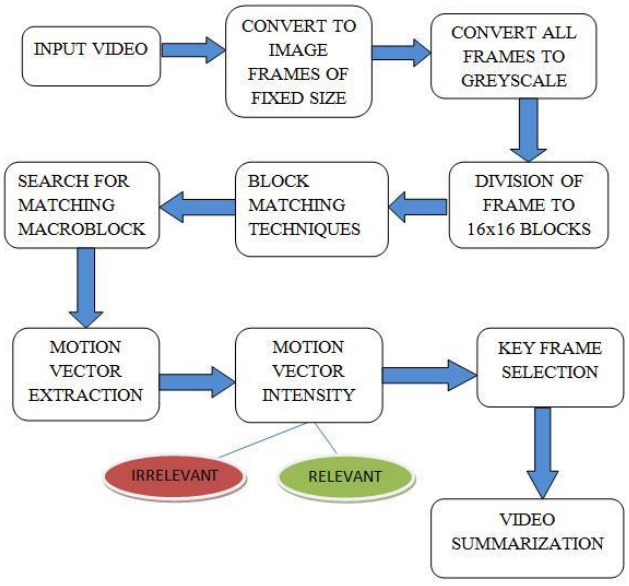

Fig -1: Proposed System

The initial process involves converting the input video into frames. After which the frames are grey scaled. Later, each frame is further divided into a fixed number of macroblocks (16x16 in this case) which facilitates the use of an individual macroblock as comparison units. The first macroblock of the first frame is then compared with the macroblocks in the

second frame to search for the closest match to the original macroblock. Comparing all macroblocks in the second frame is a tedious process and hence an astute method of selection of macroblocks is required which gives the correct match yet saves processing time. This is implemented with the use of block matching algorithms which form the crux of this system. Each block matching algorithm specifies which blocks are to be compared and in what order.

Once a block of the first frame is matched with the block of the second frame, the motion activity descriptor of the block can be established. This process is then repeated for each block of the first frame, and sum of all such motion descriptors is considered to produce the cumulative motion descriptor between the two frames. Such a cumulative motion descriptor is obtained between each pair of consecutive frames. These motion descriptors are then compared to categorize them into irrelevant and relevant. The motion descriptors signify the amount of motion present between two consecutive frames. Absence of motion signifies no or minimum difference between two frames, whereas a high motion descriptor signifies a vast difference between two frames and thus leads to the conclusion of them being key frames. Summation of all such key frames will lead to the formation of the summarized video.

3.1 Block Matching Algorithms

Block matching algorithms are essential in selecting which blocks are to be selected for comparison and the order in which they are to be traversed. They often include iterative processes which continue until the closest match to the original block is found. Based on the pattern on matching, there are multiple block matching algorithms. This study utilizes two such algorithms viz. Diamond Search and Three Step Search.

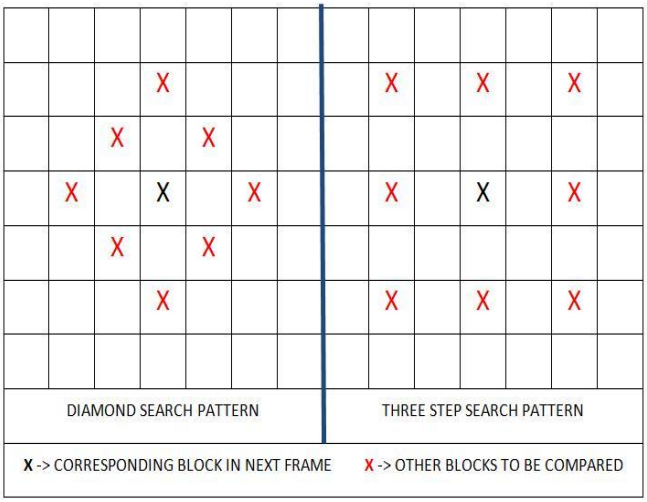

Fig -2: Block Matching Patterns

3.1.1 Diamond Search

The search pattern in diamond search is in the shape of a diamond. It consists of one block at the center and 8 blocks in a diamond pattern around it as show in Fig -2. Each of the 9 blocks from the second frame is compared with the original block from the first frame and the least cost match is found. That block then becomes the new center block and another diamond pattern is formed around it. This process is repeated until center block itself is the least cost match after which the diamond is contracted and only the immediate neighbours of the center block are checked. The closest match in this last step is selected as the result block.

3.1.2 Three Step Search

In three step search pattern, a parameter S which is known as step size is set. The center block is considered, and then 8 blocks at a distance of +/- S from the center block are selected. These blocks are compared with the original block and least cost match is selected. This becomes the new center for the pattern in the second step while the step size S is then halved. This iterative process is carried out till S = 1 wherein the closest match is then selected as the result block.

3.2 Block Comparison

Once two blocks are selected to be compared by the block matching algorithms, the cost between those two blocks has to be found. Lower the cost, higher the similarity between the two blocks whereas a high cost signifies a high difference between the blocks. The blocks are compared to find a match and thus get the resultant motion activity descriptor.

x(i,j) and y(i,j) are assumed to be the scalar displacement or motion along the X and Y axis respectively . The motion activity matrix of a frame is defined by

$$C_{mv} = \{R(i,j)\} \tag{1}$$

Where R, the resultant motion descriptor is given as

$$R_{xy}(i,j) = \sqrt{x(i,j)^2 + y(i,j)^2} \tag{2}$$

The average motion activity of each frame is given by:

$$C_{mv}^{avg} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} C_{mv}(i,j) \tag{3}$$

The frames which then fall in the high motion or relevant region are then selected as key frames and used to summarize the entire video.

4. RESULTS

This system aims at providing a summary of the original video such that when the target watches the summarized video, he/she gets the crux of the idea presented in the original video. Although the motion activity descriptors can provide high compression, precision is an important factor in how effective the summarization is. Therefore, this system works best in situations where the recording device is constant and there are infrequent scene changes. If a video includes constant scene changes, then it proves difficult to summarize it effectively. The effectiveness of this system on different categories of videos is scene from Table -1.

The parameters are calculated as follows:

Precision = No. of correctly matched frames / Desired Frames (4)

Summarization Factor = (Total Frames - Obtained Frames) / Total Frames (5)

Precision determines the accuracy of the summarized video whereas summarization factor shows to what extent the original video has been shortened. There is often a trade-off between precision and summarization factor as can be seen from Table-1.

Table 1

Videos	Total Frames	Desired Frames	Diamond Search			Three Step Search		
			Output Frames	Precision	Summarization Factor	Output Frames	Precision	Summarization Factor
Surveillance	37480	135	136	96.29	99.63	127	94.25	99.66
Documentary	42921	1793	1710	94.64	96.01	1655	92.35	96.14
Outdoor	23430	160	120	75	99.48	125	78.65	99.46
Racing	44954	970	938	96.70	97.91	927	95.59	97.93
Dance	36700	1539	1463	94.41	96.01	1440	93.56	96.07
Sunrise	36957	969	969	100	97.37	969	100	97.37
Table-Tennis	46946	576	533	92.53	98.86	527	91.36	98.87
Tennis	17878	743	709	94.61	96.03	682	91.86	96.18
Speech	44737	1637	1631	98.16	96.35	1595	97.45	96.43
Lecture	57203	1144	1125	97.20	98.03	1091	95.38	98.09
Animation	42469	344	240	69.18	99.43	213	62.08	99.49
Tornado	53997	261	255	94.25	99.52	251	96.07	99.53
Theatre	45058	1839	1791	97.17	96.02	1812	98.55	95.97
Office	39127	232	224	96.12	99.42	222	95.72	99.43
Cricket	54700	2379	2302	96.67	95.79	2326	97.75	95.74

Documentary, theatre, outdoor and sports have constant scene changes or high motion in them which leads to a higher number of key frames and hence lowers the summarization factor.

The precision is high in videos where motion can be captured effectively. In certain categories such as Animation and Outdoor where the motion is minimal and quick whereas area of consideration is large and objects are small, the precision tends to be low. Precision is higher in videos where motion is cognizable and area of consideration is smaller such as Speech, Lecture and Theatre. A noticeable exception is Sunrise which has very high summarization factor due to the fact that it has a single object, slow motion and no shot changes.

5. CONCLUSIONS

The aim of this system is to provide with a summary of a video by utilizing and capturing the motion throughout it. It was found out that precision and summarization factor are important parameters in this process and the idea was to maximize both. However, as per the above observations different categories of video produced different results. The summarization proves effective in situations having limited area and definite objects as it eases the formation of motion activity descriptors. The block matching technique used affects the process which can be seen from the results. Diamond Search has an advantage over Three Step Search where it achieves higher precision.

REFERENCES

[1]. Huayong Liu, Lingyun Pan,Wenting Meng, "Key Frame Extraction from Online Video Based on Improved Frame Difference Optimization".Communication Technology, IEEE 14th International Conference, 2012: 940-944.
[2]. Zeinab Zeinalpour, Behrouz Minaei Bidgoli, Mahmud Fathi, "Video Summarization Using Genetic Algorithm and Information Theory" Computer Conference, 14th International CSI, 2009: 158-163.
[3]. Aju Sony, Kavya Ajith, Keerthi Thomas, Tijo Thomas , Oeepa P. L., "Video Summarization By Clustering Using Euclidean Distance". Proc. International Conference on Signal Processing, Communication, Computing and Networking Technologies, 2011: 642-646
[4]. Omer Gerek, Yucel Altunbastak, "Key Frame Selection from MPEG Video Data", Proc. SPIE Vol. 3024, Visual Communications and Image Processing, 1997: 920-925.
[5]. Huayong Liu, Wenting Meng, Zhi Liu, "Key Frame Extraction of Online Video Based on Optimized Frame Difference". 9th International Conference on Fuzzy Systems and Knowledge Discovery, 2012: 1238-1242.
[6]. Sujatha C, Uma Mudenagudi," A Study on Keyframe Extraction Methods for Video Summary" International Conference on Computational Intelligence and Communication Systems, 2011: 73-77
[7]. Ebrahim Asadi, Nasrolla Moghadam Charkari, "Video Summarization Using Fuzzy C-Means Clustering". 20th Iranian Conference on Electrical Engineering, 2012: 690-694.
[8]. Bernn Erol and Fnoiizi Kossentini, "Video Object Summarization in the Mpeg-4 Compressed domain". Acoustics, Speech, and Signal Processing, IEEE International Conference, 2000:2027-2030
[9]. Shinya Fujiwara and Akira Taguchi,"Motion-Compensated Frame Rate Up-Conversion Based on Block Matching Algorithm with Multi Size Blocks" Proc. International Symposium on Intelligent Signal Processing and Communication Systems, 2005: 353-356
[10]. Anastasios D. Doulamis, Nikolaos D. Doulamis and Stefanos D. Kollias "Efficient Video Summarization Based On A Fuzzy Video Content Representation". IEEE International Symposium on Circuit and Systems,2000:301-304.
[11]. Noboru Babaguchi Kouzou Ohara Takehiro Ogura," Effect of Personalization on Retrieval and Summarization of Sports Video*" Proc. Joint Conference of the Fourth International Conference on International Communication and Signal Processing, 2003:940-944

BIOGRAPHIES:

Supriya Kamoji has received B.E. in Electronics and Communication Engineering with Distinction from Karnataka University in 2001 and M.E. from Thadomal Shahani College of Engineering, Mumbai, with Distinction. She has more than 10years of teaching experience and is currently working as an Assistant Professor in Fr. Conceicao Rodrigues College of Engineering. Mumbai, India. She is a life time member of Indian society of Technical Education (ISTE). Her areas of interest are Image Processing, Computer Organization and Architecture and Distributed Computing

Rohan Mankame is pursuing his B.E. in Computer Engineering from Fr. Conceicao Rodrigues College Of Engineering. His areas of interest are Image Processing, Artificial Intelligence and Database Management Systems.

Aditya Masekar is pursuing his B.E. in Computer Engineering from Fr. Conceicao Rodrigues College Of Engineering. His areas of interest are Database Management Systems, Data Structures and Data Warehousing.

Abhishek Naik is pursuing his B.E. in Computer Engineering from Fr. Conceicao Rodrigues College Of Engineering. His areas of interest are Data Strcutures, Core JAVA and Database Management Systems.