2016 Conference on Advances in Signal Processing (CASP)
Cummins College of Engineering for Women, Pune. Jun 9-11, 2016

Summarization with Key Frame Extraction using Thepade's Sorted n-ary Block Truncation Coding Applied on Haar Wavelet of Video Frame

Shalakha R. Badre
Student, Department of Information Technology
Pimpri Chinchwad College of Engineering
Pune, India
badre.shalakha@gmail.com

Dr. Sudeep D. Thepade
Professor, Department of Computer Engineering & IT
Pimpri Chinchwad College of Engineering
Pune, India
sudeepthepade@gmail.com

***Abstract*— Due to advance growth in videos available across the internet, it is required to navigate and handle them properly. It is essential to select only valuable and accurate information from video. Video summarization helps in acquiring essential information. Video summary produces concise and exact data of the video. With help of key frame extraction video summary can be generated. Key frames from video represent main content of video. In the proposed methodology, to extract key frame from video, haar wavelet transform with various levels and Thepade's sorted pentnary block truncation coding is used. For experimentation purpose test bed of 30 videos is used here. To measure the diversity among successive frames various similarity measures are used. Alias Canberra distance, Sorencen distance, Wavehedge distance, Euclidean distance and mean square error similarity measures are used. The Euclidean distance has given better performance. The increase in accuracy is observed till Haar wavelet of level 5, then higher levels have shown drop in accuracy.**

Keywords—Key frame; TSPBTC; Similarity measure; Video summarization; Haar transform

I. INTRODUCTION

The enormous amount of videos accessible across the globe leads to very slow processing of videos. In this case video summarization shows good performance. Video summarization gives the abstract form of large video. This abstract form is the main outline of video. In abstract form complete sequence of frames is not there, only selected frames give brief introduction of whole video. It is not compulsory that each successive frame contains different information. Most of the successive frames can have same data which is not required. This data can be neglected for the video summarization. So while summarizing the video these redundant frames are removed from summary and resulting summary consists of most informative frames. Video summarization can be used in many fields such as video skimming and film making. Set of images with audio and motion information can be termed as video skimming [1].

Key frame extraction plays integral role in video summarization. In the consecutive sequence of frames, frames which are most informative and more discriminative are taken as key frame. While considering this, each pair of consecutive frames is compared to find key frame.

As far as video summarization is considered lot of work has been done in this area. Latest approaches to extract key frames are sequential comparison based, global comparison based and reference frame based [2]. In sequential comparison based frames, frames subsequent to previously extracted key frames are sequentially compared with key frame [2].

Video consist of scenes, frames and key frames. Detailed structure of video is given in figure 1.

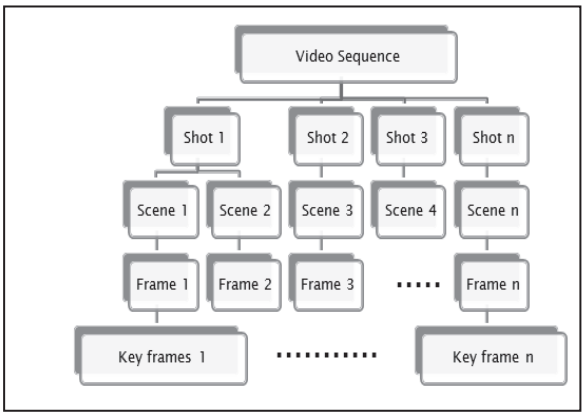

Fig. 1. Video structure in terms shots, scenes and key frames [3]

II. LITERATURE SURVEY

A. Block Truncation Coding

Block Truncation Coding (BTC) invented in 1979 was initially used for grayscale images. Block truncation coding is used in video content summarization. In block truncation coding, image is partitioned into nonlinear blocks. Block truncation coding proves to be better in color feature extraction from video and threshold is considered in BTC formulation [4]. If RGB color space is considered then for each color such as red, BTC will give two values which is

upper red and lower red. Same will be the thing for the green and blue color component. So for each frame there will be six values which are nothing but feature vector of the frame. These features can be used to extract key frame from video.

B. Haar Transform

Haar function were first derived and analyzed by Hungarian mathematician Alfred Haar in 1910. Haar function can be defined as [5]. Here the N is power of 2 and elements of Haar transform are 1, -1 and 0.

$$\mathbf{H}_N = \begin{bmatrix} \mathbf{H}_{N/2} \otimes [1 \quad 1] \\ \mathbf{I}_{N/2} \otimes [1 \quad -1] \end{bmatrix} \quad .. \ (1)$$

For ex,

$$\mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \mathbf{H}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

III. THEPADE'S SORTED PENTNARY BLOCK TRUNCATION CODING

In Thepade's sorted pentnary block truncation coding (TSPBTC), pixel values of each and every frame from video are read. Each pixel has certain value associated with that. If RGB color space is considered then for each color component pixel values are read. These values are stored in one column vector. Three columns vectors will be there for three color components. Then these column vectors are sorted in ascending order. These vectors are then divided into five parts. The average of each of these part is considered as one of the values in feature vector of respective video frame. Then each color component will have five values associated with it. In all, for each frame there will be fifteen values associated with that. These fifteen values are nothing but feature vector of the frame. These features will be used for key frame extraction.

Total intensity of red component R of m×n size can be presented in form of a single dimensional array 'SDR' having elements with indices 1 to m×n [3]. Red component four values can be formulated as follows which are shown in equations 2 to 6.

$$lR = \left(\frac{5}{m \times n}\right) \times \sum_{i=1}^{\frac{m \times n}{5}} sortedSDR(i) \quad (2)$$

$$muR = \left(\frac{5}{m \times n}\right) \times \sum_{i=(m \times n)/5+1}^{\frac{2 \times m \times n}{5}} sortedSDR(i) \quad (3)$$

$$mmR = \left(\frac{5}{m \times n}\right) \times \sum_{i=(2 \times m \times n)/5+1}^{\frac{3 \times m \times n}{5}} sortedSDR(i) \quad (4)$$

$$mlR = \left(\frac{5}{m \times n}\right) \times \sum_{i=(3 \times m \times n)/5+1}^{\frac{4 \times m \times n}{5}} sortedSDR(i) \quad (5)$$

$$uR = \left(\frac{4}{m \times n}\right) \times \sum_{i=(m \times n \times 4)/5+1}^{\frac{m \times n}{5}} sortedSDR(i) \quad (6)$$

Like that, values of green and blue component can be calculated such as in equations 7 to 11 and 12 to 16.

$$lG = \left(\frac{5}{m \times n}\right) \times \sum_{i=1}^{\frac{m \times n}{5}} sortedSDG(i) \quad (7)$$

$$muG = \left(\frac{5}{m \times n}\right) \times \sum_{i=(m \times n)/5+1}^{\frac{2 \times m \times n}{5}} sortedSDG(i) \quad (8)$$

$$mmG = \left(\frac{5}{m \times n}\right) \times \sum_{i=(2 \times m \times n)/5+1}^{\frac{3 \times m \times n}{5}} sortedSDG(i) \quad (9)$$

$$mlG = \left(\frac{5}{m \times n}\right) \times \sum_{i=(3 \times m \times n)/5+1}^{\frac{4 \times m \times n}{5}} sortedSDG(i) \quad (10)$$

$$uG = \left(\frac{4}{m \times n}\right) \times \sum_{i=(m \times n \times 4)/5+1}^{\frac{m \times n}{5}} sortedSDR(i) \quad (11)$$

$$lB = \left(\frac{5}{m \times n}\right) \times \sum_{i=1}^{\frac{m \times n}{5}} sortedSDB(i) \quad (12)$$

$$muB = \left(\frac{5}{m \times n}\right) \times \sum_{i=(m \times n)/5+1}^{\frac{2 \times m \times n}{5}} sortedSDB(i) \quad (13)$$

$$mmB = \left(\frac{5}{m \times n}\right) \times \sum_{i=(2 \times m \times n)/5+1}^{\frac{3 \times m \times n}{5}} sortedSDB(i) \quad (14)$$

$$mlB = \left(\frac{5}{m \times n}\right) \times \sum_{i=(3\times m\times n)/5+1}^{\frac{4\times m\times n}{5}} sortedSDB(i) \quad (15)$$

$$uB = \left(\frac{4}{m \times n}\right) \times \sum_{i=(m\times n\times 4)/5+1}^{\frac{m\times n}{5}} sortedSDB(i) \quad (16)$$

After calculation of these fifteen equations the feature vector will be such as [*lR, muR, mmR, mlR, uR, lG, muG, mlG, mmG, uG, lB, muB, mlB, mmB and uR*]. Here *lR* stands for lower red. *muR* for middle upper red. *mmR* for middle middle red. *mlR* for middle lower red and *uR* stands for upper red. These fifteen values are feature vector of frame.

IV. Proposed Methodology of Key Frame Extraction

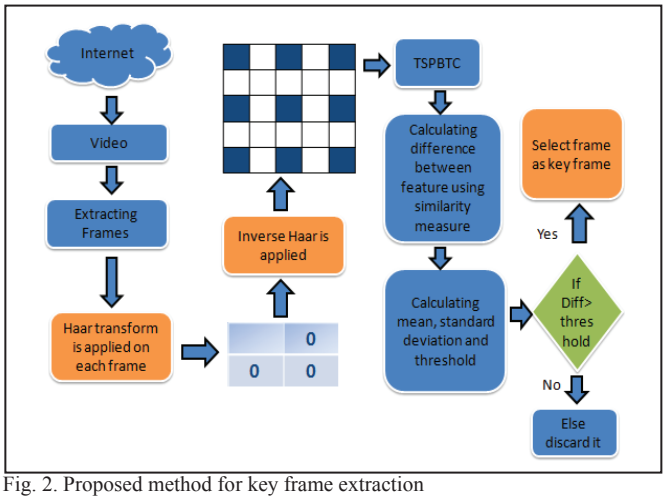

Fig. 2. Proposed method for key frame extraction

In the proposed method, video is taken from standard database called VSUMM database [6]. Video constitutes of nonlinear frames. Frames are extracted from video. On each frame from video haar transform is applied. After this process the transformed frame is divided into four parts. The left upper part is retained as it is and other three are made null. Then on that Inverse Haar is applied. After application of Inverse Haar, only alternate rows and column values are selected to get Haar wavelet level 1. Then, on the frame obtained after level 1, again whole of this process is repeated to get Haar wavelet level 2 and same is for Haar wavelet level 3 and Haar wavelet level 4,5 and 6 is followed. Then on each of the frame obtained from level 1, level 2, level 3 and level 4. Thepade's sorted pentnary block truncation coding(TSPBTC) is applied. This will give fifteen values for each frame. This is the feature vector of that frame. On these consecutive feature vectors of frame similarity measures are applied to calculate the difference between two frames. Canberra distance, Sorencen distance, Wavehedge distance, Euclidean distance and mean square error are taken here as a similarity measures. Then difference between each consecutive frame is taken which can be referred as diff. Mean of all frames differences is taken. Then standard deviation and threshold is calculated. Mean, standard deviation and threshold are given in eq no 17, 18 and 19.

$$Mean(M) = \frac{\sum_{n=1}^{N} diff(i)}{N-1} \quad (17)$$

$$S\tan dardDeviation(S) = \sqrt{\frac{\sum_{n=1}^{N-1} (diff(i) - M)^2}{N-1}} \quad (18)$$

$$Threshold(T) = M + a \times S \quad (19)$$

Here 'a' is constant. If we want to get correct key frames then following condition should be satisfied such as If (diff (i)> threshold) then output of nth frame set i+1 set as Key frame [2].

V. Similarity Measures Used

To estimate the difference between two frames, here five types of similarity measures are used which are given as follows. In this formulas R_i and S_i are consecutive frames.

A. Canberra Distance

$$CD = \sum_{i=1}^{i=n} \frac{|R_i - S_i|}{|R_i| + |S_i|} \quad (20)$$

Canberra distance was invented by Lance and Williams in 1967. The equation is given as above [3].

B. Sorencen distance

$$SD = \frac{\sum_{i=1}^{i=n} |R_i - S_i|}{\sum_{i=1}^{i=n} |R_i + S_i|} \quad (21)$$

It is used to calculate similarity between two samples.

C. Wavehedge distance

$$WD = \sum_{i=1}^{i=n} \frac{|R_i - S_i|}{\max(R_i, S_i)} \quad (22)$$

Equation 22 gives the wavehedge distance between R_i and S_i [7].

D. Mean square error distance

$$MSE = \frac{1}{N}\sum_{i=1}^{i=n}(R_i - S_i) \quad (23)$$

Mean square error is squared error distance between two consecutive frames [8].

E. Eulidean distance

$$ED = \sqrt{\sum_{i=1}^{n}(R_i - S_i)^2} \quad (24)$$

Equation no 24 calculates the Euclidean distance between two frames [9].

VI. Experimentation Environment

A. Test bed used

For the proposed technique the dataset of 30 videos are used. This database is standard which is taken from VSUMM. VSUMM database contains key frames extracted from video from different users[6].

B. Platform used

Implementation of proposed technique is developed in Matlab. System used for this purpose contain Intel core 2 duo with 2 Gb RAM. Windows 7 matlab is used as operating system.

C. Performance comparison

Accuracy [10][11]of proposed technique can be computed by using percentage accuracy formula.

$$PercentageAccuracy = \frac{ActualCorrectExtracted\ Frames}{TotalExpectedExtractionOf\ Frames} \quad ..(25)$$

Actual correct extracted frames means key frames extracted from video using proposed algorithm. Total expected extraction means key frames extracted manually [3].

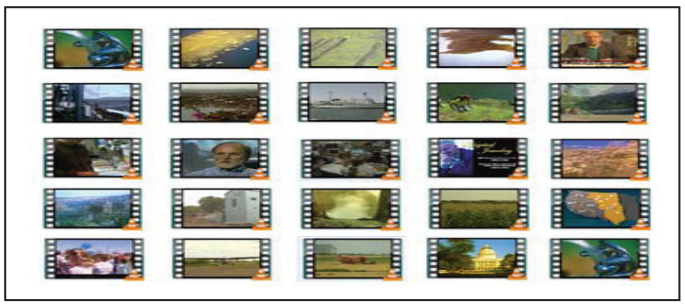

Fig. 3. Sample videos from the video dataset

VII. Result and Discussion

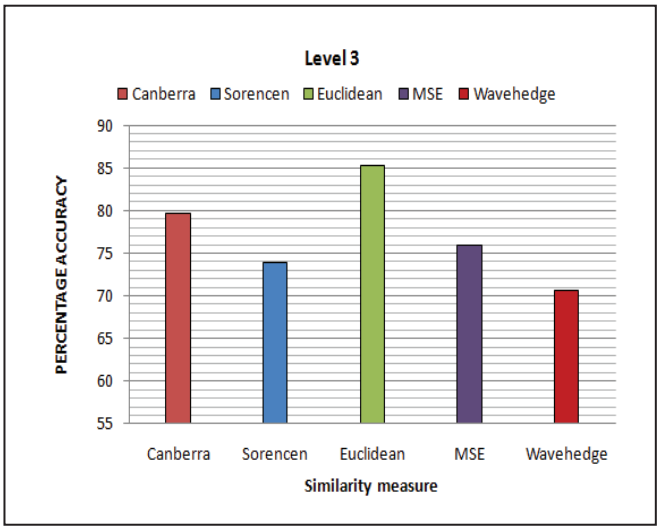

Fig. 4. Comparison of similarity measures used in proposed TSPBTC and Haar wavelet level 3 based video summarization method

Figure 4 shows that percentage accuracy for various similarity measure performed for haar wavelet level 3. In this Euclidean distance is giving better performance which is 85.30% followed by Canberra.

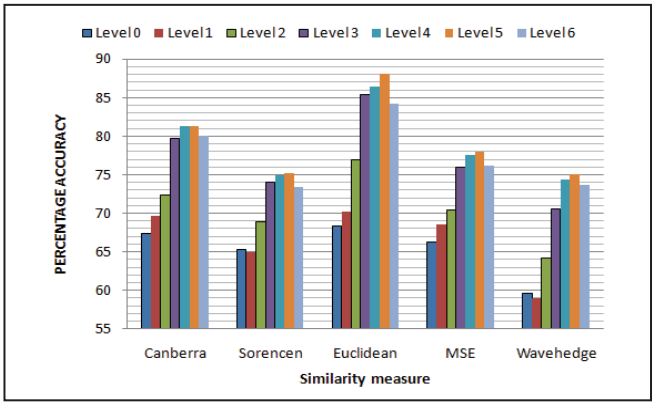

Fig. 5. Comparison of variations of proposed video key frame extraction method using assorted level of Haar wavelet for respective similarity measure

Figure 5 shows that comparison of variations of proposed video key frame extraction method using assorted level of Haar wavelet for respective similarity measure. In this Euclidean similarity measure is giving good performance which is 88.03% for level 5 followed by Canberra. After level 5 accuracy is getting decreased for all similarity measures. Accuracy is getting increased because more high energy is found there up to level 5 and it is optimal point. But on level 6 more informative content get started reducing that why accuracy gets decreased.

TABLE I. FOR PROPOSED KEY FRAME EXTRACTION METHODS OVERALL PERFORMANCE COMPARISON OF ALL VARIATIONS OF PROPOSED METHOD IS DONE IN TABLE

	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Average
Canberra	67.29	69.67	72.37	79.69	81.24	81.33	80	75.94143
Sorencen	65.29	65	68.9	74	74.89	75.2	73.35	70.94714
Euclidean	68.35	70.23	76.89	85.3	86.45	88.03	84.2	79.92143
MSE	66.23	68.559	70.37	76	77.5	78	76.2	73.26557
Wavehedge	59.6	59	64.23	70.61	74.39	75	73.68	68.07286
Average	65.352	66.4918	70.552	77.12	78.894	79.512	77.486	

VIII. CONCLUSIONS

Accessing of video in user friendly and significant way can be done with key frame extraction based on video summarization. Obtaining only key frames which contain significant information and removing redundant information summarize the video. By the use of proposed technique key frame extraction can become easy. In the proposed technique, to extract key frame from video Haar wavelet transform with various levels and Thepade's sorted pentnary block truncation coding is used. To compare consecutive frames Canberra, Sorencen, Wavehedge, Euclidean and Mean square error distance are used. Among these, Euclidean distance is performing better for level 5 which is 88.03%. Followed by this Canberra similarity measure is giving 81.33% for level 5. Overall Haar wavelet level 5 has given better key frame extraction.

REFERENCES

[1] Huayong Liu, Huifen Hao " Keyframe extraction based on improved hierarchical clustering algorithm" International conference on fuzzy system and knowledge discovery, 2014, DOI: 10.1109/TMM.2005.846906.

[2] Guozhu Liu, and Junming Zhao, " Key Frame Extraction from MPEG Video St ream", Proceedings of the Second Symposium International Computer Science and Computational Technology(ISCSCT"09) ,Huangshan, P. R. China, 26-28,Dec. 2009, pp. 007- 011, ISBN 978-952-5726- 07-7 (Print), 978-952- 5726-08-4.

[3] Dr. Sudeep D. Thepade, Pritam H. Patil "Novel visual content summarization in videos using keyframe extraction with Thepade's sorted ternary block truncation coding and assorted similarity measures " International conference on communication, information & computing technology(ICCICT), Jan 16-17 2015, DOI-10.1109/ICCICT.2015.7045726 India.

[4] YANG Shuping, LIN Xinggang," Key Frame Extraction Using Unsupervised Clustering Based on a Statistical Model", Tsinghua Science and Technology\ April 2005, 10(2): 169 – 173 Applications And Reviews, Vol. 41, No. 6, November 2011, pp no 797-819.

[5] Jian-Jiun Ding, Soo-Chang Pei, and Po-Hung Wu" Jacket haar transform" IEEE international symposium on circuits and systems 2011, DOI: 10.1109/ISCAS.2011.5937864.

[6] Sandra E. F. de Avila, Ana P. B. Lopes, Antonio da Luz Jr., Arnaldo de A. Araújo "VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method" Pattern Recognition Letters, Volume 32, Issue 1, January 2011, pages 56–68.

[7] Sung-Hyuk Cha, "Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions", International Journal of Mathematical Models and methods in Applied Sciences, Issue 4, Volume 1, 2007(300-307).

[8] Shayok Chakraborty, Omesh Tickoo and Ravi Iyer" Adaptive keyframe selection for video summarization" IEEE winter conference on applications of computer vision, 2015, DOI-10.1109/WACV.2015.99.

[9] Dr. Sudeep D. Thepade, Ashwini A. Tonge " Extraction of key frames from video using discrete cosine transform " International conference on control, Instrumentation, communication and computational technologies, 2014, DOI10.1109/ICCICCT.2014.6993160.

[10] Carles Ventura, Xavier Giro-i-Nieto, Veronica Vilaplana, Daniel Giribety and Eusebio Carasusany, "Automatic keyframe selection based on mutual reinforcement algorithm" International workshop on content based multimedia indexing, 17-19 Jun 2013, DOI: 10.1109/CBMI.2013.6576548.

[11] Hojat Yeganeh, Ali Ziaei, Amirhossein Rezaie "Novel approach for contrast enhancement based on histogram equalization" International conference on computer and communication engineering 2008, DOI: 10.1109/ICCCE.2008.4580607.