Creating Video Visual Storyboard with Static Video Summarization using Fractional Energy of Orthogonal Transforms

Ashvini Tonge[1]
Department of Information Technology
Pimpri Chinchwad College of Engineering
Pune, India

Sudeep D. Thepade[2]
Department of Computer Engineering
Pimpri Chinchwad College of Engineering
Pune, India

***Abstract*—The overwhelming number of video uploads and downloads has made it incredibly difficult to find, gather, and archive videos. A static video summarization technique highlights an original video's significant points through a set of static keyframes as a video visual storyboard. The video visual storyboards are created as static video summaries that solve video processing-related issues like storage and retrieval. In this paper, a strategy for effectively summarizing static videos using the feature vectors, which are fractional coefficients of the transformed video frames, is proposed and evaluated. Four popular orthogonal transforms are deployed for generating feature vectors of video frames. The fractional coefficients of transformed video frames taken as 25 percent, 6.25 percent, and 1.5625 percent of full 100 percent transformed coefficients are considered to form video visual storyboards. The proposed method uses the benchmark video datasets Open Video Project (OVP) and SumMe to validate the performance, containing user summaries (storyboards). These video summaries created using the proposed method are evaluated using percentage accuracy and matching rate.**

***Keywords*—Keyframe; orthogonal transform; VSUMM; video visual storyboard; video summarization**

I. INTRODUCTION

Due to the significant increase in the cases of pushing promotional videos in online drop boxes, Emails, and social network accounts of users, the users are forced to get these videos downloaded to understand the contents of the video. After seeing the video, the user often finds that he is not interested in the whole content. With the easy accessibility of Internet Services and handheld image/video capturing devices, there is a lot increase in the photos and videos in online/offline databases. But it has introduced new challenges in computer vision research, such as storage, search, and navigation, due to the huge volume of video data. There is a critical need to address these problems because of the abundance and accessibility of video data. A video content summarization aims to summarize the full video content in this situation into short video clips or groups of frames that are crucial for understanding video content. This summary is known as a visual video storyboard.

Video content summarization through storyboards enables quick browsing of a collection of sizable video datasets. Additionally, it supports associated video-related tasks like video indexing and retrieval. Nowadays, video summarization has evolved in various applications as a problem of keyframe extraction [1]. But the key frame extraction is very challenging due to the complex nature of the video. Key frame extraction, which substitutes for the most crucial elements of the movie, is one method for producing video summaries/storyboards.

The video storyboard is a quick and meaningful way of giving an abstract perspective on an entire video by creating a video summary[2]. The viewer might not have enough time to see the complete movie. At that time, the storyboards may help users to watch only the important content using these keyframes to narrate the full story of a video. These storyboards can be static [3] or dynamic [4].

Over the past two to three decades, videos have increased. Still, there isn't an ideal system that can handle the time-consuming process of creating a visual video storyboard. So, the indexing, retrieval, and storage of video are affected. All of these video-related concerns can be addressed by video storyboards. The number of approaches proposed for creating video summaries mainly focuses on feature selection techniques used for keyframe selection and evaluation with the ground truth.

The existing video summarization methods divide the whole video into shots and segments. Then it applies the feature selection process as defined in VSUMM [5][6][7], the DT triangulation method for clustering the video frames [8], Local descriptor based and temporal features based [9], diverse color space-based key frame extraction [10].

In VSUMM and DT methods, the video frames are considered in a batch of the first 25-30 frames or by interleaving sequence, thereby not including all the features in video frames. In VSUMM and DT methods, the keyframes are selected only by grouping and distance between the video frames. This leads to the loss of a few important frames in a sequence. This limitation can be overcome by using all the video frames in a video, as stated in the proposed method.

In state-of-art summarization of video, video summaries are dependent on the key frame extraction, and feature selection plays an essential role in this keyframe extraction step. Therefore many researchers have demonstrated different techniques for selecting these features for key frame extraction

[11][12]. Based on the input and output features, video summarization has two different ways: dynamic and static. A dynamic summary of the video is the abstraction of the lengthy video into a compact reel in which the scene is recreated using only keyframes, and a motion is applied [13]. A static summary of the video includes a series of static keyframes that shows the entire story of the video without motion. The choice of static and dynamic is user dependent.

The use of orthogonal transforms assures the full feature in the input frames. Orthogonal transforms are applied with fractional energy coefficients for the various applications of content-based video retrieval with performance measure as precision and recall[14]. The use of transformed features assures high energy compaction; therefore, transformed features are used in video processing. The research work presented here addresses the issue of feature selection in key frame extraction by using the transformed features and proposes a novel video summarization technique with the creation of visual video storyboards. The proposed method first segments the video into video frames, spreading all video content over multiple frames.

In most of the existing static video summarization approaches, the observed limitations are the huge size of feature vectors of video frames, unequal size of feature vectors, suitability of the features for a particular type of video dataset only, and experimental validation is done with a single dataset. Hence there is a need to have the optimal minimum size of a more robust feature vector with the ability to show analogous performance across multiple video datasets.

Depending upon the discussion above, the key contributory significance of the proposed static video summarization method is as follows:

1) The use of fractional energy of transformed video frames to produce a video summary (video visual storyboard).

2) The use of orthogonal transforms to obtain the fractional coefficients of transformed video frames.

3) Performance validation using Open Video Project (OVP) and SumMe benchmark video datasets.

The remaining part of this paper is structured as follows: Section 2 illustrates the current work. The proposed static video summarization method using fractional energy coefficients of transformed video frames is put forth in Section 3, and Section 4 describes the results with the OVP and SumMe dataset and the test bed used for experimentation. The conclusion obtained by thorough investigation and demonstration is summarised in Section 5.

II. Related Work on Static Video Summarization

Lengthy videos have a large sequence of short segments (shots) in video frames; these shots are made up of only the most essential frames (keyframes) that can be used to search and retrieve the original videos. Through these keyframes (storyboards), the videos can be understood easily. According to the literature, transformed features ensure retrieval effectiveness and reduce the calculations required for time-consuming video processing [14]. But these transformed features are extracted for content-based image retrieval, not video retrieval. It does not include a sequence of images in a query. In [15], adaptive threshold-based key frame extraction uses the MPEG -7 color layout descriptors combined with adaptive thresholding. In [16], annotation-based keyframe identification is defined with interest as a key frame identification concept. Both static video frame extraction methods use the actions in a video as a base for the shot selection. This will not apply to all videos; a few videos may be just informative or storytelling. In [17], a review presents different video summarization categories based on features, clusters, shots, and trajectories. But this study concludes with a video summarization of the region of interest problem. Every time human intervention is needed while summarizing the video. So there is a need for an automatic summary generator with minimum computations. The specific feature should be selected with automatic computations for the keyframe selection.

Orthogonal transforms, including Discrete Cosine, Kekre, Walsh, Slant, Discrete Sine, and Discrete Hartley, have been explored for content-based image retrieval (CBIR)[18]. Mean Square Error (MSE) is the similarity metric considered. This method creates an efficient image signature for each image and ensures full input feature selection. But this operation is performed on each distinct image in a dataset. In a video, many similar images, known as near duplicates, will increase the computational overhead; in such cases, using transformed features may reduce the computational complexity. Therefore in the proposed work, different orthogonal transforms are used for storyboard creation in static video content summarization.

But in [18], the transformed features are used for image retrieval; no recreation is performed here with the retrieved image. The proposed video visual storyboard creation method is explained in detail in the following section and generated video storyboards are validated using the novel performance metrics. A brief review of the techniques that support the video summarization is given in Table I.

TABLE I. Related Work Comparison of Video Summarization Techniques

Author List	Type of Features used	Dataset	Performance
Xiang et al. [16] 2020	ConvNet	VSUMM	F-score (72.1%)
Rukiye et al. [19] 2021	CNN & RNN	UCF 101	Accuracy (67.39%)
Vijay Kumar et al. [20] 2014	Discrete Wavelet Transform, Haar Wavelet based	Sports Video	Precision (0.83)
Naveed et.al. [21] 2013	Discrete Cosine Transform	Open Video Project	F-Measure (82%)
Kavitha et al. [22] 2015	Discrete Wavelet Transform	Open Video Project	F1-Score (87%)
Ajay Narvekar et al. [23] 2013	Discrete Cosine Transform	Online videos	Precision (0.78)

The work presented in [19][20][21][22], briefly compared in Table I, clearly indicates that the transformed feature gives more precision and recall than the cluster-based approach, along with the reduction in computational costs since kernel size varies. It provides a quick review of existing methods where the orthogonal transforms are used in video summarization, and CNN performance is also compared with recent work. This has given the motivation for selecting the orthogonal transforms to prove the efficiency for static video content summarization, i.e., creating the video visual storyboards.

III. Proposed Method of Static Video Summarization using Orthogonal Transforms

The proposed method of creating feature vectors for static video summarization uses fractional energy coefficients of transformed features to make a video visual storyboard. The proposed framework is shown in Fig. 1.

The Proposed framework has three steps: first, to form a transformed feature vector of all video frames using the 'T' transform; the second step is to prepare a feature vector using fractional energy coefficients; the third step is to select the number of keyframes. All these three steps are pictorially represented in Fig. 1. The above steps are elaborated in the following subsections.

A. Orthogonal Transform

Different orthogonal transforms used in this proposed method are discussed here. The 'T' transforms used in this system are defined in the form of their matrix equations.

1) Discrete Cosine Transform (DCT)

The Discrete Cosine Transform is the most widely used orthogonal transform in image processing. The *N x N* cosine transform matrix is defined as below in equation (1),

$$c(p,n) = \begin{cases} \dfrac{1}{\sqrt{N}} & p = 0, 0 \le n \le N-1 \\ \sqrt{\dfrac{2}{N}\cos\dfrac{\pi(2n+1)p}{2N}} & 1 \le p \le N-1, 0 \le n \le N-1 \end{cases} \tag{1}$$

2) Slant Transform

The Slant transform is a constant function with a one-row function and the second row is a linear function of the column index. It includes sparse matrices, reducing the computations and leading to a fast process.

The matrix equation of the Slant transform is given by equation (2),

$$S_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \tag{2}$$

3) Walsh Transform

The set of N rows denoted as W_k, for *k*= 0, 1, ……, N-1, is defined in Walsh Matrix that has distinct properties.

W_k takes on the values +1 or -1.

W_k [0]= 1 for all *k*.

W_k x W_{lt} =0, $k \neq l$ and W_k x W_{lt} has exactly *k* zero crossings, for *k =0,1, ……., N-1*.

Each W_k is either even or Odd.

4) Kekre Transform

This transform matrix is an *NxN* matrix, where the upper diagonal values are one, and the diagonal values of Kekre's transform matrix are also one, except other values below the diagonal is zero.

This matrix equation is defined using the Hadamard matrix of order N in equation (3),

$$K_{xy} = \begin{Bmatrix} 1 & , x \le y \\ -N + (x+1) & , x = y+1 \\ 0 & , x > y+1 \end{Bmatrix} \tag{3}$$

B. Feature Vector Extraction

Each video frame is resized to *256 x 256*. On each color plane video frame of size *NxN*, the 'T' Transform (alias DCT, Walsh, Slant, and Kekre Transform) is applied to extract the visual feature vector of size *NxN* as a full or 100% energy content scenario as shown in Fig. 1.

The fractional energy coefficients are computed by dividing the full features of the video frame into block sizes of 32 x 32, 64 x 64, and 128 x 128 and are taken as top left-hand side coefficients of transformed color planes of video frames, as shown in Fig. 2.

In this proposed method, the transformed coefficients are used to form the feature vector. In [24], transformed coefficients as features have shown better accuracy in the keyframe extraction. The proposed method uses these transformed video frame coefficients with a reduced number of feature vector elements.

C. Feature Vector Database using Fractional Coefficients

The proposed video visual storyboard generation method for static video summarization uses fractional energy coefficients. The diagrammatic representation of extracting fractional energy coefficients to generate feature vectors from a transformed video frame is shown in Fig. 2.

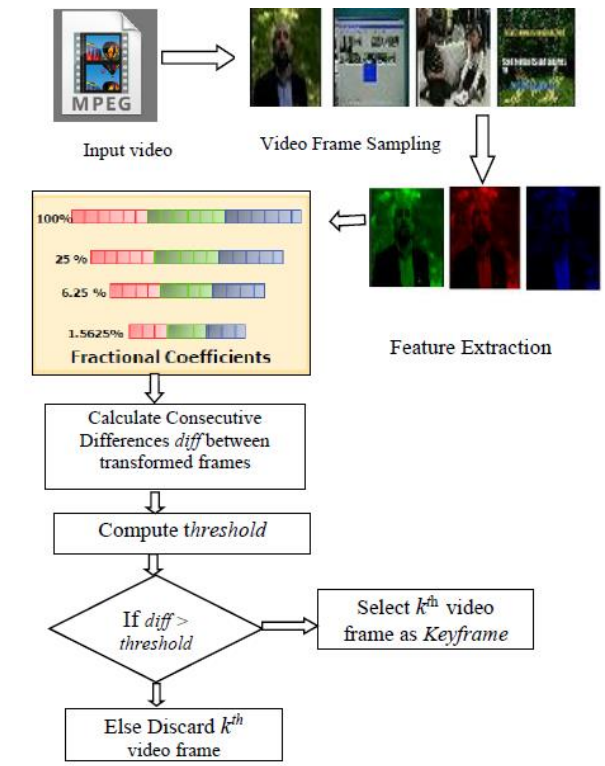

MPEG
Input video
Video Frame Sampling
Feature Extraction
100%
25 %
6.25 %
1.5625%
Fractional Coefficients
Calculate Consecutive Differences *diff* between transformed frames
Compute *threshold*
If *diff* > *threshold*
Select k^{th} video frame as *Keyframe*
Else Discard k^{th} video frame

Fig. 1. The Proposed Method of KeyFrame Extraction with Fractional Energy Coefficients of Transformed Video Frames for Visual Storyboard Creation.

If the video frame is of size *256x256*, the fractional energy coefficient proportions are taken as 25%, 6.125%, and 1.5625%, respectively, with sizes 128x128, 64x64, and 32x32. Considering high energy coefficients as feature vectors reduces feature vector size, time, and computational complexity.

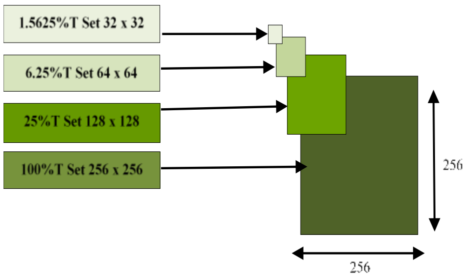


Fig. 2. Proposed Feature Vector Extraction with Fractional Energy Coefficient.

D. The Decision of Keyframe for Video Visual Storyboard

The keyframes are the significant frames that contain the maximum information in a video frame. These keyframes are selected based on the consecutive differences between two transformed video frames. These consecutive differences are then compared with the certain constant threshold calculated with standard deviation and mean. Each transformed video frame in the sequence is associated with the difference; if this difference is above the threshold, then that particular frame is selected as a keyframe; otherwise, it is discarded. In this process of keyframe selection, there might be a possibility of selecting near duplicates along with essential frames. These near duplicates are eliminated by interleaving manually or statistically.

TABLE II. VIDEO DATASET DETAILS CATEGORY WISE

Open Video Project Video Dataset			
Category	*Documentary*	*Educational*	*Lecture*
*No. of videos	15	20	20
Category	*Historical*	*Public Service*	*Ephemeral*
*No. of videos	20	20	15
Total 110 videos			
SumMe Video dataset			
Total 25 videos			

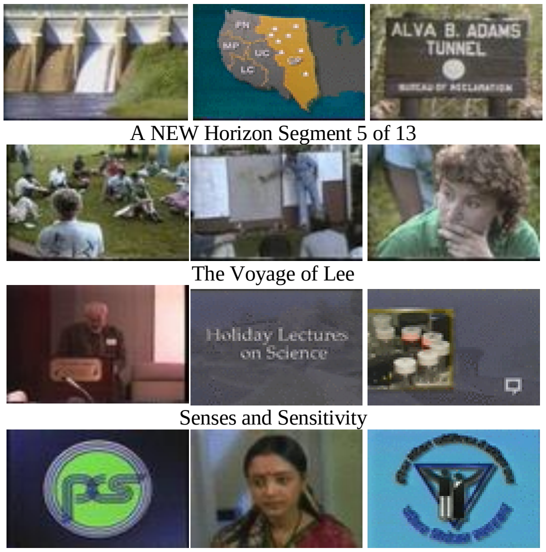
IV. RESULTS AND DISCUSSION

The implementation details and results are discussed in this section.

A. Experimental Video Testbed

Two video datasets, the Open Video Project (110 videos) and SumMe (25 videos) of various categories, are used for proposed experimentation. The videos are provided in both compatible file types (MPEG-2, MPEG-4). These video datasets are openly accessible. The OVP contains the categories of Lectures, Television, Demonstrations, and Documentary videos. The SumMe dataset includes videos of categories like Cooking, Bike polo, Base jumping, etc.

These two benchmark datasets provide video user summaries as visual storyboards for respective videos [21]. These user summaries are further used for performance comparison to evaluate the proposed static video summarization technique for creating video visual storyboards. A few video frames are shown in the following Fig. 3, with a few sample frames from the OVP and SumMe datasets. Each video length varies from less than 1 minute to 2 minutes.



A NEW Horizon Segment 5 of 13

The Voyage of Lee

Senses and Sensitivity

a) OVP- Family TV Spots around the World

b) SumMe – AirForce, Base jumping, fire Demo

Fig. 3. Few Video Frames from a) OVP Dataset and b) SumMe Dataset.

Table II shows the number of videos considered from the OVP and SumMe datasets for the experimentation of the proposed method with respective categories. The type of video in the Lecture category has slow transitions in the scene, leading to fewer keyframes in the final output of the video visual storyboard. The sudden changes in the scene will affect the number of keyframes extracted. The proposed experimentation testbed includes all types of videos with such variations.

B. *Results and Discussion*

The experimentation is performed using the above videos shown in Fig. 3. The proposed method extracts the keyframes from each video from the dataset. The obtained keyframes are compared with the already existing given ground truth in the OVP and SumMe datasets. The performance metrics matching rate and percentage accuracy are calculated to identify the exact matching frame from the given set of keyframes in the OVP and SumMe video dataset storyboard.

The number of keyframes extracted using the fractional energy coefficients of transformed video frames is evaluated using the given ground truth of videos from the OVP and SumMe datasets.

1) Performance Metric

The percentage accuracy is calculated as the ratio of the number of correctly extracted keyframes by the proposed method and the total number of keyframes given in the standard user storyboard.

In the matching rate, the matching from the given summaries with frame numbers given in ground truth is done with the keyframes obtained using the proposed method. The matching rate is calculated as the number of identical matching video frames similar to the keyframes given in the OVP video user summary. Here it is assumed that keyframes in the user summary are provided with the frame numbers from the original video frame sequence.

The performance metrics used in the proposed system are explained in equations (4) and equation (5).

The keyframes in the given OVP summary are downloaded from https://openvideo.project.com. The user summary from OVP and SumMe datasets are compared with a set of keyframes obtained through the proposed system.

The keyframes obtained are used to create a visual video storyboard. The results are summarized in Tables III and IV for the orthogonal transform using fractional energies with OVP and SumMe videos.

The percentage accuracy and matching rate are given in Tables III and IV with the detailed analysis of the proposed fractional energy-based keyframe extraction method using orthogonal transform alias DCT, Walsh, Slant, and Kekre transform.

The results show that the performance improves in the case of the proposed use of fractional energy coefficients compared to the consideration of 100% coefficients. This reduction in the sizes of the different feature vectors in the proposed method improves the accuracy of video visual storyboard creation by finding more accurate keyframes.

TABLE III. PERCENTAGE ACCURACY AND MATCHING RATE OF PROPOSED FRACTIONAL ENERGY COEFFICIENTS BASED ON KEYFRAME EXTRACTION METHOD FOR RESPECTIVE ORTHOGONAL TRANSFORMS EXPERIMENTED ON OVP DATASET

Performance using OVP dataset				
Fractional energy coefficients ▶	100%	25%	6.25%	1.526%
Discrete Cosine Transform (DCT)				
% Accuracy	76.23	75.31	75.06	76.28
Matching Rate	19.2	20	20.2	20.2
Walsh Transform				
% Accuracy	73.45	73.49	72.25	73.63
Matching Rate	16.63	17.24	17.48	16.78
Slant Transform				
% Accuracy	70.19	71.59	70.36	71.45
Matching Rate	15.82	15.24	15.63	15.89
Kekre Transform				
% Accuracy	69.38	70.39	70.72	70.49
Matching Rate	14.92	15.24	15.78	15.58

The above performance comparison, as shown in Table III and Fig. 4 indicates that the results of DCT based proposed keyframe extraction method outperform when compared to the other Walsh, Slant, and Kekre orthogonal transform-based keyframe extraction.

$$Percentage\ Accuracy = \frac{Number\ of\ correctly\ extracted\ keyframes}{Expected\ number\ of\ keyframes} \tag{4}$$

$$Matching\ Rate = \frac{Exact\ matching\ extracted\ keyframes}{Total\ number\ of\ keyframes} \tag{5}$$

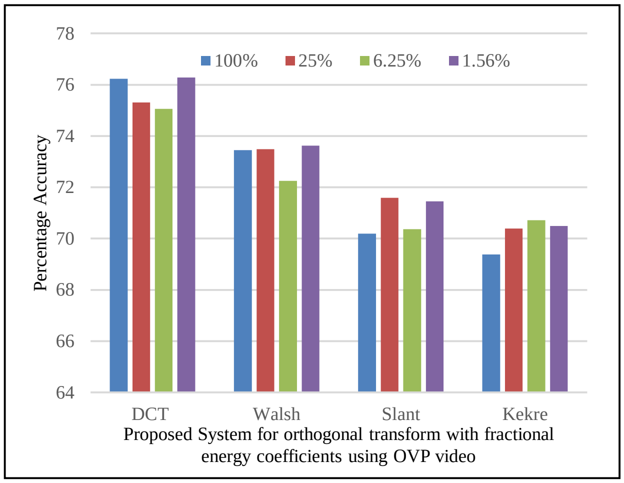


Fig. 4. Performance Comparison of the Proposed Fractional Energy Coefficients based Video Keyframe Extraction Method for Respective Orthogonal Transforms Experimented on OVP Dataset.

A similar method applies to other 'T' orthogonal transforms with fractional energy coefficients. The reduction in the feature vector is not affecting the percentage accuracy but increases the selection of keyframe matching to the given keyframe from OVP visual storyboards. Table IV and Fig. 5 show the analysis of the performance obtained by the proposed method using SumMe videos.

TABLE IV. PERCENTAGE ACCURACY AND MATCHING RATE OF PROPOSED FRACTIONAL ENERGY COEFFICIENTS BASED ON KEYFRAME EXTRACTION METHOD FOR RESPECTIVE ORTHOGONAL TRANSFORMS EXPERIMENTED ON SUMME DATASET

Performance using SumMe dataset				
Fractional energy coefficients→	100%	25%	6.25%	1.526%
Discrete Cosine Transform (DCT)				
% Accuracy	75.56	74.76	74.33	73.59
Matching Rate	18.75	17.9	17.9	17.5
Walsh Transform				
% Accuracy	73.85	73.49	72.25	73.63
Matching Rate	16.93	17.24	17.48	16.78
Slant Transform				
% Accuracy	72.56	72.03	73.14	73.24
Matching Rate	16.21	15.89	16.23	16.49
Kekre Transform				
% Accuracy	73.45	73.35	74.23	73.47
Matching Rate	16.16	16.46	15.49	16.38

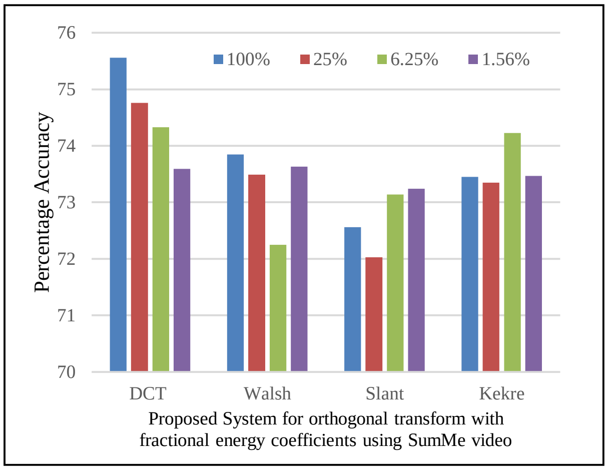
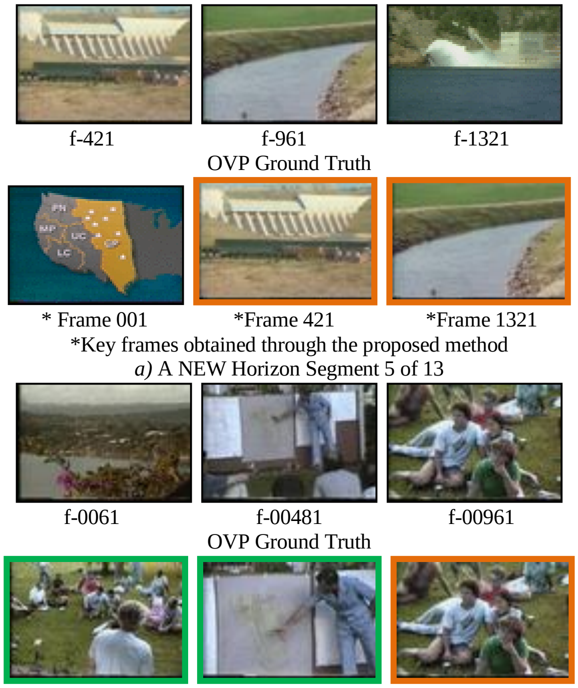


Fig. 5. Performance Comparison of the Proposed Fractional Energy Coefficients based Video Keyframe Extraction Method for Respective Orthogonal Transforms Experimented on SumMe Dataset.

2) Significance of the Proposed Method

The proposed method generates video summaries with a few keyframes displayed below in Fig. 6(a) and 6(b) to support the performance metrics discussed in this paper. The similarity can be compared with the frame numbers similar to the OVP and SumMe storyboards.



f-421 f-961 f-1321
OVP Ground Truth

* Frame 001 *Frame 421 *Frame 1321
*Key frames obtained through the proposed method
a) A NEW Horizon Segment 5 of 13

f-0061 f-00481 f-00961
OVP Ground Truth

*Frame 361 *Frame 781 *Frame 961
*Key frames obtained through the proposed method
b) The Voyage of Lee

Fig. 6. Video Frames of a Video – a) A NEW Horizon Segment 5 of 13 b) The Voyage of Lee (Highlighted Keyframes Match with OVP Storyboard).

The highlighted frames, as shown in Fig. 6 above, are the ones that are used to calculate the matching rate. The keyframes whose frame numbers match the given keyframes in the OVP storyboard are highlighted.

The relationship between the feature vector computation and the energy reduction coefficients utilized in this implementation is shown in Table V. The 6.25% feature vector space reduces the whole feature vector by 93.75% and gives similar percentage accuracy.

TABLE V. COMPARISON OF PROPOSED % FRACTIONAL ENERGY COEFFICIENTS AND REDUCTION IN % FEATURE VECTOR SIZES

Feature Vector Size	% fractional Energy Coefficients	% reduction in Feature Vector Size
$N\,x\,N\,x3$	100	0
$\frac{N}{2}x\frac{N}{2}x\,3$	25	75
$\frac{N}{4}x\frac{N}{4}x\,3$	6.25	93.75
$\frac{N}{8}x\frac{N}{8}x\,3$	1.5625	98.4375

The proposed method here is effective for storyboard generation as compared to other techniques DT [8] and VSUMM [7] in terms of computations required to process video frames. Here the dimensionality of each feature vector is reduced due to the use of fractional energy coefficients.

1) *Comparison with other Techniques*

This section compares the proposed system with existing techniques like DT [8] and OVP summary. The experiments were performed on videos downloaded from OVP and SumMe. These summaries are evaluated in percentage accuracy and compared with existing VSUMM and DT Summary ground truth. The same comparison of the proposed method performance is made with VSUMM and DT summaries using the performance metric as percentage accuracy. Tables VI and VII show this comparison.

Fig. 7 below shows the visual static storyboard comparison between OVP ground truth, VSUMM, and DT summary with the static summary obtained through the proposed method of static video summarization in the form of a set of keyframes extraction for storyboard creation. Fig. 7 shows the highlighted keyframes that match the ground truth and user summary.

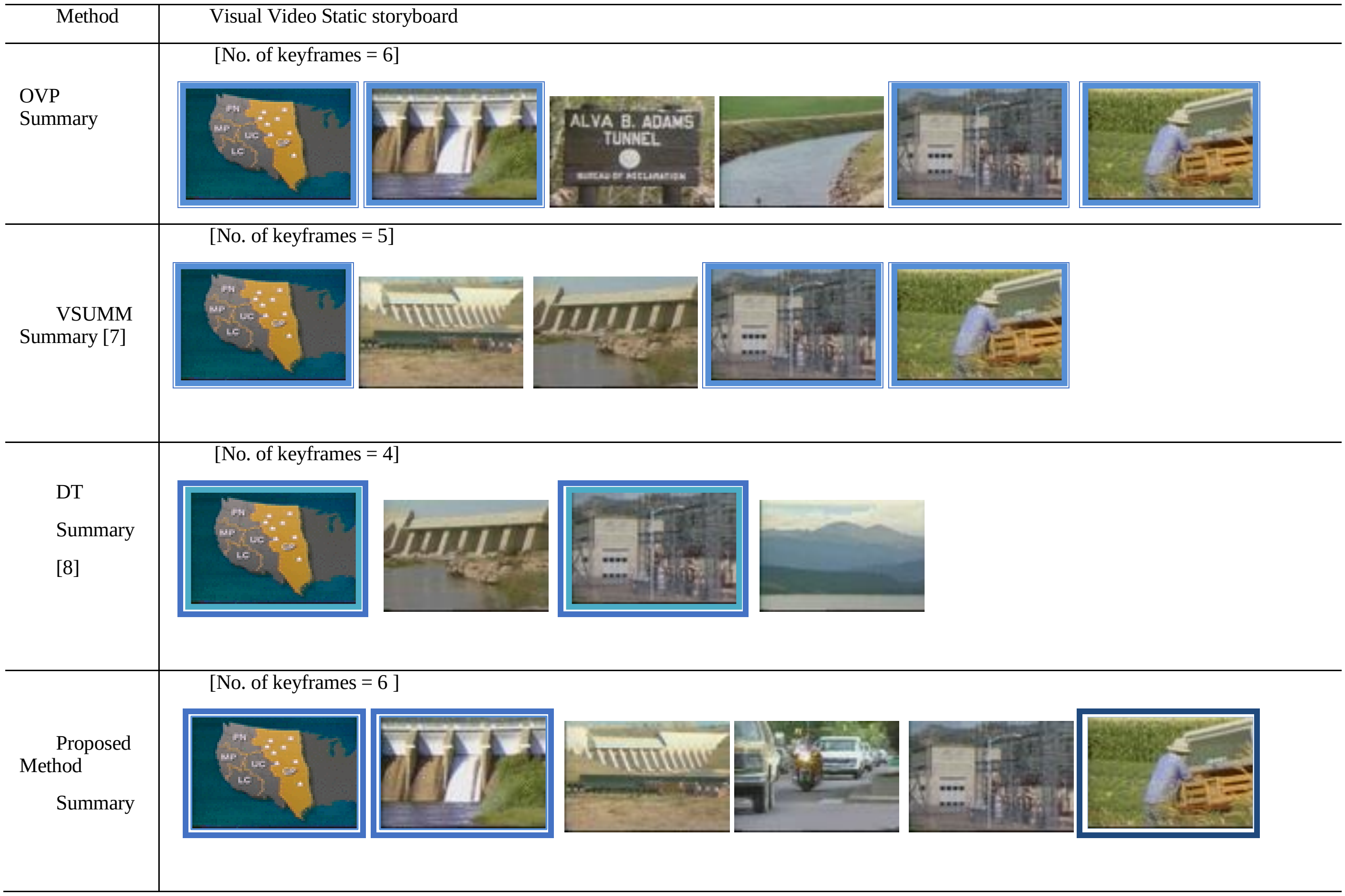


Fig. 7. Comparison of Video Storyboard Created by the Proposed Method Versus DT, VSUMM, and OVP Storyboards.

Table VI shows that the proposed system's performance is better than the DT summary using OVP videos and closer to the VSUMM summary.

TABLE VI. PERFORMANCE COMPARISON OF THE PROPOSED FRACTIONAL ENERGY COEFFICIENTS BASED VIDEO KEYFRAME EXTRACTION METHOD WITH DT[8] AND OVP VIDEO SUMMARIES EXPERIMENTED ON THE OVP DATASET

Name of the Video from (OVP)	Video length in Seconds	Percentage Accuracy (%)		
		VSUMM [7]	DT [8]	Proposed
A New Horizon Segment 5	1.59	78.57	42.86	**78.57**
A New Horizon Segment 6	1.05	80.00	60.00	60.00
A New Horizon Segment 7	0.47	87.50	75.00	**75.00**
The Voyage of Lee 15 of 21	1.15	75.00	50.00	**83.33**
The Voyage of Lee 16 of 21	1.27	71.43	42.86	42.86
The Voyage of Lee 17 of 21	2.25	76.92	61.54	46.15
Average		78.24	55.38	64.32

Comparatively, the VSUMM summaries are closer to the OVP summary, and second next better accuracy is provided by our proposed method of creating a visual storyboard. The proposed storyboard generation method performs better than the Delaunay triangulation method. The same comparison of the proposed method performance is made with VSUMM and DT summaries using SumMe videos.

Table VII shows that the proposed system's performance is better than the DT summary using SumMe input videos.

TABLE VII. PERFORMANCE COMPARISON OF THE PROPOSED FRACTIONAL ENERGY COEFFICIENTS BASED ON VIDEO KEYFRAME EXTRACTION METHOD WITH DT [8] AND SUMME VIDEO SUMMARIES EXPERIMENTED ON SUMME DATASET

Name of the Video (SumMe)	**Video length in Seconds**	**Percentage Accuracy (%)**		
		VSUM M [7]	**DT [8]**	**Proposed**
Air_Force_One	2.59	68.42	68.42	47.37
Base Jumping	2.38	78.26	47.83	52.17
Bike Polo	1.43	83.33	50.00	55.56
Cooking	1.26	84.62	61.54	**76.92**
Scuba	1.14	55.56	55.56	**77.78**
Fire Demo	0.54	66.67	66.67	66.67
Average		72.81	58.33	62.74

So the major contribution of the proposed system is that it overcomes the problem of inclusion of near duplicates due to Delaunay triangulation of clustering. Instead of that proposed system, select the keyframes from all video frames using fractional energy coefficients. In the DT method, summaries are produced for a batch of videos, whereas the proposed system processes each video one by one to include all features giving significance.

V. CONCLUSION

Video summarization faces the challenge of reducing computational complexity and retrieval accuracy due to its complex nature. This work focuses on reducing visual video frame features for static video summarization, and a new method is proposed for creating a video visual storyboard using transformed visual features with fractional energy coefficients. The transformed coefficients of color planes of the video frames are considered for finding the final feature vector as the set of fractional energy coefficients 25%, 6.25%, and 1.5625% of total coefficients using transforms alias DCT, Slant, Walsh, and Kekre. These features are used for keyframe extraction, and a set of extracted keyframes forms a video visual storyboard.

The average percentage accuracy obtained by the proposed system is 72.51 with the OVP dataset and 73.55 with the SumMe dataset. The keyframes obtained through the proposed system match with the given set of keyframes in the OVP and SumMe dataset videos. The percentage accuracy and matching rate using fractional energy coefficients are higher than using complete 100% energy coefficients used in the existing DT and VSUMM Summary. The keyframe selection done with the proposed use of fractional energy coefficients of transformed video frames for creating a video visual storyboard is better than the use of full energy content, proving the worth of the proposed method. This same method can be further extended for creating the video logs for video storage and indexing as future scope.

ACKNOWLEDGMENT

I sincerely thank my Ph.D. Guide and mentor Dr. Sudeep D. Thepade (Professor, Computer Engineering, PCCOE Pune and BOS Computer Engineering, Member SPPU Pune) for his valuable guidance in computer vision.

REFERENCES

[1] B. T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification," ACM Trans. Multimed. Comput. Commun. Appl., vol. 3, no. 1, pp. 1–37, 2007.

[2] K. Schoeffmann and O. Marques, "A Novel Tool for Quick Video Summarization using Keyframe Extraction Techniques A Novel Tool for Quick Video Summarization using Keyframe Extraction Techniques," no. March 2009.

[3] M. Furini, F. Geraci, M. Montangero, and M. Pellegrini, "STIMO: STIll and MOving video storyboard for the web scenario," Multimed. Tools Appl., vol. 46, no. 1, pp. 47–69, 2010.

[4] M. Gygli, Y. Song, and L. Cao, "Video2GIF: Automatic generation of animated GIFs from video," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 1001–1009, 2016.

[5] S. E. F. De Avila, A. Da Luz, A. D. A. Araújo, and M. Cord, "VSUMM: An approach for automatic video summarization and quantitative evaluation," Proc. - 21st Brazilian Symp. Comput. Graph. Image Process. SIBGRAPI 2008, no. June 2014, pp. 103–110, 2008.

[6] S. E. F. De Avila and A. D. A. Araujo, "VSUMM : An Approach Based on Color Features for Automatic Summarization and a Subjective Evaluation Method," XXII Brazilian Symp. Comput. Graph. Image Process. SIBGRAPI, p. 10, 2009.

[7] S. E. F. De Avila, A. P. B. Lopes, A. Da Luz, and A. De Albuquerque Araújo, "VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method," Pattern Recognit. Lett., vol. 32, no. 1, pp. 56–68, 2011.

[8] P. Mundur, Y. Rao, and Y. Yesha, "Keyframe-based video summarization using Delaunay clustering," Int. J. Digit. Libr., vol. 6, no. 2, pp. 219–232, 2006.

[9] E. J. Y. C. Cahuina and G. C. Chavez, "A new method for static video summarization using local descriptors and video temporal segmentation," Brazilian Symp. Comput. Graph. Image Process, pp. 226–233, 2013.

[10] S. D. Thepade, "Diverse Color Spa aces in Video Keyframe Extraction Technique using g Thepade' s Sorted Ternnary Block Truncation Coding with Assorted Similarity Measures," no. GCCT, pp. 256–260, 2015.

[11] M. Kogler, M. Del Fabro, M. Lux, K. Schöffmann, and L. Böszörmenyi, "Global vs. Local feature in video summarization: Experimental results," CEUR Workshop Proc., vol. 539, no. December, pp. 108–115, 2009.

[12] S. D. Thepade and A. A. Tonge, "Extraction of Key Frames from Video using Discrete Cosine Transform," in In International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014, pp. 1294–1297.

[13] S. SARMADI, "New Approach In Video Summarization Based On Color Feature," vol. 86, pp. 535–542, 2017.

[14] S. Gupta, "Content-Based Video Retrieval in Transformed Domain using Fractional Coefficients," no. 7, pp. 237–247.

[15] S. Cvetkovic, M. Jelenkovic, and S. V. Nikolic, "Video summarization using color features and efficient adaptive threshold technique," Prz. Elektrotechniczny, vol. 89, no. 2 A, pp. 247–250, 2013.

[16] X. Yan, S. Z. Gilani, M. Feng, L. Zhang, H. Qin, and A. Mian, "Self-supervised learning to detect key frames in videos," Sensors (Switzerland), vol. 20, no. 23, pp. 1–18, 2020.

[17] H. Burhan Ul Haq, M. Asif, and M. Bin Ahmad, "Video Summarization Techniques: A Review Article in," Int. J. Sci. Technol. Res., vol. 9, no. 11, pp. 146–153, 2021.

[18] H. B. Kekre, "Comprehensive Performance Comparison of Cosine, Walsh, Haar, Kekre, Sine, Slant, and Hartley Transforms for CBIR with Fractional Coefficients of Transformed Image," no. 5, pp. 336–351, 2011.

[19] R. Savran Kızıltepe, J. Q. Gan, and J. J. Escobar, "A novel keyframe extraction method for video classification using deep neural networks," Neural Comput. Appl., vol. 0123456789, 2021.

[20] V. K. D, S. K. K. L, and J. Majumdar, "Comparison of Video Shot Detection and Video Summarization Techniques," vol. 3, no. 8, pp. 829–833, 2014.

[21] N. Ejaz, I. Mehmood, and S. Wook Baik, "Efficient visual attention based framework for extracting key frames from videos," Signal Process. Image Commun., vol. 28, no. 1, pp. 34–44, 2013.

[22] J. Kavitha and P. A. J. Rani, "Static and multiresolution feature extraction for video summarization," Procedia Comput. Sci., vol. 47, no. C, pp. 292–300, 2015.

[23] A. A. Narvekar, B. E. Student, and B. E. Student, "Color Content-Based Video Retrieval Using Discrete Cosine Transform Applied On Rows and Columns of Video Frames with RGB Color Space," vol. 2, no. 11, pp. 133–135, 2013.

[24] N. Yadav, "Comprehensive Performance Comparison of Energy Compaction Techniques for CBVR with Transformed Videos," no. 07, pp. 1–7, 2016.