International Journal of Advanced Science and Technology
Vol. 29, No. 8, (2020), pp. 320-340

KFE-OPPSO: A New Key Frame Extraction Technique with Orthogonal Polynomials and Particle Swarm Optimization

Braveen M[1], Krishnamoorthy R[2]

[1]*Assistant Professor, School of Computer Science and Engineering,*
Vellore Institute of Technology Chennai, India
[2]*Professor, Department of Information Technology, University College of Engineering,*
BIT Campus, Anna University, Tiruchirappalli, India
[1]*braveenmani@hotmail.com*

Abstract

In this paper, a new Key frame Extraction technique with Orthogonal Polynomials and Particle Swarm Optimization (KFE-OPPSO) is proposed. The input video file is converted in to frames and is applied with orthogonal polynomials to derive transform coefficients. These transformed coefficients are then utilized to detect edges in a frame by identifying zero crossings in second directional derivatives. These edge detected frames are then applied morphological operation "dilate" followed with "invert" operation to yield clear and visible edges. Further, an edge movement in video frames is performed between current inverted frame and previous edge detected frame. Finally, these frames are compared with each other to detect different types of shots such as hard cuts, fade and dissolve effects in terms of Edge Change Ratio (ECR). From the detected shots, color feature is extracted with few orthogonal polynomials coefficients and then fed to Particle Swarm Optimization (PSO) for selecting optimal key frames. Experiments are carried out on 20 different videos downloaded from the standard open video project. The experimental results indicate that the proposed scheme achieves better precision, recall and f-measure rate than other existing systems in terms of key frames selection.

Keywords: *Key Frames, Orthogonal Polynomials, Particle Swarm Optimization*

1. Introduction

Now a day, the growth and usage of multimedia information is inevitable. The multimedia information may contain text, image, graphic, audio and video. Handling video information is the most challenging one, since it combines all the other information to form a single stream [1]. Therefore, accessing a video is difficult due to its length. So, need of representing video in terms of key frames became important. A video file comprises of three frames namely: Intra (I), Prediction (P) and Bi-directional (B). For easy access, video file is initially partitioned in to number of segments from which key frames are selected. These key frames provide a abstract information of the video that is highly useful for video indexing and retrieval. Treating intra frames of a video file as key frame is the simplest method followed in early days. The other common method of selecting key frames is performed by comparing consecutive frames in a shot with histogram matching [2]. Researchers paid attention towards selection of representative frames in both spatial and transformed domain. The selection of representative frames from a shot has been felt difficult, since the consecutive frames in a video shot sequence contains less variation.

2. State-of-the Art

Several key frame selection techniques to accelerate retrieval systems are found in the literature. Yannis S. Arvithis et al. [3] utilized low level motion and color information to form a feature vector. Key frames are then selected by identifying the variation in the feature vector. In addition to that, Genetic Algorithm has been utilized to minimize cross-correlation of video frames. Yueting Zhuang et al. [4] reported a key frame extraction technique with unsupervised clustering and feedback adjustment process. The clustering is performed based on the visual features such as color, texture and

shape are combination of the above. Eung Kwan Kang et al. [5] extracted Direct Coefficients (DC) to form image sequences from standard Motion Picture Experts Group (MPEG) videos and utilized accumulated histogram intersection technique to select key frames by matching with neighbor frames. Calic J et al. [6] partitioned video file into shots based on the macro blocks derived from MPEG stream and extracted key frames from the shots using difference metric with discrete contour evolution algorithm. Hun-Cheol Lee et al. [7] suggested a key frame selection scheme that follows sequential order to select set of key frames initially and then it alters the point to select the next set of key frames with time intervals. It is followed iteratively, until the final set of key frames is selected.

Tianming Liu et al. [8] developed a motion energy model that utilizes motion as a salient feature to select key frames. This motion energy model combines intensity, characteristics and dominance as a triangle. With this triangular model, the patterns are generated to segment video into shots with respect to accelerations and decelerations of motion. The frame which has motion pattern at the top of triangular model is identified as key frame. Xu-Dong Zhang et al. [9] segmented video into number of shots from which several key frames are selected with dynamic selection mechanism. Clustering is applied to categorize similar and dissimilar frames separately. Tieyan Liu et al. [10] used shot reconstruction degree as a criterion for selecting key frames. It is noted that, better the shot reconstruction through interpolation results in better key frames. Since, it maintains motion dynamics as a key point throughout selection process, optimal solution is achieved.

Markos Mentzelopoulos et al. [11] calculated entropy difference among the consecutive frames for extracting the key frames to represent a video file. Jiawei Rong et al. [12] utilized inter shot information to select key frames. Two approaches viz. calculating the difference between consecutive frames and thresholding are followed. Inter-shot and Intra-shot visual cues play a significant role here in selecting key frames. Ki Tae Park et al. [13] reported an algorithm to find key frames from significant candidate frames that are available in a video. It computes the difference among all the frames and with the help of distortion rate, the algorithm selects key frames. But, this technique consumes more time since it has to compare each and every frame of a video. Jian-Quan Ouyang et al. [14] used camera motion as a parameter to identify the key frames. This interactive scheme adopted Broyder-Fletcher-Goldfarb-Shano (BFGS) to optimize the estimation parameters that ensure proper selection of key frames.

Zhao Guang-Sheng et al. [15] computed difference among the frames with different weights. The difference value is used to detect shots and to select key frames by having first frame as a reference frame. Guozhu Liu et al. [16] segmented video into shots by histogram matching technique. From the shots, features are extracted in I, P, B frames and compared. The frame with higher most difference exceeding a threshold limit is treated to be a key frame. Hua Man et al. [17] reported the importance of clustering while multi features are extracted and represented to compare with consecutive frames. Since, each feature has its own significance this technique also focused on assigning weights for features. In this way, the frame that is very nearest to cluster center is selected as key frame. Pascal Kelm et al. [18] extracted key frames from video based on motion and camera operations. At first, videos are segmented into shots by detecting hard cuts and gradual transitions. Motion with camera zoom and pan are combined to generate weighted attention. The weighted attention curve identifies the key frame by comparing consecutive frames in a shot.

Magda B. Fayk et al. [19] reported a video abstraction scheme based on Particle Swarm Optimization (PSO). Initially, the video file is divided into segments based on time slot. From the segments, color feature is extracted. Particle Swarm Optimization utilizes color feature to identify optimal key frames both locally and globally in a video. Segmenting video file in terms of time slot will reduce computational complexity but, false selection of key frames may occur. Huiyu Zhou et al. [20] extracted audio-visual features for video summarization. Dissimilarity matrix is generated with color, motion and audio feature to compare consecutive frames for segmenting videos. Fuzzy-c means algorithm is utilized to cluster similar kind of frames based on the visual features. Gwo-Cheng Chao et al. [21] tracked multi objects and extracted foreground objects to generate a still frame termed as augmented key frame. It is augmented with representative objects, important contents like face,

license plate and motion in terms of icons. The main drawback of this system is that, representative frames can be generated only for the videos that contains moving objects. Suet-Peng Yong et al. [22] reported wild life key frame extraction scheme. The frames are segmented in to blocks to extract color and texture features with co-occurrence matrix. Semantic context feature is also extracted to monitor the sequential changes. Finally, one class classifier is deployed to extract the key frames. Naveed Ejaz et al. [23] extracted multiple features from video to estimate the frame difference between consecutive frames. It is noted that, feature significance differs for different genres. Hence, this technique assigns weights for different frames by utilizing Relevance Feedback mechanism indirectly. Color histogram, correlation and edge orientation are calculated to find difference among the frames.

Gentao Liu et al. [24] adopted SIFT algorithm to compute the visual content discontinuity values of a video frame. This technique identifies shot by applying double threshold on SIFT values and key frames are selected. Liujun Liu et al. [25] adopted Genetic Algorithm to select the representative frames. As a first step, the video file is segmented into shots and frames that are treated as initial population. With respect to fitness function, each individual frame in the population is compared to choose the key frames. Sun Shumin et al. [26] extracted color feature and utilized Artificial Fish Swarm Algorithm to generate self-organized cluster. Further, to optimize the cluster result, k-means clustering is deployed.

Naveed Ejaz et al. [27] combined the low-level features that are extracted from three color channels (R, G and B), histograms and moments of inertia to identify the candidate frames. Since, three different measures are fused to compare the consecutive frames, time consumption is very high and in case of multiple shots, redundancy may also occur. Jie-Ling Lai et al. [28] constructed static and dynamic conspicuity maps by calculating intensity and orientation along with color information. These maps are combined with fusion model to generate the key frames. Walid Barhoumi et al. [29] summarized video with the help of key frames. The key frames are identified from shots and object based event detection techniques. In addition, fuzzy segment is adopted to find the histogram similarities among the video images. But, it fails in selecting optimal key frames to represent the video. Naveed Ejaz et al. [30] fused both spatial and temporal saliency to draw an attention curve. Spatial saliency has been achieved by adopting Discrete Cosine Transformation [DCT]. Temporal gradients are used to improve the performance in selecting key frames, since optical flow features failed by missing some essential frames. Guang-Hua Song et al. [31] utilized average histogram to extract key frames from the available shots in a video file. It compares the neighbor frames with respect to the fixed threshold value and hence, there is a high possibility of eliminating some of the key frames. Further, sufficient feature descriptors to extract features are also not reported. Qing Xu et al. [32] segmented video into shots and then possibly sub shots to select the key frames. This technique uses three types of divergences such as Jensen-Shannon Divergence (JSD), Jensen-Renyi Divergence (JRD), and Jensen-Tsallis (JTD) to measure the difference between frames in terms of histograms. The video frame that attains a higher divergent value is selected as a key frame.

Since, key frame is a basic building block in designing a video retrieval system, the selection of key frames must be significant, compact and effective. But, most of the approaches follow histogram based comparison of the consecutive frames that too in spatial domain. Hence, a new key frame extraction technique in orthogonal polynomials transform domain that extracts color feature which is further subjected to Particle Swarm Optimization for selecting optimal key frames is proposed in this paper.

The rest of the paper is organized as follows. A brief introduction about the orthogonal polynomials model for the purpose of frequency domain feature extraction is revisited in section 3. In section 4, the architecture of the proposed key frame selection technique with orthogonal polynomials and Particle Swarm Optimization is presented. In sub section 4.1 and 4.2, shot detection scheme with orthogonal polynomials model and details about the proposed Edge Change Ratio (ECR) are given. The extraction of color feature and representation in terms of orthogonal polynomial coefficients is presented in sub section 4.3. In sub section 4.4, Particle Swarm Optimization for selection of key frames is adopted. To evaluate the performance of the proposed key frame extraction scheme,

standard measurements are given in section 5. Experiments and results are discussed and illustrated in section 6. In section 7, conclusion is drawn.

3. Orthogonal Polynomials Model for Color Video Frames

A linear three-dimensional (3-D) color video frame formation system may be considered around a Cartesian and color coordinates separable, blurring, point-spread operator in which the video frame I results in the super-position of the point source of impulse weighted by the value of the object f. Expressing the object function f in terms of derivatives of the frame function I relative to its Cartesian and color coordinates is very useful for analyzing the video frame in order to detect low level features. Hence, the initial requirements of the low level feature extraction in 3-D video frame may be stated as follows: Since low level features can be detected based on the local properties of the color video frame, we need to devise a local point-spread operator such that it is Cartesian as well as color coordinate separable and deblurring operator. In the case of 2-D monochrome video frames, the point spread function $M(x,\ y)$ can be considered to be a real valued function defined for $(x,\ y)\ \in\ X\ \times\ Y,$ where X and Y are ordered subsets of real values. In the case of a gray level video frame of size $(n\ \times\ n)$ where $X(rows)$ consists of a finite set, which for convenience can be labeled as $\{0,\ 1,\ \cdots\ ,\ n-1\},$ the functions $M(x,\ y)$ reduce to a sequence of functions

$$M(i,\ t)\ =\ u_i(t),\qquad i=0,\ 1,\ \cdots,\ n-1. \tag{1}$$

As shown in Equation (2) the process of analysis of 2-D monochrome video frame can be viewed as the linear (2-D) transformation defined by the point-spread operator $M(x,\ y)\big(M(i,\ t)=u_i(t)\big).$

$$\beta'(\zeta,\ n)=\int_{x\in X}\int_{y\in Y}M(\zeta,\ x)\,M(\eta,\ y)I(x,\ y)\ \ dx\ dy \tag{2}$$

Considering both X and Y to be finite set of values $\{\ 0,\ 1,\ \cdots\ ,\ n-1\}$, Equation (2) can be written in matrix notation as follows.

$$\left|\beta_{ij}^{'}\right|\ =\ (|M|\otimes|M|)^t|I| \tag{3}$$

where the point-spread operator $|M|$ is

$$|M|\ =\ \begin{vmatrix} u_0(t_1) & u_1(t_1) & \cdots & u_{n-1}(t_1)\\ u_0(t_2) & u_1(t_2) & \cdots & u_{n-1}(t_2)\\ & \vdots & & \\ u_0(t_n) & u_1(t_n) & \cdots & u_{n-1}(t_n)\end{vmatrix} \tag{4}$$

$\otimes$ is the outer product and $\left|\beta_{ij}^{'}\right|$ be the n^2 matrices arranged in the dictionary sequence. $|I|$ is the video frame and $\left|\beta_{ij}^{'}\right|$ be the coefficients of transformation.

We consider a set of orthogonal polynomials $\{\ u_0(t),\ u_1(t),\ \cdots\ ,\ u_{n-1}(t)\}$ of degrees $0,\ 1,\ 2,\ \cdots\ ,\ n-1,$ respectively. The generating formula for the polynomials is as follows.

$$u_{i+1}(t)\ =\ (t-\mu)\ \ u_i(t)-b_i(n)\ \ u_{i-1}(t)\quad for\ \ i\ \geq\ 1 \tag{5}$$
$$u_1(t)\ =\ (t-\mu),\ \text{and}\ u_0(t)\ =\ 1,$$

Where

$$b_i(n)\ =\ \frac{\langle u_i u_i\ \rangle}{\langle u_{i-1},\ u_{i-1}\rangle}\ =\ \frac{\sum_{x=1}^{n}u_i^2(t)}{\sum_{x=1}^{n}u_{i-1}^2(t)}\ \text{and}\ \mu\ =\ \frac{1}{n}\sum_{x=1}^{n}\ t.$$

Considering the range of values of t to be $t_i=i,\ \ i\ =\ 1,\ 2,\ 3,\ \cdots\ ,\ n,$ we get

$$b_i(n) = \frac{i^2(n^2 - i^2)}{4(4i^2 - 1)}, \quad \mu = \frac{1}{n}\sum_{x=1}^{n} t = \frac{n+1}{2}$$

Next, we construct point-spread operators $|M|$s of different sizes from the above orthogonal polynomials as follows.

$$|M| = \begin{vmatrix} u_0(t_1) & u_1(t_1) & \cdots & u_{n-1}(t_1) \\ u_0(t_2) & u_1(t_2) & \cdots & u_{n-1}(t_2) \\ & \vdots & & \\ u_0(t_n) & u_1(t_n) & \cdots & u_{n-1}(t_n) \end{vmatrix} \quad (6)$$

for $n \geq 2$ and $t_i = i + 1$.

The gray levels in a 2-D color video frame may be considered as function of three variables, two of which are two spatial coordinates and the third represents the color band. Hence, the linear 2-D transformation that is defined in Equation (2) for 2-D monochrome video frames can easily be extended as the linear three-dimensional (3-D) transformation defined by the same point-spread operator $M(i,\ t)(= u_i(t),\ i = 0,\ 1,\ 2,\ \cdots,\ n-1)$. This 3-D transformation for color video frame analysis is shown in Equation (7).

$$|\beta'(\zeta,\ \eta,\ s)| = \int_{x\in X}\int_{y\in Y}\int_{z\in Z} M(\zeta,\ x)M(\eta,\ y)M(s,\ z)I(x,y,z)\ dx\ dy\ dz \quad (7)$$

where $\zeta,\ \eta,\ s$ are coordinates in the 3-D transformed space and $I(x,y,z)$ is a color video frame region wherein x and $y$are two spatial coordinates and z indicates the color coordinate.

Initially, for the sake of generality, we may consider $X,\ Y$and $Z$to be finite set of values $1,\ 2,\ \cdots,\ n$. Then Equation (7) can be written in matrix notation as follows:

$$|\beta'_{ijk}| = (|M| \otimes |M| \otimes |M|)^t |I| \quad (8)$$

where $|M|$ is the point-spread operator shown in Equation (6), $\otimes$ is the outer product, $|\beta'_{ijk}|$ be the n^3 matrices arranged in the dictionary sequence, $|I|$ is the color video frame and β'_{ijk} are the coefficients of the transformation.

Difference Operators for Video Frames

In the case of R-G-B color space, the elements of the finite set Z for convenience can be labeled as $\{1,\ 2,\ 3\}$. For the sake of computational simplicity, the finite set $S,\ X$and$Y$are also labeled in the identical manner. The point-spread operator in Equation (8) that defines the linear transformation of color video frames can be obtained as $|M| \otimes |M| \otimes |M|$ where $|M|$ is computed (and scaled) from Equation (6) as

$$|M| = \begin{vmatrix} u_0(x_0) & u_1(x_0) & u_2(x_0) \\ u_0(x_1) & u_1(x_1) & u_2(x_1) \\ u_0(x_2) & u_1(x_2) & u_2(x_2) \end{vmatrix} = \begin{vmatrix} 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{vmatrix} \quad (9)$$

The set of 27 three dimensional basis operators$O_{ijk},\ (0 \leq i,\ j,\ k \leq 2)$ can be computed as follows.

$$O_{ijk} = \hat{u}_i \otimes \hat{u}_j \otimes \hat{u}_k, \quad (10)$$

where $\hat{u}_i$ is the $(i+1)^{st}$column vector of $|M|$. The operator O_{ijk} is arranged in the dictionary sequence in such a manner that it becomes the $(i \times 3^2 + j \times 3 + k) + 1^{st}$ column vector of the point-spread operator $|M| \otimes |M| \otimes |M|$ in Equation (8).

4. Proposed Key Frame Extraction Scheme

The proposed key frame extraction scheme with orthogonal polynomials comprises of three steps namely (i) Shot detection (ii) Color feature extraction and (iii) Selection of optimal key frames with Particle Swarm Optimization as shown in Figure 1.

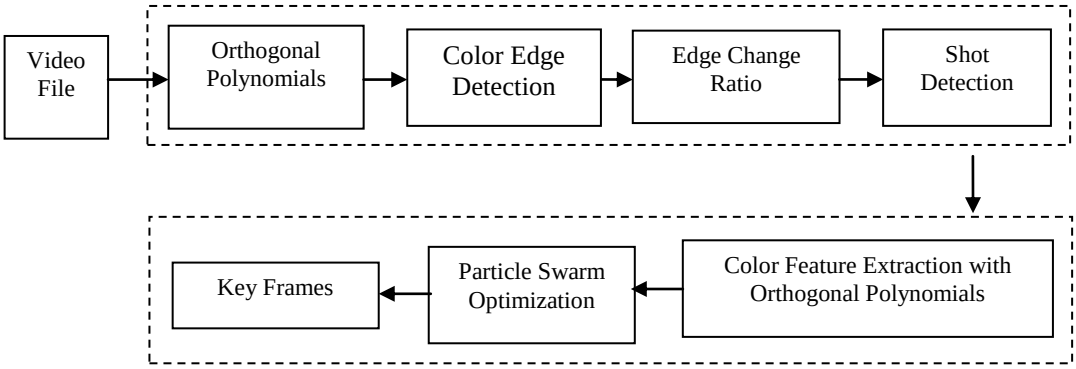

FIGURE 1. Proposed Key Frame Selection Scheme with Orthogonal Polynomials and Particle Swarm Optimization

Initially, the proposed system converts the given video into frames that are then subjected to orthogonal polynomials to extract the coefficients in transform domain. These transformed coefficients are utilized to detect edges of a video frame. The edge detected frames are subjected to morphological dilate operation and invert is performed. The proposed shot detection scheme calculates ECR from the consecutive edge detected and inverted frames. The proposed scheme is modeled to detect a shot, if ECR attains a sudden peak or low value for a particular frame. From the shots, key frames are selected to represent each shot of a video file. The proposed work extracts color feature with orthogonal polynomials model to identify the key frames of every shot. The color feature extraction scheme combines R, G and B color planes with orthogonal polynomials coefficients that takes in to consideration the individual plane as well as with interaction among these planes is a salient feature of this proposed model. The extracted color feature of the frames is then forwarded to Particle Swarm Optimization to select optimal key frames. The steps involved in proposed key frame selection scheme with orthogonal polynomials and Particle Swarm Optimization are discussed in the following sub sections.

The proposed shot detection scheme with orthogonal polynomials model detects and utilizes edges as a change ratio to identify the available shots in a video file and the same is described in this sub section. Initially, the video file is divided in to number of frames. Each frame of size $(n \times n)$ is partitioned in to $(m \times m)$ nonoverlapping blocks where $(m \leq n)$. Each block is subjected to orthogonal polynomials model to derive coefficients β' in transform domain as described in section 3. The above process is carried out for the entire frame f. For a frame f the gradient f' is modeled from few of orthogonal polynomials coefficients that are linear contrasts is estimated as given in equation (11).

$$f' = \sqrt{\beta_{010}^{'2} + \beta_{100}^{'2} + \beta_{011}^{'2} + \beta_{101}^{'2} + \beta_{012}^{'2} + \beta_{102}^{'2}} \quad (11)$$

The estimated gradient value is then compared with a preset threshold T. If $(\ gradient\ f' > T)$ then second derivative is to be calculated as given in equation (12).

$$f'' = \beta_{200}' \cos^2\alpha + 2\beta_{011}' \cos\beta \cos\gamma + \beta_{020}' \cos^2\beta + 2\beta_{101}' \cos\alpha \cos\gamma + \beta_{002}' \cos^2\gamma + 2\beta_{110}' \cos\alpha \cos\beta \quad (12)$$

where angles α, β and γ are estimated as given in equation (13)

$$\alpha = tan^{-1}\left(\frac{\beta'_{010}}{\beta'_{100}}\right)\beta = tan^{-1}\left(\frac{\beta'_{001}}{\beta'_{010}}\right) \text{ and } \gamma = tan^{-1}\left(\frac{\beta'_{100}}{\beta'_{001}}\right) \quad (13)$$

Based on second derivative value, if $f'' < 0$, then the edge point is marked at the center of a block. Similarly, the above steps are carried out for the entire frame and edges are detected with the proposed orthogonal polynomials model. The steps involved in proposed orthogonal polynomials based edge detection technique is presented.

Algorithm of Proposed Orthogonal Polynomials Based Color Edge Detection in Video Frames

Input : Color video frame in R, G and B channel
Output : Edge detected frame

Begin
Step 1: If (end of video frame), go to step 8, else
Extract a block of size (3×3)
Step 2: Obtain a column vector from the orthogonal polynomials coefficients $|\beta'_{ijk}|$ by
$|\beta'_{ijk}| = (|M|\ |I|)^t$where $M = M \otimes M \otimes M$ and $M = \begin{vmatrix} 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{vmatrix}$
Step 3: Compute the first order derivative from the linear contrasts i.e. gradient
$f' = \sqrt{\beta'^2_{010} + \beta'^2_{100} + \beta'^2_{011} + \beta'^2_{101} + \beta'^2_{012} + \beta'^2_{102}}$
Step 4: If $(gradient \geq threshold)$ go to step 5 else edge points are not marked and go to step 7
Step 5: Find the second derivative as given in Equation (12)
Step 6: If second derivative is negative, then mark central pixel as an edge point
Step 7: Go to Step 1
Step 8: Stop
End

Edge Change Ratio (ECR) for Shot Detection

In this section, a new ECR that identifies the movement of edges between consecutive frames of edge detected frames with orthogonal polynomials model is proposed. The proposed ECR detects shot with respect to abrupt and gradual changes viz. hard cuts, fade and dissolve effects. These changes are defined as:

Hard cut is an immediate transition and the gradual change may be either due to fade effects or dissolve effects. A fade occurs when the first shot disappears gradually (fade out) earlier than the second shot appears (fade in). On the other hand, dissolve is a shot change where the second shot appears while the first shot disappears gradually. Thus, having extracted the edges in a frame with zero crossings in orthogonal polynomials domain and taking in to the account of detecting shots viz. hard cuts, fade and dissolve effects, Edge Change Ratio (ECR) between two consecutive frames of a color video is presented in this section.

For this purpose, consider two consecutive frames viz. f_{n-1} and f_n. For this purpose, consider two consecutive frames viz. f_{n-1} and f_n. These frames are subjected to edge extraction with zero crossings in orthogonal polynomials domain and the edge extracted results are obtained and represented as f^e_{n-1} and f^e_n respectively. In this proposed work, mathematical morphological operation "Dilate" is then applied, followed with "Invert" operation on f^e_{n-1} and f^e_n, so as to produce clear visible edges f^i_{n-1} and

f_n^i respectively. In order to identify the movement of edges present in contiguous frames, in this proposed work, "AND" operation is performed between f_{n-1}^e and f_n^i. This is termed as entering edge for further investigation purposes. Similarly, an exiting edge frame is modeled as a result of "AND" operation between f_n^e and f_{n-1}^i. In this work, we represent the entering edge frame and exiting edge frame as f_n^{in} and f_{n-1}^{out} respectively. The entering edge pixels in f_n^{in} are fraction of edge pixels in f_n which is far away from the closest edge in f_{n-1}^{out}. Similarly, the exiting edge pixels in f_{n-1}^{out} are fraction of edge pixels in f_{n-1} which is far away from the closest edge in f_n^{in}. For the purpose of establishing a shot, in this proposed work, an ECR between two frames f_{n-1} and f_n is represented as $ECR[f_{n-1}, f_n]$, which is defined as maximum of ratio between total number of edges in entering and exiting edge frames with total count of edge pixels in edge extracted frames of f_{n-1} and f_n. This is represented as

$$ECR[f_{n-1}, f_n] = \max\left[\frac{TNE(f_n^{in})}{TNE(f_{n-1}^{e})}, \frac{TNE(f_{n-1}^{out})}{TNE(f_n^{e})}\right] \quad (14)$$

where $TNE(f_n^{in})$, $TNE(f_{n-1}^{out})$, $TNE(f_{n-1}^{e})$ and $TNE(f_n^{e})$ represent the total count of edge pixels in entering edge frame (f_n^{in}) and exiting edge frame $TNE(f_{n-1}^{out})$, edge frame and (f_{n-1}^{e}), edges of current frame (f_n^{e}) respectively.

The ECR values between two consecutive frames of a video frame are then found in raster scan fashion. In this proposed shot detection scheme, we model the presence of continuous frames with in a shot, if the difference between ECR values of continuous pairs of frame is negligible. If the ECR values of continuous pairs of frames differ substantially then a shot is detected. This is established by defining the difference between consecutive ECR values of three color frames of the given video. For example, ECR of three consecutive color frames f_{n-1}, f_n, f_{n+1} with a new index m is defined in this work as the magnitude of difference between ECR_s of consecutive pairs of frames:

$$ECR_m = |ECR(f_{n-1}, f_n) - ECR(f_n, f_{n+1})| \quad (15)$$

In this proposed work, detection of shot is modelled by finding the ratio between two consecutive ECR_m values. If the ratio between ECR_{m-1} and ECR_m converge approximately to unity, then a shot is not detected, if it converges to zero, then a shot is detected.

If any of these conditions are not satisfied, then a ratio between swapped ECR's (viz. ECR_m and ECR_{m-1}) are computed and the same convergence test is applied. The steps involved in ECR calculation is given below.

Steps Involved in Edge Change Ratio (ECR) Calculation for Shot Detection

Input : Edge detected frames
Output : Edge Change Ratio
Begin
Step 1: Count the number of edge pixels from edge detected f_n and f_{n-1} frames.
Step 2: Dilate the edge images and invert it.
Step 3: Perform "AND" operation between edge image of f_{n-1} and the image obtained from step 2 for f_n.
Step 4: Count the number of entering and exiting pixels in image obtained from step 3 to estimate f_n^{in} and f_{n-1}^{out}.
Step 5: Calculate Edge Change Ratio as given in equation (14).
Step 6: To detect shots, estimate difference between continuous pairs of frames with index m as given in equation (15).

Step 7: If ratio between ECR_m and ECR_{m-1} converges to unity, no shot is detected else shot is detected.
Step 8: Stop
End

Having detected the shots in a video, our next aim is to identify the key frames that represent the entire sequence of frames in a shot, and the same is proposed in terms of color feature that is present with the same orthogonal polynomials coefficients in the video frames. This is described in the following sub section.

Color Feature Extraction

In this sub section, color feature extraction scheme in terms of same orthogonal polynomials coefficients is proposed. The proposed color feature extraction scheme, takes in to account the individual R, G and B color planes as well as the interaction among these planes viz. between R and G, G and B, R and Bas evident from the 3-D polynomials operator defined with outer product of point-spread function M. By applying this orthogonal polynomials operator, we obtain the orthogonal polynomials coefficients as a column vector β'_{ijk}. It is proved in [38], that the mean squared amplitude responses of the finite difference orthogonal polynomials operators $O'_{ijk}s$ (except O_{000}) are uncorrelated linear contrasts per unit length. In the case of proposed orthogonal polynomials model for 3-D color images, it is further proved that $\{O_{i00}\} \cup \{O_{0j0}\} \cup \{O_{00k}\}$ represent color edges.

Extending the same notion, in this work, a color feature extraction techni que is presented. Three color features, in each of a block of a video frame are modeled with a simple linear combination of orthogonal polynomials transform coefficients that are considered to be linear contrasts of mean squared amplitude responses of finite difference operators in one direction (z-direction). It can be noted in the design of orthogonal polynomials model for 3-D color images, x and y represent the spatial coordinates and z represents color. These three color features are represented as C_R, C_G and C_B that are modeled from the orthogonal polynomials coefficients β'_{ijk} by considering the linear contrast in z-direction, with a simple multiplication factor, corresponding to the total number of color channels. That is C_R, C_G and C_B are extracted with orthogonal polynomials model coefficients β'_{ijk} as

$$C_R = \left(\frac{\left(2 * \beta'_{ij0}\right) - \left(3 * \beta'_{ij1}\right) + \beta'_{ij2}}{6} \right) \quad , i = j = 0 \tag{16}$$

$$C_G = \left(\frac{\left(\beta'_{ij0}\right) - \left(\beta'_{ij2}\right)}{3} \right) \quad , i = j = 0 \tag{17}$$

$$C_B = \left(\frac{\left(2 * \beta'_{ij0}\right) + \left(3 * \beta'_{ij1}\right) + \beta'_{ij2}}{6} \right) \quad , i = j = 0 \tag{18}$$

It is to be noted that, these color features consider not only the individual color planes but also the interaction among these planes. Having extracted these three color features locally in a block, we extend the extraction of color features globally for the entire frame of a video in terms of mean and standard deviation. That is the color feature C_R is calculated for all the blocks and mean of such C_R represented as μ_R is formed globally for the entire frame. Similarly, mean of C_G and C_B are computed and represented as μ_G and μ_B respectively. In the same way, standard deviation of the color feature C_R, C_G and C_B is calculated globally and represented as σ_R,σ_G and σ_B respectively. These six color feature values viz. μ_R, μ_G,μ_B,σ_R,σ_G, σ_B are then utilized to represent a video frame. The steps involved in proposed color feature extraction scheme are presented as an algorithm hereunder.

Algorithm of Proposed Color Feature with Orthogonal Polynomials

Input : Video frames of a shot
Output : color feature
Begin
Step 1: Input the video frames of a shot.
Step 2: Divide each frame in to blocks of (3 x 3).
Step 3: Each block is applied with orthogonal polynomials model as given in section 3.
Step 4: The color feature is calculated with β'_{ijk} as given in equation (16-18).
Step 5: Repeat step 2 to step 4 until all the blocks are encountered.
Step 6: The color feature for an entire video frame is represented by calculating mean and variance and stored as $\mu_R, \mu_G, \mu_B, \sigma_R, \sigma_G, \sigma_B$
End

The extracted color feature with orthogonal polynomials is then fed to Particle Swarm Optimization (PSO) for selecting optimal key frames to represent a shot of a video and the same is presented in forthcoming sub section.

Particle Swarm Optimization

Particle Swarm Optimization is inspired by social behavior patterns of organisms that live and interact with in large groups. In order to select optimal key frames, PSO is adopted here. In this work, both particle and swarm are assigned to be frames that are available inside a shot. Discrete PSO is used where, a particle position is represented as a vector is given in equation (19) as below:

$$p_i = p_1, p_2, \ldots, p_j \ldots, N, \quad p_j^i \in \{0, 1\} \tag{19}$$

where N is the number of frames. For a particle p_i, $p_j = 1$ if the frame j is one of the key frames to represent a shot else, $p_j = 0$. initially, the particle position is randomly chosen, and the difference between frames are calculated by comparing frames consequently with respect to the extracted color feature in terms of orthogonal polynomials coefficients and represented viz. $\mu_R, \mu_G, \mu_B, \sigma_R, \sigma_G, \sigma_B$ by the proposed scheme as discussed in the sub section 4.3. This scheme compares subsequent frames in terms of the color feature and finds the distance (D) between two frames as given in below equation (20).

$$D = \sqrt{(\mu_R - \mu_R')^2} + \sqrt{(\mu_G - \mu_G')^2} + \sqrt{(\mu_B - \mu_B')^2} + \sqrt{(\sigma_R - \sigma_R')^2} + \sqrt{(\sigma_G - \sigma_G')^2} + \sqrt{(\sigma_B - \sigma_B')^2} \tag{20}$$

$\mu_R, \mu_G, \mu_B, \sigma_R, \sigma_G, \sigma_B$ represent color feature of current frame in RGB plane and in the same way $\mu_R', \mu_G', \mu_R', \sigma_R', \sigma_G', \sigma_R'$ represent color feature of next frame in RGB plane respectively. The difference between a group of frames is nothing but, the average difference between each two successive frames in a group. Each particle has its own position p_i and a velocity represented by v_i. Each particle remembers its own best position and swarm remembers the best position of all the particles with respect to the distance (D). The velocity of the particle v_i determines how far the new position is from old. The value of particle velocity and position are updated until the best solution in key frame extraction is achieved with respect to the equation (21) as given below:

$$V_{t+1}(p,i) = w \times v_t(p_i) + c_1 \times r_1 \times LB(p,i) - p_t(p_i) + c_2 \times r_2 \times GB(i) - p_t(p_i) \tag{21}$$

where , LB is the local best solution at iteration t for particle$p$
GB is the global best solution at iteration t for swarm
p is the particle number, i represents dimension

$v_t(p)$ is velocity of the particle at time t
$p_t(p)$ is position of the particle at time t
$c_1$and c_2 are acceleration constants
r_1 and r_2 are random numbers between 0 and 1.

In this way, by utilizing the extracted color feature with orthogonal polynomials coefficients, best solution with in a group of frames (locally) is found by particle and swarm finds best solution among the groups (globally). Hence, PSO adopted here selects optimal key frames to represent a shot. To check the relevancy of the proposed key frame extraction scheme, standard measures are used and the same is illustrated in the below sub section.

5. Performance Measures

To evaluate and compare the proposed key frame selection scheme with other systems, three standard measures such as precision, recall and f-measure (Equation 22-24) are used. In this measurement, two mechanisms viz. (i) the key frames that are extracted manually (ii) the key frames are generated with the proposed technique. The precision and recall metric with respect to the above two mechanisms are as follows:

$$\text{Recall} = \frac{T_p}{T_p + F_n} \tag{22}$$

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{23}$$

where,
True Positive(T_p): a frame is chosen as key frame (both manually and technique)
False Positive(F_p): a frame is chosen as key frame by the technique but not manually
False Negative(F_n): a frame is chosen as key frame manually but not by the technique

Recall value is the probability that a relevant key frame is chosen whereas precision is the probability that a selected key frame is relevant. Recall value is high, if number of selected key frames is very less, on the other side, selection of more number of key frames increases the precision rate. Since, both are complement to each other, a combined metric termed as "F-measure" is defined.

$$\text{F-measure} = 2 \times \frac{recall \times precision}{recall + precision} \tag{24}$$

A high value of F-measure thus indicates a high value for both recall and precision rate.

6. Experiments and Results

The proposed key frame selection technique for video retrieval system with orthogonal polynomials and Particle Swarm Optimization has been experimented on 20 different genres of video that are downloaded from open-video project [41]. The details of test videos with number of frames are then presented in Table 1. Sample video frames of 'Hurricane Force – A Coastal Perspective, segment 03' that are of size (640 X 360) with pixel values in the range (0 – 255) are shown in Figure 2. The color information for these video frames are available in RGB color space. At first, to identify the shots, the video frames are divided into (3 x 3) blocks on which the proposed orthogonal polynomials model is applied as given in section 3. In this work, edges of a frame are extracted with orthogonal polynomials coefficients by gradient estimation. The estimated gradient is then compared with a threshold value T to identify the thick edges present in the video frame. In this work, a threshold value 20 is taken after rigorous experiments. Then second derivative of gradient is calculated and if it is lesser than zero, then the proposed scheme marks edge points in the center of a block. Similarly, the above process is carried out for all the blocks in a video frame and edges are detected by identifying

the zero crossings in second directional derivatives. The sample edge detected video frames with the orthogonal polynomials model for the video frames shown in Figure 2 is presented in Figure 3.

S. No	Video Name	Frames
1	Wetlands Regained, segment o3 of 8	3562
2	Technology at Home: A Digital Personal Scale	3346
3	Introduction to HCIL 2000 reports	2454
4	Ocean floor Legacy, segment 05 of 14	4665
5	The Great Web of Water, segment 01	3279
6	The Great Web of Water, segment 02	2118
7	The Great Web of Water, segment 07	1745
8	A New Horizon, segment 01	1806
9	A New Horizon, segment 02	1797
10	A New Horizon, segment 06	1944
11	A New Horizon, segment 08	1815
12	Exotic Terrane, segment 04	4797
13	The Future of Energy Gases, segment 05	3615
14	The Future of Energy Gases, segment 09	1884
15	Oceanfloor Legacy, segment 01	1740
16	Oceanfloor Legacy, segment 02	2325
17	Oceanfloor Legacy, segment 09	2106
18	Hurricane Force-A Coastal Perspective, segment 03	2310
19	Drift Ice as a Geological Agent, segment 05	2187
20	Drift Ice as a Geological Agent, segment 10	1407

TABLE 1: Details of test videos

The proposed shot detection scheme utilizes ECR to identify the shots available in a video file. Hence, from the edge detected continuous frames viz. f_n and f_{n-1} in Figure 3, mathematical morphological operation "Dilate" is performed. The dilate operation considers background pixel values on edge detected frames to superimpose the structuring element on top so that the origin of the structuring element coincides with the pixel position. If at least one such pixel value in structuring element coincides with the fore ground value in the image, then the input pixel value is set to foreground.

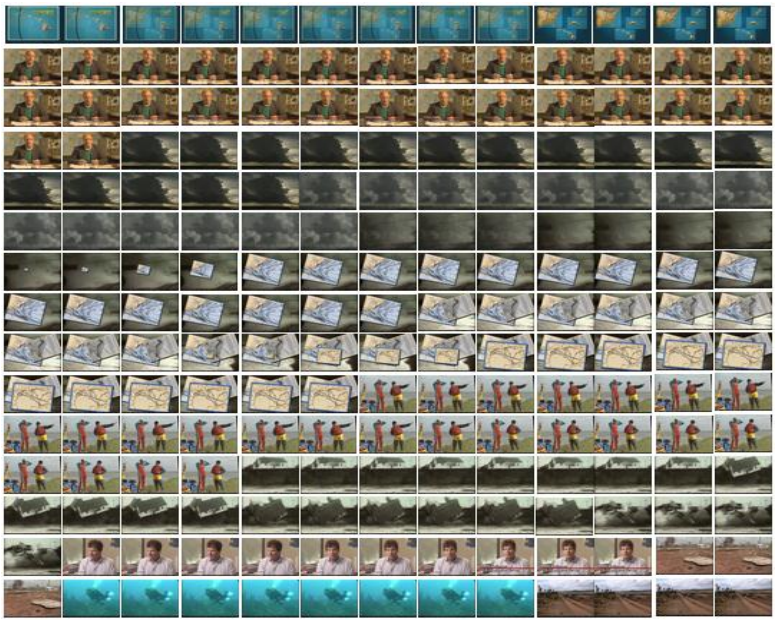

FIGURE 2. Sample Video Frames of Force – A Coastal Perspective, Segment 03

On the dilated frames, invert operation (mere complement) is applied to invert dark areas into bright areas and bright areas into dark areas. The idea behind utilizing dilation and invert operation is to attain clear and visible edges from the edge detected frames. The results of dilate and invert operation for the edge detected frames with proposed scheme for the video frames shown in Figure 3 are presented in Figure 4 and Figure 5 respectively.

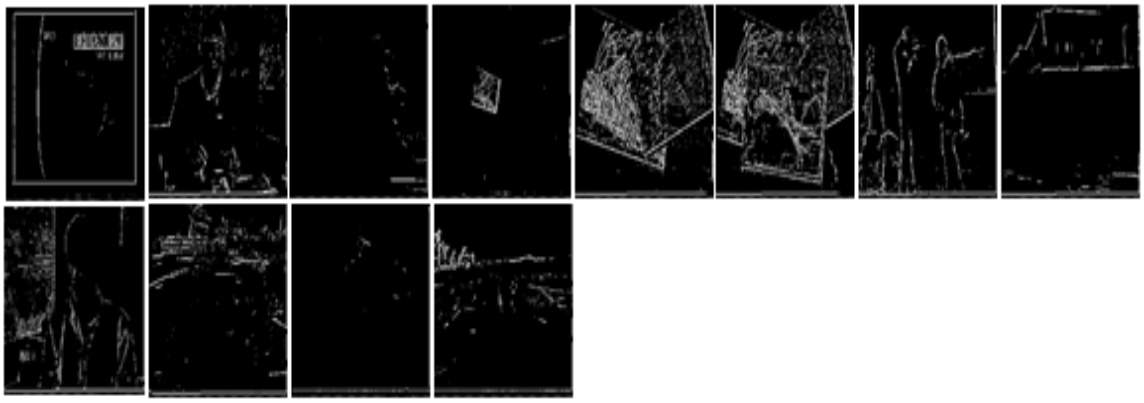

FIGURE 3. Result of Proposed Edge Detection Scheme with Orthogonal Polynomials

Now, to calculate ECR, the inverted frame of f_n and edge detected frame of f_{n-1} is carried out with AND operation from which the total count of pixels, entering and exiting pixels are calculated as described in sub section 4.2. The result of AND operation between sample edge detected frame f_{n-1} and sample inverted frame f_n of Figure 3 and Figure 5 is presented in Figure 6.

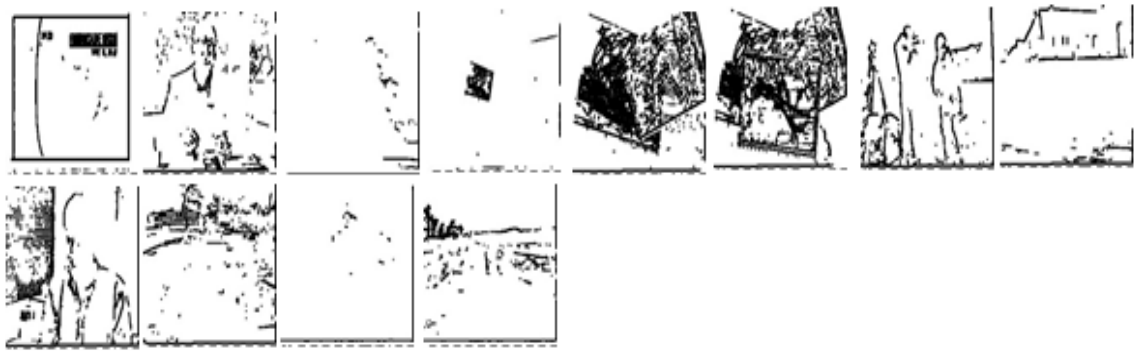

FIGURE 4. Result of Dilation for the Frames Shown in Figure 3

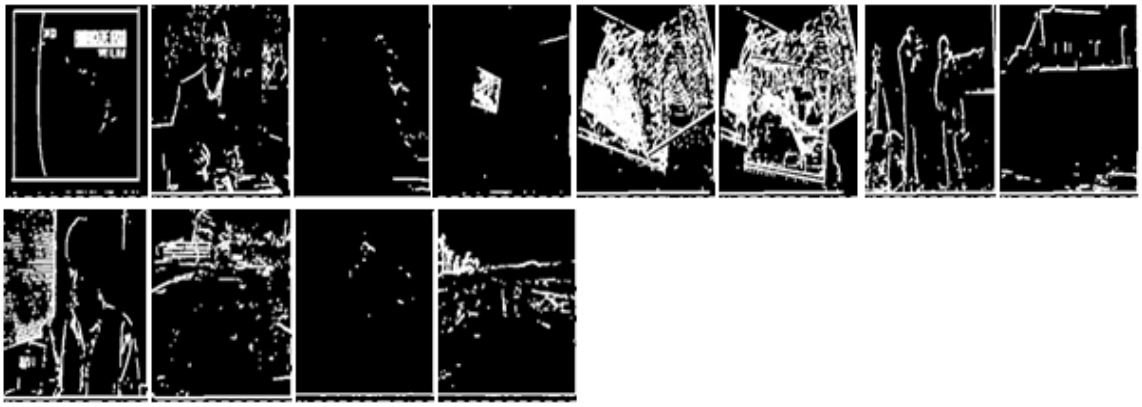

FIGURE 5. Result of Invert Operation for the Frames Shown in Figure 4

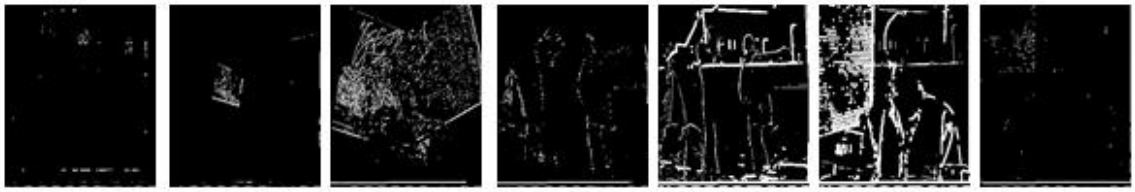

FIGURE 6. Result of AND Operation for Sample Frames Shown in Figure 3 And Figure 5

As described in sub section 4.1, the proposed system detects a shot in terms of ECR. Hence, for consecutive frames, total numbers of edge pixels, entering and exiting edge pixels are calculated. With the calculated edge pixels, the proposed system compares consecutive pairs of frames and finds the difference among the frames. In this work, it is modeled to detect a shot, if the estimated difference between the pairs of frames is substantially high. Since, ECR is capable of identifying the changes accurately, the proposed shot detection scheme successfully detects shots in terms of hard cuts, fade and dissolve effects.

Experiments are carried out on 20 types of video as given in Table 1. From which "Hurricane Video – A Coastal Perspective" that is presented in Figure 2 is chosen as a test bed. With respect to ECR, the proposed scheme detects the abrupt change among the sequence of frames and exactly 11 hard cuts are identified.

For example, the significant change has been identified in the frame *f0031* compared with first frame *f0001*. Hence, a shot is detected at frame *f0031*. Likewise, the proposed shot detection scheme finds the abrupt change on frames *f0576, f0842, f0941, f1013, f1153, f1387, f1887, f2038* and *f2160* respectively and the same is shown in Figure 7.

FIGURE 7. Results of the proposed shot detection technique for "Hard Cuts"

The proposed scheme also identifies fade and dissolves effects. The fade effects may be either fade in or fade out. In case of fade in, the occurrence of a new shot is slowly visible in the current frame. On the other hand, the visibility of a frame may get faded away and said to be as fade out. In this case, *f0006* and *f0021* has similar visual information but still due to fade effect, shot is detected on *f0006* and *f0021*.

Similarly, fade effects are identified on 16 frames viz. *f0587, f0675, f0832, f0909, f1012, f1153, f1315, f1387, f1457, f1915, f2037, f2041, f2139* and *f2161* respectively to detect shots and the same is presented in Figure 8. Also, the dissolve effects have been detected on 20 frames as shown in Figure 9 viz. *f0001, f0007, f0031, f0576, f0692, f0839, f0869, f0917, f1012, f1053, f1153, f1333, f1388, f1460, f1887, f2035, f2038, f2157* and *f2160* respectively.

FIGURE 8. Results of the Proposed Shot Detection Technique for "Fade Effects"

On the whole, the proposed shot detection technique detects 46 shots (inclusive of 11 hard cuts, 16 fade effects and 19 dissolve effects) in Hurricane Force – A Coastal Perspective – Segment'03 video that has 2310 frames and the same is presented in Table 2. Similarly, the proposed shot detection scheme with orthogonal polynomials coefficients in terms of ECR has been experimented on remaining 19 video files and number of shots that are detected is presented in Table 2.

FIGURE 9. Results of the Proposed Shot Detection Technique for "Dissolve" Effects

Video No	Total No. of Frames	Results of Proposed Shot Detection Scheme			Total No. of Shots
		Hard Cuts	Fade Effects	Dissolve	
1	3562	19	17	21	57
2	3346	18	14	13	45
3	2454	12	9	6	27
4	4665	28	20	11	59
5	3279	22	17	12	51
6	2118	6	3	3	12
7	1745	16	14	9	39
8	1806	14	17	8	39
9	1797	12	10	9	31
10	1944	11	7	7	25
11	1815	13	5	11	29
12	4797	22	16	13	51
13	3615	9	4	2	15
14	1884	10	4	3	17
15	1740	5	3	1	9
16	2325	11	3	5	19
17	2106	12	9	2	23
18	2310	11	16	19	46
19	2187	10	2	2	14
20	1407	8	3	1	12

TABLE 2: Total Number of Shots Detected with Proposed Shot Detection Scheme for Test Videos Presented in Table 1

In order to select key frames from 46 shots of video file 18, the proposed system extracts and utilizes color information that are derived with orthogonal polynomials coefficients as a feature vector. As a first step in color feature extraction, the video frames in 46 shots are initially partitioned into (3 x 3) blocks and orthogonal polynomials model is applied. The proposed color extraction scheme is performed in RGB color space and hence for a block, three color features are extracted with the proposed orthogonal polynomials coefficients β'_{000}, β'_{001} and β'_{002} in R plane, β'_{000}, β'_{001} and β'_{002} in G plane and β'_{000}, β'_{001} and β'_{002} in B plane individually and represented as C_R, C_G and C_B. Since, three planes are individually representing the color feature in each block it may increase the computation time and hence to represent the color feature for an entire frame, mean and standard deviation are estimated and represented as μ_R, μ_G, μ_B, σ_R, σ_G, σ_B. Therefore, the size of color feature is reduced to 6 for an entire frame by which the computation time for comparing adjacent frames in key frames selection is considerably reduced. To ensure optimal selection of key frames for representing shots in a video, PSO is adopted to choose a particle randomly and the difference between frames with respect to the extracted color feature is calculated. Each particle holds its own best position and swarm has best position of all the particles as described in sub section 4.4. The result of the proposed key frame selection technique with orthogonal polynomials and Particle Swarm Optimization is shown in Figure 10.

FIGURE 10. Result of the Proposed Key Frame Selection with Orthogonal Polynomials and Particle Swarm Optimization for Hurricane Force – A Coastal Perspective, Segment 03 Video

As a result of the proposed color feature extraction technique with PSO, only 10 key frames are selected to represent 46 shots. Since, fade and dissolve effects can be detected with in a hard cut, these 10 key frames are sufficient enough to represent a shot. In that way, the proposed system selected optimal key frames that is almost equivalent to 11 hard cuts (approximately one key frame for each cut).

The proposed key frame selection scheme is evaluated with standard recall (equation 28), precision (29) and f-measure (30) rates. Hurricane Force – A Coastal Perspective, segment 03 video frames are taken as input to conduct experiments for selecting key frames by existing systems. In the similar way, the proposed key frame selection scheme also utilized Hurricane Force – A Coastal Perspective, segment 03 frames. For the video frames shown in Figure 2, the proposed key frame selection scheme achieves a recall of 0.93, precision of 0.92 and f-measure of 0.92 respectively. Similarly, the proposed scheme is also evaluated with 20 different types of video frames and its performance results are presented in Table 3. In addition to that, the comparison of proposed scheme is performed with other schemes such as OV [33], DT [34], STIMO [35], VUMM [36], Naveed et al. [30].

The selected key frames with proposed scheme and existing schemes for the video frames shown in Figure 3 are presented in Figure 11. The measures of performance of these existing schemes are also calculated and incorporated in the same Table 3. From Table 3, it is evident that, the proposed system achieves higher recall, precision and f-measure value than that of other systems. With respect to that, the proposed technique achieves a recall rate of 0.93 which is equivalent to Naveed's approach. Also, a higher precision of 0.92 is achieved when compared with that of Naveed's 0.91. But, interestingly the proposed technique and Naveed approach achieves equal f-measure rate of 0.92 but with some exceptions.

No	OV [33]			DT [34]			STIMO [35]			VUMM [36]			Naveed et. Al [30]			Proposed KFE-OPPSO		
	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F
1	0.59	0.63	0.61	0.53	0.59	0.56	0.76	0.51	0.61	0.57	0.59	0.58	0.84	0.80	0.82	0.88	0.83	0.85
2	0.56	0.83	0.67	0.58	0.73	0.65	0.72	0.56	0.63	0.66	0.65	0.65	0.85	0.82	0.83	0.86	0.84	0.85
3	0.82	0.49	0.61	0.36	0.63	0.46	0.72	0.59	0.65	0.58	0.67	0.62	0.83	0.82	0.82	0.85	0.82	0.83
4	0.66	0.70	0.68	0.49	0.49	0.49	0.69	0.61	0.65	0.61	0.57	0.59	0.83	0.80	0.81	0.85	0.81	0.83
5	0.48	0.49	0.49	0.63	1.00	0.77	0.67	0.41	0.51	0.62	0.51	0.58	0.82	0.75	0.78	0.82	0.78	0.80
6	0.78	0.67	0.72	0.30	0.33	0.32	0.77	0.48	0.59	0.62	0.67	0.64	0.90	0.85	0.87	0.92	0.85	0.88
7	0.72	0.58	0.64	0.52	0.59	0.55	0.59	0.43	0.50	0.90	0.59	0.71	0.81	0.80	0.80	0.81	0.80	0.80
8	0.70	0.70	0.70	0.49	0.52	0.50	0.67	0.67	0.67	0.79	0.70	0.74	0.85	0.80	0.82	0.86	0.82	0.84
9	0.37	0.93	0.53	0.46	0.50	0.48	0.42	0.89	0.57	0.71	0.90	0.79	0.87	0.85	0.86	0.89	0.85	0.87
10	0.47	0.73	0.57	0.63	0.71	0.67	0.52	0.67	0.59	0.78	0.72	0.75	0.83	0.81	0.82	0.84	0.81	0.82
11	0.54	0.81	0.65	0.42	0.62	0.50	0.63	0.90	0.74	0.70	0.88	0.78	0.75	0.75	0.75	0.77	0.75	0.76
12	0.92	0.44	0.59	0.52	0.67	0.58	0.50	0.42	0.45	0.92	0.75	0.82	0.83	0.80	0.82	0.86	0.83	0.84
13	1.00	0.55	0.71	0.59	0.75	0.66	1.00	0.64	0.78	1.00	0.71	0.83	0.80	0.75	0.77	0.81	0.77	0.79
14	0.69	0.50	0.58	0.86	0.73	0.79	0.92	0.67	0.77	0.92	0.72	0.81	0.85	0.80	0.83	0.87	0.82	0.84
15	0.22	0.50	0.30	0.50	0.78	0.61	0.61	0.33	0.43	0.41	0.55	0.47	0.70	0.66	0.68	0.73	0.68	0.70
16	0.64	0.89	0.74	0.48	0.80	0.60	0.64	0.76	0.69	0.64	0.76	0.69	0.88	0.85	0.86	0.90	0.87	0.88
17	0.71	0.72	0.71	0.58	0.75	0.66	0.62	0.79	0.69	0.62	0.79	0.69	0.86	0.89	0.87	0.90	0.92	0.91
18	0.65	0.72	0.69	0.67	0.78	0.72	0.73	0.76	0.75	0.73	0.76	0.75	0.93	0.91	0.92	0.93	0.92	0.92
19	0.74	0.89	0.81	0.92	0.83	0.87	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.87	0.88	0.91	0.91	0.91
20	0.80	0.73	0.77	0.72	0.67	0.69	0.85	0.80	0.82	0.82	0.84	0.90	0.84	0.81	0.82	0.87	0.84	0.85
av	0.65	0.67	0.64	0.56	0.67	0.61	0.70	0.64	0.65	0.73	0.71	0.71	0.84	0.81	0.82	0.86	0.84	0.85

TABLE 3: Comparison of the Proposed KFE-OPPSO Scheme with Existing Systems in Terms of Standard Recall (R), Precision (P) and F-Measure (F)

For example, in case of video 5, the existing DT scheme achieves a precision of 1, since they have selected only one key frame. Similarly, OV scheme achieves a higher recall of 1 for video 13 and the same achieved a very precision of 0.55. Likewise, for video 9, DT, OV, STIMO achieved a low recall rate but the proposed system achieves higher rate even better than VSUMM and Ejaz schemes. On the whole, the proposed system achieves consistently a higher precision and recall values. The extracted key frames for the video 'Hurricane Force – A Coastal Perspective, segment 03' by proposed scheme various existing schemes are presented in Figure 11. It is a documentary video where, the researchers explaining about geology of coastal regions and about catastrophic effects of extreme weather on coastal erosion.

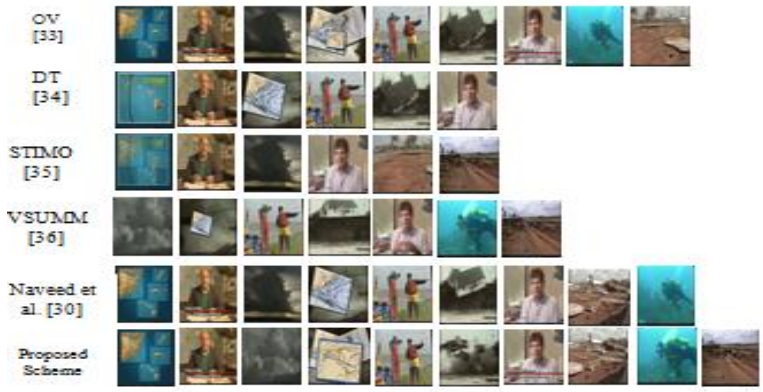

FIGURE 11. Comparison of Proposed Key Frame Extraction Technique for Video 'Hurricane Force – A Coastal Perspective, Segment 03' with Existing Systems

The ground truth key frames and extracted key frames different schemes are also presented. Now, if we compare with the ground truth frames with other techniques, it is clearly visible that, key frames 2 and 4 of OV, key frames of 2 and 6 of DT missing. On the other hand, STIMO misses a key frame of 5 and VSUMM misses frame number 8. The proposed system extracts all the key frames that has been given as a ground truth and achieves a better precision, recall and f-measure.

7. Conclusion

In this paper, a new Key Frame Extraction technique with Orthogonal Polynomials and Particle Swarm Optimization (KFE-OPPSO) is presented. The proposed technique segments video into shots and extract color feature to select key frames in the same orthogonal polynomials transform domain. Since, both the steps are carried out in single transform domain, the time complexity is considerably reduced. In addition to that, to select optimal key frames from a shot, the extracted color feature is then fed to Particle Swarm Optimization for comparing subsequent frames. The experimental results indicate that the extracted key frames with proposed scheme are highly relevant than other existing systems. As a future scope, features such as texture, edge and shape can be combined together to improve the selection of key frames.

References

[1] Hong Jiang Zhang, Jianhua Wu, Di Zhong and Stephen W. Smoliar, "An Integrated System For Content Based Video Retrieval And Browsing", Pattern Recognition, Vol. 30, No. 4, **(1997)**, Pp. 643-658.
[2] Alan F. Smeaton, "Techniques Used and Open Challenges to the Analysis, Indexing and Retrieval of Digital Video", Information Systems, vol. 32, no. 4, **(2007)**, pp. 545-559.
[3] Yannis S. Avrithis, Anastasios D. Doulamis, Nikolaos D. Doulamis and Stefanos D. Kollias, "A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases",

Computer Vision and Image Understanding, vol. 75, nos. 1 / 2, **(1999)**, pp. 3-24.
[4] Yueting Zhuang, Young Rui and Thomas S. Huang, "Video Key Frame Extraction by Unsupervised Clustering and Feedback Adjustment", Journal of Computer Science and Technology, vol. 14, no. 3, **(1999)**, pp. 283-287.
[5] Eung Kwan Kang, Sung Joo Kim and Jong Soo Choi, "Video Retrieval Based on Key Frame Extraction in Compressed Domain", IEEE International Conference on Image Processing, vol. 3, **(1999)**, pp. 260-264.
[6] Calic J and Izuier Do, "Efficient Key - Frame Extraction and Video Analysis", IEEE International Conference on Information Technology: Coding and Computing, **(2002)**, pp. 28-33.
[7] Hun-Cheol Lee and Seong-Daekam, "Iterative Key Frame Selection in the Rate-Constraint Environment", Signal Processing: Image Communication, vol. 18, no. 1, **(2003)**, pp. 1-15.
[8] Tianming Liu, Hong-Jiang Zhang and Feihu Qi, "A Novel Video Key-Frame-Extraction Algorithm Based on Perceived Motion Energy Model", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 10, **(2003)**, pp. 1006-1013.
[9] Xu-Dong Zhang, Tie-Yan Liu, Kwok-Tung Lo and Jian Feng, "Dynamic Selection and Effective Compression of Key Frames for Video Abstraction", Pattern Recognition Letters, vol. 24, no. 9-10, **(2003)**, pp. 1523-1532.
[10] Tieyan Liu, Xudong Zhang, Jian Feng and Kwok-Tung Lo, "Shot Reconstruction Degree: A Novel Criterion for Key Frame Selection", Pattern Recognition Letters, vol. 25, no. 12, **(2004)**, pp. 1451-1457.
[11] Markos Mentzelopoulos and Alexandra Psarrou, "Key-Frame Extraction Algorithm Using Entropy Difference", ACM Workshop on Multimedia Information Retrieval, **(2004)**, pp. 39-45.
[12] Jiawei Rong, Wanjun Jin and Lide Wu, "Key Frame Extraction Using Inter-Shot Information", IEEE International Conference on Multimedia and Expo, vol. 1, **(2004)**, pp. 571-574.
[13] Ki Tae Park, Joong Yong Lee, Keewook Rim and Young Shik Moon, "Key Frame Extraction Based on Shot Coverage and Distortion", Advances in Multimedia Information Processing, Lecture Notes in Computer Science, vol. 3768, **(2005)**, pp. 291-300.
[14] Jian-Quan Ouyang, Li Jin-Tao and Huanrong Tang, "Interactive Key Frame Selection Model", Visual Communication and Image Representation, vol. 17, no. 6, **(2006)**, pp. 1145-1163.
[15] Zhao Guang-Sheng, "A Novel Approach for Shot Boundary Detection and Key Frames Extraction", IEEE International Conference on Multimedia and Information Technology, **(2008)**, pp. 221-224.
[16] Guozhu Liu and Junming Zhao, "Key Frame Extraction from MPEG Video Stream", Proceedings of the International Computer Science and Computational Technology, **(2009)**, pp.7-11.
[17] Hua Man and Jiang Peng, "A Feature Weighted Clustering Based Key-Frames Extraction Method", IEEE International Forum on Information Technology and Applications, vol. 1, **(2009)**, pp. 69-72.
[18] Pascal Kelm, Sebastian Schmiedeke and Thomas Sikora, "Feature-Based Video Key Frame Extraction for Low Quality Video Sequences", IEEE Workshop on Image Analysis for Multimedia Interactive Services, **(2009)**, pp. 25-28.
[19] Magda B. Fayka, Heba A. El Nemr and Mona M. Moussa, "Particle Swarm Optimization Based Video Abstraction", Journal of Advanced Research, vol. 1, no. 2, **(2010)**, pp. 163-167.
[20] Huiyu Zhou, Abdul H. Sadka, Mohammad R. Swash, Jawid Azizi and Umar A. Sadiq, "Feature Extraction and Clustering for Dynamic Video Summarization", Neurocomputing, vol. 73, no. 10-12, **(2010)**, pp. 1718-1729.
[21] Gwo-Cheng Chao, Yu-Pao Tsai and Shyh-Kang Jeng, "Augmented Keyframe", Visual Communication and Image Representation, vol. 21, no. 7, **(2010)**, pp. 682-692.
[22] Suet-Peng Yong, Jeremiah D. Deng and Martin K. Purvis, "Wildlife Video Key-Frame Extraction Based on Novelty Detection in Semantic Context", Multimedia Tools and Applications, vol. 62, no. 2, **(2011)**, pp. 359-376.

[23] Naveed Ejaz and Sung Wookbaik, "Weighting Low Level Frame Difference Features for Key Frame Extraction Using Fuzzy Comprehensive Evaluation and Indirect Feedback Relevance Mechanism", International Journal of The Physical Sciences, vol. 6, no. 14, **(2011)**, pp. 3377-3388.
[24] Gentao Liu, Xiangming Wen, Wei Zheng and Peizhou He, "Shot Boundary Detection and Key Frame Extraction Based on Scale Invariant Feature Transform", IEEE International Conference on Computer and Information Science, **(2011)**, pp. 1126-1130.
[25] Liujun Liu, Xiaohong Wang and Shizheng Zhou, "Key Frames Extraction Algorithm Based on GA", IEEE International Conference on Computational Sciences and Optimization, **(2011)**, pp. 808-811.
[26] Sun Shumin, Zhang Jianming and Liu Haiyan, "Key Frame Extraction Based on Artificial Fish Swarm Algorithm and K Means", IEEE International Conference on Transportation, Mechanical and Electrical Engineering, **(2011)**, pp. 1650-1653.
[27] Naveed Ejaz, Tayyab Bin Tariq and Sung Wookbaik, "Adaptive Key Frame Extraction for Video Summarization Using an Aggregation Mechanism", Journal of Visual Communication and Image Representation, vol. 23, no. 7, **(2012)**, pp. 1031-1040.
[28] Jie-Ling Lai and Yang Yi, "Key Frame Extraction Based on Visual Attention Model", Visual Communication and Image Representation, vol. 23, no. 1, **(2012)**, pp. 114-125.
[29] Walid Barhoumi and Ezzeddine Zagrouba, "On-The-Fly Extraction of Key Frames for Efficient Video Summarization", AASRI International Conference on Intelligent Systems and Control, vol. 4, **(2013)**, pp. 78-84.
[30] Naveed Ejaz, Irfan Mehmood and Sung Wook Baik, "Efficient Visual Attention Based Framework for Extracting Key Frames from Videos", Signal Processing and Image Communication, vol. 28, no. 1, **(2013)**, pp. 34-44.
[31] Guang-Hua Song, Qing-Ge Ji, Zhe-Ming Lu, Zhi-Dan Fang and Zhen-Hua Xie,"A Novel Video Abstraction Method Based on Fast Clustering of The Regions of Interest in Key Frames", International Journal of Electronics And Communications, vol. 68, no. 8, **(2014)**, pp. 783-794.
[32] Qing Xu, Yu Liu, Xiu Li, Zhen Yang, Jie Wang, Mateu Sbert and Riccardo Scopigno, "Browsing and Exploration of Video Sequences: A New Scheme for Key Frame Extraction and 3D Visualization Using Entropy Based Jensen Divergence", Information Sciences, vol. 278, **(2014)**, pp. 736-756.
[33] D. Dementhon, V. Kobla and D. Doerrman, "Video Summarization by Curve Simplification", International Conference on Multimedia, **(1998)**, pp. 211-218.
[34] P. Mundur, Y. Rao, and Y. Yesha, "Keyframe-Based Video Summarization Using Delaunay Clustering", International Journal on Digital Libraries, vol. 6, no. 2, **(2006)**, pp. 219-232.
[35] M. Furini, F. Geraci, M. Montangero, and M. Pellegrini, "STIMO: Still and Moving Video Story Board for The Web Scenario, Multimedia Tools and Applications, vol. 46, no. 1, **(2010)**, pp. 47-69.
[36] S. E. D Avila, A. B. P Lopes, L. J. Antonio and A. D. A Araujo, "VSUMM: A Mechanism Designed to Produce Static Video Summaries and a Novel Evaluation Method", Pattern Recognition Letters, vol. 32, no. 1, **(2011)**, pp. 56-68.
[37] R. Krishnamoorthi and P. Bhattacharya, "Color Edge Extraction Using Orthogonal Polynomials Based Zero Crossings Scheme, "Information Sciences, vol. 112, no. 1-4, **(1998)**, pp. 51–65.
[38] R. Krishnamoorthy, "Transform Coding of Monochrome Images with Statistical Design of Experiments Approach to Separate Noise", Pattern Recognition Letters, vol. 27, no. 8, **(2007)**, pp. 771-777.
[39] Tikonov and A. N. Arsenin, "Solutions of Ill-Posed Problems," John Wiley and Sons, New York, 1977.
[40] R. Courant and D. Hilbert, "Methods of Mathematical Physics", vol. 1, Wiley Eastern, New Delhi.
[41] Www.Open-Video.Org

Authors

Dr. M. Braveen is currently working as Assistant Professor in the Department of Computer Science and Engineering at Vellore Institute of Technology Chennai, Tamil Nadu. He obtained his Ph. D in the Faculty of Information and Communication Engineering from Anna University, Chennai, in the year 2018 and M. Tech in Computer Science and Engineering from Pondicherry University in the year 2008 and B. Tech in Information Technology under Pondicherry University in the year 2006. He has published research articles in reputed International Journals and International Conferences His areas of interest include Image / Video Retrieval, Compression and Steganography.

Dr. R. Krishnamoorthy is working as a Professor in the Department of computer Science and Engineering, Anna University Chennai, Bharathidasan Institute of Technology (BIT) Campus, Tiruchirappalli. He obtained his Ph. D in Image Processing from the Indian Institute of Technology (IIT), Kharagpur in 1995 and M. Tech in Computer Science and Engineering from the Indian Institute of Technology (IIT), Kanpur, in 1992. His areas of interest include Image Compression, Image Retrieval, Image encryption and authentication, Biometrics, Software engineering, Digital Watermarking, Steganography.

ISSN: 2005-4238 IJAST