Shot Boundary Detection By Bidirectional Empirical Mode Decomposition

[1]Hina Shahnur Sayyad,[2]Prof. D. A. Jadhav,
[1]Student, [2]Professor
[1,2]Department of E&TC, Jspm's Imperial college of engineering and research, Wagholi, Pune, India.

Abstract: Shot boundary detection is the fundamental step of video retrieval, summarization, and analysis. In our paper, bidirectional empirical mode decomposition (BEMD) is presented into the shot boundary detection method. Using this method, every frame present in videos is decomposed by BEMD into a series of Intrinsic Mode Function images and a residue image. The IMF image having the lowest frequency of each frame is cast-off and an auxiliary frame is composed back by the other IMF images and the residue image. By employing these substitute frames to compute variances of adjacent frames unwanted detections are avoided which may be caused by immediate illumination changes. Further to avoid unwanted detections caused due to object motions, we have also applied a region-based object tracking method. The results proposed in this paper demonstrate that the precision of detecting shot boundaries is significantly improved.

***Index Terms* - Bidirectional Empirical Mode Distribution (BEMD), CBIR, Shot Boundary Detection (SBD), RAI Dataset, Video Frame.**

I. Introduction

The latest developments in multimedia technology, combined with a significant increase in computer performance and the progression of the Internet, have given people admittance to a huge amount of video information. Video applications, which are considerably rising, have introduced the increasing demand for ground-breaking tools and technologies for efficient video indexing, video browsing, and video retrieval. To automate all these, an excessive deal of research is being done on content-based video recovery over the last decade [1, 2]. Midst of the numerous structural levels such as frame, scene, shot, etc., shot level has been observed as suitable for content-based retrieval and browsing[3]. A video shot is defined as an arrangement of frames captured with the help of an individual camera in a single constant action in space and time [4]. Largely, it is a collection of frames that contain regular visual features, such as texture, motion, and color. The transition between shots can be gradual or abrupt, depending on this shot boundaries can be classified into two types: gradual and cut transitions. The cut transition is the typical, abrupt change wherein one frame belongs to the disappearing shot and the upcoming frame to the appearing shot. Gradual transitions are further categorized into fade-out/in, wipe, dissolve, according to the features of different editing effects [5]. In dissolve transition, the last few frames from the disappearing shot for a moment overlap with the first few frames of the upcoming shot. During this overlap, the intensity of the disappearing shot goes on reducing from normal to zero (fade out), however the intensity of the upcoming shot increases from zero to normal (fade in). Thus, during a fade transition, the vanishing shot fades out into a blank frame, and then the blank frame fades into the appearing shot. The key focus of most of the recent works is on abrupt transitions because gradual transitions are typically tougher to identify, due to the objections and/or camera in a shot. Over the previous decade, shot boundary detection (SBD) is being observed in video retrieval, summarization, pattern recognition [6].

Our paper makes use of CBIR i.e. Content-Based Image Retrieval. We are using the contents of an image rather than metadata[7]. In CBIR, the search is carried out by using color, shape, or texture feature. These are principal components in CBIR. Lots of techniques are available to retrieve an image such as Low-level feature extraction, Relevant feedback method, Semantic-Based image retrieval, etc. Shape, color, texture are low-level features. In Color feature extraction search is carried out with the help of colors in an image. The procedure is carried out by calculating the color histogram of an image. The color histogram of the query image is calculated first and it will be matched with the available image database[8].

In the shape-based image, the recovery shape does not mean the shape of the image but it refers to the shape of any particular region. Edge detection or segmentation method can be used shape-based image retrieval[9]. But in this paper, we are using the Bi-directional Empirical Mode Decomposition method to retrieve an image. This is a highly accurate method. EMD decomposes fast oscillations into slowest oscillations. This method is helpful for linear and non-linear signals. In this method, we are using multiple iterations to decompose a signal. Multiple iterations calculate the IMF of an image. IMF is a basic part of EMD. IMF stands for Intrinsic Mode Function. We are using EMD to retrieve an image so the IMF acts as Principle component. When we enter the query image first IMF is calculated after that it will be matched with the IMF of available database image.

The paper is designed as follows. Section II. Literature review, III. Proposed System, IV. Implementation, V. Results, VI. Conclusions.

II. Literature Survey

In this section, we introduce the work related to our proposed framework.

H. Y. Mark Liaoff [10] offered an innovative dissolve detection algorithm that was able to avoid the misdetection of motions with the help of the binomial distribution model to methodically define the threshold required for discerning a real dissolve from local or global motions. JesusBescos [11], offered a detection of the modification of video shot(i.e. cut) in real-time on MPEG22 online video which is capable enough of detecting unexpected transitions and all types of gradual transitions in the real world. Guillermo Cisneros [12] offered an approach that focuses on the mapping spaces of distances between frames in a new decision space through which independent thresholding of the sequence can be achieved. Liuhong Liang [13], offered an Improved Trigger

Limit Detection with the help of videotext information, in which numerous edge-based techniques are presented to detect sudden firing limits to circumvent the impact of common flashlights in videos.

Daniel DeMenthon [14] proposed a document focused on the correlation functions of video images. The detection of cut is dependent on the 2max ratio criterion in a consecutive image buffer. Kota Iwamoto and others [15], the discovery of digital video effects and wipes was suggested based on an autonomous model of image boundary line characteristics pattern that is based on a novel independent model of patterns. These models depend on the appearances of the image boundary lines that divide the two image areas in the transition frames. Jinhui Yuan [16] offered a document on a trigger limit detection method for news videos based on the segmentation and tracking of objects. It connects three key techniques: the method of comparison of partitioned histograms, segmentation of video objects, and wavelet analysis based tracking.

Yufeng Li [17] proposed an article on the Algorithm of detection of new shots based on the theory of the information. First, the attributes of the texture and color are extracted by using wavelet transform, then the difference between two consecutive frames that collide the communal information of the color characteristic and the communal information of matching of the texture characteristic is defined. The threshold is adjusted adaptively depending on the entropy of the constant frames and is not dependent on the type of shot and video. Vasileios T. Chasanis [18] presented the detection of scenes in videos using the clustering of shots and alignment of sequences. First, the keyframes were taken out using a spectral clustering method using the fast global k-means algorithm in the clustering phase and also providing an estimate of the number of the keyframes. Then, the shots are grouped into groups using only the visual similarity as a function and are labeled by following the group assigned to them.

Jinchang Ren [19] offered a document on the detection of trigger limits in MPEG videos using local and global indicators that operate directly in the compacted domain. Numerous local indicators are extracted from the MPEG macroblocks, and AdaBoost is used for the selection and merging of features. The selected characteristics are then used to classify the candidate cuts in five subspaces by pre-filtering and rules-based decision making, then the global indicators of frame similarity are examined among cut-off frames of cut candidates using the phase correlation of CC images. Priyadarshini Adhikari [20] offered a document on Video Shot Boundary Detection. This document presents the recovery of video using detection of the limit of the shot. LihongXu[21] offered a paper on a new shot detecting algorithm built on grouping. This article demonstrates a novel trigger limit detecting algorithm based on the K-means grouping. The extraction of the color feature is done first and then the differences of the video frame are defined. The video frames are segregated into numerous different sub-clusters by performing K-means clusters.

Wenzhu Xu and others [22] proposed an article on a new shot detecting algorithm based on graph theory. This article offers a trigger limit detecting algorithm based on graph theory. Video frames are segregated into numerous different groups through the realization of a theoretical graphics algorithm. Arturo Donate [23] presented the detection of shooting limits in videos using a robust three-dimensional tracking. The proposal is to extract the highlighted features of a video sequence and track them over time to estimate the limits of the shots within the video. Min-Ho Park [24] presented a paper on the detection of efficient trigger limits using characteristics based on block movement. It is a degree of break-in camera and an object/background movement for SBD is proposed focused on the grouping of two movement characteristics: the block-wise movement similarity and the modified displaced frame difference (DFD). Goran J. Zajic [25] presented a document on the detection of video trigger limits focused on multifractal analysis. Low-level features (texture and color) are pulled out from every frame in the video sequence, then concatenated into feature vectors (FV) and stored in the feature matrix. The rows of the matrix correspond to the FV of frames of the video sequence, while the columns are time series of a particular FV component.

Partha Pratim Mohanta [26], proposed an article on a model-based trigger limit detection technique that uses frame transition parameters that is focused on a formulated frame estimation scheme that uses the previous frame and the next frame. Pablo Toharia [27] proposed an article on Shot Boundary detection using Zernike moments in multi-CPU multi-GPU architectures alongside different probable hybrid combinations focused on Zernike moments. Sandip T [28] presented a document on the video summary based on keyframes using the automatic threshold and the speed of correspondence of the edges. First, the Histogram variance of each frame is computed, and then edges of the candidate keyframes are extracted by the Prewitt operator. Zhe Ming Lu [29] present a fast video trigger limit detection focused on SVD and pattern matching. It is focused on the selection of segments and the decomposition of singular values (SVD). Primarily, the locations of the firing limits and the lengths of the gradual transitions are anticipated using adaptation thresholds and most non-contour frames are cast-off at the same time.

Sowmya R [30] presented a document on Analysis and verification of summary video using Shot Boundary Detection. The analysis is focused on the difference of the block-based histogram and the Euclidean distance difference based on blocks for various block sizes. Ravi Mishra [31] proposed an article on a "Comparative study of the block matching algorithm and the complex transformation of two trees for the detection of shots in videos". This article shows an evaluation between the two detection methods concerning several parameters, such as hit rate, false rate, failure rate tested in a set of different video sequences. Wenjing Tong [32] presented a document on trigger limit detection focused on CNN and video annotation. This examination is based on TAG frames generated by a CNN model. Ahmed Hassanien [33] presented a document on the detection of large-scale, rapid, and precise trigger limits through spatial-temporal convolutional neural networks. This investigation is based on exploiting Big Data to optimize both the accuracy and speed of two large data sets.

According to K. Hemalchandram [34], nowadays content-based image retrieval is used for many applications. Color and Texture feature can be extracted through color histogram and wavelet transform. With the help of lots of methods, we can advance the performance of Content-Based Image Retrieval system. One of them is Relevance Feedback. Dacheng Tao et al. [35] Evaluated Support Vector Machine (SVM) and Traditional Kernel BDA (KBDA) based RF algorithms. Xiaoou Tang et al. [36]state that our previous approaches treat positive and negative feedback equivalently. But we know that they are not equal. The homogeneous concept was shared by positive feedback where negative feedback did not. A combination of Color-shape feature and the color-texture feature was taken by S. M. Zakariya [37]. Robust feature set for image retrieval can be provided by this combination. Said Jai-Andaloussi. Introduces Bi-dimensional Empirical Mode Decomposition for mass segmentation in image mammography.

III. Proposed System

Until now, the threshold selection problem is still involved in most shot boundary detection algorithms, and it directly affects the accuracy of detection results. However, the threshold is often derived from experiences or experiments and does not have good versatility. Bi-directional empirical mode decomposition based shot boundary detection will be one of the prominent solutions for the same.

3.1. Block Diagram

A Block diagram of the proposed shot boundary detection is as shown in Fig.1.

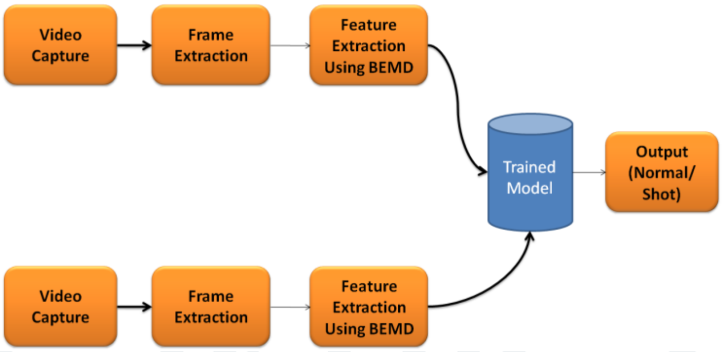

Fig. 1. Block diagram of the proposed bi-directional EMD based shot boundary system

In this paper, we are using Empirical mode decomposition to retrieve an image. As we know content-based image retrieval requires principle component so in this process we are using Intrinsic Mode Function as a principal component. Fig. 1 consists of four main blocks as shown.

A. Dataset

The TRECVID IACC.3 dataset was employed as it is with a set of predefined time-based segments. Due to this, pairs of the predefined segments can be unsystematically selected from the group for automatic formation of transitions for training purpose. In addition to that, we take segments of 3000 IACC.3 unsystematically selected videos. Moreover, segments that had less than 5 frames were not included, and from the leftover group, only every other segment was chosen, which results in selected 54884 segments.

The training examples will be generated when demanded during training by unsystematically sampling two shots and combining them with the help of random transition. Only dissolves and hard cuts were measured for training. The transition position was generated unsystematically. For dissolves, also its length was created unsystematically from the interval. The length N of every training sequence was designated to be 100 frames. The size of the input frames was fixed to 48 × 27 pixels.

To validate the models, an additional 100 IACC.3 videos (i.e., different from the training set) were manually labeled, resulting in 3800 shots. For testing, the RAI dataset [30] was considered.

B. EMD

The EMD i.e. IMF extraction procedure is carried out in this block. First, it will read an image from the Image Acquisition block then the preprocessing of an image will be performed. Then it reads input image and pre-processes the image. Then after that, it identifies local minima and maxima in an input image. Deduce a lower and an upper envelope by interpolation. Then withdraw the mean envelope from the image. The next step is to Subtract the so-obtained Intrinsic Mode Function (IMF) from the image. And finally, separate residue.

EMD is an acronym of Empirical Mode Decomposition. One wonders what is the meaning of decomposition. It means breaking down the compound process into separate constituent components. EMD is a basic part of Hilbert Hung transforms. It was first invented by Hung et al. In 1998. EMD decomposes a signal into Intrinsic Mode function. Time-Frequency information can be obtained by Hilbert Spectral Analysis of Intrinsic Mode Function & also determines the amount of variation due to oscillations at unlike time scales & time locations. By using EMD we can remove the highest frequency from a signal. Once the highest frequency removed from the signal, the same practice is applied to the residue signal to recognize the next highest frequency. A next new signal will be a residue signal which will be used for decomposition [9]. So finally we have,

$$x(t) \sum_{i=1}^{n} imf_i(t) + r(t) \tag{1}$$

Where,

$$imf_i(t) = IMF$$
$$r(t) = residue$$

C. BEMD

The complex data set can be decomposed into a finite number of diverse frequency components defined as two-dimensional intrinsic mode function (BIMF). These BIMFs are achieved through a process called sifting which carries on until the figure of extrema of maximum (or minimum) points is less than or equal to 2. Using the BEMD method, the local high-frequency oscillations present in the original data set can be extracted. The procedure of BEMD is analogous to the one-dimensional EMD. Observably, the surface interpolation of envelopes and the two-dimensional extrema detection are more complex than single-dimensional ones. If we can call $\mathrm{Ori(m,n)}$ as the original data set which needs to be decomposed into a definite number of bi-dimensional intrinsic mode functions (BIMFs) from high frequency to low frequency, $\mathrm{Ori(m,n)}$is decomposed into a chain of BIMFs and a residue in the following equation.

$$Ori(m,n) = \sum_{i=1}^{t} Bi(m,n) + \ Res\ (m,n) \tag{2}$$

where the $Bi(m,n)$ gives the i^{th} BIMF component (from high to low frequency, $B_1(m,n), B_2(m,n), \ldots \ldots, B_n(m,n)$ and gives us the residue $Res\ (m,n)$. The frequency of BIMF1 is greater than the rest of the BIMF's.

In this bi-dimensional sifting process, the nearby window is used for identifying the extrema and the multi-quadric method is employed for calculating the surface interpolation of envelopes to extract the BIMF. For every BIMF, a stopping standard may be ascertained in the sifting process.

This could be apprehended by the constraint of the size of the standard deviation (SD), which can be computed by making use of the formula given below :

$$SD_{ij} = \sum m = 1 \sum n = 1 \frac{|h_{j(i-1)}(m,n) - h_{ji}(m,n)|^2}{h^2_{j(i-1)}\ (m,n)} \tag{3}$$

The determination of SD influences the 2D sifting process. Smaller SD tends to increase the number of BIMFs. Besides, the standard deviation and mean value of the mean matrix $\mathrm{mean(m,n)}$ of the lower and upper envelopes have to be considered together.

BEMD Algorithm for Extraction of the j^{th} BIMF component:

1. Initialization: $r_0(m,n) = Ori(m,n)$ and $j = 1$is the BIMF index.
2. Initialize and make $h_0(m,n) = r_{i-1}(m,n), i = 1$.
3. Detect all the points of local maximum ($m_{max,i-1}$) and local minimum ($m_{min,i-1}$) of $h_{i-1}(m,\ n)$ respectively.
4. Compute the upper (lower) envelope of the local maximum and local minimum points, respectively.
5. Calculate the envelope mean: $m_{i-1}(m, n) = (m_{max,i-1}\ (m, n) + m_{min,i-1}\ (m, n))/2$;
6. $h_i(m,n) = h_{i-1}(m,n) - m_{i-1}(x,y), i = i+1$
7. $h_i(m,n) = BIMF_j(m,n)$ if $h_i(m,n)$matches the stopping criterion, then the j^{th} BIMF is got, or repeat steps (ii)–(vi).
8. $r_j(m,n) = r_{j-1}(m,n) - BIMF_j(m,n)$
9. Go to step (2), when the extrema number of $R_i(m,n)$ is more than 2, $j = j+1$, or decomposition is ended.

D. CBIR

CBIR stands for content-based image retrieval. In CBIR search uses the content of an image instead of metadata. CBIR system extracts characteristics from an image based on that retrieves relevant images. There are two types of feature extraction Low-level feature extraction and High-level feature extraction. Low-level features are just like shape, color, texture. By using these features retrieving an image is a very old method. Wavelet transform is used to pull-out texture feature and the color feature can be extracted through a color histogram. Next Block after EMD is a CBIR system. It will take the Intrinsic Mode Function of the Input image from the EMD block and compare it with the IMF of available database images.

E. Classification

The classification step divides the frame into normal or transition according to the BEMD features using a machine learning algorithm. The algorithms used for classification in the proposed system are as explained below.

- Random Forest

A random forest multi-way classier comprises of several trees, where each tree is grown with the help of some kind of randomization. The leaf nodes of all the trees are categorized by estimates taken from the posterior distribution with the help of image classes. Each inner node consists of a test that best divides the space of data that is to be classified. For classifying any image it is sent down every tree and further amassing the reached leaf distributions. Randomness may be injected at two points while training: while selecting the node tests and while subsampling the training data because of which each tree grows using a different subset.

The trees are binary and are created in a top-down fashion. The binary test at any node can be chosen in one of two ways: (i) randomly, i.e. data independent; or (ii) by a greedy algorithm that picks the test that best splits up the given training examples. “Best” here is measured by the information gain

$$\Delta E = -\sum_i \frac{|Q_i|}{|Q|} E(Q_i) \tag{4}$$

caused due to dividing the set Q of examples into two subsets Q_i according to the given test. Here E(q) is the entropy$-\sum_{j=1}^{N} p_j\ log_2\ (p_j)$ with p_j the proportion of samples in q belonging to class j, and $|Q|$ the size of the set. The process of choosing a test is carried out repeatedly for each non-terminal node, using only the training examples falling in that node. Whenever the node receives very few examples, or whenever it attains a given depth, the recursion gets stopped.

Assume that T stands for the set of all trees, C stands for the set of all classes, and L stands for the set of all leaves for a given tree. In the training phase the posterior probabilities ($P_{t,l}(Y(I) = c)$) for each class $c \in C$ at each leaf node $l \in L$, is found for each tree $t \in T$. These probabilities are computed as the ratio of the number of images I of class c that reach l to the total amount of images that reach l. $Y(I)$ is the class-label c for the image I.

- Gradient Boosting (GB)

GB is a combination of two methods, i.e., gradient descent method and Adaboost. It is a powerful machine learning algorithm that can perform regression, classification, and ranking. It builds the model in a forward fashion and optimizes the differential loss function. The algorithm is highly customizable for the particular application. Adaboost has an advantage as it boosts the outliers near classification boundaries. The equation for Adaboost can be given as follows:

$$H(x) = \sum_t \rho_t h_t(x) \tag{5}$$

The first step is to fit an additive ensemble $\sum_t \rho_t h_t(x)$ in a forward stage-wise manner. In each stage, a weak learner is introduced to compensate for the shortcomings of existing weak learners. In Gradient Boosting, inadequacies are identified by gradients whereas, in Adaboost, inadequacies are identified by high-weight data points. Both gradients and high-weight data points tell us how to improve our model.

Thus Gradient Boosting helps in advancing the accuracy of the classifier. Albeit Gradient Boosting is an ensemble-based technique, meaning it is focused on learning multiple models, the notion is to combine several weak models into a single, strong model. So depending upon the form of the week models used (e.g., most commonly decision trees), the final model is of the same type, i.e., in this example, a single decision tree comes out of the training process.

- Multi-Layer Perceptron

The Multilayer Perceptron neural network is a feed-forward neural network with one or more hidden layers. Cybenko and Funahashi have proven that the MLP network with a single hidden layer can estimate any continuous function to a limited accuracy. The classification performance of the MLP network will highly depend on the structure of the network and training algorithm. Here, two unlike training algorithms namely Levenberg-Marquardt(LM) and Bayesian Regulation (BR) have been used to determine the applicability of the MLP network.

The LM algorithm is used to train the MLP network because it has been proven that the LM algorithm has a much better learning rate and can keep the relative stability compared to the famous Back Propagation (BP) algorithm [40]. LM algorithm is an approximation of the Gauss-Newton algorithm, which usually offers a much-enhanced learning rate than the BP algorithm which is centered around the steepest descent algorithm [40]. The LM modification to the Gauss-Newton algorithm is shown below:

$$\Delta x = [J^T(x) J(x) + \mu I]^{-1} J^T(x) e(x) \tag{6}$$

Based on equation (6), μ is the Marquardt adjustment parameter and I is an identity matrix. When μ is small, the LM algorithm approximates the Gauss-Newton algorithm. Since the Gauss-Newton algorithm converges faster and more accurate near an error minimum, so the goal is to shift towards the Gauss-Newton algorithm as quickly as possible. Thus, the value of μ is decreased after each step unless the change in error is positive. On the other hand, when μ is large, the LM algorithm will approximate the steepest descent algorithm [40].

IV. Implementation

The flow diagram of the system is as shown in Fig.2.which explains system implementation.

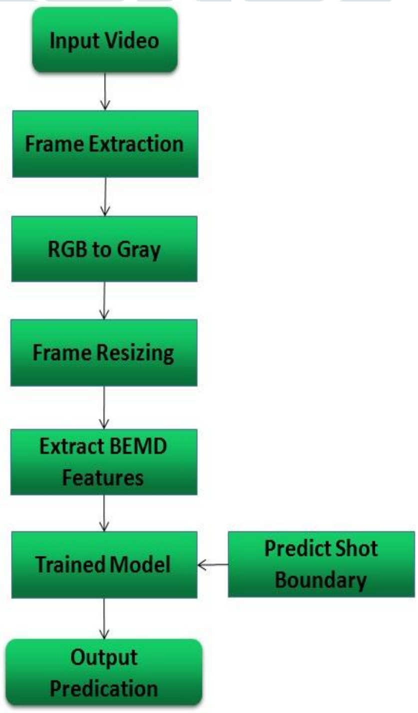

Fig.2. Flow chart of Proposed System

The implementation starts with frame extraction from the input video. The next step is to RGB be the Gray conversion of this extracted frame. After this, the frame is resized and then bidirectional empirical mode decomposition is applied and its features are extracted. Thus we the help of these extracted features we can now easily predict the shot boundary with the help of the trained boundary. This gives us the final output of shot boundary prediction.

V. RESULTS

The proposed shot boundary detection system by BEMD is implemented using the OpenCV library with the python language. The anaconda open source distribution of the python libraries is used. The image processing algorithms were developed using the open-source OpenCV library. The proposed algorithm is tested on the TRECVID IACC.3 dataset. The results are presented in qualitative and quantitative ways.

5.1. Qualitative Analysis

Qualitative analysis is the pictorial and non-statistical representation of the research. The results of the proposed system implementations have been shown below in Table I.

Table I. Qualitative Analysis

Sr. No.	Frame 1	Frame 2	Result
1			Correct Detection
2			Correct Detection
3			Correct Detection
4			Correct Detection
5			Correct Detection
6			Wrong Detection

7			Correct Detection
			Correct Detection
			Correct Detection
			Correct Detection
			Correct Detection
			Correct Detection
			Correct Detection
			Correct Detection
			Correct Detection
			Correct Detection

From the qualitative analysis of the proposed system, it is observed that the system has been able to rightly detect the transition from frame 1 to frame 2. As the experiment proposes to detect cut transitions in a video frame, we can see successful cut transitions identified in the results shown. Frame 2 images are mentioned with 'Transition Detected' as can be seen from Table 1.

5.2. Quantitative Analysis

The performance of the system will be calculated using three important metrics Sensitivity, Specificity, and Accuracy. Eq.1-Eq.3 represents them mathematically respectively.

$$Sensitivity = \frac{TP}{TP+FN} \quad (1)$$

$$Specificity = \frac{TN}{TN+FP} \quad (2)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Here, where TP is the true positive value which defines the positive sample is predicted as positive. TN is the true negative value that defines Negative detected as Negative. FP is a false positive value that defines occurred positively detected as Negative. FN is a false negative value that defines Negative detected as positive.

Table 5.1: Quantitative analysis

Video Title	Category	Sensitivity	Specificity	Accuracy
0EXCdXUN_fk.mp4	Sports	0.89	0.97	0.97
2i9mB1EQV7k.mp4	News	0.87	0.98	0.98
5L5used_AY0.mp4	Nature	0.8	0.98	0.98
7z0Bs6SHRx4.mp4	Info graphic	0.85	0.97	0.97
agu2oLm-QKA.mp4	Movie	0.83	1	1
AuzzSBzECXs.mp4	Nature	0.78	0.99	0.98

Table II represents the different types of videos such as sports, news, nature, infographic, movie, which are taken as input to the system during implementation. The values for sensitivity, specificity, and accuracy for all the types of videos are also reflected in the results. Thus we can say that this system can produce successful results for all the different types of video frame inputs.

The accuracy of the system for different machine learning algorithms on BEMD features is tabulated in Table III.

Table 5.2 Comparative analysis of the performance of the different classifier

Classifiers	Sensitivity	Specificity	Accuracy
KNN	0.74	0.93	0.83
ERT	0.87	0.98	0.92
RF	0.94	0.98	0.95
DT	0.83	0.8	0.8
GBC	0.89	0.96	0.93
MLP	0.9	0.97	0.93

The performance analysis of the different algorithms on BEMD features is shown in TABLE III, and the Random Forest algorithm shows the highest training accuracy of 95%. Also, GBC and MLP show an accuracy of 93%. Hence the RF algorithm is used for the testing.

To summarize the performance of our proposed systems with the existing ones, we would say that making use of BEMD as a local analysis method offers better performance than the existing shot boundary detecting system. BEMD can decompose images without complicated convolution processes and decomposes images into their components adaptively without using a priori basis. The superior quality of BEMD in extracting high-frequency data helps achieve efficient and satisfactory results.

VI. CONCLUSION

Shot change detection proves to be a very stimulating task. From this experiment, we can retrieve an image with the help of Bidirectional Empirical Mode Decomposition. In this method, the Intrinsic Mode Function (IMF) with the lowest frequency is removed and an effective and simple motion compensation method is adopted. As a result, the shot boundary detection method is no longer sensitive to abrupt illumination changes, object motion in shots, motion, and zoom in/out of cameras, so that the precision of detecting shot boundaries is improved. This method has the highest accuracy and data loss is less as compared to other methods.

REFERENCES

[1] Cotsaces, N. Nikiolaidis, and I. Pitas, Video shot detection and condensed representation. A review, IEEE Signal Process. Mag, Vol. 23, no. 2, pp. 28-37, Mar. 2006.

[2] Breazeale, D. Cook, D. J."Automatic Video Classification: A Survey of the Literature", IEEE Transactions on Systems, Man, and Cybernetics-Partrt C: Applications and Reviews, Vol. 38, 2008, No. 3, pp. 416–430.

[3] Lienhart, R, "Comparison of Automatic Shot Boundary Detection Algorithms. In: Storage and Retrieval for Image and Video Databases VII", Vol. 3656 of Proceedings of SPIE, San Jose, Ca, USA, January 1999, pp. 290–301.

[4] Cooper, M. Liu, T. Rieffel, E, "Video Segmentation via Temporal Pattern Classification", IEEE Transactions on Multimedia, Vol. 9, 2007, No. 3, pp. 610–618.

[5] Hanjalic, A, Shot Boundary Detection: Unraveled and Resolved? IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12, 2002, No. 2, pp. 90–105.

[6] Cotsaces, C. Nikolaidis, N. Pitas, I.: Video Shot Detection and Condensed Representation: A Review", IEEE Signal Processing Magazine, Vol. 23, 2006, No. 2, pp. 28–37.

[7] Cernekova, Z. Pitas, I. Nikou, C, "Information Theory-Based Shot Cut/Fade Detection and Video Summarization". IEEE Transactions on Circuits and Systems for Video Technology, Vol. 16, 2006, No. 1, pp. 82–91.

[8] www.wikipedia.org

[9] Alexandre Xavier Falcao, Ricardo da Silva Torres, “ContentBased Image Retrieval: Theory and Applications,” RITA, Volume XIII, Número 2, 2006.

[10] Prabir Bhattacharya, MD. Mahmudur Rahman, and Bipin C. De, "A Framework for Medical Image Retrieval Using Machine Learning and Statistical Similarity Matching Techniques With Relevance Feedback", IEEE Transaction on Information Technology in Biomedicine, Vol.11, No.1, January.

[11] H. Y: Mark Liaoff C. W. Su, H. R. Tyanf, and L. H. Chen, “A motion-tolerant dissolve detection algorithm”, IEEE 2nd Pacific-Rim Conference on Multimedia, Beijing, China, Vol.2195, Oct 2002.

[12] Jesús Bescos, “Real-Time Shot Change Detection Over Online MPEG-2 Video”, IEEE transactions on circuits and systems for video technology, vol. 14, no. 4, April 2004.

[13] Guillermo Cisneros, Jesus Bescos, Jose M. Martínez, Jose M. Menendez, and Julian Cabrera “A Unified Model for Techniques on Video-Shot Transition Detection”, IEEE transactions on multimedia, vol. 7, no. 2, April 2005.

[14] Liuhong Liang Liuhong Liang, Yang Liu, Hong Lu, Xiangyang Xue, and Yap-Peng Tan, ”An Enhanced Shot Boundary Detection Using Video Text Information”, IEEE Transactions on Consumer Electronics, Vol. 51, No. 2, MAY 2005

[15] Daniel De Menthon, “Shot boundary detection based on Image correlation features in the video”.

[16] Kota Iwamoto and Kyoji Hirata, “Detection of wipes and digital video effects based on a pattern-independent model of image boundary line characteristics".

[17] Jinhui Yuan, Huiyi Wang, Lan Xiao, Wujie Zheng, Jianmin Li, Fuzong Lin, and Bo Zhang, “a shot boundary detection method for news video based on object segmentation and tracking” IEEE transactions on circuits and systems for video technology, vol. 17, no. 2, February 2008.

[18] Yufeng Li, Zheng Zhao, “A Novel Shot Detection Algorithm Based on Information Theory", IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application.

[19] Vasileios T. Chasanis, Aristidis C. Likas, and Nikolaos P. Galatsanos, “Scene detection in videos using shot clustering and sequence alignment”, IEEE transactions on multimedia, vol. 11, no. 1, January 2009.

[20] Jinchang Ren, Jianmin Jiang, and Juan Chen, "Shot boundary detection in MPEG videos using local and global indicators", IEEE transactions on circuits and systems for video technology, vol. 19, no. 8, August 2009.

[21] Priyadarshinee Adhikari, Neeta Gargote, Jyothi Digge, “Video Shot Boundary Detection”, IEEE Transactions on Consumer Electronics, Vol.1.

[22] LihongXu & WenzhuXu, “A Novel Shot Detection Algorithm Based on Clustering” 2010 2nd International Conference on Education Technology and Computer (ICETC).

[23] Wenzhu Xu & Lihong Xu, “A Novel Shot Detection Algorithm Based on Graph Theory”.

[24] Arturo Donate and Xiuwen Liu, “Shot Boundary Detection in Videos Using Robust Three-Dimensional Tracking”. IEEE Transactions on Consumer Electronics.

[25] Min-Ho Park, Rae-Hong Park, and Sang Wook Lee, “Efficient Shot Boundary Detection Using Blockwise Motion-Based Features”

[26] Goran J. Zajic, Irini S. Reljin, and Branimir D. Reljin, “Video Shot Boundary Detection Based on Multifractal Analysis” Telfor Journal, Vol. 3, No. 2, 2011.

[27] Partha Pratim Mohanta, Sanjoy Kumar Saha, and Bhabatosh Chanda, “A Model-Based Shot Boundary Detection Technique Using Frame Transition Parameters” IEEE transactions on multimedia, vol. 14, NO. 1, February 2012.

[28] Pablo Toharia et al “on Shot boundary detection using Zernike moments in multi-GPU multi-CPU architectures”, Journal of Parallel and Distributed Computing Volume 72 Issue 9, September 2012.

[29] Mr. Sandip T. Dhagdi, Dr. P. R. Deshmukh “Keyframe Based Video Summarization Using Automatic Threshold & Edge Matching Rate” International Journal of Scientific and Research Publications, Volume 2, Issue 7, July 2012

[30] Zhe Ming Lu and Yong Shi “Fast Video Shot Boundary Detection Based on SVD and Pattern Matching," Image processing IEEE Transactions, Vol. 22, Issue: 12, Dec. 2013.

[31] Sowmya R, Dr. Rajashree Shettar “Analysis and Verification of Video Summarization using Shot Boundary Detection”, American International Journal of Research in Science, Technology, Engineering & Mathematics, 3(1), June-August, 2013, pp.82-86.

[32] Ravi Mishra, S. K. Singhai, M. Sharma, “Comparative study of block matching algorithm and dual tree complex wavelet transform for shot detection in videos”, Electronic system, signal processing, and computing technologies(ICESC), 2014 International Conference, Jan 2014.

[33] Wenjing Tong1, Li Song, Xiaokang Yang, Hui Qu, and Rong Xie, "CNN-Based Shot Boundary Detection and Video Annotation".

[34] Ahmed Hassanien, Mohamed Elgharib, Ahmed Selim, Sung-Ho Bae, Mohamed Hefeeda, Wojciech Matusik, "Large-scale, Fast and Accurate Shot Boundary Detection through Spatio-temporal Convolutional Neural Networks", Computer Vision and Pattern Recognition (cs.CV),

[35] K.Hemachandran, Manimala Singha, "Content-Based Image Retrieval using Color and Texture," Signal & Image Processing: An International Journal (SIPIJ) Vol.3, No.1, and February 2012.

[36] Xiaoou Tang and Yong Rui, Xuelomg Li, and Dacheng Tao, "Direct Kernel Biased Discriminant Analysis: A New Content-Based Image Retrieval Relevance Feedback Algorithm," IEEE Transaction on Multimedia, Vol.8, No.4, August 2006.

[37] Xiaoou Tang, Xuelomg Li, and Dacheng Tao, "Which Components are Important for Interactive Image Searching?," IEEE Transaction on circuits and systems for Video Technology, Vol.18, No.1, January 2008.

[38] Rashid Ali, S.M. Zakariya, and Nesar Ahmad, "Combining Visual Features of an Image at Different Precision Value of Unsupervised Content-Based Image Retrieval," IEEE 2010.

[39] A. Linderhed, "2-D empirical mode decompositions in the spirit of image compression," in Wavelet and Independent Component Analysis Applications IX, vol. 4738 of Proceedings of SPIE, pp. 1–8, Orlando, Fla, USA, April 2002.

[40] Kisi O., "Multi-Layer Perceptrons with Levenberg-Marquardt Training Algorithm for Suspended Sediment Concentration Prediction and Estimation," Hydrological Sciences Journal, vol. 49, no. 6, pp. 1025-1040, 2004.