

# Novel Technique for Video Content Retrieval - using SBD & KFE

Reeba John, Serin Paul, Shilpa Prakash, Vishwas V

Dept. of CSE  
TJIT, Affiliated to VTU  
Bangalore, India

Sharath P C

Assistant Professor  
Dept. of CSE  
TJIT, Affiliated to VTU  
Bangalore, India.

**Abstract-** The fundamental step in organizing large video data includes Shot Boundary Detection and Key Frame Extraction. Key Frame Extraction plays a major role in video information retrieval. Video Shot Boundary Detection segments a video and detects boundary between camera shots, which is important step for content-based video retrieval and video summarization. This paper discusses importance of Key Frame Extraction, also overcomes shortcomings of existing approaches, proposes new approach for Key Frame Extraction based on block based Histogram difference and edge matching rate. At first Histogram difference of every frame is calculated, edges of candidate key frames are extracted by Prewitt operator. At Last, edges of adjacent frames are matched. If edge matching rate is above average edge matching rate, current frame is said to be redundant key frame and is discarded. Histogram based algorithms are applicable to shot boundary detection. It provides global information about video content and is faster without any performance degradation.

## I. INTRODUCTION

The availability of cheap digital storage, led to expansion of digital video archives, thus enabling efficient browsing searching and retrieval in video database system. Initial attempts involved the use of textual data from subtitles and speech recognition transcripts, such as broadcast news retrieval, with no account of visual content of video. Traditional video indexing method, uses humans to manually tag videos with text keywords, is time consuming, and lacks speed of automation. Therefore advanced approaches, that require automatic indexing, are based on videos to provide efficient search for content-based video retrieval [1].

A video consists of sequence of images or also known as frames that are played consecutively at a speed of 20-30 frames per second. A shot is defined as consecutive frames from start to end

recording in camera. There are abrupt (discontinuous) transitions and gradual (continuous) transitions occurring between shots. Abrupt also referred as cut is an instantaneous transition from one scene to the next. Gradual transition includes fades, dissolves and wipes [1]. A fade is a gradual transition between scene and a constant image (Fade out) or between constant image and a scene(Fade in).Dissolve is a gradual transition from one scene to another while in wipe a line moves across the screen with new scene behind the line.

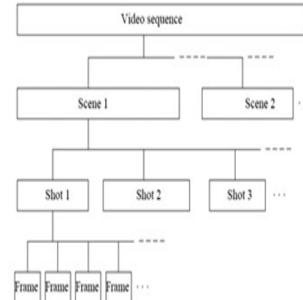


Figure 1:Overview of shot boundary detection

## A. Motivation

Recent developments in video compression technology, which includes use of digital cameras, growth of internet led to the increased usage, availability of digital video. Management of these activities requires knowledge of content of video. The algorithm "Video Content Analysis" enables us to understand or generate summary of large video material. Shot Boundary Detection also known as temporal video segmentation is one of the commonalities in applications of content analysis.

The initial step in process of video content analysis is parsing a video into temporal basic units called shots. A shot is series of video frames taken by a single camera, by zooming into a person or object. The region where there is significant content change is called shot boundaries [2].

## B. Thresholding

The most critical activity in Shot Boundary Detection process is selection of threshold in shot boundary detection step. Performance of algorithm remains in this phase. Single threshold or dynamic global threshold cannot solve the problem for video sequences, but dynamic local threshold are considered as better alternative

## II. PROBLEMS OF SHOT BOUNDARY DETECTION

### A. Detection of Gradual Transition

During video production process, first step is capturing shots by using single camera. Two shots attached together by Shot Boundary can either be abrupt or gradual. Abrupt transitions are created by

simply attaching a shot to another while Gradual transition involves dissolve, fades, wipes. Gradual transition pose problem due to amount of video editing effects. It gets harder when multiple effects are composed in case of lot of objects or camera motion. It also spread over time. Each editing effect changes from three to hundred frames.



Figure 2: Dissolve



Figure 3: Fade-in



Figure 4: Fade-out

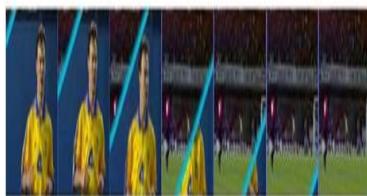


Figure 5: Wipe

### B. Flashlights

Color is primary element of video content, their representation

employ color as a feature. Continuity signals based on color feature exhibit changes under illumination changes, such as flashlights. They are identified as content change by Shot Boundary Detection tools.

### C. Object or camera motion

Visual content of video changes significantly with extreme object or camera motion and screenplay effects. Slow motion cause content change similar to Gradual transition whereas fast camera or object movements cause abrupt transition.

## III. LITERATURE REVIEW & RELATED WORK

Shot boundary detection plays an important role in video content analysis. Different kinds of features such as histogram, shape

information, motion activity, etc. have been used to detect shot boundary. Among these approaches, histogram is the popular approach and in these, pixel's space distribution was neglected. Different frames may have the same histogram.

In this view, Cheng et al[5] divided each frame into  $r$  blocks, and the difference of the corresponding blocks of consecutive frames was computed by color histogram. By adding up of all the block's difference, the difference  $D(i, i+1)$  of the two frames was obtained. Without using any blocks, the difference  $V(i, i+1)$  between two frames  $i$  and  $i+1$  was measured. Based on  $D(i, i+1)$  and  $V(i, i+1)$ , shot boundary can be determined.

An unsupervised clustering method was proposed by Zhuang et al[8]. Based on color histogram features in the HSV color space, a video sequence is segmented into video shot by clustering. For each video shot, the frame closest to the cluster centroid is chosen as the key frame for the video shot. Regardless of the duration or activity of the video shot, only one frame per shot is selected in to the video summary.

A new approach for shot boundary detection in the uncompressed image domain based on the MI and the joint entropy (JE) between consecutive video frames was proposed by Zuzana Cernekova[6]. Mutual information (MI) is a measure of information transported from one frame to another frame. It is used for directing abrupt cuts, where the image intensity or color is abruptly changed. Better results are provided by the measure of entropy, it exploits the inter-frame information flow in a more compact way than a frame subtraction.

A novel video summarization algorithm based on QR-decomposition was proposed by Ali Amiri[7]. Some efficient measures to compute the dynamicity of video shots using QR decomposition was utilized in detecting the number of key frames selected for each shots. Inorder to summarize the video shots with low redundancy, this property was utilized.

A similar approach was developed by Hanjalic et al[9] by dividing the sequence into number of cluster, and by cluster-validity analysis, find the optimal clustering. The main idea was to remove the visual redundancy among frames.

Doulamis et al[10] also developed a two step approach according to which the sequence is segmented into shots, or scene, and within each shot, frames are selected to minimize the cross correlation among frames features

### A. STATE OF ART SHOT BOUNDARY DETECTION

SBD algorithms from the literature are studied and analyzed and a brief summary of the algorithm is given below.

#### 1. Pixel-wise Difference with Adaptive Thresholding

Individual pixels from frames are compared to find out frame difference. Pair-wise comparison evaluates the differences in intensity or color values of corresponding pixels in two successive frames. In this algorithm the pixel-wise difference algorithm gives quite acceptable results with adaptive thresholding. By considering difference between the difference signal values of adjacent frames is a worthwhile approach. In practice, it is observed that it is useful to reduce the effects of scenes containing a lot of movement by

comparing the difference signal with a threshold derived from the maximum and minimum difference signals over a small aperture.

The algorithm produces false alarms, if the shot before/after the shot boundary includes high motion activity. The reason can be explained as follows: The weakness of the pixel based features is the high sensitivity to the video content. It is difficult for this algorithm to understand whether the change in the continuity signal is due to shot boundary or due to disturbances/motion. In order to enhance the algorithm, adaptive thresholding can be used. However, the high level of activity in the images around shot boundary produces a larger difference signal than expected. As a result adaptively obtained threshold is larger. A threshold that is larger than expected results in missed shot boundary.

The main disadvantage of this method is its inability to distinguish between a large change in a small area and a small change in a large area. It is observed that cuts are falsely detected when a small part of the frame undergoes a large, rapid change. For the same reason, the algorithm is not able to detect most of the flashlights.

## 2. Histogram Difference with Adaptive Thresholding

This method deals with global percentage of colors that an image contains. Here, we calculate the percentages from the bin total and compare them with adjacent frames giving a difference value. A difference above the threshold value will be classed as a shot change. Histogram difference method provides robust performance and better results compared to pixel-wise difference algorithm in case of slight illumination changes or small camera/object motion. On the other hand, global changes in the video frame such as zooming or fading effects, results in false alarms. This is an expected result, since histogram feature is sensitive to the overall content of the video. This algorithm cannot be used to detect shot boundary if there is a video-in-video effect, the transition is missed because the amount of histogram difference is small.

## 3. Edge change ratio

New edges appear for from the location of old edges and old edges disappear for from the location of new edges, during a cut or dissolve. Zabih et al[4] applied this observation to digital video segmentation. Two new type of edge pixels were identified by him

- Entering pixel: one that appear for from an existing edge pixel
- Existing pixel : One that disappear for from an existing edge pixel

It is possible to detect the CUTs and GTs by counting the entering and exiting pixels. The main disadvantages is its execution time, cut detection results can be accepted, if the algorithm could be faster however ,gradual transition results are not promising

## 4. Petersohn's Algorithm with 2-means clustering

The algorithm uses pixels, edge and histogram difference statics for detecting CUTS and GTs. The algorithm is quite fast and the system uses luminance information only and down samples all frames by a factor of 8 in x and y direction

The algorithm senses only significant changes in the video content, it can also employ pixel and histogram features together.

## 5. Segmentation method

In this method, we assume that each frame is a word and then the shot boundaries are treated as text segmentation based approaches in natural language processing can be used. The shot boundary detection process for a given video is carried out through 2 stages

- The frames are extracted and labeled with predefined labels.
- The shot boundaries are identified by grouping the labeled frames into segments.

To label frames in a video the following labels can be used: NORM FRM, PRE CUT, POST CUT, PRE GRAD, IN GRAD & POST GRADE

The method uses a support vector machine (SVM) for the purpose of shot boundary detection.

## 6. Motion-based algorithm

Here, the frames are first down sampled by a factor of 2 in both x and y direction. It also performs a preprocessing step for filtering out the non boundary frames. This algorithm is faster due to the processing step for skipping the frames which have less probability of being shot boundary. Also, utilizing down sampled images together significantly increases the algorithm speed.

## 7. Motion activity descriptor based algorithm

The motion activity is one of the motion features included in the visual part of the MPEG-7 standard. It is used to describe the level or intensity of activity, action, or motion in that video sequence. The main idea underlying the method of segmentation schemes is that imaged in the vicinity of a transition are highly dissimilar. The extraction of key frames method based on detecting a significant change in the activity of motion is used

## IV IMPLEMENTATION

The method for key frame extraction consists of three steps: Input a video and calculate the block based histogram difference of each consecutive frame. Choose the current frame as a candidate key frame whose histogram difference is above the threshold point. Extract the edges of the candidate key frames and calculate the edge matching rate of adjacent frames. If the edge matching rate is above average edge matching rate, the current frame is considered as a redundant frame and should be eliminated from the candidate key frames.

### A. Shot Boundary Detection

Let  $F(k)$  be the  $k^{\text{th}}$  frame in video sequence,  $k = 1, 2, \dots, F_v$  ( $F_v$  denotes the total number of video). The algorithm of shot boundary detection is described as follows.

#### Algorithm : Shot boundary detection

**Step 1:** Partitioning a frame into blocks with  $m$  rows and  $n$  columns, and  $B(i, j, k)$  stands for the block at  $(i, j)$  in the  $k^{\text{th}}$  frame;

**Step 2:** Computing  $x2$  histogram [14] matching difference between the corresponding blocks between consecutive frames in video sequence.  $H(i, j, k)$  and  $H(i, j, k+1)$  stand for the histogram of blocks at  $(i, j)$  in the  $k^{\text{th}}$  and  $(k+1)^{\text{th}}$  frame respectively. Block's difference is measured by the following equation:

$$D_B(k, k+1, i, j) = \sum_{l=0}^{L-1} \frac{[H(i, j, k) - H(i, j, k+1)]^2}{H(i, j, k)} \quad (1)$$

Where  $L$  is the number of gray in an image;

**Step 3:** Computing  $x2$  histogram difference between two consecutive frames:

$$D(k, k+1) = \sum_{i=1}^m \sum_{j=1}^n w_{ij} D_B(k, k+1, i, j) \quad (2)$$

Where  $w_{ij}$  stands for the weight of block at  $(i, j)$  ;

**Step 4:** Computing threshold automatically: Computing the mean and standard variance of  $x2$  histogram difference over the whole video sequence [15]. Mean and standard variance are defined as follows :

$$MD = \frac{\sum_{k=1}^{Fv-1} D(k, k+1)}{Fv - 1} \quad (3)$$

$$STD = \sqrt{\frac{\sum_{k=1}^{Fv-1} (D(k, k+1) - MD)^2}{Fv - 1}} \quad (4)$$

**Step 5:** Shot boundary detection

Let threshold  $T = MD + a \times STD$ . Shot candidate detection: if  $D(i, i+1) \geq T$  , the  $i^{\text{th}}$  frame is the end frame of previous shot, and the  $(i+1)^{\text{th}}$  frame is the end frame of next shot. Final shot detection: shots may be very long but not much short, because those shots with only several frames cannot be captured by people and they cannot convey a whole message. Usually, a shortest shot should last for 1 to 2.5 s. For the reason of fluency frame rate is at least 25 fps, (it is 30 fps in most cases), or flash will appear. So, a shot contains at least a minimum number of 30 to 45 frames. In our experiment, video sequences are down sampled at 10 fps to improve simulation speed. On this condition, the shortest shot should contain 10 to 15 frames. 13 is selected for our experiment. We formulate a “shots merging principle”: if a detected shot contain fewer frames than 13 frames, it will be merged into previous shot, or it will be thought as an independent one.

## B. Key Frame Extraction

### Algorithm: Key frame extraction

**Step 1:** Computing the difference between all the general frames and reference frame with the above algorithm:

$$D_C(1, k) = \sum_{i=1}^m \sum_{j=1}^n w_{ij} D_{CB}(1, k, i, j), k = 2, 3, 4 \dots F_{CN} \quad (5)$$

**Step 2:** Searching for the maximum difference within a shot:

$$\max(i) = \{D_C(1, k)\}_{\max}, k = 2, 3, 4, \dots F_{CN} \quad (6)$$

**Step 3:** Determining “ShotType” according to the relationship between  $\max(i)$  and  $MD$ : StaticShot(0) or DynamicShot:

$$\text{ShotType } C = \begin{cases} 1 & \text{if } \max(i) \geq MD \\ 0 & \text{others} \end{cases} \quad (7)$$

**Step 4:** Determining the position of key frame: if  $\text{ShotTypeC} = 0$ , with respect to the odd number of a shot's frames, the frame in the middle of shot is chose as key frame; in the case of the even number, any one frame between the two frames in the middle of shot can be choose as key frame. If  $\text{ShotTypeC} = 1$ , the frame with the maximum difference is declared as key frame.

## C. ELIMINATE REDUNDANT FRAMES

The candidate key obtained from the above method works well providing the main content of the given video, but there exists a small amount of redundancy, which has to be eliminated. Candidate key frames are based on histogram difference which depends on the distribution of pixel gray value in the image space. There may be redundancy in the event that two images whose content are the same but there is a great difference in the distribution of pixel gray value.

We can extract the edges of objects in the image to eliminate redundancy. Edge detection can remove the irrelevant information and retain important structural properties of the image. There are many edge detection operators that are used Robert's operator, Sobel operator, Prewitt operator and the Laplace operator etc however we extract edges of frames by Prewitt operator.



Figure 6: Edge detection of Images

## V. CONCLUSION

Shot boundary detection and key frame extraction system using image segmentation is the approach for video summarization. We have proposed block based  $x2$  Histogram algorithm for shot boundary detection. First video is segmented in frame, then

employed different weights to compute the matching difference and threshold. By using the automatic threshold, boundaries are detected. We detect various shot boundaries like Cut, Fade and Dissolve with the help of  $\chi^2$  histogram matching difference between consecutive frames and automatic threshold. Experimental results show that the proposed algorithm gives satisfactory performance for shot boundary detection. The contributions and characteristics of the proposed approach are summarized as follows:

- It is easy to implement and fast to compute. Only the  $\chi^2$  Histogram is to be found out in order to extract key frames. Thus leading to high efficiency.
- No Redundant Key frames, as Redundant Key Frames are removed using Edge detection algorithm. Thus the time increases to summarize the video.
- Detection of Zoom: Proposed algorithm is also efficient in detecting Key frames with Zoom and Object in front of camera effect.
- Higher Recall Rate and Precision Rate as compared to the algorithm proposed in [1,2].

## VI. FUTURE WORK

The approach used for efficient way of boundary detection is through  $\chi^2$  Histogram and Automatic Threshold. We can improve the performance of the algorithm by using the Graph Partition Model with Support Vector Machine as a new research in shot boundary detection. This algorithm is used in areas of computer vision and pattern recognition. Representing the data set in the form of an edge weighted graph converts the data clustering problem into a graph partitioning problem.

## REFERENCES

- [1] ZHAO Guang-sheng "A Novel Approach for Shot Boundary Detection and Key Frames Extraction" 2008 International Conference on Multimedia and Information Technology.
- [2] A. Nagasaka and Y. Tanaka, "Automatic Video indexing and full video search for object appearances, in Visual Database Systems II", Elsevier Science Publishers, pp113-117, 2002.
- [3] H.J. Zhang, A. Kankanhalli, and S.W. Smolar, "Automatic Partitioning of full motion video, in Multimedia Systems", Volume 1, pp 10-28, 2008.
- [4] Ramin Zabih, Justin Miller, Kevin Mai , "A Feature Based Algorithm for Detecting and Classifying Scene Breaks", in 3rd International Multimedia Conference and Exhibition, Multimedia Systems, pages 189-200, San Francisco, California 2010.
- [5] Y. Cheng, X. Yang, and D. Xu, "A method for shot boundary detection with automatic threshold", TENCON'02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communication, Control and Power Engineering[C], Vol.1, October 2002: 582-585.
- [6] Zuzana Cerneko, Ioannis Pitas "Information Theory-Based Shot Cut/Fade Detection and Video Summarization" in IEEE proc. in circuits and systems for video technology, VOL. 16, NO. 1, JANUARY 2006.
- [7] Ali Amiri and Mahmood Fathy "Hierarchical Keyframe-based VideoSummarization Using QR-Decomposition and Modified k-MeansClustering" in Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing, Volume 2010.
- [8] Y.Zhuang,Y.Rui,T S Huan, and S.Mehrotra,"Adaptive key frame extracting using unsupervised clustering," in Proc. Int. Conf. Image Processing, Chicago,IL,1998,pp.866-870
- [9] A. Hanjalic,"Shot Boundary detection:Unraveled and Resolved?",IEEE Transactions on Circuits and Systems for Video Technology, vol.12,no.2,pp.90-105,February 2002.
- [10] N.Doulamis,A.Doulamis,Y.Avrithies, and S.Kollias,"Video content representation using optimal extraction of frames and scenes,"in Proc.IEEE Int.Conf.Image Processing,Chicago,IL,1998,pp.875-879.
- [11] Sandip T.Dhagdi,P.R.Deshmukh,"Key Frame Based Video Summarization Using Automatic Threshold & Edge Matching Rate,"ijsrp,vol.2,Issue 7,July 2012