ISSN No. 0976-5697

Volume 3, No. 1, Jan-Feb 2012

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

A Novel Approach for Key Frames Extraction using Automatic Threshold

Mr. Sandip T. Dhagdi*
M.E. (C.E. Second Year)
Sipna COET, Amravati
sandip.yml@gmail.com

Dr. P.R. Deshmukh
Professor and Head of Computer science & IT Department,
Sipna's COET, Amravati
pr_deshmukh@yahoo.com

***Abstract*:** partitioning a video sequence into *shots* is the first step toward video-content analysis and content-based video browsing and retrieval. A video shot is defined as a series of interrelated consecutive frames taken contiguously by a single camera and representing a continuous action in time and space. As such, shots are considered to be the primitives for higher level content analysis, indexing, and classification.

This paper we analyze novel algorithm for shot boundary detection and key frames extraction. The algorithm differs from conventional methods mainly in the use of compression module and automatic threshold. Matching difference between two consecutive frames is computed with different weights. Shot boundaries are detected with automatic threshold. Key frame is extracted by using reference frame-based approach.

Algorithm uses histogram-based algorithms for finding out shot boundaries. Histogram-based algorithms are very applicable to SBD. They provide global information about the video content and are faster without any performance degradations.

Keywords: shot boundary detection; automatic threshold; key frame

I. INTRODUCTION

With the rapid advance of multimedia and Web technologies, video data in various formats are becoming available at an explosive rate. For example, based on a Yahoo! Answers post dated on June 2009, there were over 240,000,000 videos on YouTube, which is the most popular online video sharing Website. The time required to view all these videos was over 800 years. More amazingly, around 500,000 new videos were uploaded to YouTube everyday!

With such enormous video data resources, sophisticated video database systems are highly demanded to enable efficient browsing, searching and retrieval. However, the traditional video indexing method, which uses human beings to manually annotate or tag videos with text keywords, is time-consuming, lacks the speed of automation and is hindered by too much human subjectivity. Therefore, more advanced approaches such as content-based video retrieval are needed to support automatic indexing and retrieval directly based on videos content, which provide efficient search with satisfactory responses to the scenes and objects that the user seeks.

Video shot boundary detection, which segments a video by detecting boundaries between camera shots, is usually the first and important step for content-based video retrieval [1]. A video consists of a sequence of images (often being called frames), which can be played consecutively at the speed of around 20 to 30 frames per second in order to view smooth motion. To index and retrieval a video, shot boundary detection is usually conducted to segments the video into shots by detecting boundaries between camera shots.

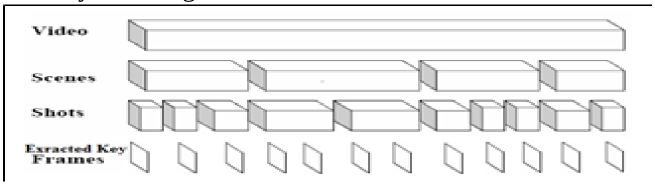

Figure 1.Overview of Shot boundary detection

A shot is defined as the consecutive frames from the start to the end of recording in a camera. It shows a continuous action in an image sequence. There are two different types of transitions that can occur between shots, abrupt (discontinuous) also referred as cut, or gradual (continuous) such as fades, dissolves and wipes [1]. The cut boundaries show an abrupt change in image intensity or colour, while those of fades or dissolves show gradual changes between frames.

a. A cut is an instantaneous transition from one scene to the next and it occurs over two frames.
b. A fade is a gradual transition between a scene and a constant image (fade out) or between a constant image and a scene (fade in).
c. A dissolve is a gradual transition from one scene to another, in which the first scene fades out and the second fades in.
d. A wipe occurs as a line moves across the screen, with the new scene appearing behind the line.

A. *Motivation:*

Recent developments in video compression technology, the widespread use of digital cameras, high capacity digital systems, coupled with the significant increase in computer performance and the growth of Internet and broadband communication, have increased the usage and availability of digital video. Applications such as multimedia information systems, distance learning, video on-demand produce and use huge amount of video data. This situation created a need for tools that can effectively categorize, search and retrieve the relevant video material.

In general, management of such activities over large collections of video requires knowledge of the "content" of the video. In particular, digital video data can be processed with the objective of extracting the information about the content conveyed with this data. The algorithms developed for this purpose, referred as "video content analysis" algorithms serve as the basis for developing tools that would enable us to understand the events and objects within the

© 2010, IJARCS All Rights Reserved

scene of a video, or generate summary of large video material or even to derive semantically meaningful information from the video [2].

The definition of "content" is highly application dependent but there are a number of commonalities in the applications of content analysis. Among others, *shot boundary detection* (SBD), also known as *temporal video segmentation* is one of the important aspects.

Parsing a video into its basic temporal units -shots- is considered as the initial step in the process of video content analysis. A shot is a series of video frames taken by a single camera, such as, for instance, by zooming into a person or an object, or simply by panning along a landscape [2]. The content is similar in shot regions. The regions where the significant content change occurs are, therefore, called shot boundaries. Since the SBD is a prerequisite step for most of the video applications involving the understanding, parsing, indexing, characterization, or categorization of video, temporal video segmentation has been an active topic of research in the area of content based video analysis.

B. Methods of Visual Content Representation:

Feature selection is the crucial step in the SBD process. The algorithms in this step can be summarized under two classes: algorithms run on the compressed domain and the algorithms run on the uncompressed domain. Following are the some of the methods for representing visual contents.

a. Pixel Based Methods:

Pixel based methods are the first and the most simple algorithms in the SBD literature. The basic idea behind pixel-based methods is that the intensity values of the pixels at the same locations of the sequential frames do not change significantly unless there is a shot boundary.

The initial pixel based algorithms [3] investigates the sum of absolute pixel intensity differences and if the difference is above a certain value a shot boundary is assigned. Even very small changes in the illumination or very small vibration in the camera can result in significant changes in total value of the pixel differences. Therefore, later algorithms count only the pixels that have changed significantly from one frame to another. If the total number of pixels that have changed is above a threshold, it is decided that there is a shot boundary between two frames.

Even with this improvement pixel based algorithms are still very sensitive to object/camera motion and illumination changes. More robust techniques use block based motion compensation and then apply the above algorithms.

Instead of pixel-wise differences, some of the literature proposes using statistics of pixel intensities between two frames. An example to this is comparing the blocks using a metric called *likelihood ratio,* which is simply based on statistical properties mean and variance.

Another method proposed by Zhang [10] against disturbances is to smooth the images by a 3x3 filter before performing pixel wise comparison.

Pixel-based methods are simple algorithms and do not require high computational power. The problem with pixel-based methods is high sensitivity to camera/object motion and disturbances.

b. Edge-Based Methods

Another feature that is proved to be useful in shot boundary detection is edges. Three edge-based features are mostly referred in the literature: Edge Change Ratio (ECR), Edge Contrast (EC) and Edge Energy.

Zabih [4] propose an edge-based technique based on the idea that during a shot transition new intensity edges are observed far from the locations of the old intensity edges. Similarly, old edges disappear far from the location of new edges. Moreover, the patterns in the appearance of new edges and disappearance of old edges are different for different types of transitions [2].

ECR algorithm employs motion compensation techniques prior to edge comparison. Therefore, this feature is robust against motion. On the other hand literature shows that ECR algorithm does not outperform histogram based algorithms in abrupt transition detection. The advantage of the ECR is that it can be used for detecting different types of transitions (i.e. cut, dissolve, fade, wipe).

Edge-based methods can be used both for abrupt and gradual transition detection. Edge-based methods require significant computational power. A method adopting edges as a feature is relatively more robust against motion but in general does not outperform histogram-based or pixel-based algorithms.

c. Motion-Based Methods:

Motion-based algorithms rely on the observation that while motion within a shot is smooth, motion between the frames that are surrounding a shot boundary tends to be abrupt. This assumption makes sense, because the motion of the objects or camera is generally smooth and continuous within the shot, which results in a continuous motion field. In contrast, abrupt changes are expected for the motion field at the shot boundary.

Motion-based algorithms in the uncompressed domain are computationally very expensive. Therefore, there are very few SBD algorithms based on motion vectors.

d. Histogram Based Methods:

Another example of a feature that is from the full pixel domain is the histogram. The reasoning is that the frames within the same shot should have similar colour histograms, while frames of different shots should have significantly different colour histograms. Earlier approaches compare gray level histograms [5] and recent methods utilize colour histogram information.

Several histogram comparison metrics are proposed in the literature. The most common techniques are: histogram difference, histogram intersection, cosine measure, Kolmogorov-Smirrov test and Chi-Square test. Research shows that histogram intersection formula performs best in the SBD area [5].

Twin-comparison is a method to detect gradual transitions using the colour histogram difference [5]. This method requires two thresholds. Abrupt transitions are detected using the higher threshold. A lower threshold is used on the remaining frames. A frame that differs from the previous frame by an amount above this threshold is declared as a potential start of a gradual transition. This frame is then compared to the subsequent frames to get the accumulated difference. During a gradual transition, this accumulated value will gradually increase. The end frame of a gradual transition is detected when the difference between consecutive frames drops below the lower threshold and the accumulated value has increased to a value that exceeds the higher threshold. If the difference between consecutive

© 2010, IJARCS All Rights Reserved

frames drops below the lower threshold before the accumulated difference exceeds the higher one, then the starting point is dropped and the search process is applied for other gradual transition candidates. Otherwise, a gradual transition is assigned [5].

As the histograms do not change with the spatial changes within a frame, histogram differences are more robust against the object motion with a constant background. However, histogram differences are also sensitive to camera motion, such as panning, tilting or zooming.

One can note that two images, which have completely different visual content, might still have similar histograms. However, research has shown that the probability of such events is low enough [5].

Similar to the pixel based methods, block based techniques can be utilized in order to improve the performance of the histogram based SBD algorithms. Histogram-based algorithms are less sensitive to object motion than the pixel based algorithms. Histogram-based algorithms are robust against global motion.

C. Thresholding:

Shot boundaries are identified based on the visual content change. Therefore, the most critical activity in the SBD process is the selection of the thresholds in any shot boundary detection step. The performance of the algorithm mainly remains in the thresholding phase. However, using a single threshold cannot perform equally well for all video sequences. Using a dynamic global threshold by extracting the overall sequence characteristic cannot solve this problem. Dynamic local thresholds are considered as a better alternative but thresholding still remains as a major problem in this area.

II. FUNDAMENTAL PROBLEMS OF SBD

Shot boundary detection (SBD) is not a new problem anymore. It has been studied more than a decade and resulting algorithms have reached some maturity. However, challenges still exist and are summarized in the upcoming sections:

A. Detection of Gradual Transitions:

During the video production process, first step is capturing the shots by using a single camera. Two consecutive shots are then attached together by a shot boundary that can either be abrupt or gradual. Abrupt shot boundaries are created by simply attaching a shot to another. While there is no modification in the consequent shots in an abrupt shot boundary, gradual transitions result from editing effects applied to the shots during attachment operation. According to the editing effect gradual transitions can be further divided into different subtypes. The number of possible transitions due to editing effect is quite high but most of the transitions fall into the three main categories: dissolve, fades (fade in, fade out), and wipes. Different types of transitions are demonstrated in the following figures:

Figure 2. Wipe Effect

Figure 3. Dissolve effect

Figure 4. Dissolve Effect

Figure 5. Fade In/Fade Out Effect

Detection of abrupt changes has been studied for a long time. On the other hand, gradual transitions pose a much more difficult problem. This situation is mainly due to the amount of available video editing effects. The problem gets harder when multiple effects are composed in the case of a lot of object or camera motion. Another reason is that the gradual transitions spread over time. Each editing effect has a different temporal pattern than the others and the temporal duration changes from three frames to hundred frames. Finally, the temporal patterns, as a result of editing effects to create a gradual transition, are very similar to the patterns due to camera/object motion. Therefore, gradual transitions remain to be one of the most challenging problems in SBD.

B. Flashlights:

Color is the primary element of video content. Most of the video content representations employ color as a feature. Continuity signals based on color feature exhibit significant changes under abrupt illumination changes, such as flashlights. Such a significant change might be identified as

© 2010, IJARCS All Rights Reserved

a content change (i.e. a shot boundary) by most of the shot boundary detection tools. Several algorithms propose using illumination invariant features, but these algorithms always face with a tradeoff between using an illumination invariant feature and loosing the most significant feature in characterizing the variation of the visual content. Therefore, flashlight detection is one of the major challenges in SBD algorithms.

C. Object/Camera Motion:

Visual content of the video changes significantly with the extreme object/camera motion and screenplay effects (e.g. one turns on the light in a dark room) very similar to the typical shot changes. Sometimes, slow motion cause content change similar to gradual transitions, whereas extremely fast camera/object movements cause content change similar to cuts. Therefore, it is difficult to differentiate shot changes from the object/camera motion based on the visual content change. Therefore, the most critical activity in the SBD process is the selection of the thresholds in any shot boundary detection step. The performance of the algorithm mainly remains in the thresholding phase. However, using a single threshold cannot perform equally well for all video sequences. Using a dynamic global threshold by extracting the overall sequence characteristic cannot solve this problem. Dynamic local thresholds are considered as a better alternative but thresholding still remains as a major problem in this area.

III. LITERATURE REVIEW & RELATED WORK:

SBD is a popular area in the video content analysis and has been studied for a long time. Research has resulted in a variety of algorithms. In this section, briefly the SBD work in the literature is reviewed.

Many approaches used different kinds of features to detect shot boundary, including histogram, shape information, motion activity. Among these approaches, histogram is the popular approach. However, in these histogram-based approaches, pixels' space distribution was neglected. Different frames may have the same histogram.

In view of this, Cheng [6] divided each frame into r blocks, and the difference of the corresponding blocks of consecutive frames was computed by color histogram; the difference $D(i,i+1)$ of the two frames was obtained by adding up all the blocks' difference; in the meanwhile, the difference $V(i,i+1)$ between two frames i and $i+1$ was measured again without using blocks. Based on $D(i,i+1)$ and $V(i,i+1)$, shot boundary was determined[1].

Zhuang [7] proposed an unsupervised clustering method. A video sequence is segmented into video shots by clustering based on color histogram features in the HSV color space. For each video shot, the frame closest to the cluster centroid is chosen as the key frame for the video shot. Notice that only one frame per shot is selected into the video summary, regardless of the duration or activity of the video shot.

Zuzana Cernekova [8] proposed 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. Mutual information is a measure of information transported from one frame to another one. It is used for detecting abrupt cuts, where the image intensity or color is abruptly changed. A large video content difference between two frames, showing weak inter-frame dependency leads to a low MI. The entropy measure provides with better results, because it exploits the inter-frame information flow in a more compact way than a frame subtraction.

Ali Amiri [9] proposed a novel video summarization algorithm which is based on QR-decomposition. Some efficient measures to compute the dynamicity of video shots using QR-decomposition was utilize in detecting the number of key frames selected for each shot. Also, a corollary that illustrates a new property of QR-decomposition. This property was utilized in order to summarize video shots with low redundancy.

Hanjalic [10] developed a similar approach by dividing the sequence into a number of clusters, and finding the optimal clustering by cluster-validity analysis. Each cluster is then represented in the video summary by a key frame. The main idea in this paper was to remove the visual redundancy among frames.

DeMenthon [11] proposed an interesting alternative based on curve simplification. A video sequence is viewed as a curve in a high dimensional space, and a video summary is represented by the set of control points on that curve that meets certain constraints and best represent the curve.

Doulamis [12] also developed a two step approach according to which the sequence is first segmented into shots, or scenes, and within each shot, frames are selected to minimize the cross correlation among frames' features.

IV. ANALYSIS OF PROBLEM

The demand for intelligent processing and analysis of multimedia information has been rapidly growing in recent years [13]. Researchers have actively developed different approaches for intelligent video management, including shot transition detection, key frame extraction, video retrieval, etc[13]. Among these approaches, shot transition detection is the first step of content-based video analysis and key frame is a simple yet efficient form of video abstract. It can help users to understand the content at a glance and is of practical value.

Many approaches used different kinds of features to detect shot boundary, including histogram, shape information, motion activity. Among these approaches, histogram is the popular approach. However, in these histogram-based approaches, pixels' space distribution was neglected. Different frames may have the same histogram. In view of this, Cheng [6] divided each frame into r blocks, and the difference of the corresponding blocks of consecutive frames was computed by colour histogram; the difference $D(i,i+1)$ of the two frames was obtained by adding up all the blocks' difference; in the meanwhile, the difference $V(i,i+1)$ between two frames i and $i+1$ was measured again without using blocks. Based on $D(i,i+1)$ and $V(i,i+1)$, shot boundary was determined.

A. Image Segmentation:

Each frame is divided into blocks with m rows and n columns. Then the difference of the corresponding blocks between two consecutive frames is computed. Finally, the

© 2010, IJARCS All Rights Reserved

final difference of two frames is obtained by adding up all the differences through different weights.

B. Attention Model:

Attention, a neurobiological concept, means the concentration of mental powers upon an object by close or careful observing or listening, which is the ability or power to concentrate [7]. Attention model means that, from the visual viewpoint, different contents are ranked based on importance. Zhuang [7] proposed a face attention model, which thought that face's size and position reflect the importance of protagonists. Correspondingly, it also reflected the importance of frames. Based on the consideration, we think that different position's pixels have different contribution to shot boundary detection: pixels on the edge are more important than others. Thus, different weights are given to blocks of different position. Both the space distribution characteristic of pixels of different gray and the different importance of pixels of different position are considered.

C. Matching Difference:

There are six kinds of histogram match. Colour histogram was used in computing the matching difference in most literatures. However, through comparing several kinds of histogram matching methods, Nagasaka reached a conclusion that x^2 histogram outperformed others in shot boundary recognition.

V. PROPOSED WORK

Project will consist of following two modules.

a. Shot boundary detection
b. Key frame extraction

Following is the explanation of each module with their algorithm.

A. Shot Boundary Detection:

Let $F(k)$ be the k th frame in video sequence, $k = 1,2,\ldots,Fv$ (Fv 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 th frame;

Step 2: Computing x^2 histogram [13] 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 kth and $(k+1)$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 x^2 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 wij stands for the weight of block at (i, j) ;

Step 4: Computing threshold automatically:

Computing the mean and standard variance of x^2 histogram difference over the whole video sequence [13]. 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 ith frame is the end frame of previous shot, and the $(i+1)$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.

Definition 1: Reference Frame: it is the first frame of each shot; General Frames: all the frames except for reference frame; "Shot Dynamic Factor" max(i): the maximum x^2 histogram within shot i;

Dynamic Shot and Static Shot: a shot will be declared as dynamic shot, if its max(i) is bigger than MD; otherwise it is static shot; $F_C(K)$: the kth frame within the current shot, k=1,2,3… FCN (K) (FCN (k) is the total number of the current shot).

B. Key Frame Extraction:

The algorithm of key frame extraction is described as follows.

Algorithm: Key frame extraction

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

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

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

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

Step 3: Determining "ShotType" according to the relationship between max(i) and MD: StaticShot(0) or DynamicShot:

$$ShotTypeC = \{ \begin{matrix} 1 if \max(i) \geq MD \\ 0 Others \end{matrix} \} \quad (7)$$

Step 4: Determining the position of key frame: if $ShotType_C = 0$, 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

© 2010, IJARCS All Rights Reserved

between the two frames in the middle of shot can be choose as key frame. If $ShotType_C = 1$, the frame with the maximum difference is declared as key frame.

VI. APPLICATIONS

A. Video Restoration:

Video shot boundary detection, which segments a video by detecting boundaries between camera shots, is usually the first and important step for content-based video retrieval. A video consists of a sequence of images (often being called frames), which can be played consecutively at the speed of around 20 to 30 frames per second in order to view smooth motion. To index and retrieval a video, shot boundary detection is usually conducted to segments the video into shots by detecting boundaries between camera shots.

The development of shot-boundary detection algorithms has the longest and richest history in the area of content-based video analysis and retrieval—longest, because this area was actually initiated some decade ago by the attempts to detect hard cuts in a video, and richest, because a vast majority of all works published in this area so far address in one way or another the problem of shot-boundary detection. This is not surprising, since detection of shot boundaries provides a base for nearly all video abstraction and high-level video segmentation approaches. Therefore, solving the problem of shot-boundary detection is one of the major prerequisites for revealing higher level video content structure. Moreover, other research areas can profit considerably from successful automation of shot-boundary detection processes as well.

A good example is the area of video restoration. There, the restoration efficiency can be improved by comparing each shot with previous ones and—if a similar shot in terms of visual characteristics is found in the past—by adopting the restoration

Video shot boundary detection, which segments a video by detecting boundaries between camera shots, is usually the first and important step for content-based video retrieval. A video consists of a sequence of images (often being called frames), which can be played consecutively at the speed of around 20 to 30 frames per second in order to view smooth motion. To index and retrieval a video, shot boundary detection is usually conducted to segments the video into shots by detecting boundaries between camera shots.

B. Video Indexing and Browsing:

Due to the recent progress in the decreasing storage costs and the growing availability of broadband data connection, digital videos are becoming widely used. However, the increasing availability of digital video has not been accompanied by an increase in its accessibility. This is due to the nature of video data, which is unsuitable for traditional forms of data access, indexing, browsing, and retrieval. Traditional forms of data retrieval are either text based or based on the query-by-example paradigm. If we want to find a clip of interest, we have to sequentially browse through the video.

This is an extremely time consuming, tedious and labor-intensive process. Therefore, the demand of new technologies and tools for effective and efficient indexing, browsing and retrieval of video data has been exacerbated by recent trends. The area of content based video retrieval, aiming to automate the indexing, retrieval and management of video, has attracted extensive research during the last years. The first step of content based video retrieval is shot boundary detection.

C. Video surveillance system:

In a video surveillance system it takes time to watch and understand lengthy video. Depending on length it will take that much of time to watch it, using shot boundary detection and key frame extraction we can summaries the video into key frames or short messages which will convey the various messages in video. It will take very less time to understand lengthy video using shot messages.

Video shot boundary detection, which segments a video by detecting boundaries between camera shots, is usually the first and important step for content-based video retrieval. A video consists of a sequence of images (often being called frames), which can be played consecutively at the speed of around 20 to 30 frames per second in order to view smooth motion. To index and retrieval a video, shot boundary detection is usually conducted to segments the video into shots by detecting boundaries between camera shots.

VII. CONCLUSION

Shot boundary detection and key frame extraction system using image segmentation is a novel approach for video summarization. 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. The extracted key frames can satisfactorily represent the content of video. In order to further improve accuracy, multimodal information to segment video and generate video abstract can be used. Multimodality-based video indexing can be future direction.

VIII. REFERENCES

[1] Naimish.Thakar "Analysis and Verification of Shot Boundary Detection in Video using Block Based χ2 Histogram Method" International Journal of Advances in Electronics Engineering.

[2] 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.

[3] A. Hanjalic, Content-based Analysis of Digital Video, Boston: Kluwer Academic Publishers, 2004.

[4] R. Zabih, J. Miller, and K. Mai, "A Feature-Based Algorithm for Detecting and Classifying Scene Breaks," Proc. ACM Multimedia 95, pp. 189-200, 1995.

[5] J. Mas, and G. Fernandez, "Video Shot Boundary Detection based on Colour Histogram," Notebook Papers TRECVID2003, 2003.

[6] 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.

[7] Y. Zhuang, Y. Rui, T. S. Huan, and S. Mehrotra, "Adaptive key frame extracting using unsupervised clustering," in

© 2010, IJARCS All Rights Reserved

Proc. Int. Conf. Image Processing, Chicago, IL, 1998, pp. 866–870.

[8] Zuzana Cerneková, 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.

[9] Ali Amiri and Mahmood Fathy “Hierarchical Keyframe-based Video Summarization Using QR-Decomposition andModified k-Means Clustering” in Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing, Volume 2010.

[10] A. Hanjalic and H. Zhang, “An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis,” IEEE Trans. Circuits Syst. Video Technol., vol. 8, pp. 1280–1289,Dec. 1999.

[11] D. DeMenthon, V. Kobla, and D. Doermann, “Video summarization by curve simplification,” in Proc. 6th Int. ACM Multimedia Conf., Bristol, U.K., 1998, pp. 211–218.

[12] N. Doulamis, A. Doulamis, Y. Avrithis, and S. Kollias, “Video content representation using optimal extraction of frames and scenes,” in Proc. IEEE Int. Conf. Image Processing, Chicago, IL, 1998, pp. 875–878.

[13] ZHAO Guang-sheng “A Novel Approach for Shot Boundary Detection and Key Frames Extraction”, 2008 International Conference on Multimedia and Information Technology.

© 2010, IJARCS All Rights Reserved