ISSN No. 0976-5697

Volume 3, No. 3, May-June 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

New Technique for Keyframe Extraction Using Block Based Histogram

Mr. Sandip T. Dhagdi
M.E. (C.E. Second Year)
Sipna's COET, Amravati, India
sandip.yml@gmail.com

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

Abstract: Shot boundary detection and Keyframe Extraction is a fundamental step for organization of large video data. Key frame extraction has been recognized as one of the important research issues in video information retrieval. 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. This paper discusses the importance of key frame extraction; briefly review and evaluate the existing approaches, to overcome the shortcomings of the existing approaches.

This paper also presents a new approach for key frame extraction based on the block based Histogram difference and edge matching rate. Firstly, the Histogram difference of every frame is calculated, and then the edges of the candidate key frames are extracted by Prewitt operator. At last, the paper makes the edges of adjacent frames match. If the edge matching rate is up to 50%, the current frame is deemed to the redundant key frame and should be discarded.Histogram-based algorithms are very applicable to SBD, They provide global information about the video content and are faster without any performance degradations.

Keywords- key frame extraction; Histogram Difference; Prewitt operator; edge matching rate

I. INTRODUCTION

With the development of multimedia information technology, the content and the expression form of the ideas are increasingly complicated. How to effectively organize and retrieve the video data has become the emphasis of the study. The technology of the key frame extraction is a basis for video retrieval. The key frame which is also known as the representation frame represents the main content of the video. Using key frames to browse and query the video data greatly reduces the amount of processing data. Moreover, key frames provide an organizational framework for video retrieval.

However, efficient access to video is not an easy task due to video's length and unstructured format. Sophisticated video database systems are highly demanded to enable efficient browsing, searching and retrieval. 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

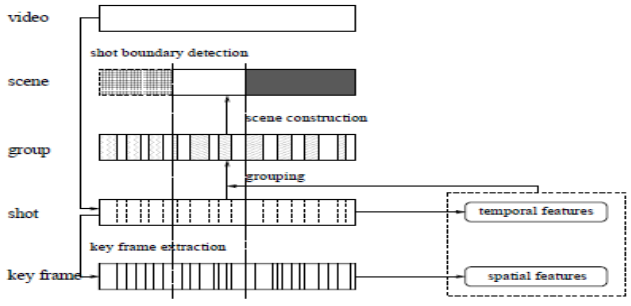

Figure 1. A Hierarchical Video Representation.

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.

II. RELATED KEY FRAME EXTRACTION TECHNIQUES

Feature selection is the crucial step in the SBD process. Following are the some of the methods for representing visual contents.

A. Shot Boundary Based Approach:

After the video stream is segmented into shots, a natural and easy way of key frame extraction is to use the first frame of each shot as the shot's key frame [3]. Although simple, the number of key frames for each shot is limited to one, regardless of the shot's visual complexity. Furthermore, the first frame normally is not stable and does not capture the major visual content.

B. Visual Content Based Approach:

Zhang propose to use multiple visual criteria to extract key frames [4].

a. Shot based criteria: The first frame will always be selected as the first key frame; but, whether more than one key frame need to be chosen depends on other criteria.
b. Color feature based criteria: The current frame of the shot will be compared against the last key frame. If significant content change occurs, the current frame will be selected as a new key frame.
c. Motion based criteria: For a zooming like shot, at least two frames will be selected: the first and last frame, since one will represent a global, while the other will represent a more focused view. For a panning-like shot, frames have less than 30% overlap is selected as key frames.

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

C. Motion Analysis Based Approach:

Wolf proposes a motion based approach to key frame extraction [5]. He first computes the optical flow for each frame, and then computes a simple motion metric based on the optical flow. Finally he analyzes the metric as a function of time to select key frames at the local minima of motion. The justification of this approach is that in many shots, the key frames are identified by stillness either the camera stops on a new position or the characters hold gestures to emphasize their importance [5].

D. Shot Activity Based Approach:

Motivated by the same observation as Wolf's, Gresle and Huang[6] propose a shot activity based approach. They first compute the intra and reference histograms and then compute an activity indicator. Based on the activity curve, the local minima are selected as the key frames.

E. Histogram Based Approach:

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 [7] 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 [7].

Twin-comparison is a method to detect gradual transitions using the colour histogram difference [7]. 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 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 [7].

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 [7].

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.

III. 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:

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.

Figure 2. Wipe Effect

Figure 3. Dissolve effect

Figure 4. Dissolve Effect

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

Figure 5. Fade In/Fade Out Effect

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 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.

IV. ANALYSIS OF PROBLEM

The demand for intelligent processing and analysis of multimedia information has been rapidly growing in recent years [8]. Researchers have actively developed different approaches for intelligent video management, including shot transition detection, key frame extraction, video retrieval, etc[8]. 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 [9] 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 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 [10]. Attention model means that, from the visual viewpoint, different contents are ranked based on importance. Zhuang [10] 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 out performed others in shot boundary recognition.

V. PROPOSED WORK

Project will consist of following three modules.

a. Shot boundary detection
b. Key frame extraction
c. Eliminate Redundant Frames

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 [8] 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:

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

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

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 [8]. Mean and standard variance are defined as follows :

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

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

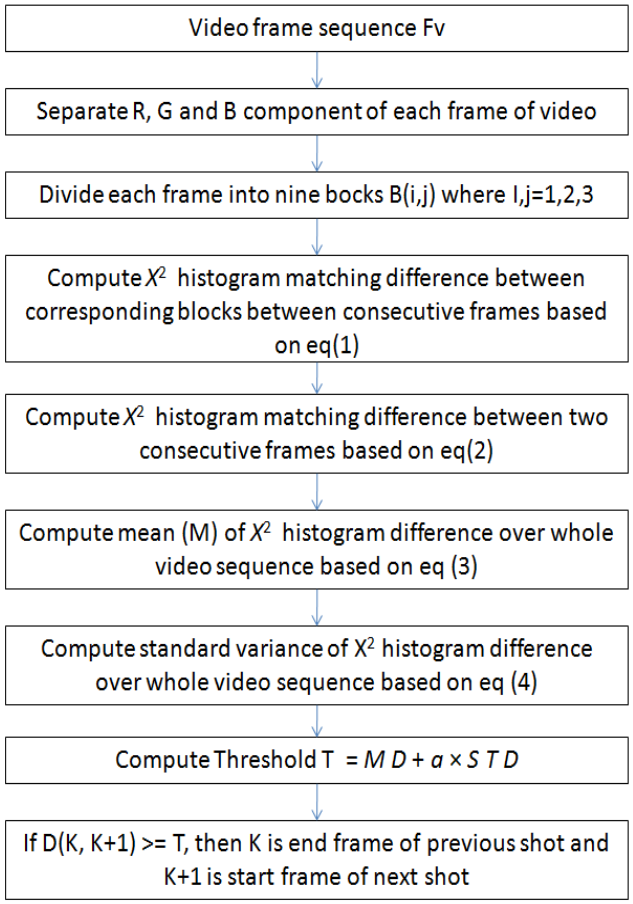

Figure 6. Flowchart of shot boundary detection algorithm.

Step 5: Shot boundary detection

Let threshold $T = M D + a \times S T D$. Shot candidate detection: if $D(i,i+1) \geq T$, the *i*th 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 *k*th 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:

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

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 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.

C. Eliminate Redundant Frames:

a. Extract Edges of the Candidate Key Frames:

The candidate key frames obtained from the above treatment do well in reflecting the main content of the given video, but exist a small amount of redundancy, which need further processing to eliminate redundancy. As the candidate key frames are mainly based on the Histogram difference which depends on the distribution of the pixel gray value in the image space, there may cause redundancy in the event that two images whose content are the same exist great difference from the distribution of the pixel gray value. For example, the substance content of images a and b in Fig.7 don 't change, but the two images are both identified as key frames as a result of the different gray value distribution, resulting in redundancy.

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

Figure 7. Redundant Frames a and b

Figure 8. Edge Images of A and B.

As edge detection can remove the irrelevant information and retain important structural properties of the image, we can extract the edges of objects in the image to eliminate redundancy. At present, there are many edge detection algorithms, which are mainly based on the differentiation and combined with the template to extract edges of images.

Edge detection operators that are commonly used are: Roberts [15] operator, Sobel operator, Prewitt operator and the Laplace operator etc. Here we extract edges of frames by Prewitt operator.

b. Eliminate Redundant Frames Based on the Edge Matching Rate:

The edge images have no difference in the distribution of the gray value. For example, the images shown in Fig.8 are the edge images of the images shown in Fig.7 with Prewitt operator and both of them are remarkably similar. So we use the edge matching rate to match the edges of adjacent frames to eliminate redundant frames. The formula for calculating the edge matching rate is as follows:

$$p(f_i, f_{i+1}) = s / n \quad (8)$$

In the formula,

$$n = \max(n_{fi}, n_{fi+1}) \quad (9)$$

$$s = \sum_{i}^{m} \sum_{j}^{n} h(i, j) \quad (10)$$

$$h(i, j) = \begin{cases} 1, v_{fk}(i, j) = v_{fk+1}(i, j) \\ 0, Otherwise \end{cases} \quad (11)$$

Where $v_{fk}(i,j)$ and $v_{fk+1}(i,j)$ are the pixel values of the position (i, j) in the frame *fk* and the frame *fk+1*, respectively. *m* and *n* indicate the height and the width of the image, *nfi* and *nfi+1* represent the number of the pixels on the edge of the frame *fi* and the frame *f* i+1 respectively. Assume the keyframe sequence as {f1,f2,f3,…….fk} (the total number of the candidate key frames is k), we make use of the following steps to eliminate redundant frames:

a) Use the Prewitt operator to extract edges of the candidate key frames and obtain their corresponding edge images.
b) Set j=2.
c) Calculate the edge matching rate *p(fj-1 , fj)* between the current frame *fj* and the previous frame *fj-1* with the formula (8). If *p(fj-1 , fj)* is up to 50%, the current frame *fi* will be marked as a redundant frame.
d) j =j+1, if j>k, go to (5). Otherwise, return to (3) and continue processing the remaining frames.
e) Remove the frames which have been marked as redundant frames from the candidate key frames. The remaining candidate key frames are the ultimate key frames. With the edge detection and edge matching, we eliminate redundant key frames, improve the accuracy rate of the key frame extraction and reduce the redundancy.

VI. 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.

VII. 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. Nagasaka and Y. Tanaka, \Automatic video indexing and full-video search for object appear-ances," in Visual Database Systems II, 1992.

[4] H. Zhang, J. Wu, D. Zhong, and S. W. Smoliar, \An integrated system for content-based video retrieval and browsing," Pattern Recognition, vol. 30, no. 4, pp. 643{658, 1997.

[5] 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.

[6] W. Wolf, \Key frame selection by motion analysis," in Proc. IEEE Int. Conf. Acoust., Speech,and Signal Proc., 19

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

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

[9] Y. Cheng, X. Yang, and D. Xu, "A method for shot boundary detection with automatic threshold", TENCON'02. Proceedings. 2002 IEEE Region 10

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

Conference on Computers, Communication, Control and Power Engineering[C], Vol.1, October 2002: 582-585.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] Kintu Patel, "Key Frame Extraction Based on Block based Histogram Difference and Edge Matching Rate" International Journal of Scientific Engineering and Technology, Volume No.1, Issue No.1

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html