# Sports Video Summarization using Priority Curve Algorithm

K Susheel Kumar [1], Shitala Prasad[2], Santosh Banwral[3], Vijay Bhaskar Semwal[4]
[1]Department of Computer Science & Engineering
[2,3]Department of Information Technology
[1]Ideal Institute of Technology, Ghaziabad, India. &
[2,3]Indian Institute of Information Technology, Allahabad, India.

***Abstract*-- The noble technique, Video summarization is process to represent the content of video in compact manner. It is basically of two type's static video summary and dynamic video skimming based. Static video summary it is the process that selects a set of salient images (called key frames) extracted or synthesized from the original video to represent the video contents. Dynamic video skimming it is shorter video of the original video made up of several short video clips it can be done in two ways. Present study mainly aims to extract the main events (highlights of the match) from cricket video. And make a short summary of the match so that it can take a small memory space as well as for fast content browsing, transmission, and retrieval.**

***Keywords: - Video compression, discrete wavelet transforms, Feature Extraction, Optical Character Recognition (OCR), and SVM***

## I. INTRODUCTION

Video enriches the content delivery by combining visual, audio, and textual information in multiple data streams. Thus it is always the favourite medium for most people and communication entities for its extraordinary expressive power. With the fast development of the network bandwidth and large-capacity storage devices, video data has become pervasive on the Internet in these days. On today's network, many companies provide video sharing services, which further speeds up the growth of the volume of Internet videos. People can retrieve and enjoy multimedia information in the form of text, image and particularly video. Moreover, individual people also begin to share their own edited videos. In 2000, a survey by PC Data showed that an estimated 57.2% of Internet users watch online video clips, and 7.3% of them edited video clips on their personal Computers [1]. In 2003, another survey by Broadband4Britain.com [2] shows that 34% of British broadband user in often enjoys video-on-demand service. On January 2006, there are more than 469 million digital media documents created globally with 315 million of them being actively accessed [3]. AltaVista [4] has been serving around 25 million search queries per day, with its multimedia search featuring over 45 million images, videos and audios. People watch more than 70 million videos on YouTube [5] daily to see first-hand accounts of current events, find videos about their hobbies and interests, and discover the quirky and unusual through the Internet. Google Inc [6] in 2000 started the world's first open online video marketplace (Google Video) where users can search for, watch and buy an ever-growing collection of TV shows, movies, Music Video, documentaries, personal productions etc. What's more, with the ability to watch and share videos world-wide through the Internet, amateurs can also capture their own moments on video and become future broadcasters [5].

## II. REVIEW OF RELATED WORK

To solve this problem, video summarization, which engage in providing concise and informative video summaries to help people to browse and manage video files efficiently, have received more and more attention in these years. Basically there are two kinds of video summaries.

### *A. Static video summary*

Static video story board, which is composed of a set of salient images (key frames) extracted or synthesized from the original video. Based on the way a key frame is extracted, existing work in this area can be categorized into three classes: sampling based, shot based, and segment based.

1) *Shot based:* A shot is defined as a video segment taken from a continuous period; a natural and straightforward way is to extract one or more key frames from each shot using low-level features such as color and motion. A typical approach was proposed in [9], where key frames were extracted in a sequential fashion via thresholding. The first frame with in the shot is always chosen then the color-histogram difference between the sub sequent frames and the latest key frame is computed. Once the difference exceeds a certain threshold, anew key frame will be extracted. One drawback of the shot-based key frame extraction approach is that it does not scale up well for long video. Figure-1 shows the shot based summary.

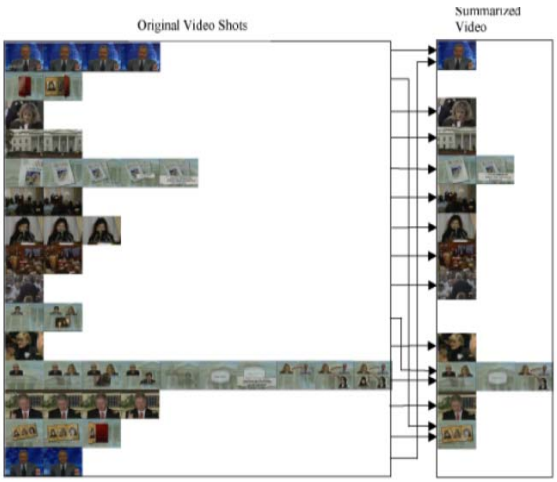


Figure 1. Shot based summary.

2) *Sampling based:* The sampling-based, where key frames were either randomly chosen or uniformly sampled from the original video. The video Magnifier [7] and the Mini Video [8] systems are two examples. This approach is the simplest way to extract key frames, yet such an arrangement may fail to capture the real video content, especially when it is highly dynamic.

3) *Segment based:* More recently, efforts have been made in extracting key frames at a higher unit level, referred to as the segment level. Various clustering-based extraction schemes have been proposed. In these schemes, segments are first generated from frame clustering and then the frames that are closest to the centroid of each qualified segment are chosen as key frames [10], [11].

4) *Dynamic video skimming:* Dynamic video skimming, which a shorter version of the original video is made up of several short video clips.

### B. Video Structure analysis

A video is composed of images and audio and textual. The images is converted into frames and the video play at least 25 to 30 frames per second. Each video also uses some video compression. Video compression refers to reducing the quantity of data used to represent video images and is a straightforward combination of image compression and motion compensation. There are four methods for compression, discrete cosine transforms (DCT), vector quantization (VQ), fractal compression, and discrete wavelet transforms (DWT).

1) *Discrete cosine transform:* Discrete cosine transform is a lossy compression algorithm that samples an image at regular intervals, analyzes the frequency components present in the sample, and discards those frequencies which do not affect the image as the human eye perceives it. DCT is the basis of standards such as JPEG, MPEG, H.261, and H.263.

2) *Vector Quantization:* Vector quantization is a lossy compression that looks at an array of data, instead of individual values. It can then generalize what it sees, compressing redundant data, while at the same time retaining the desired object or data stream's original intent.

3) *Fractal Compression:* Fractal compression is a form of VQ and is also a lossy compression. Compression is performed by locating self-similar sections of an image, then using a fractal algorithm to generate the sections.

4) *Discrete Wavelet Transforms:* Like DCT, discrete wavelet transform mathematically transforms an image into frequency components. The process is performed on the entire image, which differs from the other methods (DCT) that work on smaller pieces of the desired data. The result is a hierarchical representation of an image, where each layer represents a frequency band.

## III. PLAN OF WORK

### A. Structural Analysis for Sports Videos

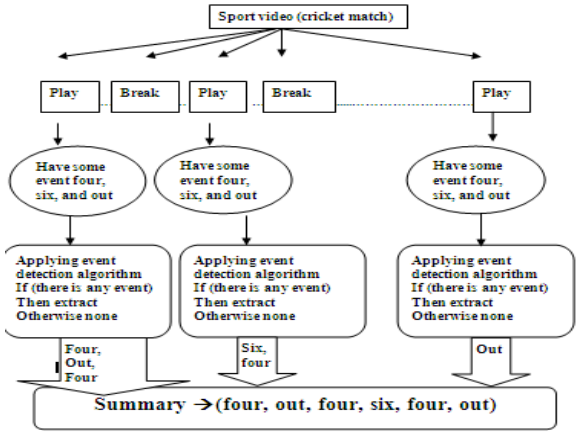


Figure 2. Cricket video summarization process.

1) *Play-Break Detection:* Sports video is composed of play events, which refer to the times when the ball is in-play, and break events, which refers to intervals of stoppage in the game. The play-break was detected by thresholding the duration of the time interval between consecutive long shots.

2) *Event Detection:* The sport highlight (cricket match) detection means to automatically extract the most interesting segments, also known as game highlights, from the full-length sports video. The highlight portions in sports video can usually be distinguished by the detection of certain low-level feature patterns, e.g. the occurrence of replay scene, the excited audience or commentator speech, certain camera motion or certain highlight event related sound, and crowd excitement.

### B. Feature Extraction

In this paper for event detection we are taking visual, audio and textual features.

1) *Visual Feature Extraction:* In image/video processing research, features such as edge, texture (grass ratio), etc are widely adopted. As features however do not possess high-level semantics, they are referred to as "Low-level" features. In sports video analysis literature, researchers usually use these low-level features to extract semantic information for high-level analysis. Umpire action always gives the important information about an event in case of cricket match. Grass ratio is defining as the ratio of play ground grass color pixel verses other pixel in each frame. In case of cricket video whenever a player hit a shot (four, six) the grass pixel ratio decreases from play ground to boundary in each frame. If the numbers of frames contain the pixel ratio from a particular threshold below and above then there may be boundary shot. In this paper I am using the grass ratio for detecting the boundary shot. The figure-3 shows the complete calculation process of grass pixel ratio per frame and on the basis of extraction visual features of cricket video.

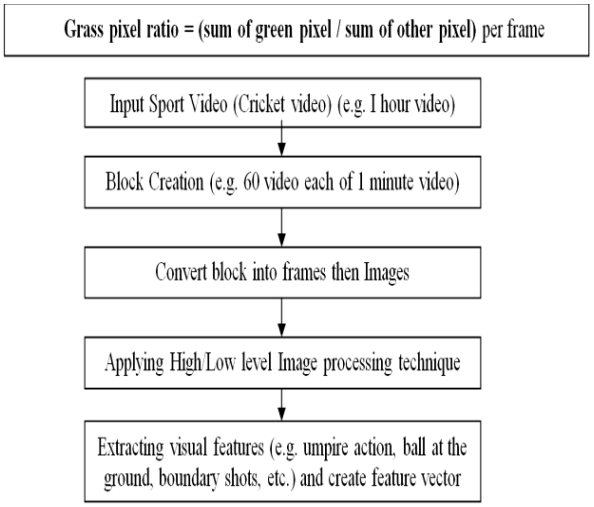


Figure 3. Shows the complete process of extraction audio feature of cricket video.

2) *Audio Feature Extraction:* There are certain game-specific audio signals such as applause, booing and loudness that are possible indicators of important events occurrence. Hence some researchers have used the audio data to identify high-level semantics. When exciting event occur, generally the crowd's cheer and commentator's speech become louder, and higher (in pitch) and less pauses occur. To localize louder clips, we used this equation to calculate the volume of each audio frame:

$$\text{Volume}=\frac{1}{N}*\sum_{n=1}^{N}|s(n)| \quad \ldots (1)$$

where N is the number of frames in a clip and s(n) is the sample value of the nth frame.

The audio features, namely, "Root mean square volume", "Zero crossing rate", 'Pitch Period", "Frequency Centroid", "Frequency Bandwidth" and "Energy Ratio" to discriminate among news reports, commercials, weather forecasts, football videos, basketball video and cricket video clips.

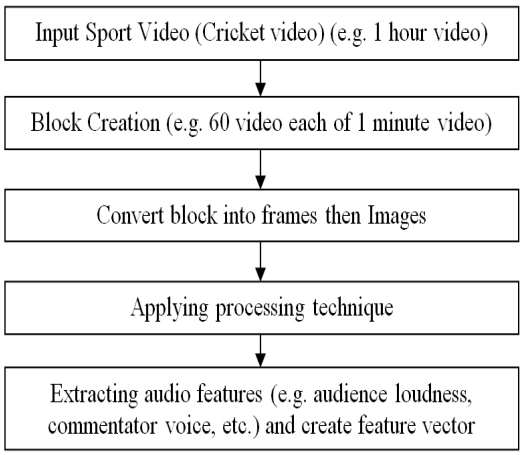


Figure 4. Shows the complete process of extraction audio feature of cricket video.

3) *Textual Feature Extraction:* In sports broadcasting there are always texts that can be recognized from the video frame. These detected texts can be divided into two classes: Scene text such as scoreboard, audio content the caption text usually contains mush writings on human clothes, etc, and Caption text that is mechanically superimposed over video frames to supplement the visual and useful information about the game, and many researchers have proposed

Techniques to detect and recognize caption text to assist sports video analysis. For example, Assfalg et al. [13] defined heuristic rules to detect caption, e.g., the caption must remain stable for a period of time for the audience to read, and caption should have high luminance contrast. Chen et al. [14] used SVM to identify single text line in video frames and perform Optical Character Recognition (OCR) for text recognition.

To verify that the detected lines are the candidates of a text display region, the system only retains the lines that follow these criteria:

1) The absolute value of *r* is less than *n* percent of the maximum *y*-axis, and

2) The corresponding *t* is equal to 90 (horizontal). This *n*-value represents threshold2, the maximum possible location of the horizontal line in terms of the *y*-axis.

The first check is important to ensure that the location of the lines is within the usual location for a text display. The second check is to ensure that the line is horizontal because there are potentially other prominent horizontal lines that can be detected from other areas besides the text display, such as the boundary between a field and a crowd. Finally, for each of the lines detected, the system checks that their location (the *r* values) is consistent for at least *m* seconds (that is, if the video frame rate is 25, 2 seconds is equal to 50 frames). We consider this *m*-value as threshold3, the

minimum period (in terms of seconds) that the lines must stay in a consistent location.

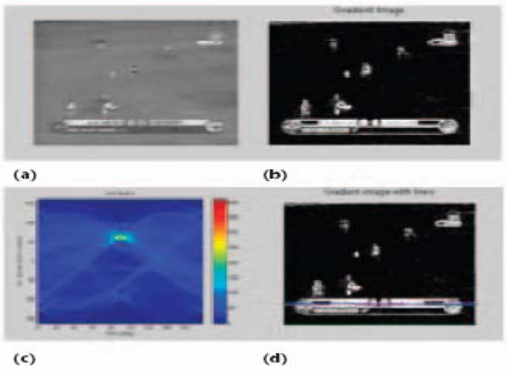


Figure 5. *How the Sobel filter and Hough transform detects text: Grayscale and resized frame, (b) Gradient image reveals the lines, (c) Peaks in Hough transform, (d) Horizontal lines detected.*

## IV. IMPLEMENTATION

In this section now I give the detailed procedure of cricket video summarization and also describe the algorithm. This proposed algorithm I will implemented in next semester for cricket video summarization. There are six step process of video summarization.

### *A. Block Creation*

In this section we have first split the long video (I hour) into small blocks of equal size (I minute) long. That is why I am first creating block of the input video and then process. For this purpose we have using mat lab.

### *B. Priority Assignment*

Each block is then assigned a priority (e.g. 0 to 9 an integer vale) based on the objects and events occurring in that block. Yet another alternative would use the audio stream and/or accompanying text associated with the video to identify the priority of the block. The priority assignment can be done automatically using object and event detection algorithms or can be done manually. In our cricket video summarization application, for example, the priority assignment could be done by using (low/ high) level image processing algorithms for events such as four, sixes and out etc. Here we process each block one by one and on the basis of features (visual, audio, and textual) find out the event in each block and assign a priority to each block.

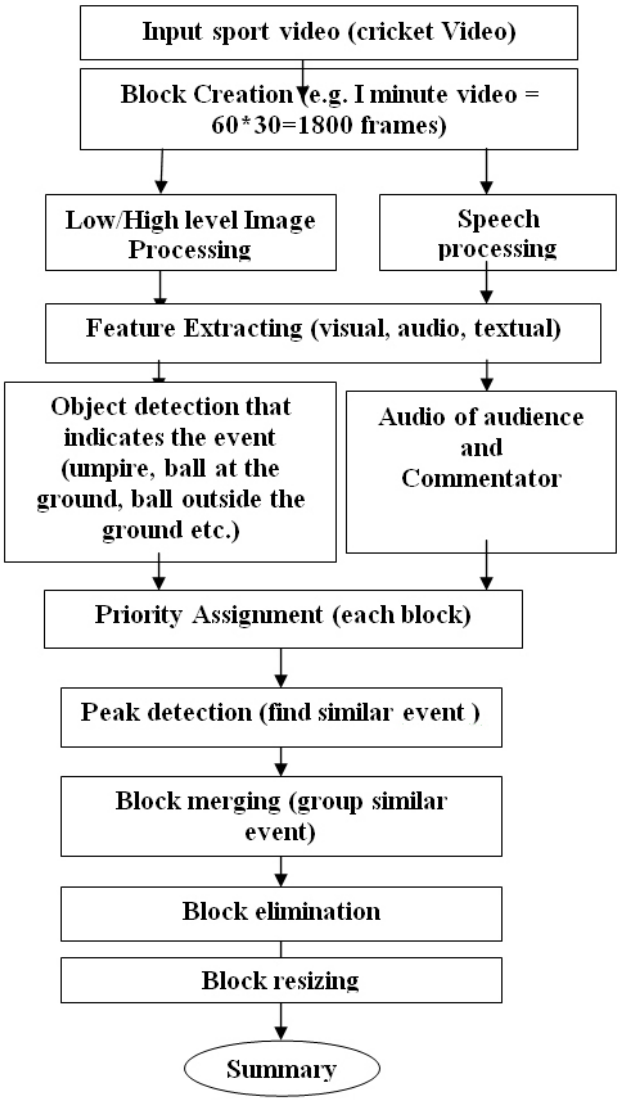


Figure 6. Block diagram of video summarization.

The table below shows all the events in the cricket, in common, played with some priority values. These events are common and are not only the events that are played. They can be always and addition to but increase the complexity of the system.

TABLE I. PRIORITY LIST OF EVENTS IN CRICKET

| S. no | Priority | Event |
|---|---|---|
| 1 | 0 | Just play |
| 2 | 1 | Umpire action for no ball |
| 3 | 2 | Umpire action for wide ball |
| 4 | 3 | Umpire action for by run |
| 5 | 4 | Out bold |
| 6 | 5 | Out lbw |
| 7 | 6 | Out caught |
| 8 | 7 | Out run |
| 9 | 8 | four |
| 10 | 9 | six |

### C. Peak Detection

After assigning a particular priority to each block. Consider a graph whose x axis consists of block numbers and whose y axis describes the priority of the blocks. We identify the blocks associated with the peaks in this graph using a peak identification algorithm.

The Peaks() algorithm we have developed can automatically find peaks in this priority curve. A peak consists of a sequence of blocks containing high priority events.

*Peaks Algorithm:*

Algorithm Peaks($v$,$r$,$s$)
  $v$ is a sequence of block-priority pairs
  $r$ is the peak width
  $s$ is the peak height
begin
  $Res := \emptyset$
  for each $j \in [r, card(v) - r]$ do
    $center := 0$
    $total := 0$
    for each $\langle b_i, p_i \rangle \in v$ such that $i \in (j - r, j + r]$ do
      $total := total + p_i$
    end for
    for each $\langle b_i, p_i \rangle \in v$ such that $i \in (j - \frac{r}{2}, j + \frac{r}{2}]$ do
      $center := center + p_i$
    end for
    if $\frac{center}{total} \geq s$ then
      $Res := Res \cup \{\langle b_i, p_i \rangle \in v \mid i \in (j - \frac{r}{2}, j + \frac{r}{2}]\}$
    end if
  end for
  return $Res$
end

Consider the 35 block sequence shown in Figure-7. We now describe how the Peaks () algorithm finds the peaks in this figure. Suppose r = 6 and s = 0.8.

Subsequently, we merge multiple adjacent blocks into one. These are cases where the same or similar events are occurring in these blocks even though the blocks were different segments produced by the video segmentation algorithm.

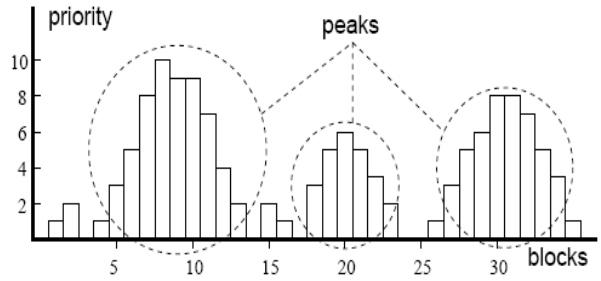


Figure 7. Result of Running Peaks () Algorithm

The *Peaks*() algorithm slides a 2*r*-wide window along a sequence of blocks, computing the total sum of block priorities in that window (*total*). It then computes the sum of block priorities in a narrower *r*-wide window in the middle of the 2*r*-wide window (*center*). When the ratio of these two sums (center/total) exceeds the threshold **S**, all blocks in the r-wide window are picked as a peak.

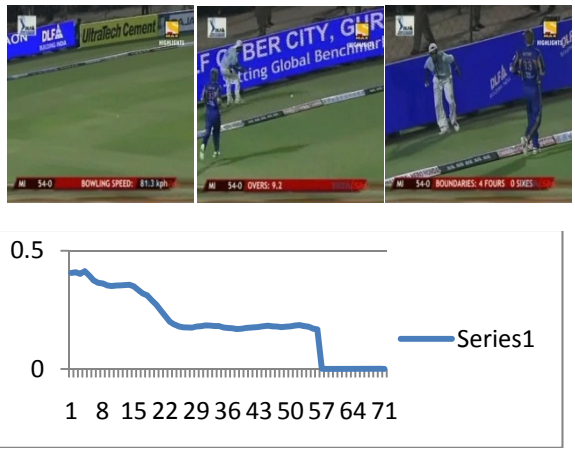


Figure 8. Boundary shot (four): the graph shows the grass ratio pixel decrease from play ground to boundary.

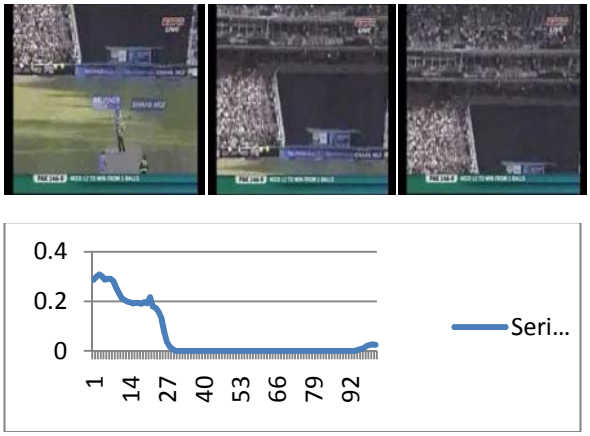


Figure 9. Boundary shot (six): the graph shows the grass ratio pixel decrease from play ground to boundary.

### D. Block Merging

The block merging algorithm uses rules to determine conditions under which multiple contiguous blocks can be merged together into a new block (whose priority equals the sum of the priorities of the blocks being merged).

*Block Merging Algorithm:*

```
Algorithm Merge(v,sim(),d)
    v is a sequence of block-priority pairs
    sim() is a similarity function on blocks
    d is the merging threshold
begin
    Res := ∅
    B := first block-priority pair ⟨ b_1, p_1 ⟩ ∈ v
    for each ⟨ b_j, p_j ⟩, ⟨ b_{j+1}, p_{j+1} ⟩ ∈ v do
        if sim(b_j, b_{j+1}) ≥ d then
            B := ⟨ B.b ⊕ b_{j+1}, B.p + p_{j+1} ⟩
        else
            add B to the tail of Res
            B := ⟨ b_{j+1}, p_{j+1} ⟩
        end
    end for
    add B to the tail of Res
    return Res
end
```

*E. Block Elimination*

We then have a block elimination algorithm which eliminates certain unworthy blocks these are blocks whose priority is too low for inclusion in the summary. This is done by analyzing the distribution of priorities of blocks, as well as the relative sizes of the blocks involved, rather than by setting an artificial threshold.

The set of blocks produced after merging is then shipped to a block elimination module. This module eliminates blocks whose priority is too low. For example, it may turn out that 10 merged blocks are returned after the block merging algorithm and these 10 blocks have a total of 5000 frames. If we want a summary consisting of just 3600 frames, we may want to re-examine whether a block of relatively low priority should be eliminated. For example, if we compute the average priority of the 10 blocks above to be 25 and the standard deviation to be 3, then we may want to eliminate all blocks with a priority under 16 (this is the classical statistical model which says that for a normal distribution, most objects in the distribution must occur within 3 standard deviations of the mean). Other statistical rules can also be used here.

*F. Block Resizing*

Finally, we have a block resizing algorithm that shrinks the remaining blocks so that the final summary consists of these resized blocks adjusted to fit the desired total length.

The block resizing component eliminates frames from the blocks in proportion to the priorities of the blocks involved.

*Block Resizing Algorithm:*

```
Algorithm Resize(v,k)
    v is a sequence of block-priority pairs
    k is the desired summary length
begin
    Res := ∅
    p_total := Σ_{⟨ b,p ⟩ ∈ v} p
    p' := 0
    k' := 0
    for each ⟨ b, p ⟩ ∈ v do
        if len(b) ≤ (p·k)/p_total then
            Res := Res ∪ {⟨ b, p ⟩}
            v := v \ ⟨ b, p ⟩
            p' := p' + p
            k' := k' + len(b)
        end if
    end for
    p_total := p_total − p'
    k := k − k'
    for each ⟨ b, p ⟩ ∈ v do
        alloc := round((p·k)/p_total)
        b' := b truncated to alloc frames
        Res := Res ∪ {⟨ b', p ⟩}
    end for
    return Res
end
```

## V. CONCLUSION

Video is getting more and more popular now than ever before, due to the rapid growth of the Internet bandwidth and the growing use of video in education, entertainment, and information sharing. Many organizations produce huge volume of video data every day. Facing the massive data volume, end users find that it is inefficient to browse a favorite video from the Internet, and the content providers have to face the tedious work of managing the ever growing video database. The urgent problem brings a lot attention to video summarization, which is a new technology intends to solve the problem by providing the people with concise and informative content presentations so that the users can quickly grasp the major contents of a video. Most video summaries goes into the following two types: static video story board, which is composed of a set of salient images extracted or synthesized from the original video, and dynamic video skimming, which is a shorter version of the original video made up of several short video clips. This thesis presents our work done on automatic video summarization.

## VI. REFERENCES

[1]. Creativepro.com. Popcast selected by mgi to offer personal broadcasting services to mgi videowave 4 users. http://www.creativepro.com/story/news/10207.html, 2000.
[2]. Broadband4Britain. Uk broadband usage survey. Http://www.net.com/pdf/BB4Britian-ir.pdf, 2003.
[3]. http://www.nielsennetratings.com/news.jsp?section=dat gi,"
[4]. http://www.altavista.com/about/default/,"
[5]. http://www.youtube.com/t/about/,"
[6]. http://video.google.com/video about.html,"
[7]. M. Mills, "A magnifier tool for video data," in *Proc. ACM Human Computer Interface*, May 1992, pp. 93–98.
[8]. Y. Taniguchi, "An intuitive and efficient access interface to real-time incoming video based on automatic indexing," in *Proc. ACM Multimedia'95*, Nov. 1995, pp.25–33.
[9]. H.J. Zhang, J. Wu, D. Zhong, and S.W. Smoliar, "An integrated system for content-based video retrieval and browsing," *Pattern Recognit.*, vol. 30, no. 4 pp. 643–658, Apr. 1997.
[10]. S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky, "Video manga: Generating semantically meaningful video summaries," in *Proc. ACM Multimedia'99*, Oct. 1999, pp. 383–392.
[11]. A. Girgensohn and J. Boreczky, "Time-constrained keyframe selection technique," in *Proc. ICMCS'99*, June 1999, pp. 756–761.
[12]. D. Zhang and D. Ellis, \Detecting sound events in basketball video archive," *http://www.ctr.columbia.edu/ dpwe/courses/e6820-2001-01/projects/dqzhang.pdf*.
[13]. Assfalg, M. Bertini, C. Colombo, A. Bimbo, and W. Nunziati, \Semantic anno- tation of soccer videos:automatic highlights identi¯cation," *Computer Vision and Image Understanding (CVIU)*, vol. 92, no. 2-3, pp. 285{305, 2003.
[14]. D. Chen, K. Shearer, and H. Bourlard, \Video ocr for sport video annotation and retrieval," *Proc. of IEEE International Conf. on Mechatronics and Machine Vision in Practice*, no. 28, pp. 57{62, 2001.
[15]. R. C. Gonzales and P. Winz. "Digital Image Processing" Addison-Wesley Publishing Company, Knoxville, Tennesee, 1987.
[16]. M. Fayzullin and V.S.Subrahmanian, M. Albanese and A. Picariello "The Priority Curve Algorithm For Video Summarization", *MMDB'04*, November 13, 2004, Washington, DC, USA.
[17]. Dian Tjondronegoro and Yi-Ping Phoebe Chen, Binh Pham "Integrating Highlights for More Complete Sports Video Summarization" Published by the IEEE Computer Society 2004.

AUTHORS PROFILE

**K Susheel Kumar,** presently working as Assistance Professor in Ideal Institute of Technology, Ghaziabad, India. He is M.Tech form Indian Institute of Information Technology, Allahabad, his major research work interst in Image Processing and Pattern Recognition

**Shitala Prasad**, presently pursing his master degree in Infornation Technology from Indian Institute of Infornation Technology, Allahabad, India. He is B.Tech. form IILM Greater Noida in Computer Scince. His majore research work intrests in Image Processing and Gesture Recognition.

**Santosh Banwral,** presently pursing his master degree in Infornation Technology from Indian Institute of Infornation Technology, Allahabad, India. He is M.C.A. form IGNOU delhi. His majore research work intrests in Image Processing and Gesture Recognition.

**Vijay Bhaskar Semwal,** present working in NEWGEN SOFT as a softwere Developer, He is M.Tech form Indian Institute of Information Technology, Allahabad, his major research work interst in Image Processing and wireless sensor network