A Survey on Different Video Scene Change Detection Techniques

Dolley Shukla¹, Chandra Shekhar Mithlesh², Manisha Sharma³

¹Shri Shankeracharya College of Engineering and Technology, Senior Associate Professor Junwani, Bhilai, Dist-Durg, Chhattisgarh 490020, India
²Shri Shankeracharya College of Engineering and Technology, Junwani, Bhilai, Dist-Durg, Chhattisgarh 490020, India, csmithlesh1987@gmail.com
³Bhilai Institute of Technology, Professor & Head (E & Tc), Durg, Dist-Durg, Chhattisgarh 490020, India, Manishsharma1@rediffmail.com

Abstract: A number of automated scene change detection methods for indexing a video sequence to facilitate browsing and retrieval have been proposed in recent years. Scene change detection plays an important role in a number of video applications, including video indexing, semantic features extraction, multimedia information systems, Video on demand (VOD), digital TV, online processing (networking) neural network mobile applications services and technologies, cryptography, and in watermarking. For copy protection, watermarking is an important technique. Recent advances in technology have made tremendous amounts of multimedia content available. The amount of video is increasing, due to which the systems that improve the access to the video is needed. A current research topic on video includes video abstraction or summarization, video classification, video annotation and content-based video retrieval. In all these applications one needs to identify shot and key frames in video which will correctly and briefly indicates the contents of video.

Keywords: Video, Video Sequence, False Detection, Detection Rate, SCD, Recall & Precision, PSNR, Macroblocks etc.

1. Introduction

Video is arguably the most popular means of communication and entertainment [3]. Video is an important component of the collaboration data streams. In order to enable filtering of on-line streams as well as effective retrieval of desired data from an archive, the video/audio streams need to have annotations or tags generated that describe the “content” of the stream[5]. To create indexed video databases, video sequences are first segmented into scenes, a process that is referred to as scene change detection (SCD) [1]. With the rapid growth of video indexing, editing and transcoding system, video analysis and segmentation become more important than ever before. Scene change detection plays an indispensable role in segmenting a long video sequence into smaller unit for further processing.

A video shot is a sequence of successive frames with similar visual content in the same physical location. Therefore, abrupt scene change can be identified when there is a large amount of visual content change across two successive frames while gradual transition is much difficult to detect since the successive frame difference during transition is substantially reduced[7].

2. Terms of Scene Detection

Image: Images may be two-dimensional, such as photographs, screen display, and as well as a three-dimensional, such as hologram. They may be captured by optical such as camera, mirror, lenses, telescope etc.

Picture Element: The pixel (a word invented from “picture element”) is the basic unit of programmable color on a computer display or in a computer image.

Video: A typical video sequence is organized into a sequence of groups of pictures (GOP). A video can be broken down in scene, shot and frames. Each GOP consists of one I (intra-coded) and a few P(predicted) and B (bi-directionally interpolated) frames as shown in Figure 1.

Figure 1: Group of Picture (GOP)

I-Picture: A 1-picture is intra-frame coded and does not depend on any previous or future frames. The bit-rate of a 1-picture depends on the visual content of the picture. If the picture is of high activity, then the bit-rate will be high.

When a scene change occurs between two consecutive I-pictures, the content of the two I-pictures will be totally different. The transition may be from high-activity to low-activity, low-activity to high-activity, or with similar activity. Therefore, the bit-rate of the I-picture following the scene change will be increased, decreased or remained similarly. This implies that we cannot guarantee to detect all cut points
if we only look for the significant change in total bit-rate of two consecutive I-pictures. However, when two pictures are of different content, they will have different local statistics even if their global characteristics are similar. Therefore, if we measure the bit-rate difference at macroblock level, we can extract the change of image content between two consecutive frames of sequence. A large change in bit-rate at macroblock level between two consecutive I-pictures means that there is a scene change between them.

P-Picture : P-picture is a forward motion-predicted frame, which depends on the previous P- or I-picture. When a scene change occurs between two consecutive P-pictures, it is difficult to make the forward prediction since the content of the current frame will be totally different from the previous one. Most of the macroblocks have so large prediction error that they will be encoded using the intra-coded mode. This will of course lead to a significant increase in the bit-rate of the current frame. Therefore, we can detect the scene change by checking whether there is a large increment in the bit-rate of two consecutive P-frames.

B-Picture : B-picture is a bidirectional-predicted frame, which depends on both the previous and future P- or I-picture. When a scene change occurs, the content of the current frame will have large difference with the previous anchor frame but have similar content with the future one. Therefore, most of the macroblocks depend mostly on the future P- or I-picture and will be backward predicted. Since the prediction can only be done mostly in one direction, the coding efficiency of the motion compensation will be lower and it will lead to the increase in the bit-rate when compared with the previous B-picture. However, such an increment will not be as significant as that occurred in I- or P-picture.

Scene : A scene is a logical grouping of shots into a semantic unit.

Shot : A shot is a sequence of frames captured by a single camera in a single continuous action. A shot boundary is the transition between two shots.

Frame: A digital video consists of frames that are a single frame consists of pixels.

Precision and Recall : "Precision" defines how reliable the detected by the algorithm is while "Recall" defines the overall performance of the algorithm. Result are reported using F-measure which is combine precision and recall in a single measure. These measure are computed by

\[
\text{Recall} = \frac{\text{correct}}{\text{correct} + \text{false}} \\
\text{Precision} = \frac{\text{correct}}{\text{correct} + \text{missed}} \\
F = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\]

PSNR : Peak signal-to-noise ratio is the important parameter of scene change detection algorithm, efficient SCD algorithm having the value of PSNR greater than 35. All of the SCD techniques gives approximately same PSNR result or increase PSNR by at least same.

3. Review on Scene Change Detection

The survey of various research papers that have contributed to solve the problem of scene change detection in video sequences, there is a growing demand of video indexing, scene browsing and retrievals in Signal Process.

The comparison and analysis of scene change detection algorithms is described in the following subsections

Ralph M. Ford et al. (1997) presented a five metrics for scene change detection in video sequences. Metric previously applied to this problem are surveyed, and are quantitatively compared to the new metrics. The proposed metrics are superior. These metrics were also applied to the detection of gradual transitions, y is a good global metric for detecting abrupt cuts. [1]

Limitations
- The F-test and vi-3 perform the best overall for abrupt cuts, but require more computation time.
- Global metric did not perform well for gradual transitions, but the statistic based metrics did.

Taehwan Shin et al. (1998) presented a hierarchical scene change detection in an mpeg-2 compressed video sequence. In this paper, we propose an efficient scene change detection algorithm for direct processing of MPEG-2 video bitstreams. The proposed algorithm utilizes the hierarchical structure of the compressed bitstreams and statistical characteristics of the coded parameters, thus greatly reducing computational requirement compared to pixel domain processing with full decompression. Occurrence of scene change is checked first in a COP level, and if the result is affirmative it is checked again in lower levels: sub-GOP and each picture. We used several metrics for different levels: variance of DC images for I-pictures, number of macroblock types for P-pictures and motion vector types for B pictures. The proposed algorithm uses such features of the video which are easily extracted by minimal decoding of the input bitstream that very efficient processing is achieved. [2]

Limitations
- In this algorithm, scene change is checked in a hierarchical fashion from GOP to sub-GOP and picture level, thus reducing much of the processing requirements.

W.A.C. Fernando et al. (1999) presented a novel algorithm for wipe scene change detection in video sequences. In the proposed scheme, each image in the sequence is mapped to a reduced image. Then statistical features and structural properties of the images were used to identify wipe transition region. Finally, Hough transform is used to analyze the wiping pattern and the direction of wiping. The algorithm is capable of detecting all wipe regions accurately even when the video sequence contains other special effects like fading, dissolving, panning the proposed algorithm can be used in uncompressed video to detect wipe regions with a very high reliability. [3]
Limitations
- Further work is required to extend this algorithm for compressed video.

W.A.C. Fernando, et al. (2000) presented a unified approach to scene change detection in uncompressed. There is an urgent need to extract key information automatically from video for the purposes of indexing, fast retrieval and scene analysis. To support this vision, reliable scene change detection algorithms must be developed. Results on video of various content types are reported and validated with the proposed scheme in uncompressed and MPEG-2 compressed video. The accuracy of the detected transitions is above 95% and 90% for uncompressed and MPEG-2 compressed video respectively. The proposed algorithm can be used in uncompressed and compressed video to detect scene changes with a high reliability. The algorithm is capable in detecting all scene changes accurately. [4]

Limitations
- Future work is required to extend the algorithm for camera movements detection within the same framework.

Wensheng Zhou, et al. (2000) proposed an On-line Scene Change Detection of Multicast Video. Network-based computing is becoming an increasingly important area of research, whereby computational elements within a distributed infrastructure process/enhance the data that traverses through its path. These computations as online processing and this paper investigates scene change detection in the context of MB one based proxies in the network. Online processing varies from traditional off-line processing schemes, where for example, the whole video scope is known as a priori, which allows multiple scans of the stored video files during video processing. The proposed algorithms do scene change detection and extract key frames from video bitstreams sent through the mbone network. Several algorithms based on histogram differences and evaluate them with respect to precision, recall, and processing latency a few effective methods of on-line video scene change detection over MBone video for the purposes of annotation/filtering. Main advantages of our algorithms is their ability to support real-time video processing on the network. Joint algorithms based on video codec characteristics for acquiring fast and accurate scene detection were carried out. Both global color histograms were extracted and block information of DCT coding was used. Algorithms were capable of satisfying online annotation needs. The algorithm was tested on different data bandwidths & gives the best performance in each case. [5]

Limitations
- Algorithms were mainly targeting real-time video scene analysis, improving processing speed and reducing latency are two more issues of concern, which makes it improper to compare the proposed algorithm with other off-line scene detection techniques, such as those mentioned in the Related Work section, where video data are processed off-line.
- In the future, they plan to study more video on-line annotations based on video content, such as key frame classification. Because of the complicated nature of video processing, they plan on using multiple workstations working in parallel to realize more sophisticated real-time annotation
- Plan to study algorithms that are tolerant of packet losses in the network.

The on-line processing system is designed to meet the requirements of real-time video multicasting over the Internet and to utilize the successful video parsing techniques available today.

Shu-Ching Chen, et al. (2002) presented a change detection by audio and video clues. Automatic video scene change detection is a challenging task. Using audio or visual information alone often cannot provide a satisfactory solution. However, how to combine audio and visual information efficiently still remains a difficult issue since there are various cases in their relationship due to the versatility of videos. Shu-ching presented an effective scene change detection method that adopts the joint evaluation of the audio and visual features. First, video information is used to find the shot boundaries. Second, the audio features for each video shot can be extracted. Lastly, an audio-video combination schema is proposed to detect the video scene boundaries. Unlike the traditional methods that first analyze audio and video data separately and then combine them, we analyze them at different phases. The audio feature extraction is based on the detected video shots, which tends to be more stable and more reliable in characterizing the audio data. The experimental results demonstrate that our method performs very well in terms of precision and recall value. [6]

Xiaoquan Yi et al. (2005) presented a Fast Pixel-Based Video Scene Change Detection. This paper proposes a simple and efficient method to detect abrupt scene change based on only pixel values. Convolutional pixel-based techniques can produce a significant number of false detections and missed detections when high motion and brightness variations are present in the video. To increase scene change detection accuracy yet maintaining a low computational complexity, a two-phase strategy is developed. Frames are firstly tested against the mean absolute frame differences (MAFD) via a relaxed threshold, which rejects about 90% of the non-scene change frames. The rest 10% of the frames are then normalized via a histogram equalization process. A top-down approach is applied to refine the decision via four features: MAFD and three other features based on normalized pixel values - signed difference of mean absolute frame difference (SDMAFD*), absolute difference of frame variance (ADFV*), and mean absolute frame differences (MAFD*) method contributes to higher detection rate and lower missed detection rate while maintaining a low computational complexity, which is attractive for real-time video applications. Method is relatively immune from sharp illumination changes, moving objects, camera motion, and other similar effects because of the well combined metrics. Furthermore, our method uses only frame pixel values without any motion estimation processes so that frugal computational complexity is maintained, which makes it very attractive for real-time video applications. [7]
Gao, J., Jiang et al. (2006) presented a PCA-based approach for video scene change detection on compressed video. An automatic, real-time detection approach to video scene change detection is presented. Owing to the high correlation of two consecutive video frames, it is proposed that only the eigenvector corresponding to the largest eigenvalue is retained in the principal component analysis (PCA) for video data. A one-dimensional PCA feature of video data is then generated from the PCA. It shows superior performance compared to the histogram feature and the pixel feature. The detection algorithm based on this PCA feature is then designed to detect both abrupt and gradual transitions. The proposed approach is tested on the TREC video test repository to validate its performance. The superior performance of the new PCA feature. Then our change detection algorithm is based on the PCA feature to detect the scene change. [8]

Limitations
- Tested on the TREC video test repository to validation for its performance.

J.-R. Ding et al. (2007) presented adaptive group-of-pictures and scene change detection methods based on existing H.264 advanced video coding information. The H.264 advanced video coding (H.264/AVC) standard provides several advanced features such as improved coding efficiency and error robustness for video storage and transmission. In order to improve the coding performance of H.264/AVC, coding control parameters such as group-of-pictures (GOP) sizes should be adaptively adjusted according to different video content variations (VCVs), which can be extracted from temporal deviation between two consecutive frames. The authors present a simple VCV estimation to design adaptive GOP detection (AGD) and scene change detection (SCD) methods by using the obtained motion information, where the motion vectors and the sum of absolute transformed differences as VCV features are effectively used to design the AGD and SCD algorithms, respectively. In order to avoid unnecessary computation, the above VCV features were obtained only in the 4 x 4 inter-frame prediction mode. The proposed AGD with SCD methods can increase the peak signal-to-noise ratio by 0.62 dB on average over the H.264/AVC operated with a fixed GOP size. Besides, the proposed SCD method reached a scene change detection rate of 98% . On the other hand, if the scene-changed frame without any detection mechanism is coded by intercoding, it will become inefficient and waste considerable computation time in motion estimation. and found that the GOP size for H.264/AVC should be appropriately decreased and increased for higher and lower VCV frames, respectively, to improve the coding performance. The method successfully utilized the existing video coding information, such as motion vectors and motion residuals, to become an effective VCV feature. Both the proposed AGD and SCD methods can effectively help to adjust a better GOP size to improve the coding performance of H.264/AVC. The proposed joint AGD and SCD schemes can increase PSNR by at least 0.62 dB compared with the H.264/AVC operated in the fixed GOP size on average. [9]

Limitations
- The proposed and some existing VCV characteristics can be further extended for adaptive search window, adaptive search methods (of three step search and

ZHANG Ji et al. (2007) presented an effective error concealment framework for H.264 decoder based on video scene change detection. In this paper, we propose an effective error concealment framework for H.264 decoder based on the scene change detection. The proposed framework quickly and accurately detects whether scene change occurs in the decoding frame, based on the detection result, both corrupted intra frames and damaged inter frames can be reconstructed by spatial or improved temporal EC (Error Concealment) algorithm. The experiment shows that, compared with the traditional error concealment method in the H.264/A VC non normative decoder, the proposed framework has better robustness and can efficiently improve the visual quality and PSNR of the decoded video. [10]

Limitations
- The final experimental results demonstrate that our framework can apparently improve the visual quality of reconstructed video sequence and PSNR.
- In the future, they will emphasize more on exploring the scene change algorithm for B frames and improve spatial or temporal EC algorithms in the new frame work.

Jens Brandt et al. (2008) presented a fast frame-based scene change detection in the compressed domain for MPEG-4 video. Detection of scene changes is an elementary step in automatic video processing like indexing, segmentation or transcoding. Video indexing and segmentation allow fast browsing without decoding the complete video. In the case of transcoding, the information about scene changes such as cuts and fades as well as about special movements like rotations or zooms in video frames is helpful to determine suitable transcoding parameters. In the compressed domain only information about DCT values as well as motion information can be used to determine such scene changes and movements. Therefore different measures were defined which use the encoded DCT values and motion vectors of each compressed frame. Based on these measures as well as on motion vector histograms, a fast approach to detect different kinds of scene changes and special movements in MPEG-4 videos in the compressed domain was proposed. In this paper we have presented a fast frame-based algorithm for compressed domain scene change detection in MPEG-4 video streams. The easy computation of the used metrics and histograms makes the whole scene change detection algorithm very fast which is very important for real time video processing. Therefore the scene change detection can only be a tool to indicate a stronger or weaker correlation towards one or the other scene movement. However, the evaluation results show that the algorithm detects a high number of scene changes and movements. Based on these promising results Jens Brandt developed a method to use the information about detected scene changes and special movements within a video stream for determining suitable transcoding parameters. [11]
Limitations
- The highly statistical nature of the scene change detection results in a statement that by design cannot be absolutely correct.

Hakan Haberdar et al. (2010) presented a disparity map refinement for video based scene change detection using a mobile stereo camera platform. This paper presents a novel disparity map refinement method and vision based surveillance framework for the task of detecting objects of interest in dynamic outdoor environments from two stereo video sequences taken at different times and from different viewing angles by a mobile camera platform. Preliminary disparity images were refined based on estimated disparities in neighboring frames. Segmentation is performed to estimate ground planes, which in turn are used for establishing spatial registration between the two video sequences. Finally, the regions of change are detected using the combination of texture and intensity gradient features. We present experiments on detection of objects of different sizes and textures in real videos. The framework is evaluated on 4 real video sequences and the experimental results show that the proposed framework is able to detect changes/objects of different sizes and textures, and is able to ascertain if there are no changes in the scene, thereby minimizing false positives. [12]

Limitations
- This method need for future work and it will focus on relaxing our assumption on known temporal alignment and improving the spatial registration module.

Ufuk Sakarya et al. (2010) presented a video scene detection using graph-based representations. One way of realizing this step is to obtain shot or scene information. One or more consecutive semantically correlated shots sharing the same content construct video scenes. On the other hand, video scenes are different from the shots in the sense of their boundary definitions; video scenes have semantic boundaries and shots are defined with physical boundaries. In this paper, we concentrate on developing a fast, as well as well-performed video scene detection method. Our graph partition based video scene boundary detection approach, in which multiple features extracted from the video, determines the video scene boundaries through an unsupervised clustering procedure. For each video shot to shot comparison feature, a one-dimensional signal is constructed by graph partitions obtained from the similarity matrix in a temporal interval. After each one-dimensional signal is filtered, an unsupervised clustering is conducted for finding video scene boundaries. We adopt two different graph-based approaches in a single framework in order to find video scene boundaries. The proposed graph-based video scene boundary detection method is evaluated and compared with the graph-based video scene detection method presented in literature. [13]

Limitations
- Some issues are considered into account for future research directions. Different video features can be used in order to improve the results and; moreover, content varying based changes in the multiple features may increase the scene detection performance.

Ankita P. Chauhan1 et al. (2013) presented Hybrid Approach for Video Compression Based on Scene Change Detection. In order to fulfill the requirement of limited channel bandwidth and of growing video demand like streaming media delivery on internet, and digital library, video compression is necessary. In video compression, temporal redundancy between adjacent frames is removed with block based motion estimation algorithms. Video represents a sequence of frames captured from camera. Scene is a series of consecutive frames captured from narrative point of view. In this paper we present an effective scene change detection method for an uncompressed video. We have divided frames in to blocks and applied a canny edge detector in consecutive frames. Count no of pixels (ones) in each block and compare it with consecutive frames. If scene change happens then number of pixels per block will change, based on that change we can detect scene change in consecutive frame. presented a hybrid approach, in which we have used scene change detection along with block based motion estimation algorithms (BME) to compress video.In this paper we have presented a simple approach with high accuracy. To reduce the number of computations, hybrid approach along with scene change detection have been implemented.[14]
Limitations

- Improvement in its basic inputs would result in better performance of all these techniques.

4. Conclusion

The above survey of various researchers of different algorithms of Scene Change Detection shows that the two Main problem associated with exiting algorithms i.e. threshold-dependent algorithms and color Histogram Method. The threshold-dependent algorithm gives false detection with scenes involving fast camera or object motion. Color Histogram Method having some advantages like Efficient representation, Easy computation, Global color distribution, Insensitive to Rotation, Zooming Changing resolution and Partial occlusions. Bu some disadvantages also like Ignorance to spatial relationship, Sensitive in illumination changes and Chooses illumination-insensitive color channels. The method avoided the false alarms by using the validation mechanism. It also proves that the statistical model-based approach is reliable for gradual scene-change detection. A very high detection rate is achieved while the false alarm rate is comparatively low. For better output detection, a scene change detection algorithm should be proposed with the properties low false alarming,high robustness in gradual transition, camera fabrication, special events and so on . For a considerably good quality of scene change detection (SCD) algorithm PSNR ranges should be >40, Recall & precision rate > 95% and change detection rate >98%.

References


Author Profile

Chandra Shekhar Mithlesh received the B.E. degrees from government Engineering College jagdalpur in Electronics & Telecommunication with first Class and currently perusing his master degree in Communication Engineering from Shri Shankaracharyar College of Engineering and Technology., Bhilai., Chhattisgarh.