Gradual Transition Detection Algorithms in Video Segmentation: A Survey

Salim A. Chavan, Sudhir G. Akojwar

Abstract— A large number of shot boundary detection, or equivalently, transition detection techniques have been developed in recent years. The shot boundary detection includes abrupt and gradual shot change detection. Gradual shot change detection is more challenging task than abrupt shot change detection. Thus aim of this paper is mainly to focus on gradual boundary detection. Previous researches related to gradual shot change detection yields different methods such asTwin comparison method, Feature-based detection, DCD method etc. This survey paper discuss various algorithms used for gradual shot change detection. Performance measures such as precision and recall, for testing different algorithms are also considered

Index Terms— Shot boundary, gradual transition, fade in, fade out, dissolve, wipe and video segmentation.

1 INTRODUCTION

DIGITAL video is becoming an increasing common data type in the new generation of multimedia databases. Many broadcasters are switching to digital formats for broadcasting, and some of them already have a significant amount of video materials available in digitalformats for previewing. Improved compression technologies and increased Internet bandwidth have made a webcasting a real possibility. The ever growing amount of digital videos poses new challenges, both of the storage and access, as vast repositories are being built at an increasing pace. A key step for managing a large video database is to segment the video sequences into shots. Video segmentation makes the video data more manageable by imposing on it a hierarchy. It also forms the first step to understanding video content by dividing it into shots on which content analysis can be performed. This segmentation process is generally referred to as shot boundary detection. A shot is a sequence of frames generated during a continuous camera operation and represents a continuous action in time and space. Video editing procedures produce abrupt and gradual shot transitions. A cut is an abrupt shot change that occurs in a single frame. A gradual change occurs over multiple frames and is the product of fade-ins, fade-outs, or dissolves (where two shots are superposed). Figures1 and 2 show examples of abrupt and gradual changes.

- *Salim A. Chavan is currently pursuingPhD degree program in electronics engineering inSGB Amravati University, India, PH-9763702575. E-mail: salimsahil97@rediffmail.com*
- *Sudhir G. Akojwar is doctorate from VNIT, India, PH-9422835577. E-mail: sudhirakojwar@rediffmail.com*

There have been tremendous works reported in past few years on shot boundary detection in the literature. Earlier works concentrate mainly on an abrupt cut. Therefore, recent related works geared towards gradual shot boundary detection. The detection of gradual changes is more difficult than that of abrupt cuts. This is because a difference sequence is temporally well separated for cuts, whereas, it is not at any time of the sequence for gradual changes.

Fig. 1 Abrupt shot changes (cuts)

For the evaluation of the detection algorithm, the precision and recall can be computed using following formulae.

Precision = [Ncorrect/(Ncorrect+Nfalse)] *100

Recall = [Ncorrect/(N_{SCD}=Ncorrect+Nmissed)]*100

where, $N_{Correct}$ is the number of correctly detected frames, N_{False} is the number of falsely detected frames, N_{Missed} is the number of missed frames, and N_{SCD} is total number of frames where transitions are occurred.

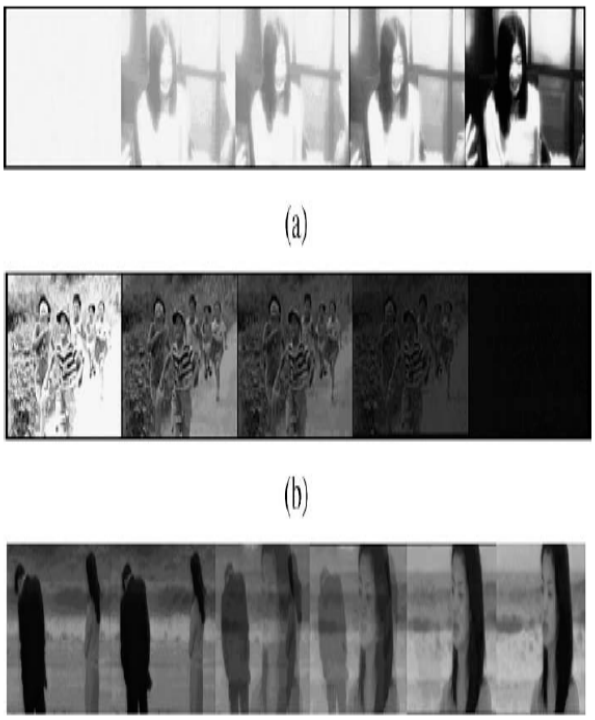

Fig. 2 Gradual shot changes (a) fade-in; (b) fade-out; (c) dissolve

2 Related Researches on Gradual Transitions

2.1 Twin Comparision Method

Zhang et al. proposed this method. It is the first attempt to detect and classify abrupt and gradual changes. In the approach, dual threshold values are applied to the difference of intensity histogram in order to detect gradual transitions. The method requires two thresholds: higher one, Th, for detecting cuts and a lower one, Tl, for detecting gradual transitions. First, the threshold Th is used to detect high discontinuity values corresponding to cuts, and then the threshold T1 is applied to the rest of the discontinuity values. If a discontinuity value is higher than T1, it is considered to be the start of the gradual transition. At that point, the summation of consecutive discontinuity values starts and goes on until the cumulative sum exceeds the threshold Th. Then, the end of the gradual transition is set at the last discontinuity valued included in the sum. However, one of the major problems in this approach is that many false positives can be generated when thresholds are not properly assigned. Implementation of this algorithm shows average precision of 77.2% and recall rate is 82%. [1]

2.2 Plateau Detection Method

Yeo and Liu [2] noted that the comparison based solely on successive frames will not be adequate for the detection of gradual transitions. They used the difference between a current frame and a following kth frame. It obtains first the sequence of delayed inter-frame distances $\{D_i i^1 k = d(x_i(i,) x_i(i+k))\}$. If we choose k greater than the length of the gradual transition, the sequence $\{D_i^k\}$ exhibits a plateau of maximal width. A significant plateau at location i is characterized by a sequence of similar values $D_j^k, j = 1-s, \ldots, 1+s$, which are consistently higher than the preceding or successive values. The value of s is proportional to the difference between k and the transition length. The method applies to linear and nonlinear gradual transitions; it is the shape of the rises and falls at the plateau boundaries.

2.3 Intensity Variance Method

In a compressed domain, an intensity variance of successive frames is used to detect gradual changes. This method proposed by Meng et al.exploits the DCT DC coefficients and motion vectors. Theoretically, as most dissolves show aparabolic shape, the authors tried to use the depth and width of that curve. However, in actual cases, due to the noises and motions in a video, the graph is not sufficiently pronounced. [4]

2.4 Chromatic Video Edit Model

Song et al. proposed a chromatic video edit model for gradual transitions is built based on the assumption that discontinuity values belonging to such a transition form a pattern consisting of two piece-wise linear function of time, one decreasing and one increasing. Such linearity does not apply outside the transition area. The authors search for close-to-linear segments in the series of discontinuity values by investigating the first and second derivative of the slope in time. A close-to-linear segment is found if the second derivative is less than a prespecified percentage of the first derivative. [3]

2.5 Feature Based Detection

This algorithm is based on calculating edgechange fraction in temporal domain. During a cut or a dissolve, new intensity edges appear far from location of old edges. Edge pixels that appear/disappear far from existing edge pixels are considered as entering/exiting edge pixels.Cuts, fades, and dissolves can be detected by counting the entering and exitingedge pixels, while wipes can be detected by looking at their spatial distribution. [5]

The algorithm is based on the following steps:

1. Frames Ft and Ft+1 are aligned using a global motion compensation algorithm.
2. Edges are computed by applying the canny algorithm to a smoothed version of the frames.
3. The binary edge maps are dilated by radius r, so that the condition on the mutual distance of edge pixels can be easily verified by set intersection.

4. The fraction of entering edge pixels rin and exiting pixels rout are computed. Shot changes are detected by looking at the edge change fraction r = max (rin, rout). A cut leads to a single isolated high value of r while the other scene breaks lead to an interval where r's value is high. During a fade-in the value rin is much higher than rout. The reverse happens for fade-outs. A dissolve is characterized by a predominance of rin during the first phase and rout during the second phase.The technique works properly also on heavily compressed image sequence. This approach presents high accuracy, but it takes a large amount of computation time.

2.6 Luminance Histogram Difference Curve Method

Truong et al. [6] tried to improve cut detection accuracy byutilizing an adaptive threshold computed from a local window on the luminance histogram difference curve. Also, based on the mathematical models for producing ideal fades and dissolves, the existences of these effects were examined. In that procedure, constraints on the characteristics of frame luminance mean and variance curves were derived to eliminate false positives caused by camera and object motions during gradual transitions

2.7 Detection based on spatio-temporal distribution of the macro block types

It performed dissolve detections based on the spatio-temporal distribution of themacro block types in MPEG-compressed videos. The ratio of forward macroblocks in the B-type frames and the spatial distribution of forward/backward macro blocks is utilized for detecting dissolve changes. After finding such sequence of frames two heuristic rules are applied:

1. The global color distributions of the frames at which the dissolve starts and terminates are very different.
2. The duration of a dissolve transition is typically more than 0.3 s.[7]

2.8 Machine Learning Approach

A novel dissolve detection algorithm using machine learning and multi resolution concept was proposed. The approach is less concerned about actual features used for dissolve detection, but more with ageneral framework for recognizing gradual transitions. First, a huge number of dissolve examples are created from a given video database using a dissolve synthesizer. Then these examples are used to train a heuristically optimal classifier whichis then employed in a multi-resolution search for dissolves of various durations.[8]

2.9 DCD Method

Variance, gradient magnitude, and double chromatic difference (DCD) of image sequence were used for dissolve detection. The firststep of the DCD segments the video into non-overlapping categories of "potential dissolve'' and "non-dissolves'' using edge-based or pixel-based statistics. The second step of the DCD detector uses this segmentation to define one synthetic dissolve per potential-dissolve segment, beginning and ending at the first and last frame of the segment, respectively. From these starting and ending frames, the center frame of a synthetic dissolve is formed and compared to the intervening footage. If the shape of the comparison error over time is parabolic shaped, thepotential-dissolve segment is accepted. Implementation of this algorithm showsaverage precision of 74.3%.and recall rate is65.3%.[9]

2.10 Localized Edge Block Method

This algorithm is proposed byHun-Woo Yoo, Han-Jin Ryoo& Dong-Sik Jang.The proposed algorithm is based on the fact that most of gradual curves canbe characterized by variance distribution of edge information in the frame sequences. Average edge frame sequence is obtained by performing Sobel edge detection. Features are extracted by comparing variance with those of local blocks in theaverage edge frames. Those features are further processed by the opening operationto obtain smoothing variance curves. The lowest variance in the local frame sequenceis chosen as a gradual detection point. Experimental results show that the proposed method provides 87.0% precision and 86.3% recall rates for selected videos. Experimental results obtained by implementing this algorithm are found to be encouraging, but it is worth stressing some problems encountered. Camera motion, object motion, and extensive content change withinthe shot should be considered for high performance. The method of handling these problems along with the proposed algorithm could yield better results. Future work includes more testing on different types of videos and efforts on shottransitions on the basis of camera motion.[10]

2.11 Wavelet Transform Method

Li Yufeng, Yang Yinghua, Li Guiju had propsed a wavelet transform method in which wipe transition was considered as vital mode of gradual transition. In the intended idea, each frame of color sub-image and edge sub-image are decomposed with the aid of Db-4 wavelet transition. To minimize the noise influence effectively, the color sub-image is divided into 8*8 pixel blocks and a Gaussian mode is used to amend the threshold dynamically in detecting the potential wipe transition. The intended method is tested on various sequences of news, sport, MTV, film and other videos. Upon the summarization and comparison of the results obtained it reveals that 92% recall with 96% precision was attained which is much greater than that of the previous one's which is mainly featured for both color-feature extraction and edge feature extraction application. This approach is used for detecting the straight lines at the centre using Hough transform and the threshold were also found dynamically. Therefore this method is also reliable for detecting the wipe transition reliability in uncompressed video. In future this method can be extended to compressed video as well. [11]

2.12 Slipped Window Method

This algorithm is proposed by Tuanfa Qin, JiayuGu, Huiting Chen, Zhenhua Tang in which fast shot boundary was detect-

ed based on K-step slipped window by taking the sufficient portion of edit-feature and low feature video-shots. In this method the distance of the images of both the ends were calculated and selects candidate fragments that comprises of shot-boundaries by adaptive threshold. The efficiency of this method will be expressed by comparison of calculating amounts and comparison of accuracy. This method is capable of improving the accuracy of shot detection and to minimize the amount of calculation. The parameters, values and formulae obtained in these approach posses a very high common ability.[12]

2.13 Temporal Video Segmentation

Chen Yinzi, Wang Kongqiaoet al. proposed a temporal video segmentation method based upon detection of shot abrupt transition and gradual transition, which then takes conditions of user terminals into account and produces various kinds of summarization for the individuals. From the experimental results it was evident that the video summarization could able to meet the browsing requirements of the user which in turn allows the user to yield a better and a joyful browsing experience. The future task of the proposed method is to enhance the method by picking the threshold automatically by machine learning method.[13]

2.14 Correlation Metric Method [14]

Abdul Hameed proposed a frame work in which a set of selection of representative key frames are done which are useful in summarizing the content of entire video into an abstract for creation of an efficient indexing. Then the videos are classified into its constituent variable size blocks preceded by the selectionof better possible similarity metric. The intended segmentation algorithm is to detect shot transitions of any kind that are present in the video sequence. It was evident from the results that the precision metric ranges from 75% to 100%, on the other hand recall metric ranges from 80% to 100%. This method will make use of motion estimation based upon a correlation metric. The results obtained through this approach will prove the effectiveness of the approach. [14]

2.15 CFAR Content-based video retrival Algorithm

In this paper, a new cut detection algorithm with constant false-alarm ratio (CFAR) is proposedfor video segmentation. In this method, a theoretical threshold determination strategyusing the non-parameter based CFAR processing technique is developed to achieve a controllableprecision as well as an evaluative recall performance for video cut detection. Simulationresults show that this algorithm leads to very good detection performance of a precision of about 90%. [13]

3 Conclusion

Earlier works concentrate mainly on abrupt cut detection but as gradual shot boundary detection is more challenging task, this paper reviewed different algorithms related to gradual shot boundary detection. We focused on 15 different methods for gradual shot boundary detection. Among different methods wavelet transform method is more effective among all the methods with precision 96% and 92% recall rates for selected videos. In average wavelet transformation method showed best performance (96%) followed by CFAR (Content-based video retrieval) algorithm with a precision of about 90% then thelocalized edge block method has the precision 87.0% and 86.3% recall rates for selected videos and twin comparison (77.2%), Correlation metric method ranging from 75% to 100%and the DCD (74.3%). The major issues in gradual transition detection are distinguishing gradual effects from camera and object motion.

4 Now from the literature it was clearly evident that the wavelet transformation algorithm yields qualitative and enhanced results, on the other hand CFAR algorithm also seems to be worthy as well. On combining both wavelet transformation & CFAR algorithms we can obtain an outcome which will be capable of improving the detection of the transition effects and reduce all other noise influencing factors existing in the camera and object motion videos. The results yielded will also have a great feasibility in obtaining an innovative solution for the problems that are encountered in the present day scenario in detecting the gradual transition effects to a greater extent.

Acknowledgment

The authors wish to thank Dr. Krishna Warhade for his support and motivation Also authors extend thanks to JDIET and DBNCOET, Yavatmal, India for providing necessary research facilities.

References

[1] Zhang H.J, Kankanhalli A, Smoliar S.W, TanS.Y,"Automatic partitioning of full motionvideo". ACM Multimedia Systems 1(1):10–28,1993.

[2] YeoB.L, Liu B "Rapid scene analysis on compressed video". IEEE Trans Circuits SystemVideo Techno 5(6):533–544, 1995.

[3] Song S, Kwon T, Kim W,"Detection of gradual scene changes for parsing of video data". Proc IS&T/SPIE, vol 3312. pp 404–413,1998

[4] Meng J, Juan Y, Chang S.F, " Scene change detection in a MPEG compressed videoSequence", Proc. SPIE/IS&T Symp. Electronic Imaging Science and Technology: Digital Video Compression: Algorithms and Technologies, vol 2419. pp 14–25, 1994.

[5] Zabih R, Miller J, Mai K, "A feature-based algorithm for detecting and classifying production effects" Multimedia Syst7(2):119 128,1999

[6] Truong BT, Dorai C, VenkateshS " New enhancements to cut, fade, and dissolve detection processes in video segmentation", ACM International Conference on Multimedia pp 219–227, 2000.

[7] Jun SB, Yoon K, Lee HY, " Dissolve transition detection algorithm using spatio-temporal distribution of MPEG macro-block types", ACM International Conference on Multimedia 391– 394,2000.

[8] Lienhart R,"Reliable dissolve detection. Storage and Retrieval for Media Database SPIE 4315:219–230, 2001.

[9] Lu HB, Zhang YJ, Yao YR, "Robust gradual scene change detection", International Conference on Image Processing 3:304–308,1999.

[10] Hun-Woo Yoo& Han-Jin Ryoo& Dong-Sik Jang, "Gradual shot boundary detection",Multimedia Tools Applications,pp283–300,2006.

[11] Li Yufeng, Yang Yinghua, Li Guiju, "A Novel Wipe Transition Detection Method Based on Multi-Feature", 2010 Third International Conference on Knowledge Discovery and Data Mining, pp451-454-2/10

[12] Tuanfa Qin, JiayuGu, Huiting Chen &Zhenhua Tang, "A Fast shot-boundary detection based on K-step slipped window", pp-190-195IC-NIDC2010.

[13] Chen Yinzi, Wang Kongqiaoet al, "A Temporal Video Segmentation and Summary Generation Method Based on Shots' Abrupt and Gradual Transition Boundary Detecting", International Conference on Communication Software and Networks, pp978-0-7695-3961-4/10.

[14] Abdul Hameed, "Video Shot Detection by Motion Estimation and Compensation", International Conference on Emerging Technologies, pp978-1-4244-5632-1/09

[15] Tie-Yan Liuet. al. "A new cut detection algorithm with constant false-alarm ratio for video segmentation", J. Vis. Commun. Image R. 15 (2004) 132–144.