A Survey of Recent Work in Video Shot Boundary Detection

Costas Cotsaces, Marios A. Gavrielides, and Ioannis Pitas *

Department of Informatics, University of Thessaloniki, Thessaloniki 54124, Greece,
pitas@aiia.csd.auth.gr, http://www.aiia.csd.auth.gr

Abstract. A video shot is defined as a continuously imaged temporal segment of a video. Since the semantic content of a video is largely based on its production (imaging) process, the complete segmentation of a video into shots is a fundamental first step for most kinds of semantic video processing tasks. Here we present a classification of shot boundary detection algorithms, including those that deal with gradual shot transitions. For lack of space we give neither a general introduction to the shot change detection problem, nor a detailed look at specific algorithms.

1 Classification of Shot Boundary Detection Algorithms

1.1 Features Used

Almost all shot change detection algorithms reduce the large dimensionality of the video domain by extracting a small number of features from each video frame. These are extracted either from the whole frame or from a subset of it, which is called a *region of interest* (ROI). Such features include:

1. *Luminance/color.* The simplest feature that can be used to characterize a ROI is its average grayscale luminance [1]. This, however, is susceptible to illumination changes. A better choice is to use some statistics of the values in a color space [3, 6].
2. *Luminance/color histogram.* A richer feature for a ROI is the grayscale or color histogram. It is quite discriminant, easy to compute and mostly insensitive to translational, rotational and zooming camera motion. For the above reasons it is widely used [2].
3. *Image edges.* An obvious choice of feature is edge information in a ROI [4, 21, 19]. Edges can be used as is, be combined into objects or used to extract ROI statistics. They are invariant to illumination changes and most motion, and they correspond somewhat to the human visual perception. Their main disadvantage is computational cost, noise sensitivity and high dimensionality.
4. *Transform coefficients (DFT, DCT, wavelet).* These are a classic way to describe the texture of a ROI [13]. The DCT coefficients are also present in MPEG encoded video streams or files. Their greatest problem is that they are generally not invariant to camera zoom.

* This work has been supported by the European Union project DELOS: a Network of Excellence on Digital Libraries, G038-507618

5. *Other features.* A number of other features are used in the literature, such as the color anglogram [7].
6. *Multiple features.* Many algorithms extract several types of features either to use them in combination or for subsequent processing and analysis. [5, 20, 15, 14, **?**].

1.2 Spatial Feature Domain

The size of the region from which individual features are extracted plays a great role in the performance of shot change detection. A small region tends to reduce detection invariance with respect to motion, while a large region tends to miss transitions between similar shots.

1. *Single frame pixel per feature* Some algorithms use a single frame pixel per feature. This feature can be luminance [11] , edge strength [4] or other. However, such an approach results in a very large feature vector and is very sensitive to motion.
2. *Rectangular block* Another method is to segment each frame into equal-sized blocks, and extract a set of features per block [13, 3, 6, 15]. This approach is invariant to small camera and object motion. By computing block motion it is possible to enhance motion invariance, or to use the motion vector itself as a feature.
3. *Arbitrarily shaped region* Feature extraction can also be applied to arbitrarily shaped and sized regions [16]. This exploits the most homogeneous regions, enabling better detection of discontinuities. Object-based feature extraction is also included in this category. The main disadvantage is high computational complexity and instability due to the complexity of the algorithms involved.
4. *Whole frame* The algorithms that extract features from the whole frame at once [20, 2, 9] have the advantage of being very resistant to motion, but tend to have poor performance at detecting the change between two similar shots.

1.3 Temporal Domain of Continuity Metric

Another important aspect of shot boundary detection algorithms is the temporal window of frames which is used to perform shot change detection. These can be one of the following:

1. *Two frames* The simplest way to detect discontinuity is to look for a high value of the discontinuity metric between two successive frames [20, 13, 3, 8, 7, 23, 22]. However, such an approach fails when there is significant variation in activity among different parts of the video, or when certain shots contain events that cause short-lived discontinuities (e.g. photographic flashes).
2. *N-frame window* The most common technique for alleviating the above problems is to detect the discontinuity by using the features of all frames within a temporal window [9, 13, 5, 3, 1]. This is either by computing a dynamic threshold against which a frame-by-frame discontinuity metric is compared or by computing the discontinuity metric directly on the window.

3. *Entire current shot* Another method is to compute one or more statistics for the whole shot and to check if the next frame is consistent with them, as in [4, 10, 6, 15, 2]. But if there is variability within shots, statistics computed for an entire shot may not be representative of its end.
4. *Entire video* The most thorough method is to take the characteristics of the whole video into consideration when detecting a shot change, as in [9, 14]. Again, the problem is that the video can have great variability within and between shots.

1.4 Shot Change Detection Method

1. *Thresholding* This means comparing the computed discontinuity value with a constant threshold [4, 10, 2]. This method only performs well if video content exhibits stationarity with time, and only if the threshold is adjusted by hand.
2. *Adaptive Thresholding* The obvious solution to the problems of the simple thresholding is to vary the threshold depending on the average discontinuity within a temporal domain, as in [17, 20, 13, 1, 15, 18].
3. *Probabilistic Detection* A rigorous way to detect shot changes is to model the pattern of specific types of shot transitions and perform optimal a posteriori shot change estimation, presupposing specific probability distributions for shots. This is demonstrated in [3, 6].
4. *Trained Classifier* A radically different method for detecting shot changes is to formulate the problem as a classification task, with the classes being "shot change" and "no shot change" [9, 14].
5. *Heuristics* A number of authors use various domain-specific heuristics for the detection of different transition types [13, 5, 12].
6. *User interaction* If automatic procedures fail, cut detection in ambiguous cases can be resolved by user input [23].

References

1. P. Campisi, A. Neri, and L. Sorgi. Automatic dissolve and fade detection for video sequences. In *Proc. Int. Conf. on Digital Signal Processing*, July 2002.
2. Z. Cernekova, C. Kotropoulos, and I. Pitas. Video shot segmentation using singular value decomposition. In *Proc. 2003 IEEE Int. Conf. on Multimedia and Expo*, volume II, pages 301 – 302, Baltimore, Maryland, USA, July 2003.
3. Alan Hanjalic. Shot-boundary detection: Unraveled and resolved? *IEEE Trans. on Circuits and Systems for Video Technology*, 12(2):90 – 105, February 2002.
4. W.J. Heng and K.N. Ngan. An object-based shot boundary detection using edge tracing and tracking. *Journal of Visual Communication and Image Representation*, 12(3):217 – 239, September 2001.
5. Chung-Lin Huang and Bing-Yao Liao. A robust scene-change detection method for video segmentation. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(12):1281 – 1288, December 2001.
6. Dan Lelescu and Dan Schonfeld. Statistical sequential analysis for real-time video scene change detection on compressed multimedia bitstream. *IEEE Trans. on Multimedia*, 5(1):106 – 117, March 2003.

7. W. K. Li and S. H. Lai. Integrated video shot segmentation algorithm. In *Storage and Retrieval for Media Databases*, volume 5021 of *Proc. of SPIE*, pages 264 – 271, January 2003.
8. Ze-Nian Li, Xiang Zhong, and Mark S. Drew. Spatialtemporal joint probability images for video segmentation. *Pattern Recognition*, 35(9):1847 – 1867, September 2002.
9. Rainer Lienhart. Reliable dissolve detection. In *Storage and Retrieval for Media Databases 2001*, volume 4315 of *Proc. of SPIE*, pages 219–230, January 2001.
10. X. Liu and T. Chen. Shot boundary detection using temporal statistics modeling. In *Proc. 2002 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume 4, pages 3389 – 3392, May 2002.
11. A. Nagasaka and Y. Tanaka. Automatic video indexing and full-video search for object appearances. In *Proc. of the IFIP TC2/WG 2.6 Second Working Conf. on Visual Database Systems II*, pages 113 – 127. January 1995.
12. Jeho Nam and Ahmed H. Tewfik. Dissolve transition detection using b-splines interpolation. In *Proc. 2000 IEEE Int. Conf. on Multimedia and Expo*, volume 3, pages 1349 – 1352, July 2000.
13. Sarah Porter, Majid Mirmehdi, and Barry Thomas. Detection and classification of shot transitions. In *Proc. of the 12th British Machine Vision Conf.*, pages 73 – 82, September 2001.
14. Yanjun Qi, Alexander Hauptmann, and Ting Liu. Supervised classification for video shot segmentation. In *Proc. 2003 IEEE Int. Conf. on Multimedia and Expo*, volume II, pages 689 – 692, Baltimore, Maryland, USA, July 2003.
15. J.M. Sánchez and X. Binefa. Shot segmentation using a coupled Markov chains representation of video contents. In *Proc. Iberian Conf. on Pattern Recognition and Image Analysis*, June 2003.
16. Juan Maria Sanchez, Xavier Binefa, Jordi Vitria, and Petia Radeva. Local color analysis for scene break detection applied to tv commercials recognition. In *Proc. Third Int. Conf. on Visual Information and Information Systems*, pages 237 – 244, Amsterdam, The Netherlands, June 1999.
17. B. T. Truong, C. Dorai, and S. Venkatesh. New enhancements to cut, fade, and dissolve detection processes in video segmentation. In *Proc. of the eighth ACM Int. Conf. on Multimedia*, pages 219 – 227, Marina del Rey, California, USA, November 2000.
18. Sungju Youm and Woosaeng Kim. Dynamic threshold method for scene change detection. In *Proc. 2003 IEEE Int. Conf. on Multimedia and Expo*, volume II, pages 337 – 340, Baltimore, Maryland, USA, July 2003.
19. H. Yu and G. Bozdagi. Feature-based hierarchical video segmentation. In *Proc. Int. Conf. on Image Processing*, volume 2, pages 498 – 501, October 1997.
20. Jun Yu and M. D. Srinath. An efficient method for scene cut detection. *Pattern Recognition Letters*, 22(13):1379 – 1391, November 2001.
21. R. Zabih, J. Miller, and K. Mai. A feature-based algorithm for detecting and classification production effects. *ACM Multimedia Systems*, 7(1):119 – 128, January 1999.
22. Chengcui Zhang, Shu-Ching Chen, and Mei-Ling Shyu. Pixso: A system for video shot detection. In *Proc. Fourth Pacific-Rim Conf. On Multimedia*, volume 3, pages 1320 – 1324, December 2003.
23. Rong Zhao and William I. Grosky. Video shot detection using color anglogram and latent semantic indexing: From contents to semantics. In *Handbook of Video Databases: Design and Applications*, chapter 15. CRC Press, September 2003.