Indian Institute of Technology, Bombay at TRECVID 2005

M. Nithya Neela Sawant Deepali Singla Sharat Chandran
Computer Science & Engineering Department
Indian Institute of Technology, Bombay
India 400076
http://www.cse.iitb.ac.in/~{mnitya,neela,deepali,sharat}

Abstract

This year, Indian Institute of Technology, Bombay participated in TRECVID 2005 in the task of shot detection. We observe that, even though large number of available shot detection methods perform well under normal conditions (as evinced by the results at Trecvid for the last couple of years), they fail when there are deliberate or inadvertent lighting variations. This might result in labeling some frames as shot-break, where there is no actual shot-break (false positives). We have developed a method to reduce the number of such false positives significantly. Our primary focus is on hard cuts. Our method works in two passes. The first pass predicts approximate locations of shot-breaks in the video sequence. This is done using wavelet analysis of correlation values between successive frames in a video sequence. The second pass consists of detection and removal of false positives from predicted shot breaks of the first pass.

Keywords: *Video Shot Detection, Correlation, Wavelet*

1 Introduction

The shot boundary system developed by ViGIL, IIT Bombay mainly focuses on robust detection of hard cuts. Some situations which might lead to false positives are,

1. Illumination changes due to inter-reflections, user-driven light changes, flash photography. For example, the sun filtering through leaves creates unpredictable lighting changes.
2. Camera effects like zooming and tilting, shaky handling of amateur video footage, fast object and camera motion.
3. Special effects like fire, explosion, screen split

   It is a great challenge to reduce the false predictions and create a robust system. This paper describes our shot boundary detection system which mainly focuses on robust detection of hard cuts, while reducing false positives generated from situations like the above.

In section 2, some of the previous approaches are discussed. Section 3 presents our approach to shot detection. In Section 4, Trecvid result of our system is presented. Concluding remarks are made in the final section.

2 Previous Approaches

The main focus of our work is to reduce the false detection of shot breaks due to sudden changes occurring in a shot like flash photography, reflection from sunlight, jerky camera motion or fast illumination changes. There have been many problem specific attempts to solve these problems. A generic solution has been lacking. In this section we analyze some current approaches for shot detection.

Methods using pixel difference [4, 10], variants of color histograms like color ratio histograms [3] and edge change ratio [9] perform well on simple videos. Fraunhofer HHI's shot boundary system [6], uses combination of pixel difference, histogram difference and edge detection, performs well on real-time videos. But these methods happen to be computationally expensive. The temporal slice method [5] detects shot breaks nicely, and distinguishes between the wipe, motion and color changes using the structural information of the temporal slice. But this may give false alarms for camera motion, as it has the same structural pattern as dissolve. Correlation based methods [8, 7] compute correlation in time or frequency domain and use a threshold to detect shot-break. IBM cue video [2] is another good method for detecting cuts and graduals.

Although these current techniques give decent recall, they are prone to false positives occurring through special cases mentioned above, which reduces the precision. Our approach to increase the precision, is explained in the next section.

3 Our Approach

In this section, we present our shot detection system which is robust to illumination changes, camera effects, fire, explosion, and other special effects. We handle shot detection in two passes. In the first pass, the similarity between successive frames is computed and probable shot breaks are predicted. In the second pass, we improve this prediction by detecting false positives.

The features of the predicted shot breaks are studied to reduce the false positives. In this section, we briefly describe the steps to reduce false positives.

3.1 Similarity Computation

The similarity between two consecutive frames is computed using a normalized mean centered correlation.

The correlation between two frames f and g is computed as,

$$\text{correlation} = \frac{\sum_i (f(i) - m_f)(g(i) - m_g)}{\sqrt{\sum_i (f(i) - m_f)^2}\sqrt{\sum_i (g(i) - m_g)^2}}$$

where m_f and m_g are the mean values of frame f and g respectively. A high correlation signifies similar frames, probably belonging to the same scene, whereas a low value could be an indication of a shot break.

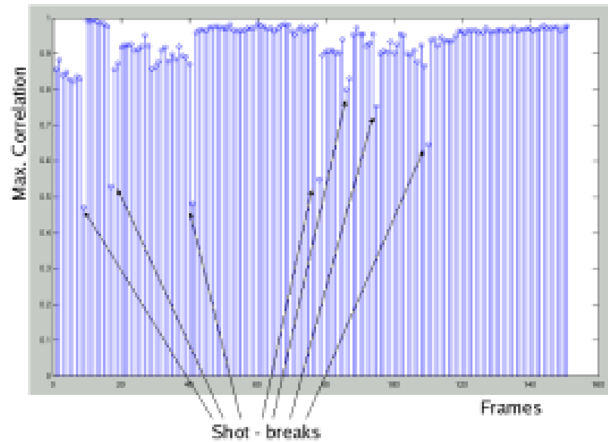


Figure 1. Sample Correlation Sequence in which simple thresholding will fail

3.2 Shot Prediction

The maximum correlation values between successive frames are plotted in Figure 1. The locations of shot breaks as identified by human annotator, are also indicated. From this diagram, it is also clear that putting an empirical value as threshold to detect shot breaks will not work. A delicate shot break, like at frame 85 would mostly be missed if a hard threshold is put.

To overcome this difficulty, we consider the continuity of correlation values rather than the correlation values themselves, as an indicator of a shot. This is achieved using wavelet analysis. We have experimented with different wavelet transforms to detect this continuity and have observed that Morlet wavelet gives a good discrimination between actual shot breaks and false positives.

Morlet wavelet [1] is a complex sine wave, localized with a Gaussian (bell shaped) envelop as shown in the Figure 2. The Morlet wavelet equation used in our computation is,

$$\psi(t) = Ce^{(\frac{-t^2}{2})}\cos(5t)$$

Morlet wavelet detects hard cut effectively. A Hard cut is detected by looking for PPNN pattern (two positive values followed by two negative values) in the

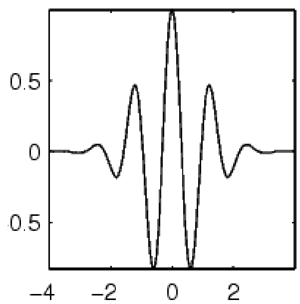


Figure 2. Morlet mother wavelet

lowest scale, and then looking for high values in other scales in and around the hard cut. This wavelet doesn't help in detecting gradual transitions, but detects hard cut accurately. Our method is less sensitive to noise in the correlation signal, as Gaussian function in the Morlet wavelet smoothen the data.

Consider the correlation sequence shown in Figure 3. The diagram shows a fluctuation in the correlation values from frames 215 up to 420. Out of these, frames 215 and 387 look like possible candidates for shot breaks. However, only frame 215 is an actual cut and frame 387, if detected would be a false positive.

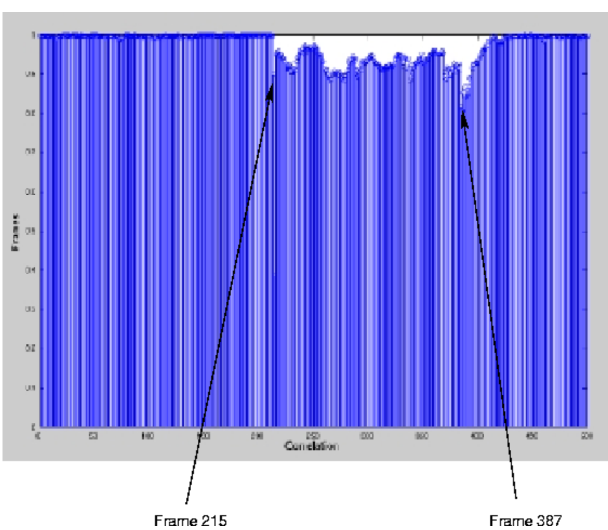


Figure 3. Sample Correlation Sequence

The corresponding Morlet wavelet transform in Figure 4 shows that only frame 215 is detected as hard cut and 387 is not.

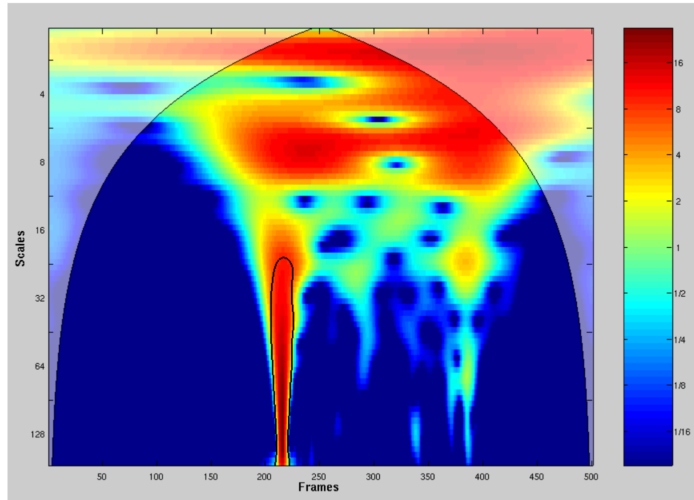


Figure 4. WT Morlet transform of sample correlation sequence

3.3 Reduction of False Positives

The major step, after detecting possible locations of shot breaks is improving on the prediction by analyzing the frames around predicted shot breaks in greater detail. Following measures are used for the same,

1. For the predicted frames, cross-correlation is computed by moving one frame over the other. It results in good correlation even in case of fast motion, either of camera or of object.

2. Due to random lighting variations, the gray-scale value of successive frames in a shot might differ considerably. The false positives resulting from this are reduced by passing the frames through median filters and taking correlations.

3. We handle the low correlations resulting from sub shots by dividing the frame into 3 x 3 sub-frames and then taking the correlation of corresponding sub-frames. In case of sub-shots, we still expect a majority of the 9 correlation values from one frame to be significantly similar to corresponding values in second frame. We developed this measure after the Trecvid submission.

In the second pass, we select the best correlation values generated using the above measures and rerun the process of computing wavelet correlation and detecting discontinuities with these new values. The final results are predicted by taking the intersection of shot breaks from both the passes. This allows us to safely ignore the new values introduced in the second pass.

4 Trecvid Results

In Trecvid 2005, our method achieved a recall of 0.716 and precision of 0.689 for hard cut. Our system was successful in identifying shot-breaks and reducing false alarms for changes occurring within the shot.

In Figure 5, the first sequence contains a flash photograph at third frame. Thought the color of the frame is changed completely, our method successfully eliminated this false positive. In the second example, though there is fast camera motion, our method didn't produce a false positive. In the third example, our method could handle tilting of the camera.

Our method was not designed to handle subshot elimination. It also failed for shots of small duration. However, we were able to handle a few challenging problems from the Trecvid data set successfully.

A Few examples of video sequence, where our system fails is shown in Figure 6. As we don't handle subshots, the first example is detected as a shot break. In the second example, as we did not distinguish between hard-cuts and fades, our method wrongly detects a few fades as hard-cuts.

5 Conclusion

We have presented an effective method which works well even in case of low quality video and varying illumination. It does not require any kind of manual tuning. It is also faster than existing approaches like temporal slice analysis.

The system is still in development state. We have focused mainly on hard cuts. The identification of gradual transitions part is under development, and we are currently developing a method for robust detection of gradual cuts.

References

[1] John J. Benedetto and Anthony Teolis. A Wavelet Auditory Model and Data Compression. *Applied and Computational Harmonic Analysis*, 1(1):3–28, December 1993.

[2] J.R. Kender et al. Ibm research trecvid 2004 video retrieval system. *TREC Video Retrieval Evaluation Online Proceedings, TRECVID*, 2004.

[3] B.V. Funt and G.D. Finlayson. Color constant color indexing. *Pattern Analysis and Machine Intelligence, IEEE*, pages 522 – 529, May 1995.

[4] Chun-Shien Lu. Characterization of abrupt/gradual video shot transitions as unsmoothed/smoothed singularity. *Multimedia Signal Processing, IEEE Workshop*, pages 202–205, 2002.

[5] C. W. Ngo, T. C. Pong, and R. T. Chin. Detection of gradual transitions through temporal analysis. *Computer Vision and Pattern Recognition, IEEE Conference*, pages 36 – 41, June 1999.

[6] C. Petersohn. Fraunhofer hhi at trecvid 2004: Shot boundary detection system. *TREC Video Retrieval Evaluation Online Proceedings, TRECVID*, 2004.

[7] S. S. Porter, M. Mirmehdi, and B. T. Thomas. Video cut detection using frequency domain correlation. *Proceedings of the 15th International Conference on Pattern Recognition, IEEE*, pages 413–416, 2000.

[8] Theodore Vlachos. Cut detection in video sequences using phase correlation. *IEEE SIGNAL PROCESSING LETTERS*, pages 202–205, 2000.

[9] Ramin Zabih, Justin Miller, and Kevin Mai. Feature-based algorithms for detecting and classifying scene breaks. *Third ACM Conference on Multimedia*, pages 189 – 200, November 1995.

[10] H.J. Zhang, A. Kankanhalli, and S.W. Smoliar. Automatic partitioning of full-motion video. *Multimedia Systems*, pages 10 – 28, 1993.

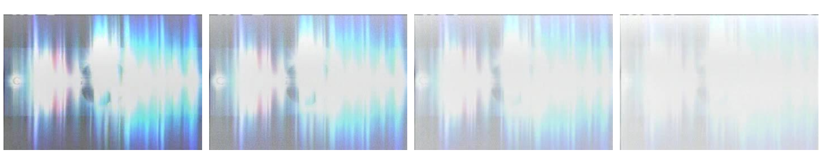
(a) Flash light

(b) Fast camera

(c) Camera tilt

Figure 5. Some challenging problems from Trecvid test video, where our method successfully eliminated false positives

(a) Sub shots



(b) Fade out

Figure 6. Some video sequence from Trecvid test video, where our method falsely detected shot-breaks