Shot Boundary Detection in Football Video Management System

Sanparith Marukatat

Image Laboratory
National Electronics and Computer Technology Center (NECTEC)
112 Thailand Science Park, Phahon Yothin Road
Pathumthani 12120, Thailand

Abstract. Today, video has become an important part in multimedia data which is broadcasted through various networks. Shot boundary detection is a fundamental task in the video processing system. This paper presents a shot boundary detection technique for football video. The detector is based on color histogram with adaptive threshold chosen by the entropic thresholding technique. This allows detecting both cut and gradual transition in the video. A special attention is taken to identify wipes among detected gradual transitions. This system is evaluated on more than one hour of football video. The obtained results are encouraging. An analysis of detection errors is also presented. This can give a guideline for further investigation of shot boundary detection.

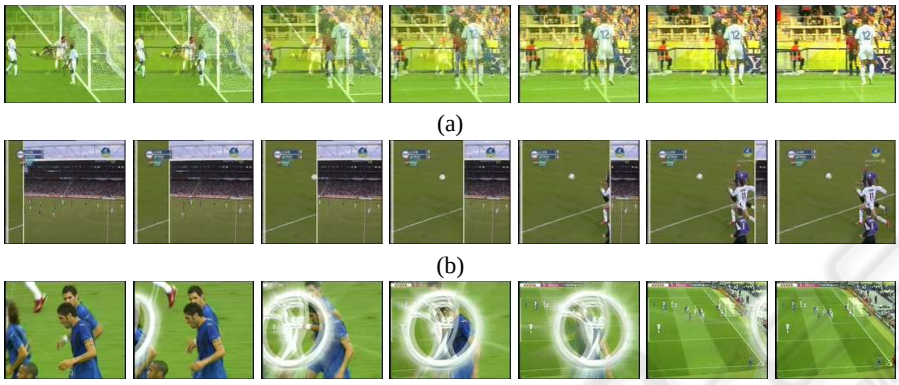
1 Introduction

Today, video, especially sport video, has become an important part in multimedia data which is broadcasted through various networks. With the advance in compression and transmission techniques, user can receive more and more video data. Video management system is then necessary to assist user in exploring their video collection. In this paper, we are interested in football video which represent a large volume of broadcasted sport video in many countries.

A fundamental step in every video analysis (indexing, retrieval or summarization) is shot boundary detection. Shot is defined as a group of frames which are filmed from the same camera. The transitions between shots can be divided in two main categories: abrupt and gradual transition. Abrupt transition, also referred to as a cut, happens when there is a complete change of shot over two consecutive frames. This is the common transition used in video editing process especially in live reports and in sport events. Gradual transition happens when the change spans over a larger number of consecutive frames. Dissolve and wipe are two types of gradual transition which are often found in common video. During dissolve the intensity of disappearing shot gradually decreases from normal to zero while the intensity of appearing shot increases from zero to normal. During wipe transition, both shots coexist in different spatial regions, and the region occupied by the appearing shot grows until it entirely replaces the other [2]. It should be noted that in some sport event, wipe is accompanied by the logo of that event. We

Marukatat S. (2007).
Shot Boundary Detection in Football Video Management System.
In *Proceedings of the 7th International Workshop on Pattern Recognition in Information Systems*, pages 207-214
DOI: 10.5220/0002421902070214

will use the term *logo-wipe* to denote this special kind of wipe. Both wipe and logo-wipe are usually used in transition between a normal play and a replay sequence. Hence they can be a key indicator in event detection module. Figure 1 (a), (b), and (c) show examples of frames during dissolve, wipe and logo-wipe respectively.



(a) (b) (c)

Fig. 1. Examples of images during dissolve (a), wipe (b) and logo-wipe (c).

This paper deals with the detection of both cut, and gradual transition in football video. After reviewing some related works on this subject in Section 2, Section 3 presents our shot boundary detection module. Sections 4 and 5 present our experimental result and the conclusion respectively.

2 Related Works on Shot Boundary Detection

While cut can be reliably detected using some low level features (e.g. pixel, histogram, edge, etc.) the gradual transition detection is still an open issue. Several algorithms have been proposed to deal with gradual transition. In [12] frame differences with value between two thresholds were accumulated and gradual transition was declared when this accumulated score exceeded the higher threshold. In [11] the authors proposed the so-called edge change ratio to detect cut, dissolve as well as fades transition, i.e. dissolve toward a monochrome image (fade out) or from this monochrome image (fade in). The authors argued that these transition effects have their characteristics in the edge change ratio time series. In [6], the author reported that many dissolves do not show the desired characteristics and remain undetected by this technique then proposed a similar measure called edge based contrast. Indeed, during gradual transition, the disappearing shot lose its contrast leading to the reduction of strong edge in favor of the weak edge. As consequent, the authors have designed this measured to accentuate the different between strong edge and weak edge. However, for football scene, the rare strong edges found in the image usually correspond to the line on the field. Therefore, this measure can not reliably detect dissolve transition in our problem.

In [4, 3, 10] the authors supposed that the dissolve transition follows a simple linear transform from the disappearing shot toward the appearing shot. Under this assumption,

it can be proved that the variance curve during the dissolve will have a parabola form. The authors proposed to analyzed this variance curve in order to identify the candidate dissolve region. Unfortunately, in our preliminary experiment, we have found that the variance curve on football video exhibits a parabola form event on non dissolve area. This is mainly due to motion contained in the video.

Another approach to gradual transition detection is based on machine learning tools like SVM [9, 7, 1, 8]. In these works, the authors used SVM to combine multiple features in order to classify if a frame is part of cut or dissolve or not. In [9] the authors used frame difference based feature along with the likelihood of current camera motion as feature in their system. In [7] SVM was used with the so-called variance projection function features. [1] proposed an SVM-based cut detection using color histogram, Zernike moments, Fourier-Mellin moments, projection histograms, and phase correlation method features. In [8], a dozen of SVMs were used in a 2-stage classification system working with more than 100 features to be extracted. These techniques reached high recalls and precisions but with large overhead on features extraction. Moreover, for a task dependent as in our case, we believe that a more simple technique should be adopted. In this work, we investigate the use of histogram based difference with adaptive threshold in detecting both cut and gradual transition.

3 Proposed Shot Boundary Detection

This work is based on histogram different between frames in order to detect shot boundary. Subsection 3.1 describes the features used for cut and gradual transition detection. Subsection 3.2 describes how to choose an appropriate threshold for each video. In Subsection 3.3 we describe how to deal with large motion which is normally present in the football video.

3.1 Histogram Based Frame Difference

We suppose that all transitions (both cut and gradual) happen between two shots with different color distributions. To detect shot transition, color histogram is used to measure the difference between frames. The histogram difference between two frames F_1 and F_2 is given by:

$$d(F_1, F_2) = 1 - \frac{1}{WH} \sum_{i=1}^{n} \min \{Hist(F_1, i), Hist(F_2, i)\} \quad (1)$$

where W and H are width and height of each frame, n is the total number of bins in the histogram and $Hist(F_1, i)$ is the count associated with the bin i in the histogram of frame F_1.

Our cut detector relies on this histogram based difference between two consecutive frames. For gradual transition like dissolve the difference between consecutive frame is relatively small. Hence comparison should be done between frames a certain step apart. As consequent, for gradual transition detection, we compute the histogram different between frame $t + w$ and frame $t - w$, where w is the window size determined experimentally. This *skipped-frame difference* is used as feature to determine if frame t is part of gradual transition or not.

3.2 Entropic Thresholding

The two thresholds T_{cut} and $T_{gradual}$ will be used to detect cut and dissolve respectively. Finding common thresholds for every video seems not to be realistic. However, we believe that for a single video, we can choose appropriate thresholds for cut and for gradual transition detection. First, we notice that shot boundaries are only a small part in a video. Therefore a large number of frames will be concentrated on low frame difference values and only a small number of frames will have high difference values. This is similar to document binarization problem where large number of pixels is concentrated on white value that is the background and only a small number of pixels have black value. Entropic thresholding has been applied with success to document binarization [5]; hence it should be able to handle this threshold selection problem as well.

The basic idea is to select the threshold which yields maximum entropy for the two sets namely the set of values lower than this threshold and the set of values higher than this threshold. Let $P_1, P_2, \ldots, P_m$ be a histogram of values we considered, e.g. frame difference or skipped-frame difference, with m bins. For each bin i we compute

$$H_{low}(i) = -\sum_{j=1}^{i} \frac{P_j}{Q_i} \log \frac{P_j}{Q_i} \tag{2}$$

$$H_{high}(i) = -\sum_{j=i+1}^{m} \frac{P_j}{1-Q_i} \log \frac{P_j}{1-Q_i} \tag{3}$$

with $Q_i = \sum_{j=1}^{i} P_j$. The entropic threshold T_{ent} is chosen as the mid value of the $i*$ bin given by

$$i* = \arg \max_{i=1,\ldots,m} \{H_{low}(i) + H_{high}(i)\}. \tag{4}$$

The threshold T_{cut} is selected by applying this entropic thresholding technique on the set of consecutive frame differences. A cut is declared whenever a consecutive frame difference is higher than T_{cut}. In analogeous manner, the threshold $T_{gradual}$ is selected by applying this entropic thresholding technique on the set of skipped-frame differences. A gradual transition is declared whenever a skipped-frame difference is higher than $T_{gradual}$.

3.3 Filtering High Activity Areas

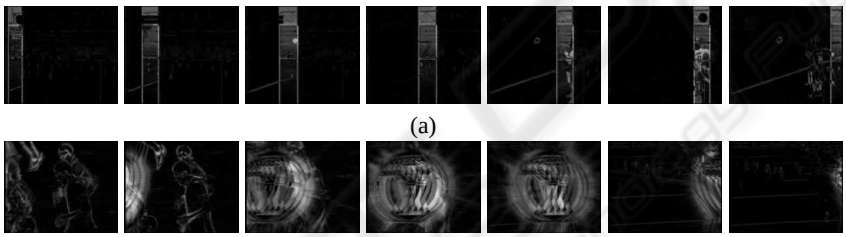
The skipped-frame difference can be used to detect gradual transition area but unfortunately it also yields high value for sequences containing large motion or high activity. Not only are the gradual transitions detected in these areas not reliable but also the detected cuts. It is then necessary to filter out the shot boundaries detected in these areas.

Usually, the high activity areas contain higher frame difference value than normal but of course lower than that of cut transition. A simple heuristic to detect these large motion areas is based on another entropic threshold on frame differences. Indeed, the frame differences which are higher than T_{cut} are first filtered out. Then another entropic threshold, denoted as T_{ha}, is selected using the remaining frame differences. The frame t whose frame difference is higher than T_{ha} is considered as part of a high activity area.

High activity area is supposed to be at least 5 frames long. Cuts and gradual transitions which correspond to the change from high activity area to another high activity area are considered as not reliable and are removed.

3.4 Wipe and Logo-transition Identification

Usually, in normal wipe, the new shot first appear on the left side of the screen then it enlarge toward the right side or vice versa. Thus, if we consider the pixel-based difference between any consecutive frames in these transition areas, we should see a group of pixels with large difference moving either from left to right or from right to left. Figure 2 (a) and (b) show examples of the pixel-based difference during wipe and during logo-wipe presented in Figure 1 (b) and (c) respectively. In this work, wipe is first detected as a gradual transition. Then for every detected gradual transition area, we use the variation of the abscissa of the center of mass from pixel-based difference between consecutive frames as feature to detect wipe.



(a)

(b)

Fig. 2. Examples of pixel-based difference during wipe (a) and logo-wipe (b) presented in Figure 1 (b) and (c) respectively.

4 Experiments

Five football videos were used in these experiments. The first and second videos are from the match between France and Italy in final FIFA world cup 2006 in DVD quality. The first one is the debut of the match including scenes of players entering the stadium and singing the national anthems. The second one is during the match play including the goal scene. The other 3 videos are recorded from TV broadcasting in lower quality. These 3 videos correspond to 3 different matches in different stadiums, thus present different field colors, different crowds, as well as different commercial boards along the field. Figure 3 present examples of image from these 5 videos. The shot boundaries in these videos are manually labeled. The Table 1 summarizes the statistics of these 5 videos.

For these experiments, RGB colors pace was used with 8x8x8 bins histogram. The window of 5 frames was used to compute the skipped frame difference. In these experiments, all video images were first resize to 180x120 before computing the histogram.

(1) (2) (3) (4) (5)

Fig. 3. Examples of images from five videos.

Table 1. Number of frames and duration in videos used in these experiments.

video	#frames	duration	#cut	#gradual	#wipe
1	40268	00:14:12	93	112	26
2	40306	00:14:12	200	60	30
3	32816	00:21:36	233	28	24
4	22055	00:14:31	63	30	0
5	21896	00:14:25	100	29	0
total	157341	01:18:56			

To evaluate the performance of our system, we measure the classical recall and precision for both detected cut and gradual transition. In this work, a detected gradual is considered as correct if it overlaps at least 10% with a true gradual transition segment.

Tables 2 and 3 present the result of cut and gradual transition detection from five videos. From these results, we may see that the cut detection can be done with average recall up to 95.7% while having the average precision of 96.3%. This is encouraging results compared to the performance of cut detection reported in other works. For gradual transition, lower recall and precision were obtained, i.e. 86.9% and 61.4% respectively.

Table 2. Cut detection results.

video	ground truth	correct	miss	false	recall	precision
1	93	92	1	1	98.92	98.92
2	200	192	8	0	96	100
3	233	229	4	2	98.28	99.13
4	63	60	3	5	95.24	92.31
5	100	90	10	9	90	90.91

While the gradual transition's recall was acceptable, the obtained precision was too low. In order to get better idea about the behavior of the system, we analyzed the video 5 where the lowest precision was obtained. The principal error in video 5 occured when the camera followed some player who walked pass different backgrounds. In this case, the color distribution in the image slowly changes just like during dissolve. The second types of error happened in close up shots when the focused player was occluded by some other player. This will cause similar effect as a wipe. Figure 4 (a) and (b) show some examples of these two principal causes of error.

Table 3. Gradual transition detection results.

video	ground truth	correct	miss	false	recall	precision
1	112	108	14	18	88.52	85.71
2	60	50	10	41	83.33	54.95
3	28	26	2	20	92.86	56.52
4	30	25	5	14	83.33	64.1
5	29	25	4	30	86.21	45.45

(a) (b)

Fig. 4. Examples of two principal errors that happens in video 5.

For wipe identification, we obtained 96.15%, 91.67% and 56.52% from videos 1, 2, and 3 respectively. The first two videos 1 and 2 used logo-wipe instead of normal wipe. As the size of logo was fairly large, the detection task was made easier. For video 3 where usual wipe was used, the identification fail especially when the wipe was used between shots containing high motion. We believe that the proposed wipe identification technique can be modified to better handle the normal wipe transition.

5 Conclusion and Future Works

This paper presents our shot boundary detection system for football video. The color histogram is used with automatically selected thresholds by the entropic thresholding method. This system reaches a good recall and precision for cut. For gradual transition, moderate recall and precision are obtained. This is due to some errors which frequently happen in close up shots. Our future works will include mechanism to deal with these errors.

References

1. G. Camara-Chavez, M. Cord, S. Philipp-Foliguet, F. Precioso, and A. de Albuquerque Araújo. Robust scene cut detection by supervised learning. In *EUSIPCO*, Firenze, Italy, 2006.
2. C. Cotsaces, N. Nikolaidis, and I. Pitas. Video shot detection and condensed representation. *IEEE Signal Processing Magazine*, pages 28–37, March 2006.
3. W. A. C. Fernando, C. N. Canagarajah, and D. R. Bull. Fade and dissolve detection in uncompressed and compressed video sequences. In *Proceedings of the 1999 International Conference on Image Processing (ICIP '99)*, volume III. IEEE Computer Society, 1999.
4. A. Hampapur, R. Jain, and T.E. Weymouth. Production model based video segmentation. *Multimedia Tools and Applications*, 1(1), 1995.

5. J. N. Kapur, P. K. Sahoo, and A. K. C. Wong. A new method of gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29:273–285, 1985.
6. Rainer Lienhart. Comparison of automatic shot boundary detection algorithms. In *Image and Video Processing VII 1999, Proc. SPIE*, 1999.
7. J. Ling, Y.-Q. Lian, and Y.-T. Zhuang. A new method for shot gradual transition detection using support vector machine. In *Proceedings of the Fourth International Conference on Machine Learning and Cybernatics*, pages 5599–5604, 2005.
8. K. Matsumoto, M. Naito, K. Hoashi, and F. Sugaya. Svm-based shot boundary detection with a novel feature. In *Proceedings of the Fifth International Conference on Machine Learning and Cybernatics*, pages 1837–1840, 2006.
9. Y. Qi, A. Hauptmann, and T. Liu. Supervised classification for video shot segmentation. In *IEEE Conference on Multimedia & Expo (ICME'03)*, 2003.
10. Jing-Un Won, Yun-Su Chung, In-Soo Kim, Jae-Gark Choi, and Kil-Houm Park. Correlation based video-dissolve detection. In *Proceedings of the International Conference on Information Technology: Research and Education (ITRE)*, pages 104– 107, August 2003.
11. R. Zabih, J. Miller, and K. Mai. A feature-based algorithm for detecting and classifying production effects. *Multimedia Systems*, 7(2):119–128, 1999.
12. H. J. Zhang, A. Kankanhalli, and S. W. Smoliar. Automatic partitioning of full-motion video. *Multimedia Systems*, 1(1):10–28, 1993.