# Online Play Segmentation for Broadcasted American Football TV Programs *

Liexian Gu[1], Xiaoqing Ding[1], and Xian-Sheng Hua[2]

[1] Department of Electronic Engineering, Tsinghua University, Beijing, China
{lxgu, dxq}@ocrserv.ee.tsinghua.edu.cn
[2] Microsoft Research Asia, Beijing, China
xshua@microsoft.com

**Abstract.** This paper proposes an online segmentation scheme for broadcasted American football TV programs. The TV video streams are online partitioned into a series of play-break patterns with a relatively short delay. All the detected plays form an event-based summarization which contains complete information about the original match. The main contributions of the proposed scheme are twofold. First, an online adaptive field-color model is employed to deal with feature variations which may caused by court configurations and time-varying lighting conditions. Second, a temporal pattern based model is constructed to capture intrinsic features of football plays, which considerably increased the performance of the proposed scheme. Effectiveness of our approach is demonstrated by experiments on real recorded TV programs.

## 1 Introduction

Due to large potential applicable demands, sports video analysis has been an active research topic in recent years. Key techniques include highlights detection, automatic summarization, program structure parsing, etc. A typical application scenario can be described as follows: A busy-working football fan who will miss a live-broadcasted NFL game can get it recorded using personal video recorder (PVR) or other set-top devices. However when he has time to watch the recorded match, he may only want to enjoy the highlights or a shorter but complete summary, skipping those less interesting portions such as game pause, foul and various breaks. The so-called "smart playback" function has been approved appealing to end users with benefits of not only saving time but also bringing the feeling of being in control of what they watched[1].

In this paper we focus on broadcasted American football programs. According to the rules of football matches[2], a broadcasted game is generally composed of a series of attacking attempts with time-consuming breaks between them. Broadcasters often fill these breaking slots with close-ups, full-field views or audience scenes. Typically a complete NFL game will cost three hours but only about one hour is covered by real actions (which means that teams are truly

* This work was performed at Microsoft Research Asia.

playing in the field). Those active segments, called plays, convey all contents for a football game in a considerable compressed manner. Hence we work on algorithms to automatically segment plays from broadcasted football video and concatenate them to form a compact summary of the whole game.

There have been a number of previous works for various types of sports videos based on different models. In [1], TV baseball game highlights were extracted on set-top devices using audio track features and probabilistic fusion. Xu et al.[3] employed HMMs to characterize motion pattern features, classifying basketball video into 16 basic events. Several probabilistic algorithms based on generic cinematic features to detect soccer goal events and basketball playbreaks were presented in [4][5]. However, little efforts have been made in football domain. Babaguchi[6] proposed an inter-modal collaboration scheme to detect football highlights through analyzing closed caption stream and a personalized retrieval and summarization system for sports video based on semantic metadata was presented in [7]. The most relevant research was reported by Li et al. in [8][9], where the start points of plays were determined by the scenes that players lined up; however, this model cannot cover a variety of plays which may started without such distinct pictures (some detailed analyses on this algorithm will be given in Sect. 2). Furthermore, none of the aforementioned approaches have the capability of online (means in real time or real time with a short period of delay) segmentation, which is important for many applications such as online skipping, namely, skipping forward or backward directly to previous or next interesting segment when watching buffered live TV programs.

The rest of this paper is organized as follows. Based on observed characteristics of various broadcasted football games and analyses on difficulties of existed algorithms, we propose a temporal pattern based online play detection scheme with adaptive field color models in Sect. 2. Performance evaluations are presented in Sect. 3, followed by conclusions in Sect. 4.

## 2 Our Approach

It is observed that in football games all plays are played in the field, which is green or near green and marked off by latitudinal stripes. Those field lines must be clearly discernable to the players and audiences since they measure the distance of ball advancing in each bout. Generally there are global views of the stadium, audience and other non-field frames in breaking intervals. Hence field scene is a strong indication of "play" events in football video. Nevertheless the dominant color of the field, green, is heavily depends on different courts. Moreover, it may vary significantly due to lighting conditions even in the same field since a football game often lasts three or more hours (for example, may from the afternoon to the evening). In [8][9], the green color of a specific game was calibrated using all potential field frames before the detection process. Such method lacks online detection capability and cannot adapt itself to the color drift along the game. To overcome this problem, in our system, the definition of green color is online dynamically updated. Furthermore, we model plays with their intrinsic temporal variations to make the detection scheme robustly applicable to various broadcasting patterns.

Diagram of our play segmentation system is shown in Fig. 1. Football video frames are fed into the adaptive color adjustment module and the play detection module simultaneously. As aforementioned, though the dominant color of the field is green, but the hue, saturation and brightness (in HSV color space) of this green color are not the same for different courts and even continue changing in the same court. Accordingly, the green color here is defined as a relatively wider (looser) interval in the color space in the initial stage and we will keep it updating dynamically to a relatively narrow interval when time is proceeding. Details of the adaptive field model and the play detection scheme are discussed in Sect. 2.1 and 2.2 respectively.

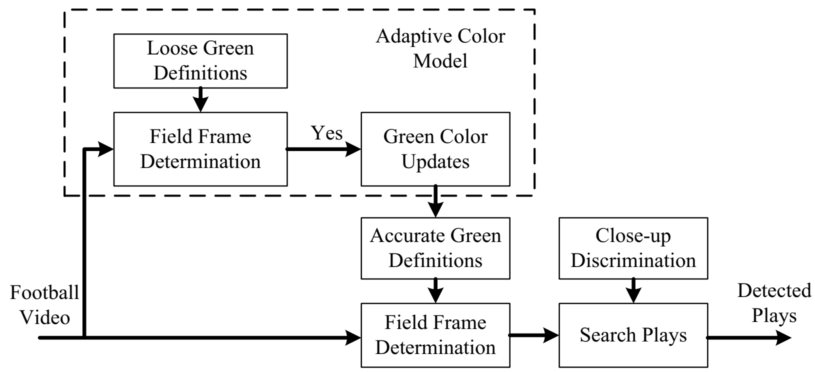


**Fig. 1.** Block diagram of our play-segmentation system for American football video

### 2.1 Adaptive Field Color Model

An authentic field frame is determined if it contains dominant green pixels and sufficient field lines. Field lines can be declared if there are more than two lines in a green background area, which should be sufficiently long and nearly parallel with each other. Although some field scenes cannot satisfy such strong constraints and are abandoned, those selected frames are sufficient to dynamically learn the accurate definition of the dominant color of the current game field in a sensitive and stable manner.

Our online adaptive green color adjustment procedure is described as follows:

1. For every incoming frame, count green pixels by the loose definition of "green". If green is not the dominant color of this frame, skip it, and then process the next one. Otherwise mark those green pixels with 1 and others are set to 0, and thus a green masked image is created.
2. Detect edges using Canny detector within the dilated green mask.

3. Find straight lines using Hough transform in the edge image.
4. If detected lines satisfy the aforementioned conditions, current frame is declared to contain fields. The color histogram of the green mask is cumulated into a buffer.
5. Seek those maximum dominant bins in accumulated histogram. The corresponding color range is the updated definition of field color.

Figure 2 shows an example where a frame is determined as field frame.

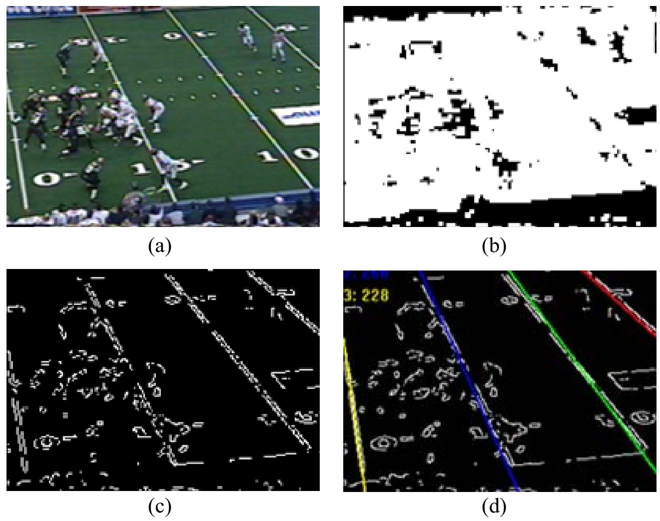


**Fig. 2.** An example of field frame selection by online green color learning procedure. (a) A sample frame image. (b) Dilated green mask image. (c) Masked edge image. (d) Detected field lines

### 2.2 Temporal Pattern Based Play Segmentation

In [8][9], the detection scheme for the start-of-play frames depends on the observation that all players line up to prepare the play while a side camera is capturing the scene, producing a picture similar to Fig. 3(a). Then combinational rules of dominant field color, field lines, clothes color, player lines and camera motion are employed to determine it. However, this is not the only pattern in broadcasted TV games. A play can be started with a medium close-up view of players captured by an end camera, as shown in Fig. 3(b), especially in replaying events and field goals. And place kicks are always started from end-zones in a global view, as shown in Fig. 3(c). In these situations, green color is not dominant and lines are often too weak to be correctly detected. Therefore the rule-based approach in a single frame would fail to recognize such plays.

(a) (b) (c)

**Fig. 3.** Various types of start-of-plays. (a) exhibits distinct visual pattern of a dominant green region with field lines. (b) is a typical replay. (c) is the first frame of a place kick

Though the start frame often does not contain sufficient field regions, a football play always contains typical field frames or segments in the middle. For instance, a place kick play begins with a frame image at low green ratio; however, after the ball is kicked out, the scene is usually switched to a view taken by a side camera to show events in the field, which can be detected as field frame rather easily. Such a typical temporal pattern depicted by the green-ratio curve and the number-of-lines curve is shown in Fig. 4. Then plays can be detected heuristically by searching these distinct patterns. Employing intrinsic correlations in temporal domain, it will be capable of detecting various types of football plays, as presented below.

1. Search segments including sufficient consecutive field frames.
2. From each segment, find the frame backwards where the green-ratio value has fallen beneath a predefined threshold. It is regarded as the start frame of this candidate play.
3. Take the following camera break as the end of current play because after a play the camera usually switches to capture non-action scenes.

Additional steps should be performed on those candidate plays to filter out false alarms. Close-up frames of certain players in the field also produce high green ratio values. From characterization of longest black runs in both horizontal and vertical directions, a frame containing heavy black blobs is recognized to be a close-up and should be removed from candidate plays. Sample images are shown in Fig. 5.

An event-based summary of a football game is then obtained through concatenating all detected plays. We have implemented a prototype system with which users can browse plays and randomly navigate among them efficiently for recorded TV football games.

## 3 Performance Evaluation

We have tested our play segmentation algorithms on three football video segments recorded from TV programs. The total length is about three hours. Videos are all encoded by MPEG-1 with 320×240 frame resolution. There are complex

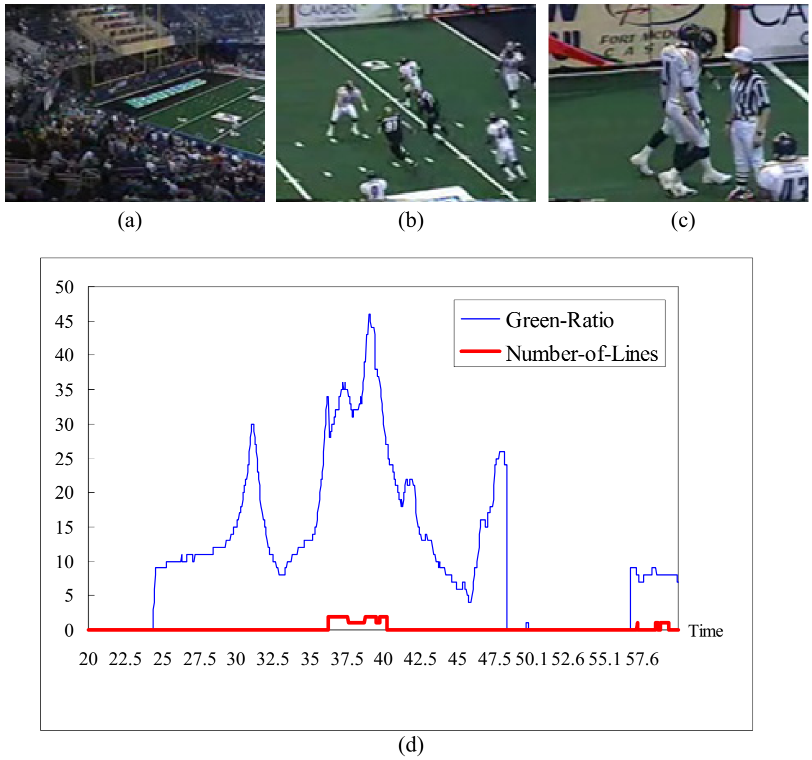


**Fig. 4.** The temporal pattern of a place kick play from 24s to 48s in a test video. (a) is the start frame of the play at 24s where green pixels are not dominant because it is captured from a global view camera. (b) is a middle frame at 37s where green pixels are dominant and field lines are distinct. (c) is the end frame at 48s. (d) depicts the temporal pattern of the whole play, where the segment from 36s to 40s can be firstly determined as field segment and then triggers the play detection module


**Fig. 5.** Samples of close-up detection. (a) A close-up frame. (b) Masked image to detect whether there are heavy black blobs

plays due to various broadcasting styles in our test videos, such as replays from different views and reviewing plays excerpted from historical games, which pose great challenges to detect it correctly and precisely. Detected plays are compared with manually labeled ground truth. To measure detection performance, two commonly used metrics, *recall* and *precision*, are defined as follows:

$$Recall = \frac{N_c}{N_c + N_m}, \qquad Precision = \frac{N_c}{N_c + N_f},$$

where $N_c$ represents the number of correctly detected plays, $N_m$ is the number of missed plays and $N_f$ denotes the number of false detections.

Our experimental results are listed in Table 1. Compared with the system developed in [8][9], it achieved 0.68 recall, 0.87 precision in this test video set. It failed to detect those plays which started without distinct pictures of players lining up, for instance, it missed starts of all kickoff events. However, they could be detected successfully by our algorithms because the temporal pattern based play model handled these situations effectively. Therefore considerable improvements in recall rates were obtained.

**Table 1.** Test results of play segmentation

| Video | Ground Truth | Detected Plays | Recall $R$ | Precision $P$ |
|---|---|---|---|---|
| 03-16-I.mpg | 70 | 89 | 0.857 | 0.700 |
| 03-23-I.mpg | 88 | 106 | 0.750 | 0.708 |
| 03-23-II.mpg | 88 | 112 | 0.796 | 0.661 |
| Average | – | – | 0.801 | 0.690 |

Most missed plays were those replaying shots captured from a different angle with low green-ratio and weak lines all the while so that no field scene segments could be determined. The precision was a bit lower because some non-action behaviors in the field are misclassified into plays.

## 4 Conculsions

We have presented an automatic online play segmentation and summarization system for broadcasted American football videos. Two major innovative contributions over previous approaches are the online field color learning scheme and the temporal pattern based modeling of football plays. Experiments on real-recorded programs have given promising results. We are investigating situations where the play model breaks down to further improve the detection performance. Moreover, it is more challenging to classify football plays into semantic classes such as Touch Down, Field Goal, Pass, Fumble, etc, which will be our future work.

## References

1. Rui, Y., Gupta, A., Acero, A.: Automatically Extracting Highlights for TV Baseball Programs. Proc. ACM Multimedia. (2000) 105–115
2. American Football: Rules, http://www.allfootballs.com/afrules.html
3. Xu, G., Ma, Y.-F., Zhang, H.-J., Yang, S.: Motion Based Event Recognition Using HMM. International Conference on Pattern Recognition. (2002)
4. Ekin, A., Tekalp, A. M.: Generic Play-Break Event Detection for Summarization and Hierarchical Sports Video Analysis. IEEE International Conference on Multimedia and Expo. (2003)
5. Ekin, A., Tekalp, A.M., Mehrotra, R.: Automatic Soccer Video Analysis and Summarization. IEEE Transactions on Image Processing. 12(7) (2003) 796–807
6. Babaguchi, N.: Towards Abstracting Sports Video by Highlights. IEEE International Conference on Multimedia and Expo. (2000)
7. Babaguchi, N., Ohara, K., Ogura, T.: Effect of Personalization on Retrieval and Summarization of Sports Video. Proc. Fourth IEEE Pacific-Rim Conference on Multimedia. (2003)
8. Li, B., Sezan, M.I.: Event Detection and Summarization in Sports Video. IEEE Workshop on Content-Based Access of Image and Video Libraries (Joint with CVPR2001)
9. Li, B., Sezan, M.I.: Event Detection and Summarization in American Football Broadcast Video. Proc. SPIE Storage and Retrieval for Media Databases. vol. 4676 (2002) 202–213