ABSTRACT
We propose an automatic soccer video summarization engine which relies on an improved algorithm for the detection of replay shots which delineate interesting events. Video shots are first detected using dominant color and histogram intersection methods. Replay shots are detected using an improved technique, then processed through a set of mid-level descriptors (goal-mouth and score board with other cinematic features) and finally fed into a rule-based classifier. The proposed summarization has been experimentally tested on 6 hours from 10 videos in total. The proposed logo-based replay detection technique achieves 100% recall with 96.3% precision. Interesting events such as goals are detected with 100% recall and 100% precision, attacks with 91.7% recall and 86.7% precision) and other events such as (fouls, free kicks etc) with 90.8% recall and 95% precision.

Index Terms— soccer video summarization, event detection in soccer video, replay detection

1. INTRODUCTION

Sports attract many people around the globe especially soccer. With rapidly evolving telecommunication technologies, people may not have the time to watch most of the content being broadcasted. Video content summarization presents itself as a promising solution. Through automated video content analysis, the user could be presented with a summary containing the most important events such as goals, attacks and other events. Hence, summarization is an essential part in several applications such as video retrieval and browsing.

In this paper, we present a soccer video summarization technique. The output summary is an aggregate of clips containing a set of interesting events from a user perspective, namely attacks, goals, free kicks and fouls. While we do not claim novelty in the main approach, we outline the main contributions as: a) a fast replay detection method based on logo detection where the optimization comes from looking through the first 30-frames for each shot rather that looking through all the frames, b) an improved shot-boundary detection and c) a detailed performance analysis for each of the summarization stages.

The paper is organized as follows. In Section 2, we present an overview of the proposed system. Section 3 discusses the preprocessing stage involving dominant color extraction and modified shot boundary detection (based on [1]). Section 4 presents the event detection stage which includes replay detection, shot classification, goalmouth and score board detection. Section 5 gives our experimental results. Finally Section 6 draws conclusions and discusses our future work.

2. RELATED WORK

Our proposed soccer video summarization system relies on event detection which has been studied in the literature by many researchers (some exceptions exist as in [6], where authors produce summaries without detection of events). Events are usually detected through defining a set of features (low-level or mid level) and then building a machine model for each trained model (or simply devising some rules in rule-based event detection techniques) [1, 2, 3, 4, 5, 7]. Examples events used in the literature are goals and attacks as in [10] where authors use goalmouth detection combined with audio energy feature. In other work [11], authors detect goals based on audio and visual keywords. A particular case of important event is the slow-motion replay indicating an important or exciting event [9][12]. On the modeling side, Ekin [1] constructs comprehensive sports video summarization and analysis using a structural-semantic video model. In [2], C. Huang et al. introduce semantic analysis based on Bayesian and dynamic Bayesian networks to identify a group of events (goals, penalties, cards and corners). In [3], detecting goal event employs audio track feature based on a Hidden Markov Model. In [8], the authors use GMM for play field detection.

Given the prevalent use of logos as a video editing technique, we focus in this paper on proposing an enhanced and fast method for the detection of logos. We study the problem of extracting logos using a combination of low-level features. Then, we use a rule-based approach for modeling cinematic features to classify events into goals, attacks and other events which are aggregated to produce an output summary.
2. SYSTEM OVERVIEW

The system processes the soccer’s game video and finally outputs a summary which contain the exciting events. The flow chart of the proposed system is shown in Fig. 1 which has two stages: a) preprocessing and b) event detection.

Fig. 1 Flow chart of the proposed system

3. PREPROCESSING STAGE

The goal of this stage is to segment video into shots. First, we detect the dominant color in the video frame, and then use a modified version of the shot boundary detection algorithm in [1] to output video shots based on dominant color derived features.

3.1. Dominant Color

Each sport has its own playfield with a handful of dominant colors. In soccer, the field is usually green as it is made from grass in most of the cases. Dominant color extraction is challenging due to effects on the playfield such as shadow, lighting and other environmental factors. In the literature, several color spaces have been used for the dominant color detection including HSV and RGB. We have experimented with both, and opted to use RGB than HSV proposed in [1], for its efficiency and satisfactory results as shown in Fig. 2. We define a color range that covers the different variations of the green color of the playfield. The range of values are experimentally determined as follows 0<R<150, 94<G<255, 0<B<100.

Fig. 2 Dominant Color examples

3.2. Shot Boundary Detection

Shot is defined as a sequence of frames captured by a single camera in a single continuous action in time and space [16]. Our shot boundary detection technique follows the approach in [1] but using RGB dominant color. We use two set of features to detect a shot: 1) the difference between dominant colored pixel ratios of two frames (Gdistance) and 2) the difference in color histograms based on HSV color space (Hdistance). We define these thresholds to detect the boundaries with three different ranges as shown in Table 1. Furthermore, we introduce a step of 8 frames as in [1] to convert a gradual transition into a cut transition. This enables us to generate shots where the logo is the starting frame (as usually a logo appears for a couple of frames as a gradual transition).

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gd ≥</td>
<td>0.5</td>
</tr>
<tr>
<td>Hd ≥ &amp; Gd ≥</td>
<td>0.8 &amp; 0.2</td>
</tr>
<tr>
<td>Hd ≥</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Table 1: Thresholds adjusted from soccer video

These thresholds have been experimentally selected to maximize the chance that logo will be treated as a single shot entity. The proposed shot boundary detection is able to separate different type of views (medium, close up, and logo) with the condition that any shot must be longer than a predefined time duration (experimentally taken as one second) to prevent the occurrences of very short irrelevant shots. These thresholds are working on most of videos but there were some non-noticeable failure cases due to fast camera transitions.

4. EVENT DETECTION

This stage takes as input the shots from the shot detection phase and flags them with one of predefined event types. In this paper, we use the following classes of events: a) replay, b) shot depicting scoreboard, c) shot depicting a goalmouth,
4.1. Replay Detection using Logo Based Detection

In most soccer videos, exciting events are often replayed to emphasize an important segment with a slow-motion pattern or logo appearance for one or several times. Replay detection can be done in two main ways: a) detecting slow motion pattern and b) detecting logos which have been recently used as an editing effect before and after a replay. Unfortunately, many difficulties face the detection of slow motion shots, because slow-motion shots have different speeds from one broadcaster to another to the extent that there are non-slow motion shots that have the same characteristics of slow motion shots. Hence we prefer to use logo based replay detection. The logo detection algorithm is shown in Fig. 3.

![Fig. 3 Logo based replay detection algorithm](image)

For speed purposes, the logo detector operates on the first 30 frames of a given shot. Then it binarizes the entire frame using an intensity threshold of 128. After binarization, we look at the percentage white pixel in a frame (“the white ratio”). A logo candidate is signaled to occur if the white ratio is > 55% (experimentally selected). To reduce false positives, we proceed to a second step: for each championship, we manually extract the logo image from which we derive the RGB color mean as shown in Table 2. We compare the RGB color mean for the extracted frames with the manually extracted logo color mean to indicate the appearance of logo. This procedure may fail for other types of broadcasts but most of modern broadcasts use logo-based replays.

<table>
<thead>
<tr>
<th>Frame Type</th>
<th>Color Mean Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>G &gt;= 0.5 &amp; G &lt; 0.75</td>
</tr>
<tr>
<td>Medium</td>
<td>(G&gt;=0.35 &amp; G &lt; 0.5)</td>
</tr>
<tr>
<td>Close-up</td>
<td>G &lt; 0.35 &amp; B &lt; 0.87</td>
</tr>
<tr>
<td>Audience</td>
<td>G &lt; 0.35 &amp; B &gt;= 0.87</td>
</tr>
</tbody>
</table>

where G represents the percentage green pixel in a frame (“the green ratio”).

4.2. Shot Classification

Cinematographers classify a shot into three classes long, medium and close-up. Most papers [1, 7, and 14] consider close-up and audience as a single shot class. In this paper, we present a novel technique to separate the close-up from the audience class, enabling us to define or refined rule-based event detectors as in Section 4.4. The classifier used for the four classes above is based on manually tuned thresholds as follows.

4.3. Score board and Goal-Mouth Detection

The score board provides information about the game and players thus provides an important cue for event detection as was used in [2, 13]. Our score board detector relies on identifying the caption that appears at the bottom part of the frame and remaining present for a minimum duration (experimentally found to be 5 seconds). An abrupt temporal intensity at the bottom part of a video frame is used to detect the appearance of the caption which is manually adapted. As for the goal-mouth detector, we first note that most of the exciting events occur in the goal-mouth area which can be selected as highlighted candidates [14]. For goal-mouth detection, first, we detect the goal post and its crossbar by searching for the white color [15], and then look for an intersection point between them.

4.4. Attack, goal and other interesting events detection

Based on the previously detected events, we proceed to combine these built hypotheses to form higher-level events such as attacks, fouls and goals as shown in Fig. 4. In a goal event, the duration between two replay logos must be within a certain time range (experimentally found to be no less than 16 seconds and no more than 40 seconds). Wherein the attack event, the duration between two replays. Logos must be no less than 10 and no more than 30 seconds. Finally, for other events (fouls, booking, injury and offside), the duration must be no less than 4 and no more than 15 seconds.
5. EXPERIMENTS

To evaluate the system performance we have used more than 6 hours of soccer games in total including 5 matches from UEFA 2008, 4 matches from Africa Championships League (FCL) 2008 and 1 match from EURO 2008. Table 3 shows the results of goals, attacks and other events for one hour manually taken from 4 soccer matches. The system achieved promising performance that’s due to shot detection accuracy as shown in Table 5. Table 4 shows comparative results for using logo based detection rather than slow-motion replay detection which gives a subjectively estimated poor performance of 60% for the summarization process.

<table>
<thead>
<tr>
<th>Match Name</th>
<th>Goals</th>
<th>Attacks</th>
<th>Other events</th>
<th>FN/FP goals</th>
<th>FN/FP attacks</th>
<th>FN/FP other events</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCL</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>0/0</td>
<td>0/1</td>
<td>1/0</td>
</tr>
<tr>
<td>FCL</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>UEFA</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0/0</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>EURO</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
</tbody>
</table>

Recall 100% 91.8% 90.8%
Precision 100% 86.7% 95%

<table>
<thead>
<tr>
<th>Match Name</th>
<th>#Logo Shots</th>
<th>#Slow-motion Shots</th>
<th>FN/FP logos</th>
<th>FN/FP Slow-motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCL</td>
<td>24</td>
<td>12</td>
<td>0/1</td>
<td>4/1</td>
</tr>
<tr>
<td>FCL</td>
<td>10</td>
<td>5</td>
<td>0/0</td>
<td>2/1</td>
</tr>
<tr>
<td>UEFA</td>
<td>20</td>
<td>10</td>
<td>0/0</td>
<td>3/2</td>
</tr>
<tr>
<td>EURO</td>
<td>8</td>
<td>4</td>
<td>0/1</td>
<td>3/2</td>
</tr>
</tbody>
</table>

Recall 100% 80.7%
Precision 96.3% 55.8%

6. CONCLUSION AND FUTURE WORK

In this paper, we presented a soccer video summarization engine using enhanced logo detection mechanism to highlight the most exciting events. Our experimental results show the high accuracy achieved in detecting these events. As far as the future work is concerned, we aim at incorporating multimodal cues in the event detection phase, such as audio and speech features of commentator and audience. We also plan to build on the proposed system for summarization to perform soccer video analysis (such as providing coaches with information on a team’s tactic).

7. REFERENCES