CHAPTER 3

PROPOSED FRAMEWORK FOR SPORTS VIDEO SUMMARIZATION

Traditionally sports videos appeal to a large number of audiences. Hence it has become important to automatically extract the useful semantic information from the sports videos to facilitate the users accessing requirements. Recently sports video analysis has emerged as a hot research topic for multimedia researchers due to its high commercial potentials. In the processing of sports videos, detection at the object level has been achieved using play field, ball, overlay text, and jersey numbers. The main objective of this chapter is to propose the video summarization framework for fast and robust extraction of audio-visual and text features for sports video summarization. The proposed framework simultaneously supports both generic and scalable sports video processing these results from classification of the visual features.

Sports videos form very significant multimedia content which is viewed globally by large crowd on the internet and mobile devices. Since sports videos are lengthy, most viewers are interested to watch particular segments of sports video of interest within short time rather than entire video. Watching a live game is exciting, but when it comes to browsing the action that had already taken place, it is a time consuming effort. The major challenging problem is to efficiently manage the sports video contents of important events. This inspired us to compress long sequence of sports video into a more compact representation through a video summarization process.

The aim of this research work is to propose a framework to summarize the sports videos using audio-visual and text features. The visual features have been used to obtain play field color shots and non-play field color shot. In the proposed work the grass and clay colored play field have been considered. Accordingly the play field color shots are further classified into grass color play field shot (GCPFS) and clay color play field shot (CCPFS).

In this Section, an extensive sports video analysis framework that employs audio-visual and text features for sports video summarization has been proposed. The flowchart of the proposed scheme is illustrated in Figure 3.1. As per the flowchart the summarization scheme is characterized by five
Figure 3.1 The flowchart of the proposed framework.
major steps namely shot boundary detection, key/representative frame extraction and decoration, video representation and summarization. Other than these five main steps intermediate blocks proposed are shot classification and shot importance measure, Play field color shot, Non-play field color shot, visual analysis, audio and text analysis, and representative frame extraction. There are various schemes available in the literature for classification but the scheme proposed in this work is different from the presented methods so far used in the state of the art work for video summarization. Most of the viewers are interested to watch particular segments of sports video of their interest to entire video. So we concentrate only on play field shots rather than non-playfield shots. Non-play field shots consist of audience, display panels, close up shots of faces, or advertisements. This non-play field shots collection is also useful to commercial people for example the company people can see how their particular advertisements are being displayed during play break time of the game and may compare with other advertises displayed. This helps company to improve their advertisement and to create or enhance more beautiful presentation than others. This shows significance of the shot depends on the viewer’s interest. However in our approach we have considered interest of sports video viewers and concentrated on analysis of play field color shots. For audio-visual and text analysis we have considered various lawn tennis tournaments, cricket, soccer and golf videos. For representative frame decoration we have used representative frames of news video existing on internet and sports frames which we have extracted during analysis. In this chapter we have explained the architecture of the proposed framework of the video summarization techniques and the scalability that results from the classification of visual information.

The Chapter is organized as follows. Section 3.1 presents various shot boundary detection techniques, and shot transition effects. Section 3.2 illustrates shot to frame conversions for various sports videos with examples, Section 3.3 elaborate key technologies for sports video shot classification to achieve fast retrieval and browsing video. Section 3.4 contributes audio and text analysis, Section 3.5 throws light on multimodal analysis. The representative frame decoration and video summary of sports video is discussed in Section 3.6.
3.1 Shot boundary detection

Shot boundary detection is the first indispensable step in any video processing. It is not efficient (some times not possible) to process a video clip as a whole. It is beneficial to first decompose the video clip into shots and do signal processing at the shot level. Hence shot detection is important. As discussed in Chapter 2 automatic shot boundary detection technique can be classified into pixel based, statistics based, transform based, feature based and histogram based shot boundary detection. The histogram based approach is the most popular amongst all.

Shot-boundary detection algorithms exploit the correlation between the frames belonging to the same shot by computing frame-to-frame differences in some features and seeking for a large difference value that should indicate the start of a new shot. In general, large differences due to object and camera motion are not considered as shot-boundaries and two types of shot-transitions are defined as 1) instant and 2) gradual transitions. Figure 3.2 shows an example of cut. A completely new view is shown instantaneously in the former whereas special editing effects occur between two shots during gradual transitions. The life-spans of gradual transitions are in the order of multiple frames. Figure 3.3 illustrates an example of wipe, dissolve and gradual transition frames with their life spans. Hence, the resulting frame-to-frame differences may not be as acute as they are for instant transitions, which are also known as cuts. Because high accuracy in shot-boundary detection is crucial for efficient video analysis, shot boundary detectors should be able to distinguish the occurrence of gradual transitions from object and camera motions and achieve robust detection rates for instant transitions.

Therefore to achieve high accuracy in shot boundary detection for efficient video analysis, we have used classic software Ulead Video Studio 6 SE DVD for shot detection. Figures 3.3 and Figure 3.4 shows the shots detected for the Wimbledon and French open lawn tennis tournaments using Ulead Video Studio 6 SE DVD.
3.2 Shot to frame conversion

In order to perform the frame level processing, it is essential to convert video into a sequence of frames. The next essential step after detecting the shot boundary is shot to frame conversion. We have used video to JPEG converter which is available freeware to achieve the said task of obtaining the sequence of frames for play field color shot. Figure 3.5 to Figure 3.13 shows the conversion of Wimbledon (shot w1), French open (shot f2), soccer (shot soc30s3), cricket (CrtGibb62), Golf (shot Golf1), audience and advertisement shots to frames. The Wimbledon shot w1 consists of 312 frames, French open shot f2 containing 96 frames, soccer shot soc30s3 encompassing 98, cricket shot CrtGibb62 covering 98, Golf shot Golf1 having 121, Wimbledon...
audience shot is having 63, French open audience comprising 76, whereas advertisement shots consists of 15 and 53 frames.

Figure 3.4 Various Detected Shots of Wimbledon tennis tournament.
After obtaining the sequence of frames we have constructed frame wise color histogram in order to classify video shots into play field color shots and non-play field color shots and a measure of these shots importance based on the respective game rules.

Figure 3.5 Various detected Shots of French Open tennis tournament.
Figure 3.6 Conversion of Wimbledon shot w1 to frames.
Figure 3.7 Conversion of French Open shot f2 to frames.
Figure 3.8 Conversion of soccer shot soc30s3 to frames.
Figure 3.9 Conversion of cricket shot CrtGibb62 to frames.
Figure 3.10 Conversion of Golf shot Golf1 to frames.
Figure 3.11 Conversion of Wimbledon audience shot w13 to frames.
Figure 3.12 Conversion of French open audience shot f14 to frames.
Figure 3.13 Conversion of advertisement shot to frames.
Figure 3.14 Conversion of advertisement shot to frames.
3.3 Shot Classification

Video shot classification is one of the key technologies to achieve fast retrieval and browsing video. So, in this chapter we proposed to construct color histogram to classify video shots into play field color shots and non-play field color shots. After classification we present a measure of shot importance based on respective game rules and clustered only play field color shots as important shots. Playfield color shot based sports video summary is potentially effective for browsing purposes because viewers will not miss any important events although they skip most of the break scenes. It is due to the fact that most highlights are contained within play scenes. To classify video into play field color shots and non-play field color shots we employed global color histogram of every frame in the shot.

Frame wise color histogram gives the number of times a particular color has occurred in the frame. These histograms are defined as $h_0, h_1, h_2 \ldots h_i$ and given by the expression.

$$H(k) = \sum_{i=0}^{N-1} h_i$$ (3.1)

where,

- $H(k)$ is histogram of frames in play field shot,
- $h_i$ is $i^{th}$ frame histogram,
- $N$ is number of frames in the play field shot.

For obtaining the color histogram first play field color shot frames that is RGB images are converted to HSV (Hue Saturation Value). Hue values are thresholded to specific values to obtain the color histogram. In our experiment we considered eight colors that are white, black, red, green, yellow, magenta, blue, and cyan colors. So we will get eight-channel histogram for any play field frame. We are considering for analysis lawn tennis tournaments, cricket, soccer, and golf with their respective dominant field colors. Hence we thresholded the hue values for these colors. Hue value thresholded for red color playfield is greater than 0.9167 and less than or equal to 0.083, for green color playfield it will be greater than 0.083 and less than or equal to 0.25, and for blue color it is greater than 0.5833 and less than or equal to 0.75. With these thresholded values computed histograms of all frame sequences in the shots. Based on obtained histogram and color component values shots are classified into
play field color shot and non-play field color shot. Based on the respective sports rules and viewers interest in the events, the play field shot importance is decided. Further play field color shot are classified into two key classes grass color play field color shot and clay color play field color shot as shown in Figure 3.38.

Figure 3.15, Figure 3.22, Figure 3.29, Figure 3.36 and Figure 3.37 illustrates sample results of play field color shot frames and their color histograms for Wimbledon, French Open, Soccer, Golf and Cricket respectively. These histograms show the frame wise dominant color component. Frame wise color component values of the shot are used to define mean value of each color component of PFCS. Playfield color shot based sports video summary is potentially effective for browsing the sports video because viewers are interested to look at the events only on the and not on the non-playfield shots. Furthermore PFCS are classified into two main classes 1. Grass color play field shots and 2. Clay color plays field shots which is helpful to distinguish between these two main playfield. Figure 3.16, Figure 3.23, Figure 3.30 shows sample results of the dominant color of the play field shots of Wimbledon, French Open and Soccer shots. Whereas Figure 3.17 to Figure 3.19, Figure 3.24 to Figure 3.26 and Figure 3.31 to 3.33 demonstrates the red, green and blue color component distribution in the Wimbledon, French Open and Soccer. Figure 3.20, Figure 3.21, Figure 3.27, Figure 3.28, Figure 3.34 and Figure 3.35 shows the variation of mean and standard deviation values in the shots w1, f2 and soc30s respectively. In case of Wimbledon tennis variation of mean and standard deviation is approximately same. Whereas in French open lawn tennis tournament when mean value is increasing there is decrease in standard deviation value and vice versa. We can tell that in case of green play field color shots standard deviation value is varying approximately same as mean value and in case of brown colored play field shots mean and standard deviation values are inversely proportionate to each other. Dominant colored frame from each play field color shot have been extracted as one of the key frame for creation of video summary, which is presented in Figure 3.18, Figure 3.24 and Figure 3.32. Figure 3.38 illustrates the shot classification of sports video into Grass color play field shots and Clay color play field shots.
Figure 3.15 (a) to (f) Sample results play field color shot frames and their color histograms for Wimbledon shot w1.
Figure 3.16 Mean histogram of the sequential frames of shot w1.

Figure 3.17 Distribution of red color component in shot w1.
Figure 3.18 Distribution of green color component and extraction of dominant color frame in shot w1.

Figure 3.19 Distribution of blue color component in shot w1.
Figure 3.20 Plot of frame number versus mean values for w1.

Figure 3.21 Plot of frame number versus STD values for w1.
Figure 3.22 (a) to (f) Sample results play field color shot frames and their color histograms for French open shot f2.
Figure 3.23 Graph of dominant color detection of the shot f2.

Figure 3.24 Distribution of red color component and extraction of dominant color frame in shot f2.
Figure 3.25 Distribution of green color component in shot f2.

Figure 3.26 Distribution of blue color component in shot f2.
Figure 3.27 Plot of frame number versus mean values for w1.

Figure 3.28 Plot of frame number versus STD values for f2.
Figure 3.29 (a) to (f) sample results play field color shot frames and their color histograms for soccer shot Soc30s31.
Figure 3.30 Mean histogram of the sequential frames of shot Soc30s31.

Figure 3.31 Distribution of red color component in shot Soc30s31.
Figure 3.32 Distribution of green color component and extraction of dominant color frame in shot Soc30s31.

Figure 3.33 Distribution of blue color component in shot Soc30s31.
Figure 3.34 Plot of frame number versus mean values for Soc30s31.

Figure 3.35 Plot of frame number versus STD values for Soc30s31.
Figure 3.36 (a) to (d) sample results play field color shot frames and their color histograms for Golf shot Golf1.
Figure 3.37 (a) to (f) sample results of play field color shot frames and their color histograms for cricket shot CrtGibb62.
Figure 3.38 Classification of play field color shot into two key classes: grass color play field shot and clay color play field shot.
3.4 Audio and text analysis
The sound effects in the video shot plays very significant role in video summarization. The dominant value of these audio signals such as volume and energy has been used to select the key frames which is a empirically selected measure. The location of the text box corresponding to the score and player names varies with the tournaments. However last fame of play field shot of lawn tennis tournament gives the updated score of the corresponding shot. Use of these shots has been elaborated in Chapter 5.

3.5 Multi-modal analysis
We have integrated audio-visual and on-screen text features for more robust and extensive sports video analysis. Because all multi-modal features in the proposed framework are compute-easy, so they are used for summarization. We use visual features to classify video shots to play field color shot and non-play field color shot. For every play field color shot three key frames are extracted to create video summary. The first key frame is chosen based on dominant value of audio energy and volume, second selected as dominant colored frame, and third as the last frame of play field color shot as key caption frame for the shot.

3.6 Representative frame decoration and video summary
Representative frames are access points to video contents. We proposed a novel generic approach of representative frame decoration which decorates the frames for representation. For every play field shot extracted frames are decorated and these are used to create video summary. The video summary is a collection of a set of static representative frames. The representative frame is a static representation of the video contents and can be used to express an outline of the video. The sample results are as shown in Figure 3.38.
(a) Original video representative frames

(d) Unsharp mask filtered video representative frames.

Figure 3.39 Results of news video representative frame decoration.

***