CHAPTER 1
INTRODUCTION

The videos form very significant multimedia content which is viewed globally by large crowd on the internet and mobile devices. Human beings prefer to watch videos of news, movies, sports, dramas, medical, teaching, meetings and so on. However because of time constraints it is not possible to watch longer video documents. Consequently needs to generate concise video summary that provide important and necessary information within short span of time at anytime and anywhere. Since the videos are lengthy, most viewers are interested to watch particular segments of the video of interest rather than the entire video. Video abstraction is a technique that abstracts video content and represents it in a compact manner. There are basically two types of video abstraction: Video summarization and video skimming. Video summarization is a process that selects a set of salient images called key frames to represent the video content. Video skimming represents the original video in the form of a short video clip. Video abstraction forms a key ingredient in a practical video content management system, as generated key frames and skims provide users an efficient way to browse or search video content. With the increase in demand of digital video, this process has became an indispensible component to any practical content management system. A video summary can be displayed without the worry of timing issues. Moreover, extracted key frames could be used for content indexing and retrieval. A summary sequence provides a small taste of entire video, while a highlight contains only the content that may appear interesting to viewers such as the movie trailer. Watching a live video is interesting, especially live sports video is exciting, but when it comes to browsing the action that had already taken place, it is a time consuming effort. The most challenging problem is to efficiently manage the sports video contents of important events. The phenomenon inspired us to compress long sequence into a more compact representation through a summarization process. Recently, this area has gained much attention of the researchers and hence considered as the potential problem in this research work.

The rest of the chapter is organized as follows. Section 1.1 highlights the need for video summarization. Section 1.2 illustrates the motivation for
selecting the said area for carrying out this research work. Section 1.3 throws light on the problem statement considered in this thesis. Section 1.4 focuses on the proposed methodology for achieving the desired task of video summarization. Section 1.5 elaborates the main objectives in this thesis. Section 1.6 presents a brief summary of our contributions in the said area and finally organization of the thesis is addressed in Section 1.7.

1.1 Need for video summarization
In today’s society the growing demand of digital video information initiates the basic need for video summarization. The rapid developments in various affordable technologies for multimedia content capturing, data storage, high bandwidth, high speed of transmission and the multimedia compression standards have resulted in a rapid increase of the size of digital multimedia data collections. This caused the greatly increase in the availability of multimedia contents to the common user. Advanced multimedia technology is generating a huge amount of image and video collections; its understanding has been attracting increased attention in recent years. The multimedia data communicated must be information-oriented rather than the actual content itself (web-browsing, video-on-demand). With the development of high-speed internet, high-capacity storage and compression standards people are quickly collecting a growing amount of available video information. Video images are voluminous, redundant, and their overall contents cannot be captured at a glance. Compact yet meaningful abstraction of multimedia data can serve as an effective index table, allowing users to browse through large amounts of data in a non-linear fashion. The drastic advances in multimedia and internet applications need more effective techniques for video retrieval and summarization. Therefore, it becomes necessary to develop techniques for automatic detection of semantically important events in the video. The summarization of video helps in indexing, browsing and retrieving the video of important events.

Moreover the demand for video summarization originates from the viewing time constraints as well as communication and storage limitations. The ever-growing availability of multimedia data creates a strong requirement for efficient tools to manipulate and present the video data in an effective manner. Recently, techniques for automatic video content summarization have been in increasing demand because of its commercial potential,
especially for home video applications. Hence one of the key technologies required for efficient access and management of video library is video summarization. In order to achieve efficient and effective video summarization various techniques are required which will effectively extract important information from the video data while removing the redundancy. Due to proliferation of digital video and drastic advances in the adoption of video production devices, such as camcorders and digital cameras, along with the advances in video compression and the increasing usage of the internet and wireless communication such as 3G and 4G standards, have enabled to create the need for more effective techniques for video summarization, browsing and retrieval. Consequently summarization is an essential requirement for achieving a more compact and interesting representations of video contents.

1.2 Motivation
Increase in production of digital videos and rapid developments in the adoption of video production devices like digital cameras and handy cams, advances in the video compression, the increasing usage of the internet, wireless communication, shifting of reliable networking towards mobile and wireless access facilities have enabled to create the need for more effective techniques for video summarization, browsing and retrieval. In recent times, we can find video content in digital libraries (e.g. internet), personal collections, social networks (e.g. YouTube), and web sites of media networks (e.g. BBC, CNN, and ABC), video on-demand services, optical storage discs (e.g. DVD, Blue ray discs), digital television broadcasting and an endless list of sources. Consequently, users are overwhelmed with a huge and increasing amount of video information, which often becomes very difficult to manage video contents for search and retrieval of specific content. The video is a time consuming media as it requires time to be visualized. The most unique characteristic of a video document is its ability to convey a rich semantic presentation through the synchronized audio, visual and text presentations over a period of time. So, in this thesis we focus on audio, visual and text feature based sports video summarization/browsing. Sports video has been selected as the primary application in this thesis because they attract viewers worldwide, thereby maximizing the potential benefits of using a video summarization system. As compared to other video types, sports video poses
some unique challenges, such as: 1) Each sports genre has context-dependent characteristics such as special game structure and camera views, 2) Sports video is recorded without controls over the script and setting/environment, therefore the temporal structures are difficult to predict and background noises cannot be avoided, 3) Sports video is broadcasted with different styles of editing effects (e.g. slow motion replay and text displays) depending on the broadcasters, 4) Sports video can be used for various purposes such as entertainment, performance analysis and will be helpful to professional players and coaches to retrieve video segments in a meaningful manner.

The existence of aforementioned multimedia facilities, worldwide interest of viewers in sports video and its challenges motivated us to work on the video summarization. Hence, this thesis presents a research work based on an integrated multi-modal approach for sports video summarization. With an increase in feature-based analysis and extraction of key frames, these systems are becoming usable and efficient in retrieving perceptual content. By combining specific features extractable from multiple (audio-visual and text) modalities, generic structure and specific events can be detected and classified.

1.3 Problem statement

This research work addresses video summarization as main objective. Nowadays digital video plays an important role in every day’s life and due to widely used low cost storage media, the volume of digital video tends to be very large and variety of available video data makes the search and retrieval of content a more and more difficult task. The amount of information generated in today’s society is growing exponentially. The search and visualization efforts point towards a waste of time as many useless video must be downloaded and visualized prior to get the desired video. Video images are voluminous, redundant and their overall contents cannot be captured at a glance. It is essential to help user to provide more compact, interesting video content with narrow bandwidth. There has been a great need for techniques which support non linear, meaningful, fast browsing and further content based video retrieval that is video summarization techniques are of great importance.
In this thesis sports video summarization from huge sports video data has been considered in general and an integrated multi-modal approach for video summarization in particular.

1.4 Proposed methodology
The audio-visual and text feature based techniques are proposed to create video summary. With the available advanced software techniques can be used to obtain the shot boundaries and the frame sequence in the particular shot. For visual analysis we intended to use color attribute to classify video shots and its significance among the available features such as color, edge, texture, shape, motion and spatial information. For audio analysis of the videos Sound effects (announcer’s excited speech, cheer, applause) magnitude per shot has been planned to utilize. On screen text from the score-boxes in play field shots aimed to use for textual analysis. The representative frame decoration scheme using filters like unsharp masking filter is proposed to increase contrast and sharpness of key frames.

1.5 Objectives
In order to achieve the main objective of integrated multi-media approach for sports video summarization we have subdivided it into various sub-objectives as follows:

1. To decompose sports video into shots using standard software like Ulead Video Studio 6 SE DVD, pinnacle etc.
2. Convert each video shot into frames using very recent techniques like video to JPEG converter.
3. To classify sports video into Play Field Color Shot (PFCS) and Non Play Field Color Shot (NPFCS).
4. To extract key frames based on visual features for video summarization.
5. To extract key frames based on audio features for video summarization.
6. To extract key frames based on audio-visual and text features for video summarization.
7. To decorate representative frames using filter like Unsharp mask filter.
1.6 Summary of Contributions

This thesis focuses on discovering the techniques for shot classification, feature extraction, discrimination and representative frame decoration applied for sports video summarization. Feature extraction and semantic annotation are the main requirements for constructing a compact representation. This creation of compact representation is the most important aspect to show the success of the video summarization method in terms of satisfying users/applications’ requirements.

The main contributions of this research work for novel video summarization methods for sports video application based on multimodal features have been summarized as follows:

- A novel video summarization method for sports video applications using audio-visual and text feature extraction and analysis. (This proposed work has been published in International Journal of Computer Applications (0975-8887), DOI 10.5120/5803-8126, Vol. 42-No.19, pp 28-35, March 2012. Impact factor of the journal is 0.835).

- Considered a widespread survey of video summarization techniques. (The literature survey has been published in International Journal of Computational Intelligence and Telecommunication Systems (IJCITS) ISSN: 2229-3078, Vol. 2, no.1, pp 47-55, Jan-June 2011).

- We have proposed a new shot importance measure scheme based on lawn tennis tournament rules and dominant color of the play field colored shot. The play field color shots are segments of interest and have been used to recognize the tennis tournament class. The effectiveness of the proposed scheme is demonstrated over various lawn tennis tournaments like Wimbledon, French Open and US Open video by objective performance metrics. It is potentially effective for browsing purposes because viewers will not miss any important events. (This work have been presented and published in proceedings of IEEE International Conference on Computational Intelligence and Computing Research (2011 IEEE ICCIC), held at Kanyakumari, 2011).

- Furthermore play field colored shots have been classified into grass (green) colored play field shot (GCPFS) and clay colored play field shot (CCPFS). (The work has been submitted to an International Journal and under the process of review).
• We have proposed a dominant audio energy based key frame extraction for sports video summarization. (The work has been presented and published in proceedings of International Conference on Advancements in Electronics and Power Engineering (ICAEPE’ 2011), held at Bangkok, Thailand, 2011).

• We have also proposed to analyze only play field color shots which serve as accurate localization of events on the field. From every play field color shot extracted key frames using dominant colored frame in the shot and dominant audio volume and energy related frame of the shot for video summarization. This scheme is applicable to various sports video like Cricket, Soccer and Golf. (The work has been submitted to an International Journal and under the process of review).

• We have introduced a novel concept of Representative Frame Decoration (RFD), in which the essential operation is to decorate the representative frame of a video shot. Our approach employs an unsharp mask filter to enhance contrast and sharpen edges of the elements without increasing noise or blotch. Representative frames are vital and attractive key components in providing Internet users a way to swiftly browse a video clip in different levels of detail, without the need to view entire video clip. This scheme can be applicable to various sports video and news video. (The work has been presented and published in proceedings of IEEE International Conference on Communications and Signal Processing (ICCSP), held at Kerela, India, 2011).

1.7 Organization of the Thesis
The remaining chapters of this thesis are outlined as follows:
In Chapter 2, we have studied the general architecture of video summarization scheme. This scheme consists of input video stream, shot boundary detection, key frame extraction, video representation and video summarization. Input video stream may be scripted video (news videos, dramas & movies) or unscripted video (sports video, teaching video and surveillance video). Shot boundary detection is the first essential step in video processing. The key frames can be extracted using audio-visual and textual information conveyed in a video stream. Audio features (whistle,
excitement, applause, cheering, music, speech, speech with music, noise),
text information (from closed captions, text detection) and object features
that is spatial or spatio-temporal features (referee, players, object that used
in games, color, object motion, intensity) can be used to extract key frames.
Extracted collection of prominent images from a video sequence is used for
content summarization. Also an extensive discussion on the current
achievements concerning the components of video indexing is provided. The
next discussion focuses on comparing visual processing, audio processing,
textual processing, and other features. This discussion provides a motivation
for one of the most promising approaches.
Chapter 3 provides an overview of the framework for a sports video
summarization system. The videos considered for analysis are lawn tennis
tournaments (Wimbledon, French open), soccer, cricket, and golf. Video shot
classification is one of the key technologies to achieve fast browsing and
retrieval. Based on visual analysis these are classified into play field color
shots and non-play field color shot. Moreover play field color shots are
classified into grass color play field shot and clay color play field shots.
Dominant colored frame of every play field color shot have been extracted to
create video summary. The discussion aims to guide readers into the
subsequent chapters in the thesis.
Chapter 4 describes the developed scheme to solve small beautiful problem
of discrimination of shots among various lawn tennis tournaments. This is
achieved using color histogram of frames in the play field color shots.
Dominant colored frame of every play field color shot have been extracted to
create video summary. For evaluation of the performance of the play field
color shots of lawn tennis tournaments we have used parameters recall rate
and precision rate. The tournament videos used for analysis are French
open, Wimbledon and US open.
Chapter 5 expresses a generic scheme for extraction of dominant audio
volume and energy values and the corresponding visual frame as
predominant frame in the shot. This predominant frame has been extracted
as key frame to create video summary. The last frame of every play field color
shot of lawn tennis tournament contains text describing the scores and team
or player names. This frame also extracted as salient frame for the particular
play field shot and used to create video summary.
Chapter 6 introduces a block of Representative Frame Decoration (RFD) scheme before video representation in video summarization. This proposed scheme uses a classic and mature linear unsharp masking filter for enhancement of contrast and sharpness. Moreover video summarization steps have been elaborated in this Chapter. Finally conclusions, applications and future work on proposed approach have been articulated in Chapter 7.

***