Semantic Segmentation and Event Detection in Sports Video using Rule Based Approach

Dr. Sunitha Abburu

Professor& Director, Department of Computer Applications
Adhiyamaan College of Engineering,
Hosur, pin-635109, Tamilnadu, India.

Summary
The paper addresses two main problems of sports video processing: semantic segmentation and event detection. The theme is domain specific approach which exploits the typical characteristics of cricket video to design the most effective approach for the semantic segmentation and event detection which supports, efficient and effective retrieval of video scenes. Cricket video has been selected as the primary application, because they attract viewer worldwide and the complexity of the game is high. This paper proposes a novel hybrid multilayered approach for semantic segmentation of cricket video and major cricket events detection. The approach uses low level features and high level semantics with the rule based approach. The top layer uses the DLER tool to extract and recognize the super imposed text and the bottom layer applies the game rules to detect the boundaries of the video segments and major cricket events. The proposed model has been implemented, tested and the results are promising. Future work has been discussed at the end.

Key words: *Semantic segmentation, Event detection, Rule based approach, Text recognition*

1. Introduction

The volume of digital video has grown tremendously in recent years, and has become a major information storage and exchange media. Multimedia plays a very important role in computing and communication environments, with various applications in entertainment, advertising, distance learning, tourism, distributed CAD/CAM, GIS, sports etc. The extensive embracing of videos is due to its ability to convey rich semantic through synchronized audio, visual and textual information. This trend has resulted in the emergence of numerous multimedia repositories that require efficient storage. The stored multimedia data poses a number of challenges in the management of multimedia information, including data and knowledge representation, video segmentation, indexing, retrieval, intelligent searching techniques, information browsing and query processing.

To sustain an ongoing rapid growth of video information, there is an emerging demand for a sophisticated content-based video indexing system. However, current video indexing solutions are still immature and lack of any standard [1]. Viewers are becoming more demanding in terms of quality, quantity, the ability to present the information and efficient ways to intelligently analyze lengthy multimedia objects. As user requirements and their information demands are changing continuously. Content-based video retrieval system using color-texture-shape based image analysis techniques are generally inefficient when users need to search for particular sports events. Sports video has been chosen as a main source. There are many reasons why analyzing sports video is important and necessary. Sports videos are widely spread over various multimedia repositories and appeal to large global audiences, hence sports video analysis and composition techniques are highly sought-after. Only small portions of the video will contain or convey the semantics or interesting events will be spread over a few frames and not the entire video. The ability to automatically identify and manage interesting segments from video documents is important and necessary [2]. Hence the ability for semantic segmentation, automatic detection of events, efficient indexing and video object retrieval techniques are highly desirable. Semantic segmentation and automatic event detection is possible in sports video due to well defined broadcasted sports video structure and domain rules of the game. The ability to automatically identify important semantics from lengthy sports videos is a key requirement in many important applications such as sports video analysis, indexing [3] [4] [5] sports highlight generation [6] [7] etc. Extracting semantics from sports video is very challenging as it is still unclear how human perceive concepts. In contrast to existing work, this paper aims to develop more generic event detection technique and semantic segmentation for the broadcast sports video using multi layer analysis model with rule based approach.

The paper is organized as follows: Section 2 presents related research work. Section 3 describes semantic

Manuscript received October 5, 2010
Manuscript revised October 20, 2010

segmentation and cricket event detection. Section 4 contains a practical implementation. Finally conclusions and future works are drawn in section 5.

2. Related Research work

Domain knowledge is required to capture the semantics of the video. A simple approach to manage the semantic content of multimedia objects is to annotate an image/video with text. Annotations are free text descriptions. The multimedia objects may be fully annotated or partially annotated. The performance of the retrieval system depends upon the amount of information available in the database about the multimedia objects. Automatic annotation may not attain extremely high accuracy with the present technology.

2.1 Video Segmentation

The effectiveness and efficiency of the systems depends upon the indexing techniques. Indexing performed on the whole video stream would be too coarse, at the same time indexing based on each frame often does not contain any important information. Video objects indexing can be performed on a group of sequential frames called as video objects. Video segmentation has been an important and challenging issue for many video applications. Generally there are two different video segmentations approaches, (i) Shot-based video segmentation and (ii) Object based segmentation.

Shot Based Video Segmentation: In typical broadcast video, a shot is a sequence of frames captured by a single camera in a single continuous action. It is the most basic unit of video data. Broadcast sports video often intertwines different shot types. Shot is a sequence of video frame which have a similar characteristics. Shot boundaries (or cuts) can be detected by looking for abrupt changes in the video image from one frame to the next, a process called shot detection. A shot is only a group of sequential frames with similar characteristics. Therefore it does not actually correspond directly to semantic content. However, this handles only one tiny aspect of the semantics of the video. A shot is seldom autonomous its meaning is affected by the other shots. Shot detection tells us nothing more than the length of the shot. Shot is not sufficient for a proper understanding of its semantics. To overcome this problem, scenes need to be extracted. A scene is composed of nearby shots that share common special, temporal, perceptual or semantic content [8]. Scene detection is generally more difficult because to group the shots into scenes, the content and semantics of the video must be understood.

Object based segmentation: An object refers to a semantic real world entity. In object based segmentation shots are replaced by temporal object segments, which may refer to the life span of objects. Object based video segmentation is to decompose one video shot into objects and background [9]. [10] presents a system for audio segmentation and classification, which can segment and classify the sports audio stream into speech, non-speech audio stream.

2.2 Event Detection

An approach to summarize a cricket game video using highlights extraction process is presented in [11]. The video is first segmented into shots. Key frames are extracted from each shot and low level features are extracted from the key frames. Features are used to extract views or states. The states and their transitions in the cricket game are represented by a Hidden Markov Model, based on which the game highlights are extracted. To create a video highlight [12] selects interesting shots by modeling excitements of the game. For this purpose, a video is first segmented into shots and classified as play and non play shots. Score of the game is automatically extracted from video frames. To select interesting shots, which should be included to the highlight video, excitement of the game is estimated from the variation of game scores. Kang et al in [13] proposes a new two-level framework to analyze high-level structure of video and to detect useful events automatically based on visual keywords. Kolekar M.H. and Sengupta S [14] detected events and classifies the events using audio and Hidden Markov Model. Bertini et al [15] presents a system for automatically detecting and analyzing complex player actions in moving background sports video sequences, aiming at action-based sports videos indexing and providing kinematic measurements for coach assistance and performance improvement. Mao [16] integrates text feature (scoreboard caption) and video feature (shot transition) in an effective and efficient way. The scoreboard caption which represents game status is extracted with a novel detection and recognition scheme, which includes two simple processes of scoreboard location and digit recognition

3. Semantic Segmentation and Cricket Event Detection

The primary aim of this paper is to propose novel techniques for:

- Semantic segmentation of sports video into video objects

- Detection of highlight event like score and dismissal using rule based approach.

Automatic semantic segmentation and major events like wicket, score detection, these issues can be addressed by a hybrid multilayered approach. Top layer analyze the video using the low level features to extract the high level semantics conveyed as super imposed text on video. The extracted superimposed text semantics facilitates to identify the key frames, where major events took place. Bottom layer explores the semantics with the rule based approach to classify the video frames in to video objects and to detect the events like score and wicket see fig 1.

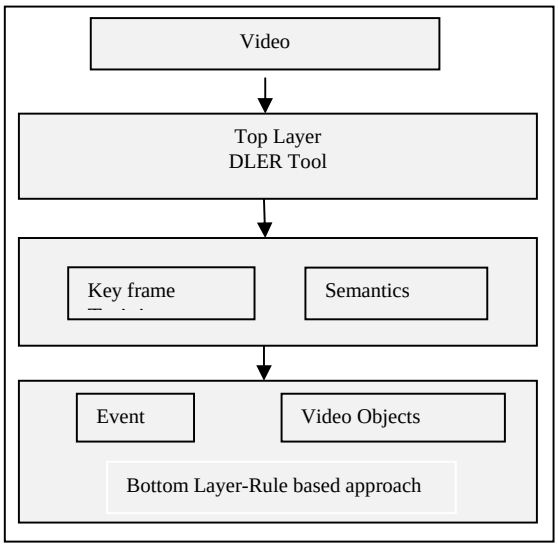


Fig. 1 Multilayered Approach

Efficiency of the retrieval system depends up on the semantic video indexing techniques. Semantic video indexing requires effective video semantic extraction and storage. Semantic annotation helps the retrieval system to support semantic search of the video objects. Semantic information extracted from a video object not only conveys information about that video object, but at the same time, it becomes semantic carrier for the next video object. In the real world, all video objects are logically related. To represent the complete semantics of the video, a logical connection must be considered, which is maintained by rule based approach. Some information content must be carried for the subsequent video objects. Hence, complete semantics of a video/video object can be extracted into two different levels. The first level of information that is content can be extracted directly from a video object either using super imposed text or using audio. The second level of information can be gathered from the previous video objects, which act as semantic carriers for the current video object, see Fig.2. To identify and represent complete semantics of a video, semantic information extracted from various video objects must be combined and maintained.

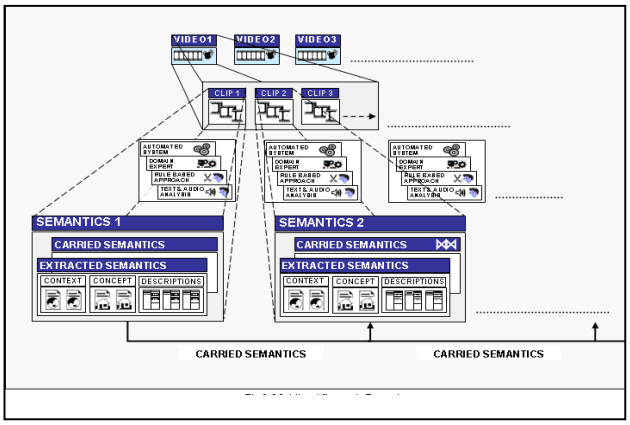


Fig. 2 Multilayer Semantic Extraction

For example, for every over, bowler details, over details etc must be maintained up to the end of that over and complete match. Batsman details must be carried for all video objects up to the dismissal of the batsman. Cumulative details like the match score, player score, number of wickets etc., must be maintained up to the end of the match. Context information like match statistics, player details his achievements till date, to maintain such information, super imposed text is used.

A semi-automatic method to generate annotation for cricket videos and an automated tool- *DLER*, to extract the semantics of cricket video are proposed in [17]. The *DLER* tool provides a fast and robust approach for text Detection, Localization, Extraction, and Reorganization in video frames, which is flexible and customer friendly. The *DLER* integrates all the pre-processing steps and the OCR steps in to a single unit. The annotator can pick the ROI, increase or decrease the threshold, contrast, brightness or inverse the image based on the type of the broadcasted video. The broad casted cricket video follows a specific pattern to display the game information like the ball/over number, score/wickets etc. The super imposed text which is displayed on broadcasted cricket video usually on a plain background and at a specific region. This feature mostly reduces the complexity of text detection and recognition. Text displayed on the screen in cricket sports video attains positional notation, score, events, wickets, balls etc, information appear in a specific location in a specific format. Generally the score and other information are displayed on the bottom ¼ or top ¼ of the frame region [18]. The TV program editor would impose the text in the broadcast cricket video based on the importance of the event or special records, editor may super impose the text with different font in different style in different

locations. Texts displayed at different regions have different semantics. Other important text information can appear in other regions of the video frame, which conveys important semantics. Text displayed on the screens of broad casted cricket video attains positional notation. Text displayed at a specific location attains a specific notation like see fig 3, 125-8, represents the current score as 125 and the number of wickets taken so far are 8.Similarly specific notation like 4/29, represents 4th ball in 29th over. Change in the ball number indicates the beginning of a new ball scene. The proposed approach identifies the frame in which the ball number change has been detected and indicates the boundaries of the new video object or a new video segment. Similarly, change in the score or wicket numbers detects major events. The top layer of the proposed model uses the DLER tool to extract the semantics of the game. The bottom layer analyzes the extracted semantics and applies the rule based approach to detect the segment boundaries or the major events.

Fig. 3 Score and Wicket fields

4. A practical implementation

4.1 Semantic Segmentation

Hierarchal semantic segmentation of sports video using rule base approach with text extraction techniques is adopted. Every domain has its own specific structure. Based on this specific structure of domain, video can be segmented in to various video objects and generated the identity for those video objects. Semantic segmentation of sports video is proposed in the paper. The structure of cricket game is used to segment the cricket video into various video objects see fig 4. The basic logical unit in cricket game is one ball. In a broadcast sports video text is displayed on the screen, which gives ball, over, score, wickets, players details. The segmentation algorithm uses superimposed text, extracted directly from the video frame. And the video object contains various scenes of one complete meaningful ball video seen. The segmentation process partitions the cricket video into the video object by applying rule based approach. A video object is a sequence of meaningful video frames that have their own concepts and concept attributes describing the semantic content. In cricket game, video object is one complete set of scenes associated to different activities of a ball. (e.g., bowling, batting, fielding, audience reaction). Change in the ball number in video frames indicates the beginning of a new ball scene. The change in the video frames in the ball number is used for semantic segmentation of the cricket video.

The algorithm work as follows, given *n* detected shots, the proposed key-frame extraction can be achieved in the following steps. Initially the first frame f1 of each shot is always chosen as the first key-frame and stored in the key-frame set. Semantic similarity of the recognized text between the next coming frame and the key-frames in the key-frame set is computed. Change in text detects coming frame as a key-frame. Otherwise, just input the next following frame. A full key frame set can be estimated after looping through all the remaining frames inside the same shot.

Algorithm: Detection of key frames:

For ∀ frames fi
{
If (detected: increment in the value of text at ball number region)
key-frame set = {f1 fj} / one video object.*
}

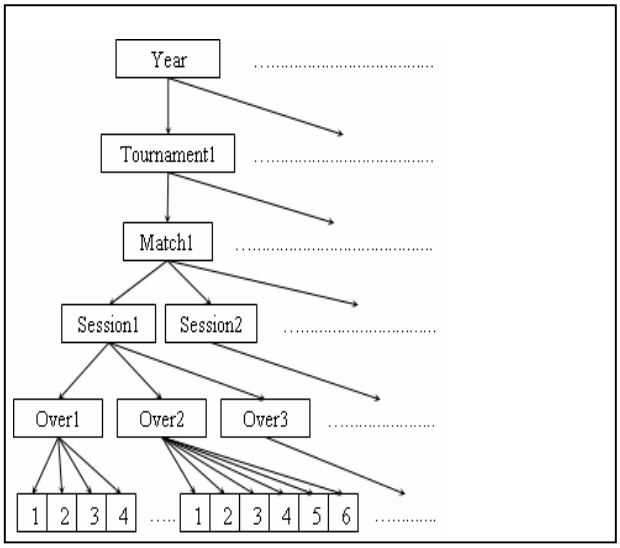


Fig. 4 Specific structure for Cricket Domain

For every cricket video object, the identity is generated as follows:

Video Object Identity = {Y, C, T1T2, M, S, O, B}
Y = year of the match
C = tournament identity
T1T2 = match identity
M = match type
S = session identity
O = over identity, which ranges from
{1, 2…49, 50}

B = ball identity, in a particular over
{1, 2 ...5, 6}
By analyzing the video object identity information like tournament, match, year etc., can be extracted .

4.2 Detection of Major Cricket Events

The rule-based approach uses domain knowledge to define rules to achieve semantic video segmentation and event detection. The bottom layer adopts rule based approach to identify the events like score, dismissal and context details. The semantics extracted by the top layer are analyzed by the bottom layer to detect the events and other semantics of the video. Text displayed at specific location has specific meaning especially in case of cricket video. Based on the position of the text in score board the semantics of text can be easily detected. E.g. India 125-8 the first element represents score and second represent dismissal. Change in value of score detects score event. Change in wicket field detects the dismissal event.
The generic rules can be formed as follows:

Algorithm: Detection of dismissal event:

If (detected: increment in the value of text at wicket region)

{

{it is a dismissal}

{no. of dismissals is reduces by 1 for the batting team}

{credit goes to bowler and bowling team}

{Striker / non-striker is replaced with new batsman}

}

Algorithm: Detection of score event:

If (detected: increment in the value of text at score region)

{

ballscore = new team score - old team score

replace the old team score with new team score

if (extras >= 1) {

ballscore = ballscore – extras

}

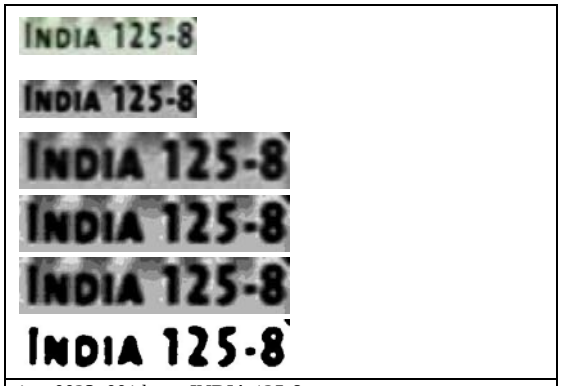
ballscore is added to the striker

If ((ballscore % 2) != 0) {

Striker is replaced with non striker

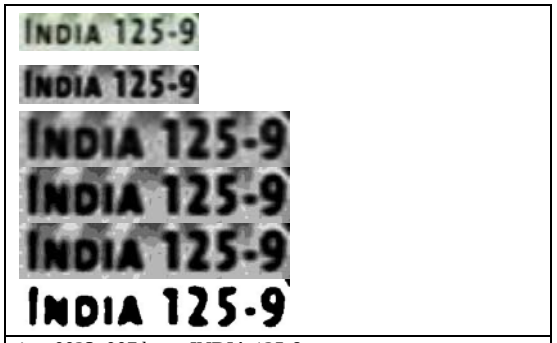
}

A snap shot of experimental results to detect the wicket event is shown in fig 5&6.



icc_0023_001.bmp: INDIA 125•8
icc_0023_002.bmp: INDIA 125.8'
icc_0023_004.bmp: INDIA 125•8

Fig. 5 Event Detection using rule based approach



icc_0023_007.bmp: INDIA 125•9
icc_0023_008.bmp: INDIA 125•9
icc_0023_009.bmp: INDiA 125.9

Fig. 6 Event Detection using rule based approach

5. Conclusion and future work

Semantic segmentation and automatic event detection in sports video are challenging tasks. Lot of research is going on to find novel and efficient techniques for semantic segmentation and sports event detection. This paper proposes a novel hybrid multilayered approach for semantic segmentation of cricket video and automatic detection of major cricket events like score and dismissal using the rule based approach. The whole process of video segmentation and event detection has been done in two layers. Top layer uses the super imposed text as a primary source for the semantic extraction, DLER tool has been used to extract and recognize the super imposed text and the extracted semantics are been passes on to the bottom layer which are combined with the rule based approach to detect the boundaries of the video segments and the major events of the cricket game. The proposed approach has been implemented, tested and the results are promising. Further research could be conducted on automatic detection of cricket actions like batting, bowling, fielding styles using multimodal approach. Audio can be another

source of the video semantics and more accurate semantics can be extracted by multimodal processing which will help to detect the actions.

Acknowledgments

This work has been partly done in the labs of Adhiyamaan College of Engineering where the author is currently working as a Professor& Director in the department of Master of Computer applications. The author would like to express her sincere thanks to Adhiyamaan College of Engineering for their support rendered during the implementation of this module.

References

[1] D. Tjondronegoro, Y. Chen, and B. Pham, \Content-based video indexing for sports applications using multi-modal approach," *Proc. of ACM MultiMedia'05, Doctoral Symposium*, pp. 1035-1036, 2005.

[2] S. F.Chang,The holy grail of content-based media analysis," *IEEE Trans. On MultiMedia*, vol. 9, pp. 6-10, 2002.

[3] Haojie Li; Jinhui Tang; Si Wu; Yongdong Zhang; Shouxun Lin," Automatic Detection and Analysis of Player Action in Moving Background Sports Video Sequences" , IEEE Transactions on Circuits and Systems for Video Technology, Volume: 20 , Issue: 3 , 2010 , Page(s): 351 – 364.

[4] Yang Li; Yafei Zhang; Jianjiang Lu; Ran Lim; Jiabao Wang, " Video Analysis and Trajectory Based Video Annotation System", 2010 Asia-Pacific Conference on Wearable Computing Systems (APWCS), Publication Year: 2010 , Page(s): 307 – 310.

[5] Raman, V.; Sumari, P.; Abdullah, R., Pilot Prototype Analysis: Ongoing Framework for Semantic Extraction Retrieval in Tennis Sports Video, Second International Conference on Computer and Electrical Engineering, 2009. ICCEE '09. Volume: 2 Publication Year: 2009 , Page(s): 619 – 623.

[6] Gwang-Gook Lee; Hyeong-ki Kim; Whoi-Yul Kim,"Highlight generation for basketball video using probabilistic excitement", IEEE International Conference on Multimedia and Expo, 2009. ICME 2009,Page(s): 318 – 321.

[7] Bertini, M.; Del Bimbo, A.; Torniai, C.; Grana, C.; Vezzani, R.; Cucchiara, R.," Sports Video Annotation Using Enhanced HSV Histograms in Multimedia Ontologies " , 14th International Conference on Image Analysis and Processing Workshops, 2007. ICIAPW 2007, Page(s): 160 - 170.

[8] Y. Fu, A. Ekin, A.M. Tekalp, and R. Mehrotra, "Temporal segmentation of video objects for hierarchical object-based motion description", IEEE Trans. On Image processing, Vol. 11, no. 2, pp. 135-145, Feb. 2002.

[9] Lijie Liu, Guoliang Fan, "Combined Key-Frame Extraction and Object-Based Video Segmentation ", IEEE Transactions on Circuits and Systems for Video Technology, vol.15, no.7, pp 869-884 July 2005.

[10] Jun Huang; Yuan Dong; Jiqing Liu; Chengyu Dong; Haila Wang," Sports audio segmentation and classification ", IEEE International Conference on Network Infrastructure and Digital Content, 2009. IC-NIDC 2009, Page(s): 379 – 383.

[11] Namuduri, K.," Automatic extraction of highlights from a cricket video using MPEG-7 descriptors", First International Communication Systems and Networks and Workshops, 2009. Comsnets 2009, Page(s): 1 – 3.

[12] Gwang-Gook Lee; Hyeong-ki Kim; Whoi-Yul Kim," Highlight generation for basketball video using probabilistic excitement", IEEE International Conference on Multimedia and Expo, 2009. ICME 2009, Page(s): 318 - 321 .

[13] Y. Kang, J. Lim, Q. Tian, and M. Kankanhallis, Soccer video event detection with visual keywords," *Proc. of IEEE PCM'03*, 2003.

[14] Kolekar M.H and Sengupta S, "Hierarchical structure for audio-video based semantic classification of sports video sequences", Visual Communications and Image Processing 2005. Proceedings of the SPIE, Volume 5960, pp. 401-409 (2005).

[15] Bertini, M.; Del Bimbo, A.; Torniai, C.; Grana, C.; Vezzani, R.; Cucchiara, R.," Sports Video Annotation Using Enhanced HSV Histograms in Multimedia Ontologies " , 14th International Conference on Image Analysis and Processing Workshops, 2007. ICIAPW 2007, Page(s): 160 - 170.

[16] Mao-Hsiung Hung Chaur-Heh Hsieh ,"Event Detection of Broadcast Baseball Videos", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 18, No. 12, December 2008, Page(S): 1713 – 1726.

[17] Sunitha Abburu , "Multi Level Semantic Extraction for Cricket Video by Text Processing" ,accepted for publication in International Journal of Engineering Science and Technology (IJEST),Vol , no ,Oct 2010.

[18] Xingquan zhu ,xindong wu, Ahmed k. Elmagarmid, zhe feng, Lide wu, "Video Data Mining : Semantic Indexing and event Detection from the Association Perspective", IEEE Transactions on Knowledge and Engineering: vol 17, no.5, May 2005.

Sunitha Abburu: Working as a Professor & Director, Department of Computer Applications, Adiyamaan College of Engineering, Tamilnadu, India. She received MCA from Osmania University, M.phil and Ph.D from Sri Venkateswara University, A.P, India. She is having 13 years of teaching experience and 3 years of industrial experience.