WICKET FALL CONCEPT MINING FROM CRICKET VIDEO USING A-PRIORI ALGORITHM

Mahesh Goyani[1], Shreyash Dutta[2], Gunvatsinh Gohil[3] and Sapan Naik[4]

[1]Department of Computer Engineering, Sardar Patel University, Anand, India
mgoyani@gmail.com
[2]Student Member, IEEE, IEEE CS
shreyashdutta@gmail.com
[3]Department of Computer Engineering, Gujarat University, Gandhinagar, India
gunvantsinh@gmail.com
[4]GEC, Gujarat Technological University, Gandhinagar, India
sapan_say@yahoo.co.in

ABSTRACT

In this paper, we propose an algorithm to detect semantic concepts from cricket video. In our previous work, we have proposed key frame detection based approach for semantic event detection and classification. The proposed scheme works in two parts. In first part a top-down event detection and classification is performed using hierarchical tree. In second part, higher level concept is identified by applying A-Priori algorithm. In part 1, key frames are identified based on Hue Histogram difference at level 1. At level 2, logo transitions classify the frames as real-time or replay. At level 3, we classify the real time frames as field view, pitch view or non field view based on thresholds like Dominant Soli Pixel Ration (DSPR) and Dominant Grass Pixel Ration (DGPR). At level 4, we detect close up and crowd frames based upon edge detection. At level 5a, we classify the close up frames into player of team A, player of team B and umpire based upon skin colour and corresponding jersey colour. At level 5b, we classify the crowd frames into spectators, player's gathering of team A or player's gathering of team B. In part two, labels are associated with each frame event, which is used as input to A-Priori algorithm for concept mining. Results at the end of paper show the robustness of our approach.

KEYWORDS

Histogram, Dominant Grass Pixel Ratio, Dominant Soil Pixel Ratio, Concept Mining, A-Priori Algorithm

1. INTRODUCTION

Video is the collection of continuous frames, displayed at some specific rate (normally 25 fps). Compared to image, video occupies much more space. And most of the time, much of the portion of Sport videos, security videos is not of one's interest. The quantity of data produced by recording sports videos need filtration and summarization. Due to long duration of the video, it is quite cumbersome process to index some particular event of the video. One has to go through each frame to find out some specific event from it. In recent years sports video analysis has become a widely research area of digital video processing because of its huge viewership and commercial importance [1], [2]. Because of the enormous difference in sports videos, sport specific methods show successful results and thus constitute the majority of work. Some of the genre specific researches have been done in soccer (football) [3], [4] tennis [5], cricket [6], basketball [7], volleyball [8], etc. less work is observed for genre-independent studies [8], [9].

In viewership and fan following cricket is next to soccer. Major cricket playing nations are India, Australia, Pakistan, South Africa, Zimbabwe, Bangladesh, Sri Lanka, New Zealand, and England. In spite of its huge viewership cricket has not obtained its share in the research community [6], [10]. Cricket video analysis is far more challenging because of the complexities

of game in itself. Cricket is a game of variable factors as compared to other famous sports such as soccer, basketball, tennis etc. Some of these are its various formats (tests, one days and the new popular addition -T20), dynamic playing Conditions (field area, pitches etc), Day and day/night matches (cause illumination related problem) and duration. Even the shortest version of cricket i.e.T20 is played for approximately 3 hours which is greater than soccer (approximately 90 minutes) and hockey (approximately 70 minutes). We propose a key frame detection approach for minimizing the computation time which is imperative as the amount of data is huge. We have concentrated our work on T20 matches as they offer lots of events in short duration as compared to ODI's. The same work can be extended to ODIs as well.

Structure of this paper is as follows. Section II shows the overall flow of the work carried out. Section III shows the classification and detection algorithms (*Previous work*). Section IV the experimental results and Section V concludes the paper with related future work.
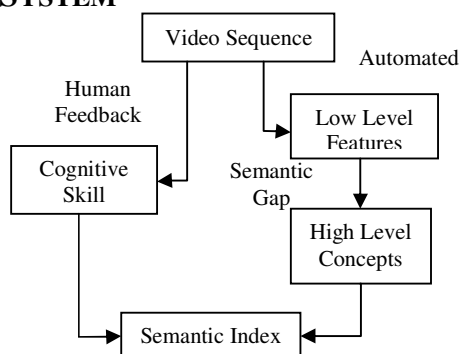
2. OVERVIEW OF THE SYSTEM

Figure 1. Block Diagram for semantic indexing

As shown in figure 1, High level concept mining from given video is two levels hierarchical process. At first level, meaningful events are identified associated with video using low level features. *Event* means it is the scene in the video having some semantic meaning attached to it. And at second level, higher level concepts are recognized based on previous level results. *Concept* means it is a collection of a temporally ordered set of events. For cricket, events like, real time video, reply, pitch view, field view, close up classification, crowd classification are extracted in first level. Collections of such events are used later on for extracting the concepts like wicket fall (balled out, catch out, run out, stumped out etc), boundary (fours and sixes), run milestones (Half century or century) etc. There is always a semantic gap between low level feature and high level concepts which needs to be filled up. Human feedback is also very much important because of his cognitive skill to identify any event at a glance. Because of dynamic nature of cricket, human interaction is always needed at first level for better concept mining.
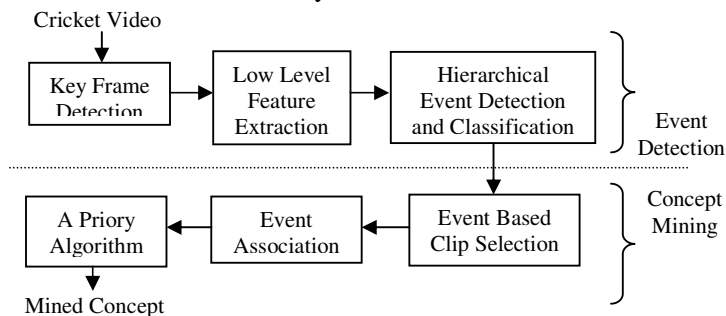

Figure 2. Block Diagram Proposed System

Figure 2 describes the complete flow of our proposed work. In first part key frames are detected, which reduces the processing time. From key frames, using image processing algorithms, events are detected. In second part, based on these events, video segment clips are segmented and processed separately. Labels are associated with frames and these labels are used as an input to A-priory algorithm to find out meaningful concept. Proposed approach is discussed in detailed in section III and Section IV. Most of the work in section III is based on our previous work discussed in [11]. So some of the secondary things we have not discussed here.
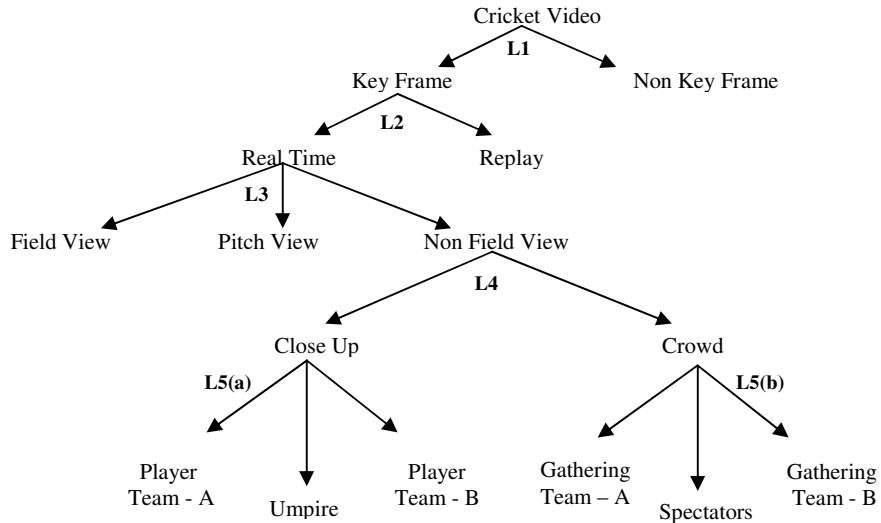
3. DETECTION AND CLASSIFICATION

Figure 3. Hierarchical framework followed for detection and classification.

3.1. Level 1: Key Frame Detection

Any video frame sequences have certain properties like some frame sequences are either same or have a little difference between them as shown in the figure 4, thus using classifiers to classify all such frames is computationally inefficient. Suppose a sequence of same (S) and different (D) frames are:

$$S\ S\ S\ D\ S\ S\ S$$

For such a sequence of frames there would be a spike in the HHD plot with a rising edge at frame 4 and falling edge at frame 5, we have to consider only the rising edges to decide on key and non key frames. As shown in Figure 4, Frame #580 is classified as a key frame. Detail algorithm is discussed in our previous work mentioned in [11].
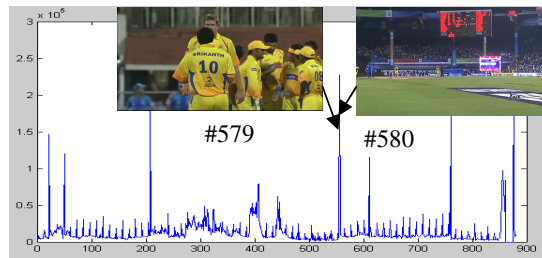

Figure 4. Hue Histogram Difference (HHD) plot for key frame detection.

3.2. Level 2: Reply / Real Time Segment Detection

To detect replays from sports video, some early works focus on the characteristics of them such as motion vector [6] and replay structures [12]. However, these methods are not robust enough to be suitable for various kinds of sports video replay detection because replays in different sports video are various and compiled in different manners and can hardly be represented by such simple features. Therefore the recent approach is to detect the accompanying logo effect of the replays in sports videos to acquire the replay segmentations. [13], [14]. A replay segment is always sandwiched between two logo transitions or flying graphics which last for 8-15 frames as shown in figure 5. We have used the logo colour property to detect logo frames or flying graphics and thus detect replay and live segments. Detail algorithm is discussed in our previous work mentioned in [11].
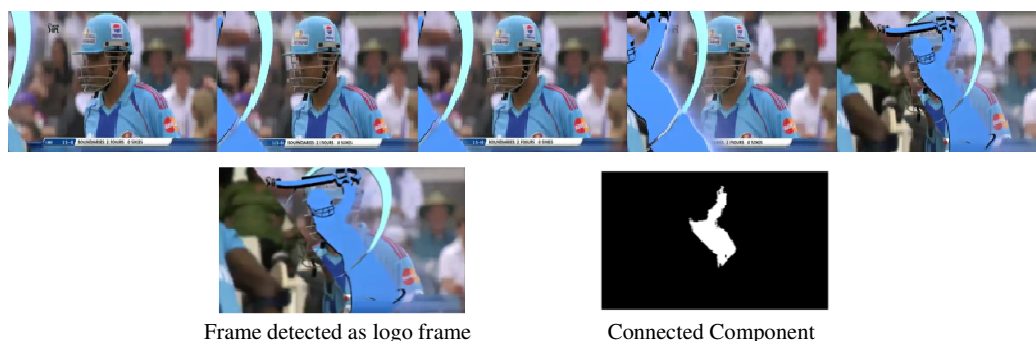

Frame detected as logo frame   Connected Component

Figure 5. Example of a logo transition and detection.

3.3. Level 3: Pitch View / Field View Detection

Pitch view and Field view can be detected using the colour properties.

3.3.1. Pitch View

A pitch in a cricket field can be easily distinguished by its color, using this property we calculate the DSPR to decide on the pitch view and non pitch view. We plot 256-bin histogram of the hue component of the frames. We pick up the peaks of hue histogram of these images. As shown in Figure 6(a), we observed peak at bin $k = 23$ and value of the peak is 15100 for the particular image of size 270× 520. By testing 50 images, we observed that the soil color peak occurs between bin $k = 20$ to $k = 25$. The summation of peaks of the histogram gives number of the pixels of the soil in the image (*xs*). From this, we compute the *DSPR* as *xs/x*, where *x* is the total number of pixels in the frame. We observed *DSPR* values vary from 0.05 to 0.1 for the pitch view. For non-pitch view image shown in Figure 6(b), we observed peaks belongs to the bins other than $k = 20$ to $k = 25$ and *DSPR* value is very small. As the soil color of pitches in a cricket field change from venue to venue change in bins is also possible so it is important to compute the histogram first then observe the bins.
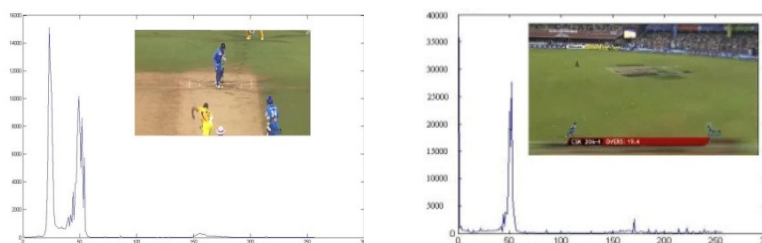

Figure 6. Hue histogram of (a). Pitch view (b). Field View

3.3.1. Field View

A field view image can be distinguished by its grass color as in, using this property we calculate the DGPR [15] to decide the field view and non field view images. We plot 256-bin histogram of the hue component of these images. We pick up the peaks of hue histogram of these images. As shown in Figure 6(b), we observed peak at bin $k = 56$ and value of the peak is 5090 for the particular image of size 270×520. By testing 50 images, we observed that the green color peak occurs between bin $k = 48$ to $k = 58$. The peak of the histogram gives number of the pixels of the grass in the image. We call this number as xg. From this, we compute the *DGPR* as xg/x, where x is the total number of pixels in the frame. We observed *DGPR* values vary from 0.25 to 0.5 for the field view. For non-field view image shown in Figure 6(a), we observed peak belongs to the bins other than $k = 48$ to $k = 58$ and *DGPR* value is very small. As the grass color of pitches in a cricket field change from venue to venue, change in bins is also possible so it is important to compute the histogram first then observe the bins. Detail algorithm is discussed in our previous work mentioned in [11].

3.4. Level 4: Close up / Crowd Detection

At this level, we try to detect close up and crowd frames from the non pitch / non field view frames detected at level 3. From our observations we have seen that close up or crowd frames are shown frequently whenever an exciting event occurs such as when a wicket falls, close up of batsman and bowler, then view of spectators and the player's gathering of fielding team are certainly shown. The edge detection is performed by finding the maximum gradient value of a pixel from its neighbouring pixels. If the maximum value of gradient satisfy the threshold than the pixel is classified as an edge pixel [16].
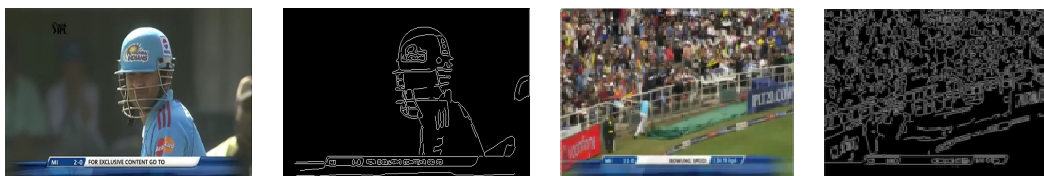

Figure 7. (a) Close up image. (b) Edges (a), (c) Crowd image. (d) Edges (c).

We have used canny edge detection algorithm to distinguish between close up and crowd frames, close up frames as shown in Fig. 8(b) has lesser edge pixel density then the crowd.

3.5. Level 5a: Close Up Classification

In this level, we will be classifying the close up frames by using the skin color and jersey color information. Because of race, face geometry, helmet, caps etc. factors the face detection is not an easy task [17]. Figure 8 shows some of the close up frames extracted from cricket videos. It can be observed that skin blocks are mostly detected in the blocks 6, 7, 10, 11. According to the face block location we will select the corresponding jersey block for close up classification [15]. Skin color block to jersey color block corresponding table is calculated by rigorous testing on various close up images. Detail algorithm is discussed in our previous work mentioned in [11].

Figure 8. Some representative close up (a), (b) Close up of batsman. (c) Close up of Umpire.

3.5. Level 5b: Crowd Classification

In this level we try to classify the crowd frames detected at level 4 using the jersey color information into the following three classes: Players Gathering of team A, Players gathering of team B and spectators. Since fielders gather on the field after exciting events such as wicket fall and third umpire referrals classifying them is important. Spectators are also shown at all important junctures. Detail algorithm is discussed in our previous work mentioned in [11].

Figure 9 (a, b): Players gathering of teams. (c): Spectators. (d): Jersey color detection of (b)

4. SEMANTIC CONCEPT MINING

In section 3, we have determined the frame labels using low level features. Now its turn to find out the concept associated with the set of events. After first level processing, we have labelled frames with the events like pitch view, field view, player of team A, player of team B, Umpire, Spectators and so on. Table 1 shows the label associated with individual such events.

Table 1. Label associated with events

Sr. No	Label	Associated Event
1	Vp	Pitch View
2	Vf	Field View
3	Pa	Player of Team A
4	Pb	Player of Team B
5	U	Umpire
6	Ga	Gathering of Players of Team A
7	Gb	Gathering of Players of Team B
8	S	Spectators
9	R	Reply

It's very obvious that during any excitement event camera will move over different parts of the ground and so various events are formed. By proper association of these events, we can extract out the concept like wicket fall. If there is a wicket fall of team A, than player gathering of team B, spectators, Close up of player of team A, umpire etc events will be detected repeatedly in very short span of video.

Entire video segment is divided in to real time small segments; because there is no meaning to process the reply scene as it is already processed in real time clip. Key frames are extracted and reply segments are skipped from further processing. Remained real time segments are applied as

input and key frames are classified. Using table 1, we associate specific label to each classified frame. For the robustness, we have kept the threshold of consecutive 25 frames. It means, if frames with same class label does not have count greater than 25 in single shot, than it would not be assigned any label for further processing. On label sequence, we are applying A-Priori algorithm to find out meaningful concepts like wicket fall or boundary and some more. Figure 10 shows the flow of complete concept mining process.
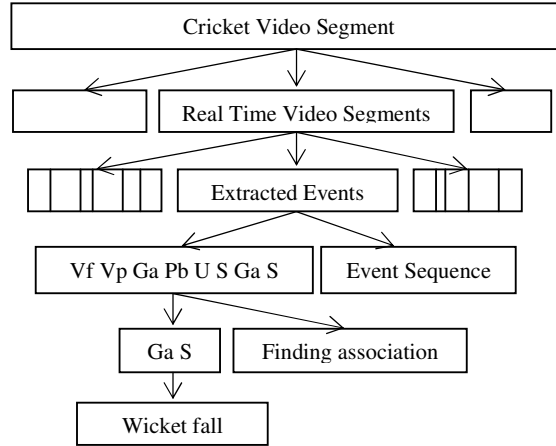

Figure 10. Typical example of concept mining

Wicket fall concept result is shown for video V1, match between Chennai Super King versus Mumbai Indians. During first part of processing, we have been able to index each frame with some class label. We have analysed about 700 frames of V1 which contains fall of wicket. The frames event detection time line is shown in figure 12. Frame #1 to frame #154 are classified as pitch view, frame #155 to frame #193 is not classified as any of the event and so on.
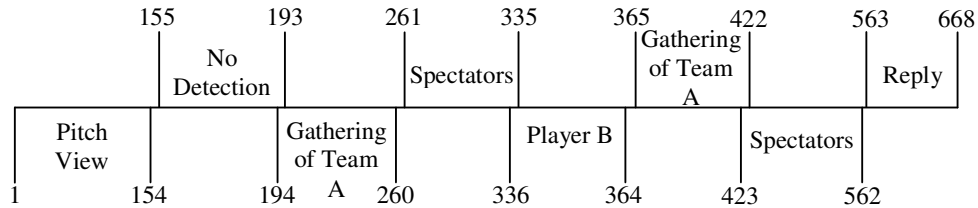

Figure 11. Time Line of Events Happening in V1

Event to label association is shown in table 1. According to that association, above time line gives following video event sequence:

Vp Ga S Pb Ga S R

This event sequence is given as an input to well known A Priori algorithm for concept mining; We have kept support count value 2.

Table 2. Frequent item set generation by A Priori algorithm

Level 1 Item Set (C1)	Level 1 Item Set After Support Count Check (L1)	Level 2 Candidate Generation (C2)	Level 2 Item Set (C2)	Level 2 Item Set After Support Count Check (L2)
{Vp} 1 {Ga} 2 {S} 2 {Pb} 1 {R} 1	{Ga} 2 {S} 2	{Ga S} {S Ga}	{Ga S} 2 {S Ga} 1	{Ga S} 2

In last iteration of the algorithm, we are getting the patter GaS, it indicates player gathering of team A and spectators are the frequent patterns which happens in the video segment. GaS means it's a wicket fall for team B.

5. RESULTS AND CONCLUSIONS

Rigorous testing has been carried out to test the above algorithms. To achieve cent percent accuracy is highly improbable because of the inherent inconsistencies in cricket videos in itself. We have concentrated our work on T20 matches as they offer lots of events in short duration as compared to ODI's.

Table 3. Experimental Database

Cricket Video	Cup	Match(A vs. B)
V1	IPL 2008	Match 8 : CSK vs. MI
V2	IPL 2008	Match 36: MI vs. CSK.
V3	T20 World Cup 2007	Final match : India vs. Pakistan

We have defined the events as scenes in the video with some semantic meaning (i.e. labels from a semantic hierarchy) attached to it based on the leaf nodes shown in Figure 3. Events are extracted as the leaf nodes of the level- 2 to 5 of hierarchical tree. The events are replay, batsman, bowler/fielder, spectator, fielder's gathering, pitch view, field view.

Table 4. Statistics of frame detection and classification

Features	Segment -1	Segment -2	Segment-3
Key Frame	56	104	108
Non Key Frame	601	778	688
Replay	108	21	325
Real time	549	861	471
Pitch View	154	28	95
Field View	0	179	150
Close up	30	46	332
Crowd	364	506	254

Our Key Frame detection based approach shows excellent detection accuracy and also results in saving of processing time. The classifiers exhibit better detection and classification ratio at various levels. Our algorithms are able to extract out events as shown in table 4 from cricket videos. Snapshots in fig. 12 show some of the classified frames. Caption in top of the frame shows the detected event for that particular frame. Video indexing is very resource heavy as it consumes a lot of CPU and hence time, the following table shows comparison between three

methods tested by us for indexing a 900 frame video. Table 6 shows that with our approach we have reduced processing time by 60 percent.

Table 6. Processing time analysis with various parameters

Sr. No.	Parameter	Indexing Time
1	No parameter	~35Min
2	Look-Ahead of 2	~20Min
3	Key frame detection	~15Min

Figure 13 shows the events detected for wicket fall for team B. As a result of A Priori algorithm, we are getting *GaS* as a frequent pattern. It is obvious that gathering of players of team A and spectators are displayed frequently on wicket fall.

Figure 12. Some of the classified frames (Part I: Event Detection)

Figure 13. Wicket fall event sequence (Part II: Concept Mining)

REFERENCES

[1]   A. Kokaram, N. Rea, R. Dahyot, M. Tekalp, P. Bouthemy, P. Gros and I. Sezan, "Browsing sports video: trends in sports-related indexing and retrieval work", in IEEE Signal Processing Magazine, vol. 23, no. 2, pp.47-58, 2006.

[2]   Y. Li, J. Smith, T. Zhang and S. Chang, "Multimedia Database Management Systems, in Elsevier Journal of Visual Communication and Image Representation", pp. 261-264, 2004.

[3]   J. Wang, E. Chng, C. Xu, H. Lu, and Q. Tian, "Generation of personalized music sports video using multimodal cues". In IEEE Transaction on Multimedia, 9(3), 2007.

[4]   M. Baillie and J. M. Jose, "Audio-based event detection for sports video. In Lecture Notes on Computer Science", 2728,2003.

[5]   G. Zhu, Q. Huang, C. Xu, L. Xing, W. Gao, and H. Yao, "Human ehaviour analysis for highlight ranking in broadcast racket sports video", IEEE Transactions on Multimedia,Vol. 9(6), 2007.

[6]   M. H. Kolekar and S. Sengupta, "Event-importance based customized and automatic cricket highlight generation", In IEEE Int. Conf. on Multimedia and expo, 2006.

[7]   C. Xu, J. Wang, H. Lu, and Y. Zhang, "A novel framework for semantic annotation and personalized retrieval of sports video", in IEEE Transactions on Multimedia,Vol. 10(3), 2008.

[8]   L. Duan, M. Xu, Q. Tian, C. Xu, and J. Jin, "A unified framework for semantic shot classification in sports video", in IEEE Transactions on Multimedia,Vol. 7(6), 2005.

[9]   A. Hanjalic, "Generic approach to highlight extraction from a sport video", in ICIP, 2003.

[10]   P. Xu, L. Xie, S. Chang, A. Divakaran, A. Vetro, and H. Sun, "Algorithms and system for segmentation and structure analysis in soccer video", in IEEE ICME 2001.

[11]   Mahesh Goyani, Shreyash Dutta, Payal Raj, "Key Frame Detection Based Semantic Event Detection and Classification Using Hierarchical Approach for Cricket Sport Video Indexing", CCSIT, SPRINGER, LNCS in Communications in Computer and Information Science, Vol. 131, pp. 388-397, Proc. of International Conference on Computer Science and Information Technology, Bangalore, India, 2-4 Jan '11.

[12]   H.Pan,P.Beek,M.Sezan, "Detection of slow-motion replay segments in sports video for highlights generation", ICASSP2001, Salt Lake City, UT, May 2001.

[13]   X.Tong,H.Lu, "Replay detection in broadcasting sports video", Proceedings of the Third International Conference on Image and Graphics. 2004.

[14]   H. Pan, B. Li, and M. I. Sezan: Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions, Proc. IEEE, ICASSP, 2002.

[15]   M.H. Kolekar ,K. Palaniappan S. Sengupta, "Semantic event detection and classification in cricket video sequence", Sixth Indian Conference on Computer Vision, Graphics & Image Processing.

[16]   Y.Jacob Vetha Raj, M.Mohamed Sathik, K. Senthamarai Kannan, "Hybrid Image Compression by Background and Non Edged", The International Journal of Multimedia and its Applications, Vol. 2, No.1, pp. 32-41, February -2010.

[17]   Mahesh Goyani, Brijesh Joshi, Gitam Shikkenawis, "Acceptance/Rejection Rule Based Algorithm For Multiple Face Detection in Colour Images", International Journal of Engineering Science And Technology, Vol. 2(6), pp 2148-2154, July 2010.

Authors

Mahesh Goyani received his Bachelor degree in 2005 from Veer Narmad South Gujarat University, India. He received his master degree in field of Computer engineering in 2009 from Sardar Patel University. He is working as a Lecturer at Department of Computer Engineering, GCET, Sardar Patel University, Anand, India. His research interest includes Image Processing, Computer Algorithms, and Artificial Intelligence. He has published fourteen papers at national and international level. He is the pioneer of Image Processing Research Group at GCET. He is working as a reviewer for two international journals.