Mining Entertainment Video Content Structure and Events towards Efficient Access and Scalable Skimming

Umakant Ahirwar[1], Deepak Bhatnagar[2]

Abstract— With the ever-growing digital libraries and video databases, it is increasingly important to understand and mine the knowledge from video database automatically. The media and entertainment industries, including streaming audio and digital TV, present new challenges for managing and accessing large audio-visual collections. In this paper, a shot ontology description based for the football match video. Shot ontology is inferred by shot manipulations that includes shot detection, shot type classification, score board detection and motion statistics .This video content management system provided event feature manipulations at multiple levels: signal, structural, or semantic in order to meet user preferences while striking the overall utility of the video. The experiment results showed that our proposed methodologies could correctly detect interested events, long shots, and close-up shots and also achieved the purpose of video indexing and weaving for what user preferences.

Index Terms— Close-up Shots, Long Shots, Multimedia Data Mining, Shot Ontology.

———————————— ◆ ————————————

1 INTRODUCTION

ORGANIZATIONS with large digital assets have a need to retrieve meaningful information from their digital collections. Applications such as digital libraries, video-on demand systems, and interactive video applications introduce new challenges in managing large collections of audiovisual content. To help users find and retrieve relevant video more effectively and to facilitate new and better ways of entertainment, advanced technologies must be developed for indexing, filtering, searching, and mining the vast amount of videos. Motivated by these demands, many video research efforts have been made on exploring more efficient content management systems. A simple framework is to partition continuous video frames into discrete physical shots and extract low-level features from video shots to support activities like searching, indexing or retrieval [1].

Multimedia data is being acquired at an increasing rate due to technological advances in sensors, computing power, and storage. Multimedia Data Mining is the process of extracting previously unknown knowledge and detecting interesting patterns from a massive set of multimedia data Video is rapidly becoming one of the most popular multimedia due to its high information and entertainment capability. It also consists of audio, video and text together.

Video mining is a process which can not only automatically extract content and structure of video, features of moving objects, spatial or temporal correlations of those features, but also discover patterns of video structure, objects activities, video events, etc. from vast amounts of video data without little assumption about their contents. Many video mining approaches have been proposed for extracting useful knowledge from video database. Finding desired information in a video clip or in a video database is still a difficult and laborious task due to its semantic gap between the low-level feature and high-level video semantic concepts. Video data mining can be classified in following categories, such as pattern detection, video clustering and classification and video association mining.

A video database contains lot of semantic information. The semantic information describes what is happening in the video and also what is perceived by human users. The semantic information of a video has two important aspects. They are (a). A spatial aspect which means a semantic content presented by a video frame, such as the location, characters and objects displayed in the video frame. (b). A temporal aspect which means a semantic content presented by a sequence of video frames in time, such as character's action and object's movement presented in the sequence. To represent temporal aspects, the higher-level semantic information of video is extracted by examining the features audio, video, and superimposed text of the video. The semantic information includes the detecting trigger events, determining typical and anomalous patterns of activity, generating person-centric or object-centric views of an activity, classifying activities into named categories, and clustering and determining the interactions between entities. The temporal aspect of videos prevents the efficient browsing of these very large databases. Many efforts are conducted to extract the association between low-level visual features and high-level semantic concepts for image annotation [4].

In this paper, we proposed a video content management system that integrated three main processing phases: Shot Ontology Definition, Feature Extraction, and Video Indexing. In shot ontology definition phase, we gathered low level statistics in football match video and then to derive the temporal descriptors in event occasions: shot change, shot sequence, court color, and shot type. With comparing the features in different shots, we used the histogram to calculate several shots which segmented to different shot from one video film [2][3][9]. In the video indexing phase, we recorded the related shot events and time in video sequence respectively.

For example, the scoring events, the long shot and the close-up shot would be detected and then to analysis the visual feature for shot classification.

2 RELATED WORK

During the past few years, many researchers have addressed video database issues from different perspectives, and various experimental systems have been implemented.

"Video Data Mining: Semantic Indexing and Event Detection from the Association Perspective" proposed by "Xingquan Zhu" in 2005 presenting a knowledge based video indexing and content management framework for domain specific videos(using basketball video) and provide a solution to explore video knowledge by mining association from video data[1].

"WVTDB- A semantic contect based video database system on the world wide web" proposed by "Haitao Jiang" in 1998 describes the design and implementation of a web based video database system (WVTDB) that demonstrates his research on video data modeling, semantic content based video query, and video database system architechure[5].

"Videocube: a nobel tool for video mining and classification" proposed by "Jia Yu Pan and Christos Faloutsos" in 2002 proposed a new tool to classify a video clip into one of n given classes (e.g. "news", "commercials"). And automatically derive a "vocabulary" from each class of video clips using "independent component analysis"[6].

"Classminer :mining medical video conent structure and events towords efficient access and scalable skimming" by "Ahmed K .Elmagarmid" in 2002. They introduce a video content structure and event mining framework for efficient video indexing and access by using video shot segmentation and key frame selection[7].

"Smart Video text: A video data model based on conceptual graphs" by "A.K. Elmagarmid, F.Kokkoras" in 2002. use the conceptual graph knowledge representation formalism to capture the semantic associations from text annotations of video data[8].

"Video data mining: mining semantic pattern with temporal constraints from movies" by "Kimiaki Shirahama, Kaichi ideno" in 2005.they extract semantic patterns from a movies and sequential patterns by connecting temporally close and strongly associated symbols and propose a parallel data mining method to reduce the computational cost[10].

"Indexing and teaching focus mining of lecture videos" by "YuTzu Lin, Bai Jang Yen" in 2009.proposed and indexing and teaching focus mining system for lecture videos. Which use an edge based shot change detection algorithm for constructed of slide structure[11].

- [1]*Umakant Ahirwar is currently pursuing masters degree program in computer science and engineering from college of science and engineering,Jhansi, in Uttar Pradesh Technical University, India*
  E-mail: ahirwar1984@gmail.com
- [2]*Deepak Bhatnagar is currently head of departmens of computer science and engineering in college of science and engineering ,jhansi ,India,*
  E-mail:bhatnagar.deepak@rediffmail.com

"Mining similarities for clustering web video clips" by "Shouquin Lin Ming Zhu" in 2008. They propose an approach to cluster similar web searched video based on video visual similarities mining[12].

"Mining High level features from video using associations and correlations" by "Lin Lin and Mei Ling Shyu" in 2009.They presenting a high level feature detection framework using AKM (Association Rule Mining) technique with the correlations among the feature value pairs and developed an association rules mining algorithm[13].

"Sport video mining with mosaic" by "Tao Mei, Yu fei Ma" in 2005. They proposed a generic approach to key event as well as structure mining for sport video analysis by using mosaic ,is generated for each shot as the representative image of shot content[14].

3 SHOT ONTOLOGY

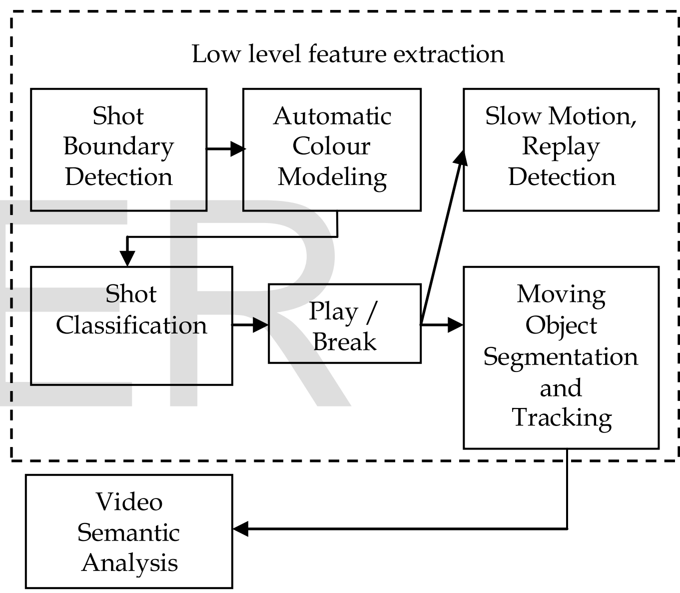

Fig3.1. Video Shots Processing Procedure

Video shot processing procedure is illustrated in figure3.1. The low level feature detection processing contained shot boundary detection, automatic color modeling, and shot classification processes. Play-break detection was proposed by A. Ekin [15].

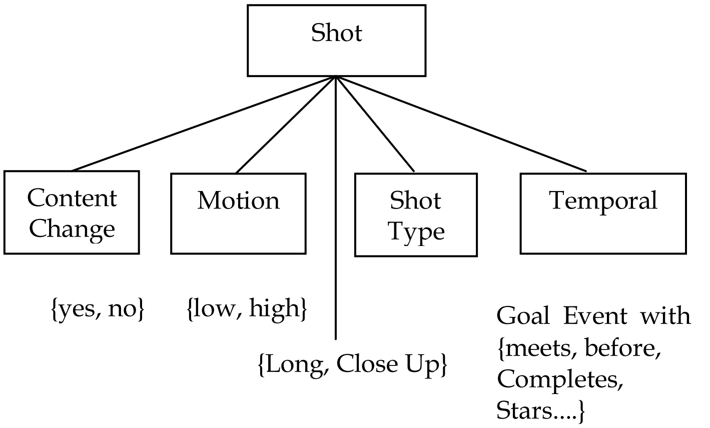

Fig3.2.An Attributes Table Example of Football Sport Shot Ontology.

Cinematic feature is considered by uses shot type and shot length features. Slow-motion replays are generated by slowing the frame rate of the playback of the recorded event. This essentially causes a single frame to be repeated several times [16]. Moving object segmentation and tracking was derivative by motion vector that was proposed by V. Mezaris, et.al [17]. They consider the region temporal and the trajectory- base region merging to segment the objects. And then to use the intermediate- level descriptors to define the attributes for shot ontology. The attributes table of the shot ontology was the basic intension of the ontology. According to the attributes, we could establish a shot ontology for the football sport as shown in figure 3.2.

In a sports game, the director took the picture with switched shot according to the game course, such as there was a series of steals, fast breaks and then kicked into the goal in a football game. The director would compose a splendid short film with a series shot switching which contained the scoring shot, the cheering of the spectators, and the hugging of the players, etc. So we considered the process of the shot switching a semantic representation consequently.

There was two kinds of shots will be analysis: Long Shot and Close-Up Shot. The Long Shot was taking the picture for the matter of a game, and the Close-Up Shot was focused on the object which the spectators should pay attention to in a game. It could be a person, or an article instead. And we named the Long Shot and the Close Up Shot as 'cmeL' and 'cmeO' respectively. Also, we listed some terminology in a football game, such as free throw (ft), pass(pass), slam dun (dk) and so on. The narrative in football video was shown on table 3.1.

TABLE 3.1.
THE ABEREVIATION IN FOOTBALL GAME

Football Events	Abbreviation
Direct Free Kick	dfk
Foul	ful
Goal	gol
Indifferent (don't care)	ind
Corner kicks	ck
Goal kicks	gk
Pass ball to other	pass
Indirect Free Kick	ifk
Penalty kick	pk
Throw-ins	ti
Overhead kick	ok
Cross	cs
Dribbling	db
Full timeout	stop
Long Shot	cmeL
Close Up Shot	cmeO

In addition to the statistics of shot time, we also aimed some narratives of the breaking event in a football game, such as goal(gol), foul(ful), and Indifferent (don't care, ind) and we had a statics anaysis result from the 'Motion', 'Score Board / SB Change', and 'Shot Type' respectively (As shown in table 3.2).

TABLE3.2.
THE ATTRIBUTE TABLE FOR THE SHOT ONTOLOGY

	Goal	Ful	Stop	Ft	Ind
Motion	High	High	High	Low	-
Shot type (cmeL/ cmeO)	cmeL	cmeL	cmeL	Both	cmeO
Sbchange (yes/no)	Yes	Both	No	Yes	No
Time(s)	>tg	-	-	-	<ti

4 METHODOLOGY FOR SHOT CLASSIFICATION

In a football game, the splendid pictures of a game would not only be a goal picture for the auditorium. But the goal shot could comprise some splendid frames of a game. Therefore, we could define the descriptors of a 'Shot Type' to be a category of the shot change consequently. And the shot could

ISSN 2229-5518

be divided into 'Long Shot' and 'Close Up Shot' which was indicated as Fig4.1 and Fig4.2.

Fig4.1.Close-up shot

Fig4.2.Long shot

There were many classifications for the Long shot and Close-up Shot theoretically [9,16,17,]. We could classify it by the background color of the field and the skin color. The definitions were described as following:

Long Shot: This should be decided by the subtraction from the court color and skin color. It would be the Long Shot if the absolute value was small than ∂ and the ratio of skin color was small than σ .

Close Shot: This should not contain any court color or skin color was greater than the court color. It would be the Close Up Shot if the absolute value was greater than ∂ and the ratio of skin color was greater than σ .

4.1 Classification Algorithm

Rc is the ratio of court color pixels, Rs is the ratio of skin color pixels
cmeL is define for long shot, cmeO is define for close shot

```
if Rc=0‖Rs>=Rc
    Shot Type = cmeO
End
If (Abs(Rc-Rs)> ∂)
    If Rs > σ
        Shot Type = cmeO
Else
    Shot Type = cmeL
   End
Else
    Shot Type = cmeL
end
```

The classification was decided by the ratio of the skin color and the court color.

Temporal Relation: After the compiling statistics result of our game, we got a event sequence list for the Long Shot and Close Up Shot which shown as Fig 4.1(a).

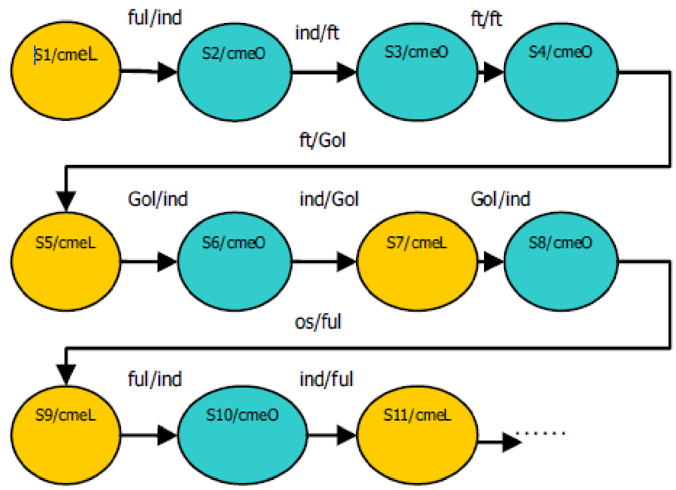

Fig. 4.1(a) The Sequence Of Types And Events Indexing

The S1 was defined as a Long Shot. When a foul(ful) occurred, the indifferent (ind) event raised accordingly. And a serial of field goal or some indifferent event would be followed consequently. We could simplify the similar type of shot as followings:

Step 1: It could be simplified to the same one sequence if the status of current shot was same as next shot. But the event of current shot should be retained.

Step 2: Repeated the previous step 1 until all sequences were simplified completely.

Step 3: Expanded the simplification to two sequences for one sequent unit. It could be simplified to one sequent unit if the status of current sequence was same as next sequence.

Step 4: Repeated the previous step 3 until all sequent units were simplified completely.

It was easier to realize a football game from the simplification of a sequence shot, such as the relations between Shot events and Shot types. Also eliminate the very small period shots by the detection of timing events. It would be much help for enhancing precision by removing these short time periods after the statistic analysis considerably.

5 SHOT INDEXING

The content indexing of a football game was the shot ontology itself. We had defined every shot descriptor through its low-level characteristics on the previous section. There were different authority offered by the shot ontology in Changing was the highest authority ratio value, due to these events could distinguish the whole game processing into game fragments from the definition of shot ontology substantially. Generally, it might be compressed to one hour video from a normal two hours film by a verified method that was known as play back film detection experimentally. And all the terminologies of the sport event happened during the game defined from the application level of the ontology.

We also could draw out a ball game event by the related searching that marked the characteristic of each video period to the time-line axis by the knowledge of shot ontology definitively. The flow of the shot indexing was described as followings (refer to fig 4).

1. The first step of shot indexing is the shot detection; this process separated a sequential video into several different shots occasionally. We utilize the colour histogram for the determination of shot separation. After the shot detection for the cutting separately, we record the beginning frame, ending frame and the time/number of each shot in a database accomplished consequently.

2. Since we got the total frame number of each shot, we could sort those shot with its frame number Close Up Shot, attractive sidelight activities or field goal. And the court colour would take most portions on each game.

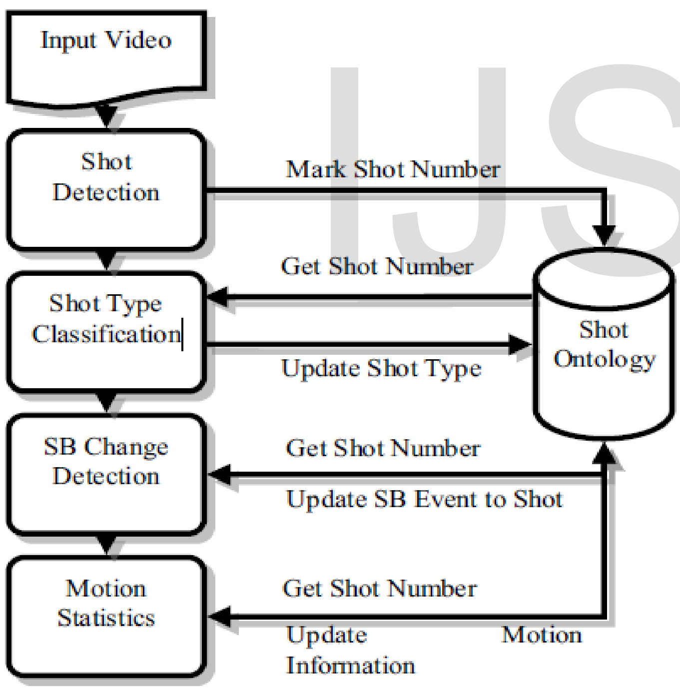

Fig 5.1. Shot Indexing Process

We set the Shot Classification and Score Board Changing was the highest authority ratio value, due to these events could distinguish the whole game processing into game fragments and non-game fragments from the definition of shot ontology substantially. Generally, it might be compressed to one hour video from a normal two hours film by a verified method that was known as play back film detection experimentally. And all the terminologies of the sport event happened during the game defined from the application level of the ontology.

6 IMPLEMENTATION RESULTS

We also could draw out a ball game event by the related searching that marked the characteristic of each video period to the time-line axis by the knowledge of shot ontology definitively. The flow of the shot indexing was described as followings (refer to figure 5.1). The efficiency of scoring segments detection was evaluated by the Precision Rate and Recall Rate simultaneously.

Precision = correct / (correct + false)

Recall = correct / (correct + miss + false)

The Correct represented the number of correct detection, and the False meant the number of misjudgments. The Precision Rate could be estimated for the rate of misjudgment entirely. The Miss would be a score board change event without any detected actually. Therefore, the calculation model of Recall could be evaluated the accuracy of the scoring event detection.

Table6.1.

SCALBLE VIDEO SKIMMING AND SUMERIZATION RESULTS

Events	Selected Number	Detected Number	True Num ber	Pr	Re
Presenta tion	15	16	13	0.81	0.87
Dialog	28	33	24	0.73	0.85
All Shot	39	32	21	0.65	0.54
Average	82	81	58	0.72	0.71

7 CONCLUSION

This paper defined a theory of shot ontology, and analyzed the descriptors of shot ontology with various low-level characteristics. We also marked the timing of a football game video film with the skin color, court color and score board changing event for the space relationship and inferred the scoring segments from the timing relation of shot types. Eventually, there was an excellent result to support our research. We could have an excellent detection of the scoring event for the video frame in a football match video game by our indexing procedure essentially. And the detection of other course events will be strengthened on our next development stage. Meanwhile, the more convenient weaving of indexing and the authority function would be provided for the system browser. This would allow users to have a interactive viewing capability ultimately.

8 ACKNOWLEDGMENT

In the course of developing this paper, I have been deeply indebted to my guide Mr. Deepak Bhatnagar head of department in the College of science and Engineering, for their knowledge and guidance that have helped shaped this paper to its present form. Without his kindest effort and arrangement, this paper will not be able to complete.

REFERENCES

[1] Xingquan Zhu, Xindong Wu, Ahmed K. Emagarmid, "Semantic Indexing and Event Detection From the Associative Perspective" IEEE Transaction, vol-17, no.5, may 2005.

[2] M.R. aphade, et.al "Modeling semantic concepts to support query by keywords in video," in Proc. IEEE Int. Conf. on Image Processing ICIP '02, vol. 1, NY, Sept. 2002, pp. 145-148.

[3] Min Xu, et.al. " Event Detection in Basketball Video Using Multiple Modalities" Information,Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on Vol.3, 15-18 Dec. 2003, pp.1526 – 1530.

[4] Vijaya kumar and Nedunchezhian. R "RECENT TRENDS AND RESEARCH ISSUES IN VIDEO ASSOCIATION MINING" The International Journal of Multimedia & Its Applications (IJMA) Vol.3, No.4, November 2011.

[5] H. Jiang and A.K. Elmagarmid, "WVTDB—A Semantic Content-Based Video Database System on the World Wide Web," IEEE Trans. Knowledge and Data Eng., vol. 10, no. 6, pp. 947-966, Nov./Dec. 1998.

[6] J. Pan and C. Faloutsos, "Video Cube: A Novel Tool for Video Mining and Classification," Proc. Int'l Conf. Asian Digital Libraries (ICADL), pp. 194-205, 2002.

[7] X. Zhu, J. Fan, W.G. Aref, and A.K. Elmagarmid, "ClassMiner: Mining Medical Video Content Structure and Events Towards Efficient Access and Scalable Skimming," Proc. ACM SIGMOD Workshop, pp. 9-16, 2002.

[8] F. Kokkoras, H. Jiang, I. Vlahavas, A. Elmagarmid, E. Houstis, and W. Aref, "Smart Video Text: A Video Data Model Based on Conceptual Graphs," ACM/Springer Multimedia Systems, vol. 8, no. 4, pp. 328-338, 2002.

[9] J. M. Corridoni, A Del Bimbo, "Structured representation and automatic indexing of movie information content", Pattern Recognition, Vol.31, No.12, 1998, pp.2027-2045.

[10] Kimiaki Shirahama, Koichi Ideno and Kuniaki Uehara "Video Data Mining: Mining Semantic Patterns with temporal constraints from Movies" Proceedings of the Seventh IEEE International Symposium on Multimedia (ISM'05) 2005.

[11]Yu-Tzu Lin,Bai-Jang Yen,Chia-Hu Chang,Huei-Fang Yang,Greg C. Lee "Indexing and Teaching Focus Mining of Lecture Videos" 2009 11th IEEE International Symposium on Multimedia.

[12] Shouqun Liu Ming Zhu Quan Zheng "Mining Similarities for Clustering Web Video Clips" 2008 International Conference on Computer Science and Software Engineering.

[13] Lin Lin and Mei-Ling Shyu "Mining High-Level Features from Video using Associations and Correlations" 2009 IEEE International Conference on Semantic Computing.

[14] Tao Mei, Yu-Fei Ma, He-Qin Zhou, Wei-YingMa, Hong-Jiang Zhang "Sports Video Mining with Mosaic" Proceedings of the 11th International Multimedia Modelling Conference (MMM'05) 2005.

[15] A. Ekin, et.al"Generic Play-Break Event Detection for Summarization and Hierarchical Sports Video Analysis" IEEE International Conference on Multimedia and Expo, 2003, pp. 169-172.

[16] H. Pan, et.al "Detection of slow motion replay segments in sports video for highlights generation" in IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP'01),Salt Lake City,Utah,May 2001,vol.3, pp.1649–1652.

[17] V.Mezaris, et.al "Video object segmentation using Bayes-based temporal tracking and trajectory-based region merging" IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 6, Jun. 2004, pp. 782-795.