Video Event Mining and Content Management System Using Shot Ontology Description

DONG-LIANG LEE[1], LAWRENCE Y. DENG[2], YI-JEN LIU[3] and NICK C. TANG[3]
[1] Department of Information Management, St. John's University
[2]Department of Computer Science and Information Engineering, St. John's University
[3] Department of Computer Science and Information Engineering, Tamkang University
499, Sec. 4, Tam King Road, Tamsui, Taipei, Taiwan, R.O.C

Abstract: Since the mass growing amount of sports video has been produced, how to analysis and to make event mining in video content management issues are become more and more important. In this paper, we developed a shot ontology description based for the basketball video. Shot ontology is inferred by shot manipulations those included: 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.

Key-Words: Shot Ontology, Content management, Event Mining, Video Indexing, Video Weaving

1 Introduction

In Sports video, due to the hue amount of complexity and rich information, it is difficult to preview and query a user interest video segmentation. More and more researchers are addressed in segmentation, content summarization, annotation and indexing within the video and then to develop a video management system recently years[1] [2] [3]. Unlike the text retrieval technologies, video and the vioce content got less semantic definition. But with the growing network facilities and the mass storage tenologies, the related content-based indexing, retrieving, and managing technologies development become very important issues. Virage [4] (Virtual Information Retrieval Image Engine) provided the content-based image retrieval facilities, it was based on color(color layout, composition), texture, and the object boundary structure information to serve the visualization management. Virage not only provided image retrieval facilities but also served the video retrieval functions. QBIC [5] (Query by Image Content) was developed by IBM Almaden Research Center. It was the first image database retrieval system. It provided a color simularity comparison and was very suit for a scenic photo retrieving. Photobook[6] was proposed by MIT multimedia labortary. It contained three sub-book: Appearance book, Shape book, and texture book. It provided different retrieving algorithm and also got the more closely to different domaination. VisualSEEK [7] is developed by Image and ATV Lab of Columbia University. It provided the image and video query system by using

Besides the dynamic and static images, the vocal data included connote plentiful informtation, and it certainly need a perfect management mechanism. In order to solve the mining problem of the multimedia content, it was brung out the MPEG-4 and MPEG-7. Those were focused on the description and the definition of the content of video and audio. It was defined by the data structures and algorithms with elasticity, extensibility, multi-layers and straightforward. Through the format definiens of the MPEG series, a user could search, filter, and define the video/audio data efficiently. In the definition of the content of video sicne the realeasing of the MPEG-4. It described the data uint in each video frame with the video object plan basically instead of the frame itself. And the detailed definition for the contents of the video were focused on the varied low-level descriptors. In this paper, we would implement this objective with the characteristic value in various definition of the MPEG series.

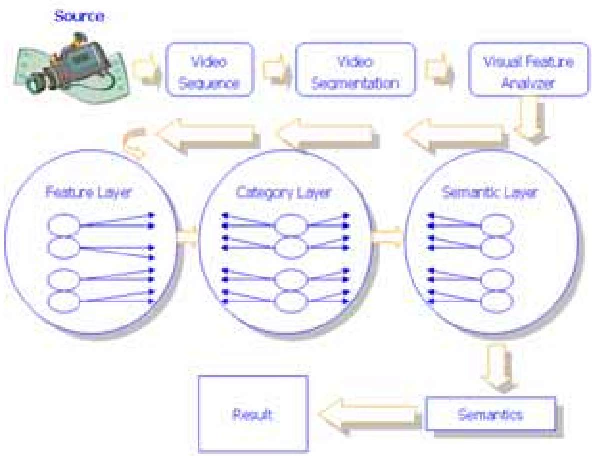

Figure 1: Video Content Management System Architecture

In this paper, we proposed a video content management system, that integrated three main components: Shot Ontology Definition, Feature Extraction, and Video Indexing component. In shot ontology component, we gathered statistics in basketball sport 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 final film indexing step, we marked the shot content and the event time on the time axis after the characteristics of the film was grabbed occasionally. Afterwards, the scoring events, the long shot and the close-up shot would be gragged with the Film Event Index System which was developed by our system.

The correlated research work would be explained and defined on chapter 2, including the introduction of the video file format, the definition of the ontology, and the related analysis of the video film. The chapter 3 would illustrate the system implementation and the methods for extracting features. And the chapter 4 came out a conclusion and the future research eventually,

2 The Video Indexing

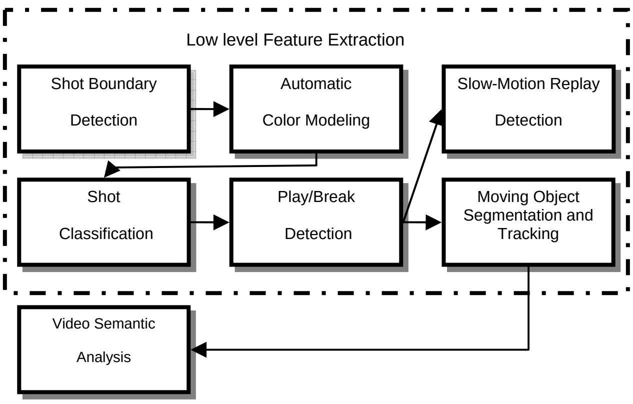

Figure 2. Abbreviation in basketball event

The attribute table of the shot ontology was the basic intension of the ontology. According to the attribute of that table, we could establish an shot ontology for the basketball sprot (refer to Figure 2 and 3).

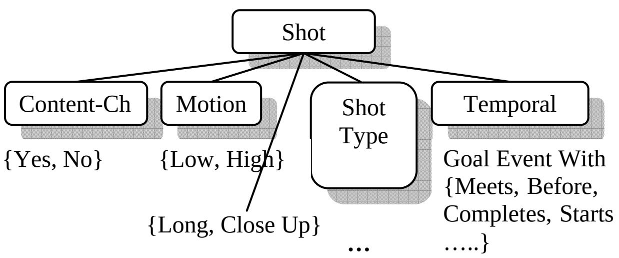

Figure 3. Shot Ontology

2.1 Shot Ontology

The major different between the sports program and the regular TV program was the scriptwriter, the process of a game and its result was unpredictable. That was the reason why the sports game was so attractive to the sport fans. When it comes to 'game', meaning competiting, there must be certain rules for game, and also there was necessary to have some buzzword to assist the definition while it occurred in a game course. Such as the essential of scoring was to throw the basketball into the basket frame in other court in a game, and the accumulation of the points in a whole game was the key factor to decide the winner eventually. We could classify the 'points' in the followins: 2-points, 3-points, free thows, and so on. Therefore, we could conduct a possible sequent order of a game with the varied buzzwords, and the process could be in a continuous series of dribbling, passing, three-pointing and goaling finally (as shown in Fig. 4).

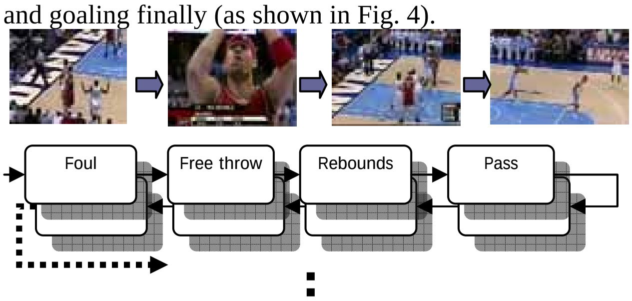

Figure 4. The possible sequent order of a game

●Merging of the events series

In a sprots TV program, 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 soccer 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 semantical representation consequently.

In this paper, 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 basketball game, such as free throw (ft), pass(pass), slam dun(dk) and so on. The narrative in basket ball was shown on table 1. Some other definition was indicated the unnecessary shooting frame in a game, such as the action of the coach, a close-up shot for a particular player or the spectators on the auditorium area, which we would name as 'ind'.

The abbreviation of a basketball game could be listed as Table 1, and the serials of the game course could be shown as Table 2.

Table 1. The Abbreviation in basketball game

Basketball Event	abbreviation
Jump ball	jb
Foul	ful
Goal	gol
Indifferent (don't care)	ind
Free throw	ft
Dribble	db
Pass ball to other	pass
Lay up	lup

Mid-range jumpers	mrj
Three pointer	tp
Dunk	dk
Rebounds	rb
Steal	st
Full timeout	stop
Long Shot	cmeL
Close Up Shot	cmeO

Table 2. The Serials list of a basketball game course

Serial	Time	Sequence	Type
1	0:02:34-00:02:35	jb	cmeO
2	0:02:43-00:02:45	jb	cmeL
3	0:03:28-00:03:42	db->pass->db->pass->db->pass->mrj->gol	cmeL
4	0:03:43-00:03:47	ind	cmeO
5	0:03:47-00:03:59	db->pass->db->mrj->rb->ful	cmeL
6	0:04:02-00:04:04	ind	cmeO
7	0:04:05-00:04:17	pass->pass->pass->pass->mrj->ful	cmeL
8	0:04:17-00:04:29	ind	cmeO
9	0:04:29-00:04:43	ft	cmeL
10	0:04:55-00:04:58	ind	cmeO

In addition to the statistics of shot time, we also aimed some narratives of the breaking event in a basketball 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 (refer to Table 3).

Table 3. The attribute table for the shot ontology

	gol	ful	stop	ft	ind
Motion	High	High	High	Low	-
Shot Type (cmeL/cmeO)	cmeL	cmeL	cmeL	Both	cmeO
SB Change (Yes/No)	Yes	Both	No	Yes	No
Time(s)	$>Tg$	-	-	-	$< Ti$

●Shot Classification

In a basketball game, there were much higher frequent goaling than a soccer game normally. The splendid pictures of a game would not only be a goaling picture for the auditorium. But the goal shot could comprised 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 be divided into 'Long Shot' and 'Close Up Sho' which was indicated as Figure 5 and 6.

Figure 5. Long Shot

Figure 6. Close Shot

There were many classifications for the Long shot and Close-up Shot theoretically[9,11,12,13]. We could classified 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 .

The valur of ∂ and σ were the experience value which could be adjusted by the accuracy. In this study, we got a best classication result by the setting of ∂ =25 amd σ =10. And the algorithm of the classification was descripted as following:

```
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
```

Since the classification was decided by the ratio of the skin color and the court color, we could define the descriptors for the shot ontology certainly(refer to Figure 7).

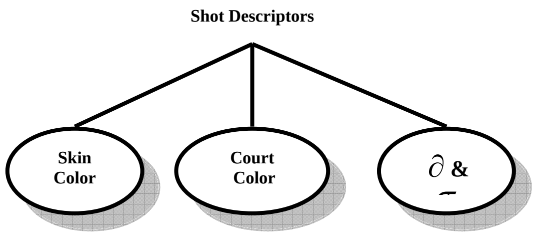

Figure 7. Descriptors of Shot classification

●Temporal Relation

After the compiling statistics result of four games, we got a event sequence list for the Long Shot and Close Up Shot whcih shown as Figure 8.

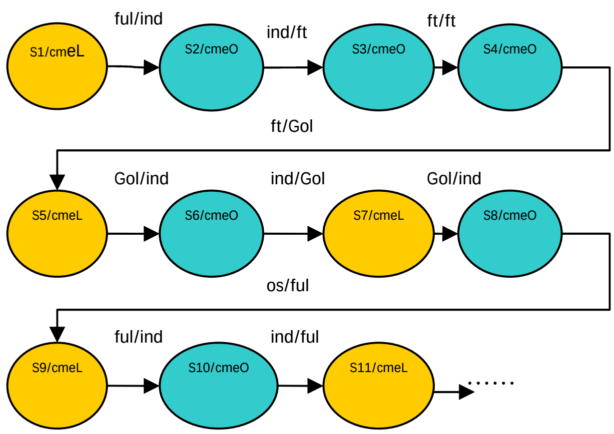

Figure 8. The sequence of Shot types and events

For example on the sequence S1 and S2: 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 indiffenent 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 simplifed completedly.

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 simplifed completedly.

For example, it could be simplified to same sequence if a Close Shot occurred after a previous Close Shot (refer to Figure 9).

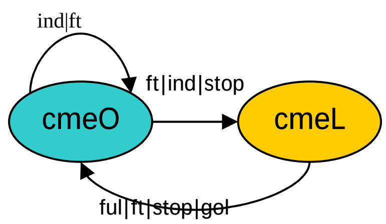

Figure 9. The simplified sequence of Shot types and events

It was more easy to realize a basketball game from the simplification of a sequence shot, such as the relations between Shot events and Shot types. Also, we could 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.

●Shot Indexing

The content indexing of a basketball game was the shot ontology itself. We had defined every shot descriptors through its low-level characteristics on the previous section. There were different authority offered by the shot ontology in the practical 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 durning 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 figure 10).

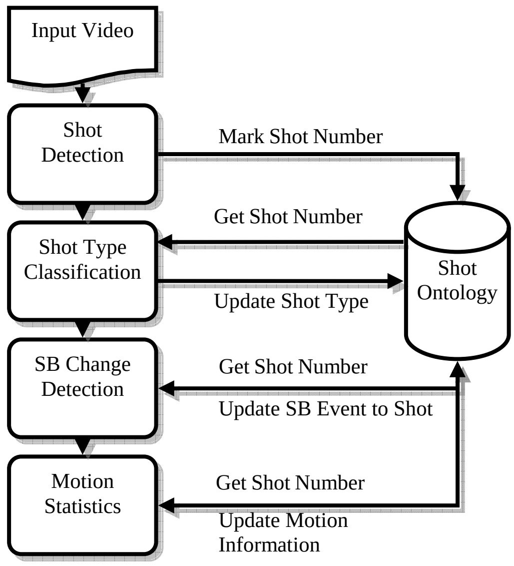

Figure 10. Shot Indexing Process

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 color 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, and picked the longest cut for the sample of court color. We also found that the video frames of regular game would be longer than the others which included the Close Up Shot, attractive sidelight activities or field goal. And the court color would take most portions on each game (Refer to Figure 11).

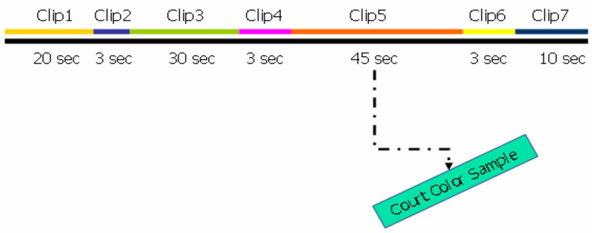

Figure 11. The longest time period for the Court Color Sample

```
for row = 1:row length
  for column = 1:column length
    pos1Hue = Hue(row,column);
    pos2Hue = find(TmpVote == pos1Hue);
    if pos2Hue > 0
      TmpVoteCount(pos2Hue) = TmpVoteCount(pos2Hue)+1;
    else
      j = j+1;
      TmpVote(j) = pos1Hue;
      TmpVoteCount(j) = TmpVoteCount(j)+1;
    end

    pos1intensity = intensity(row,column);
    pos2intensity = find(TmpVote2 == pos1instensity);
    if pos2intensity > 0
       TmpVoteCount2(pos2intensity) = TmpVoteCount2(pos2intensity)+1;
    else
      j = j+1;
      TmpVote2(j) = pos1instensity;
      TmpVoteCount2(j) = TmpVoteCount2(j)+1;
    end
```

3. We could get the value of court color through the Court Color Sample. This value would be used to decide the shot type from the proportion of skin color for each shot. And we took the first frame of each shot for the identified terms of shot type. Then we could update the table of shot detection after the shot type was collected continuously.

●Transformed RGB value to *IReBy*

(log-opponent)：

$I = L(G)$

$Rg = L(R) - L(G)$

$By = L(B) - [L(G) + L(R)]/2$

L(x) operation is defined as $L(x) = 105 * \log10(x+1)$

●We had set the window size of 8 and filtered out the noisy signals of *I*, and made a absolute value of the subtraction with the original *I* and filtered *I*. Afterward, we could get the most appropriate texture of skin. Finally, we could detect a suitable skin texture in our experiment when the value of *I* was greater or equal to 5 effectively

The transformation from IrgBy to HSV, we could capture the color range of skin with Hue and Saturation in our study. And the equation of transformation shown as folloings:

$$hue = (arctangent^2(Rg,By))$$

$$saturation = sqrt(Rg^2 + By^2)$$

skin map corresponds to a pixel

110<=hue<=180 and 0<=saturation<=130

3 Implementation Results

Table 4. The experiment value from test films

Video File	Total Clips	cmeL	cmeO	Goal
Cavaliers vs. Denver	476	146	330	109
Dallas vs. Houston.	492	151	341	111
Miami vs. Los Angeles	504	175	329	115
Detroit vs. Miami	498	156	342	107

We could evaluate the result of Long Shot and Close Up Shot with a detected Precision Rate experimentally. The efficiency of scoring segments detection was evaluated by the Precision Rate and Recall Rate simultaneously.

$$\Pr ecision = \frac{correct}{correct + false}$$

$$recall = \frac{correct}{correct + miss + false}$$

The *Correct* represented the number of correct detection, and the *False* meant the number of misjudgement. The Precision Rate could be estimated for the rate of misjudgement entirely. The *Miss* would be a score board change event without any detected actually. Therefore, the calculation model of *Recall* could be evaluted the accuracy of the scoring event detection. The detection result of Long Shot and Close Up Shot was shown on table4, and the detection result of scoring event was selected with the number of frame interval ω =5 and 10 which was shown on table4.

Table 5 The result of Precision Rate and Recall Rate

Video File	$\omega = 10$		$\omega = 5$	
	Precision Rate	Recall Rate	Precision Rate	Recall Rate
Cavaliers vs. Denver	86.3%	90.6%	89.7%	92.2%
Dallas vs. Houston.	81.4%	87.8%	86.9%	89.4%
Miami vs. Los Angeles	83.2%	89.2%	87.6%	90.7%
Detroit vs. Miami	70.3%	76.5%	85.7%	88.2%

Table 6. The detection result of scoring events

Video File	Precision Rate($\partial = 25$ and $\sigma = 10$)
Cavaliers vs. Denver	93.2%
Dallas vs. Houston.	92.3%

Miami vs. Los Angeles	85.5%
Detroit vs. Miami	93.8%

We also got an acceptable result of the accuracy in shot classification. There was a better Precision Rate and Recall Rate when we decreased the interval time between each frame on the detection of the scoring events. More or less, there was some more consuming time occurred on the calculation. Therefore, it was necessary to take a better choice on the balance of the time cost and the Precision Rate apparently.

4. Summary

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 basketball 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 had an excellent detection of the scoring event for the video frame in a basketball 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.

References:

[1] M.R. Naphade, et.al "A factor graph framework for semantic video indexing," IEEE Trans. on Circuits and Syst. for VideoTechnol., vol. 12, no. 1, pp. 40-52, Jan. 2002.

[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, pp. 145-148, Rochester, NY, Sept. 2002.

[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] J.R. Bach, et.al "The Virage image search engine: An oprn framework for image management," in Proc. SPIE Conf. On Vis. Commun,and Image Proc.

[5] M. Flicker, et.al "Query by Image and Video Content: The QBIC System," IEEE Computer, Vol.28, No9, pp.23~32, 1995

[6] A.Pentland, et.al "Photobook: Content-based manipulation of image databases," International Jounal of Computer Vision 1996.

[7] J. R. Smith and S,F, Chang, "Visual SEEK: a fully automated content-base image query system," Proceeding of ACM International Conference Multimedia, boston, MA, pp.87~98, November. 1996

[8] Blobworld, http://elib.cs.berkeley.edu/ photos/blobworld/

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