MINING INFORMATION OF ATTACK-DEFENSE STATUS FROM SOCCER VIDEO BASED ON SCENE ANALYSIS

Shuqiang Jiang[1], Qingming Huang[1, 2], Wen Gao[1,3]
[1](Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)
[2](Graduate School of Chinese Academy of Sciences, Beijing 100039, China)
[3](Institute of Digital Media, Peking University, Beijing 100871, China)
Email:{sqjiang, qmhuang, wgao}@jdl.ac.cn

ABSTRACT

Video Content is always huge by itself with abundant information. Extracting explicit semantic information has been extensively investigated such as object detection, structure analysis and event detection. However, little work has been devoted on the problem of discovering global or inexplicit information from the huge video stream. As an implementation in this topic, this paper proposes a solution to mining the statistical global attack-defense status information from soccer video by scene analysis. Semantic scene information of playfield detection, view classification, midline detection and global motion are extracted as the mid level information, and then they are fed into the finite state machine based status mining model to generate the statistical results, which will be of much usefulness for users. Experimental results reveal the feasibility of the method and more research work on the topic of discovering high-level inexplicit information from video are expected.

1. INTRODUCTION

Techniques to understand digital video content becomes a hot research topic in recent years. Their main focus is on structure analysis [1], event detection [2], highlight summary [3], et al. As video contents normally take a long period of time to play and occupy a large volume of bytes to store, mining useful statistical information from the whole video content to help users have a better understanding of it is a relatively difficult task and has not been fully explored yet. Compared with object-, structure- or event-based video semantic information, statistical mining information is hard to be acquired by human labeling.

Data mining is the process of finding useful patterns or extracting previously unknown knowledge from a massive set of data [4]. Different from textual information which has been studied for a long time, video contents have special characteristics: they are continuous sequences with temporal relations among them, and each video segment normally contains abundant information itself. Due to the semantic gap between human perception and computer-centered low-level features, many video understanding techniques such as shot boundary detection and event extraction remain open problems. Thus investigation on video mining is at its early stage as mining solutions normally need the extracted semantic information. In fact, the current status of video mining is still at the pre-processing stage, such as video clustering [5]. The main motivation of video mining is to find undiscovered knowledge from the stream based on visual and audio cues. The knowledge may typically include structure information within a video clip or association information among various clips, as well as trend information based on the analysis for a massive size of video set. Xie and Chang [6] provide a solution to find temporal structures of sports video by an unsupervised multi-scale statistical Hidden Markov method without employing domain knowledge. To find topic associations among news videos, the authors in [7] propose a tool called “GeoPlot” to provide the information of correlation of news events occurrence. Mining undiscovered video information is usually based on apparent semantic concept extraction techniques in video clips, such as object concept, scene concept and event concept. In the authors’ mind, video mining need a long way to go to explore the full potential of this topic as most of the apparent video understanding problems are still unsolved yet.

In this paper, we propose a solution to discover the global attack-defense information of soccer video based on scene analysis. In recent years, extensive research efforts [8] have been devoted to sports video (especially soccer video) content analysis and applications due to their wide viewership and high commercial potentials. Roughly three levels of information are involved within soccer video analysis: (1) object level [9] - such as object detection and tracking; (2) scene level [10] - such as scene classification and clustering; (3) event level [11] - such as event detection and highlight summary. While match status analysis is also very useful for users, especially longtime sports fans and professional sports person, this work will investigate on the problem of finding global statistical match status information of

This work was supported by National Hi-Tech Development Program (863 Program) of China under Grant: 2006AA01Z117 and National Hi-Tech Development Program (863 Program) of China under Grant: 2006AA01Z315

soccer video that is not as explicit as object, scene or event information. The flowchart of the proposed framework is illustrated in Fig.1. Given an input soccer video clip, playfield detection is first performed on each frame to extract the playfield regions as the fundamental cue for further analysis. Supervised pattern classification method is used to categorize each frame into defined views such as global views, local views, infield views and outfield views, etc. On each of the infield global view frames, midline is detected by Hough transformation and region analysis. Then view type information, midline information as well as the global motion information is input to the Finite State Machine (FSM) based attach-defense mining engine to generate the final statistical match status information.

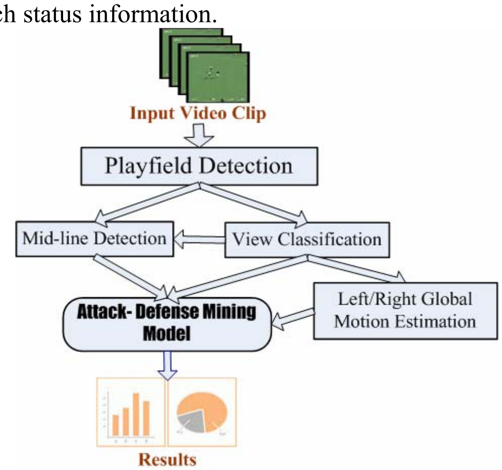


Figure 1: Flowchart of the proposed attack-defense mining scheme

The proposed mining scheme is based on the results of basic scene analysis. Compared with other video types, soccer video has relatively better structure and more domain knowledge, so research effort has been devoted to this kind of video and more semantic scene information could be extracted from it. We could use this semantic information for higher level analysis and make mining undiscovered statistical information possible.

2. SCENE ANALYSIS AND MID-LEVEL INFORMATION EXTRACTION

In this section, some scene analysis techniques will be discussed, including playfield detection, semantic view classification, mid-line detection and global motion estimation.

2.1. Playfield Detection

Playfield plays the fundamental role in analyzing soccer video as it conveys basic information of the match course. The color of playfield is generally uniform and playfield often occupies dominant color region. We use this knowledge to extract playfield by applying the Gaussian Mixture Model (GMM) based method [9] to adapt to various kinds of playfield.

2.2. View Classification

For each of the soccer frame, we can semantically assign a view label (VL) to it. Based on the playfield segmenting result, we use a framework of the hierarchical soccer view classification [11], which includes the following views: infield views and outfield views, global views, local close-up views, etc.

2.3. Mid-line Detection

For each categorized global view soccer video frame, we use a white pixel detection accompanied with Hough transformation method to find middle lines. Given the global view frame I, we first segment the playfield regions based on the playfield detection results by region analysis method. Non playfield pixels such as players and lines are integrated to the region R_{IPF}, which will be the next-step processing target. Then white pixels $wp(R_{IPF})$ in R_{IPF} are detected in this region by the following equation:

$$wp(R_{IPF}) = \{val(p) > 1.4 \times \frac{\sum_{q \in I} val(q)}{N} | p \in R_{IPF} \}\}$$

where p,q are pixels in the frame, $val(p)$ is the value of p in V channel in HSV color space. Hough transformation is performed as the last step of middle line detection. Experimental results on several soccer video clips are satisfactory to fulfill further tasks in this work.

2.4. Global Motion Estimation

In a soccer video stream especially for global views, global motion normally represents the attack direction in the match. Global motion is calculated by estimating model parameters including motions in three aspects: zooming, rotation and translation, and they are obtained by standard global motion estimation algorithm from the MPEG compressed content. We extract the left/right translation parameters to analyze the attack-defense status.

3. ATTACT-DEFENSE MINING MODEL

3.1. Attack-defense analysis in soccer video

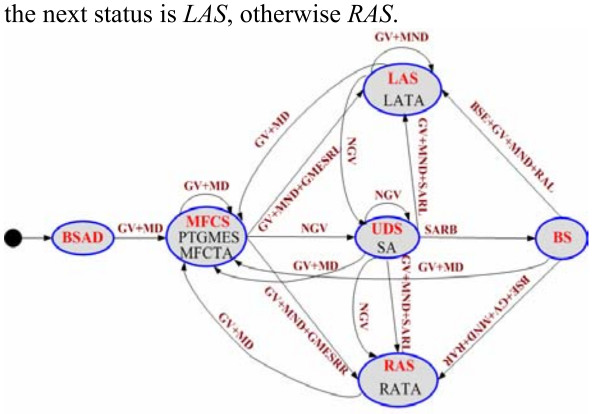
Compared with the event detection and highlight summarization, attack-defense analysis in soccer is also important, The broadcast soccer video stream has the following characteristics: (1) Most of the video shots are global views to make the audience obtain more information of the match; (2) Video camera only captures the important match position, and the motion of camera always follows the motion of the ball; (3) The midfield is an important area that the two teams want to take possession of, and the spatial transition of this area always represents the change of the attack-defense status. Based on the above observations, we perform the statistical analysis of attack-defense mining by an FSM method.

3.2. The proposed mining model

According to the above analysis, we apply the extracted scene information including view labels, detected midlines and global motions to construct the mining model, which is based on Finite State Machine (FSM). As illustrated in Fig. 2, the first state is to begin with statistic attack/defense. Three important statuses are middle field contest (*MFCS*), left attack (*LAS*, the right team are in the attacking status and) and right attack (*RAS*). The other status is break time with game course stopping, and this is not the main focus in this work because our goal is to analyze the attack and defense in the match. As the left attack and right attack are symmetrical, only the left attack status is discussed here. For each soccer game, the beginning time is at the middle field area. So the model next moves to the state of *MFCS* and accumulates the *MFCS* time. As the camera motion may change from left->right to right->left, the global motion, in the mean time, the sequential global motion estimation is analyzed and computed as the past time global motion estimation accumulation (*PTGMEA*) as an influential factor to decide what the next status will be when *MFCS* stops. If the global views are detected and middle line is not detected, the status moves to *LAS* or *RAS* according to the global motion estimation statistical results, which are decided empirically by the following equations:

$GMESR = 0.2*PTGMEA + 0.8*CGME$

where *CGME* represents the current left/right GME value from the frame of middle line detected to the frame of middle line not detected. If the results are left (*GMESRL*), the next status is *LAS*, otherwise *RAS*.



BSAD: *Begin Statistic Attack/Defense;*
MFCS: *Middlefield Contest Status;*
LAS: *Left Attack Status;* ***UDS***: *Undecided Status;*
RAS: *Right Attack Status;* ***BS***: *Break Status;*
PTGMES: *Past Time Global Motion Estimation Accumulate*
MFCTA: *Middlefield Contest Status Accumulate*
LATA: *Left Attack Time Accumulate*
RATA: *Right Attack Time Accumulate*
SA: *Status Analysis;* ***GV***: *global Motion;*
MD: *Middleline Detected;*
MND: *Middleline Not Detected*
GMESRL: *Global Motion Estimation Statistic Result Left*
GMESRR: *Global Motion Estimation Statistic Result Right;* ***SARB***: *Status Analysis Result Break*
BSE: *Break Status End;* ***RAL***: *Region Analysis Left*
RAR: *Region Analysis Right*

Figure 2: The proposed statistical mining model

When the frame views are not global views, the model moves to an undecided status and begin to analyze what the status will be according to the available scene clues. Due to space limitation, the detailed discussion of the mining model will not be provided. Fig. 2 provides more information.

The input of the mining model is the temporal sequential video frames, and the output is the statistical match status information to tell at what time which team is in the attack/defense status and the global statistical data of the whole match.

3.3. Discussion

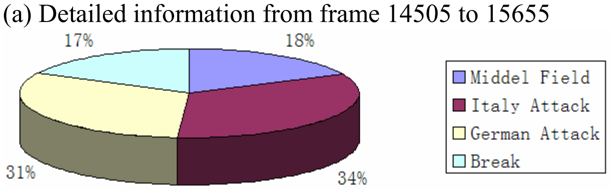
As we know, when people watch TV to enjoy soccer games, lots of information is available to help them judging the status of a match besides the scene information discussed in this paper, and human brain is strong enough to understand the match with its past experience and personal likes and dislikes. This proposed model will use computer power to acquire the global match information by applying the mid-level semantic clues, which the soccer audiences are difficult to acquire and will be of much help for them to sum up the game. Based on the limitation of current techniques, many semantic extraction methods from image or video streams are not very satisfactory. The scene semantic detection methods used in this model are relatively robust by taking the full advantage of domain knowledge in the soccer area. Experimental results will be provided in next section to validate the reliability of the method.

4. EXPERIMENTAL RESULTS

In this implementation, samples for playfield training are 100 frames from the first 100 second with one frame per second. Every five frame are selected to perform view classification and midline detection, as well as global motion estimation. This could save the processing time without interfering with the mining result, as 5 sequential frames are normally very similar and are only 1/5 second duration. Based on the above discussed mining model, information of statistical attack-defense status is accumulated from the beginning frame to the end. There are four statuses in this work: middle field time, team A attack, team B attack and break time. Experimental results of the Italy-German 1/4 final match in FIFA 2006 are illustrated in Fig. 3.

In Fig. 3(a), a detailed illustration from frame 1455 to 15655 is provided, which is accordance with the ground truth observation. Fig. 3 (b) provides the global statistical information of the whole half match. In this half match, the score result is 0:0. After the game, a famous soccer commentator said: In the first half of the mach, the struggle of Germany to control the middle field is inferior to Italy, so they want to keep stable to reduce the attacking of Italy. And the ball control time of Italy team is longer. It can be observed that mining results of this work approves with the above comments and it could be of more help to the users to understand the game. We have tested the method on several other matches in FIFA 2006, and the results are satisfactory. The following problems exist to affect the mining results: (1) The break time analysis is not very accurate (*SARB*); (2) Replay is not incorporated to the mining scheme. Future work will deal with the problems to make the analysis results more reliable.

(a) Detailed information from frame 14505 to 15655



(b) Statistical results of attack-defense status of the first half match of the game

Figure 3: Experimental results in the first half match of the Italy German 1/4 final match in FIFA2006

5. CONCLUSION

In this paper, we propose a mining scheme for soccer video to obtain statistical attack-defense status based on the result of scene analysis such as view classification, midline detection and global motion estimation. A finite state machine method is used to perform the mining task. Some information has not been included in the model such as the replay shots which will affect the final results. This problem is our next step work in the near future. Besides, more research work on the topic of discovering high-level inexplicit information from various kinds of video types is expected.

6. ACKNOWLEDGEMENT

This work was supported in part by National “242” project (2006A09), “Science 100 Plan” of Chinese Academy of Sciences (99T3002T03), and the Natural Science Foundation of Beijing, China (4063041).

7. REFERENCES

[1] Hong Lu, Beng Chin Ooi, Heng Tao Shen, Xiangyang Xue, Hierarchical Indexing Structure for Efficient Similarity Search in Video Retrieval, IEEE Transactions on Knowledge and Data Engineering, Volume 18, Issue 11, pp. 1544-1559 Nov. 2006

[2] D.A. Sadlier,N.E. O'Connor, Event detection in field sports video using audio-visual features and a support vector Machine, IEEE Transactions on Circuits and Systems for Video Technology, Volume 15, Issue 10, pp:1225 - 1233 Oct. 2005

[3] Ying Li, Shih-Hung Lee, Chia-Hung Yeh,Kuo C.-C.J., Techniques for movie content analysis and skimming: tutorial and overview on video abstraction techniques, IEEE Signal Processing Magazine, Volume: 23, Issue: 2, pp: 79- 89, Mar. 2006

[4] JungHwan Oh, JeongKyu Lee, Sae Hwang, Video Data Mining: Current Status and Challenges. Encyclopedia of Data Warehousing and Mining. (A book edited by Dr. John Wang). Idea Group Inc. and IRM Press. 2005

[5] Wallapak Tavanapong, Junyu Zhou, Shot Clustering Techniques for Story Browsing, IEEE Transaction on Multimedia, Vol.6, No.4, pp517-527, August 2004

[6] L. Xie, S.-F. Chang, A. Divakaran and H. Sun, Unsupervised Mining of Statistical Temporal Structures in Video, in Video Mining, Azreil Rosenfeld, David Doremann, Daniel Dementhon Eds, Kluwer Academic Publishers, 2003

[7] J. Pan and C. Faloutsos. GeoPlot: Spatial Data Mining on Video Libraries. Proc. Int'l Conf. Information and Knowledge Management, pp.405~412, 2002.

[8] Anil Kokaram, Niall Rea, Rozenn Dahyot, A. Murat Tekalp, Patrick Bouthemy, Patrick Gros, and Ibrahim Sezan, Browsing Sports Video: Trends in sports-related indexing and retrieval work, IEEE Signal Processing Magzine, pp.47-58, March, 2006

[9] Shuqiang Jiang, Qixiang Ye, Wen Gao, Tiejun Huang. A New Method to Segment Playfield and Its Applications in Match Analysis in Sports Video. ACM Multimedia (ACM MM 2004), New York, NY USA, Oct.10-16, 2004

[10] Ling-Yu Duan, Min Xu, Qi Tian, Chang-Sheng Xu, Jin J.S., A unified framework for semantic shot classification in sports video, Volume: 7, Issue: 6, pp. 1066-1083, Dec. 2005

[11] Qixiang Ye, Qingming Huang, Wen Gao, Shuqiang Jiang. Exciting Event Detection in Broadcast Soccer Video with Mid-level Description and Incremental Learning. Proceedings of the 13th ACM International Conference on Multimedia (ACM MM 2005), Singapore, pp455-458, Nov.06-11, 2005