SHOT AND SCORING EVENTS IDENTIFICATION OF BASKETBALL VIDEOS

Chung-Lin Huang, Huang-Chia Shih, and Ching-Lun Chen
Department of Electrical Engineering
National Tsing Hua University
HsinChu, Taiwan, R. O. C.
clhuang@ee.nthu.edu.tw

ABSTRACT

This paper presents a novel semantic-oriented video analysis system for the basketball game videos. Based on Bayesian Belief Network (BBN), it may bridge this gap between the low-level features describing image/video structure and the high-level knowledge. We apply the Support Vector Machine (SVM) to identify and track the ball, the shooter, and the basket as the low-level features. Based on these features, our BBN framework can identify four categories of shot event such as short shot, medium shot, long shot, free throw, and the scoring event. In the experiments, we demonstrate that our system may interpret the video shots in terms of four different shot events and one scoring event effectively.

1. INTRODUCTION

Multimedia information retrieval has become one of the major research topics for entertainment. Understanding the video content is major issue for accessing the highly prevalent multimedia information. Sports video happens to be one of the most popular ones. Xu *et al.* [1] presents a basketball event detection method by using multiple modalities. Instead of using low-level features, the proposed method is built upon visual and auditory mid-level features, i.e. semantic shot classes and audio keywords. Nepal *et al.* [2] identified goal segments in a basketball video using five temporal goal models which are constrained by the observation of crowd cheer, scoreboard display and change in direction.

Many model-based approaches have been proposed for event identification in sports video. Sports event detection based on Hidden Markov Model (HMM) has been proposed in [3]. Shih *et al.* [4] employed the Multi-level Semantic Network (MSN) which is Bayesian architecture for sport video content understanding and analysis provides a potential tool for accessing and browsing video database on a semantic basis. Chao *et al.* [5] proposed a Dynamic Bayesian Network (DBN) framework to extract the occurrence of highlights.

To detect and track the ball is the major issue of most of the sports video analysis. Some researchers desire to obtain the position of the ball for soccer video [6]. It can play a crucial role in analyzing video structure, ball possession, event detection, etc. Here, we apply a support vector tracking (SVT) [7] to track ball, the player, and the basket net in the basketball spots video. SVT integrates the SVM classifier [8] with an optical-flow-based tracker. Instead of minimizing an intensity difference function between successive frames, SVT maximizes the SVM classification score. Based on the low-level evidence, we may use the BBN to find the high-level semantics. Our system can identify

four different shot events, *i.e.*, short shot, medium shot, long shot, and free throw, as well as the scoring event. Finally, we will show some experimental results to illustrate the effectiveness of our system.

2. EXTRACTION OF GAME-SPECIFIC FEATURES

In the basketball video, the basic semantics consists of locations of the ball/shooter/basket. We detect five kinds of basketball-game-specific feature including ball released, ball-ring contact, basket-orientated ball moving direction, ball inside basket net, and camera motion. The flow diagram of preprocessing and feature extraction is shown in Fig. 1.
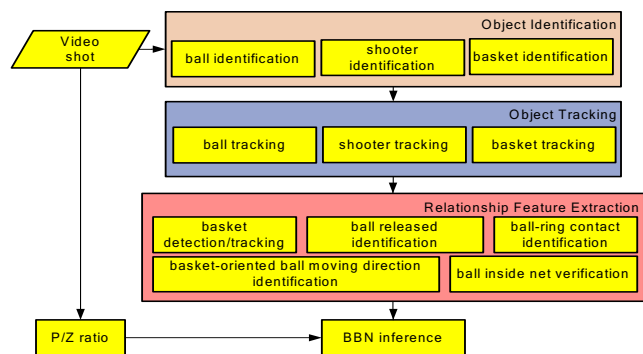

Figure 1. The flow diagram of preprocessing.

2.1. Object Tracking and Identification

First, we apply SVM classifier to identify the ball/shooter/basket and then use the SVT to track the shooter, the ball, and the basket in basketball videos. We detect the basket by using edge information, and track the human/ball by using the color histogram and luminance feature. We employ a coarse-to-fine processing in detection and tracking the objects.

a) Ball detection and tracking

Ball detection is a coarse-to-fine approach. Coarse-processing searches for the possible location of the ball in a larger spatial area, whereas the fine-processing finds the location of the ball within a smaller search area. Once the tracking fails, it can be recovered by next coarse-to-fine process. In every frame, we apply the pixel-wise ball tracking by checking the SVM score of neighboring candidates of the identified bounding circle. Fig. 2(a) shows the SVM score distribution, whereas Fig. 2(b) indicates the tracked ball from the initial position of previous frame (dashed bounding circle) to the best position of current frame (solid bounding circle).
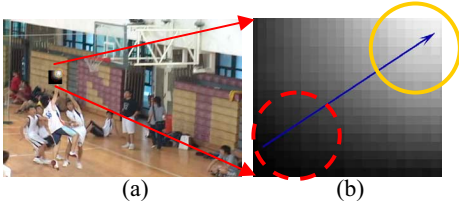
(a)   (b)
Figure 2. SVM score distribution

b) Shooter Identification and tracking

In basketball game video, the players in the two different teams are always dressed in two different colors. We pre-trained two distinguished SVM model by the color histogram of the clothes as the feature vector. Once the SVT locates the ball, it may continue searching the area around the ball to find the coarse location of the shooter. To identify the size of the shooter, we assume the size of image of the shooters can be quantized into 10 different scales and the aspect ratio=0.3. Once the location of the shooter is identified, we use the fine-processing to track the shooter by using color-based SVM tracking.

c) Basket detection and tracking

We find that the edge information of basket is obviously different from background region such as audience, player, and field etc. With the corresponding edge map, we rescale the bounding box to normal size as the input feature for SVT. Due to the basket appears in the upper half of the image frame, we only check the candidate positions in the upper half of the image frame using coarse-processing (see Fig. 3(a)). To find the best position of basket, we apply the fine-processing around the located bounding box in the previous frame and search for the best SVM score. The tracking result is shown in Fig.3.
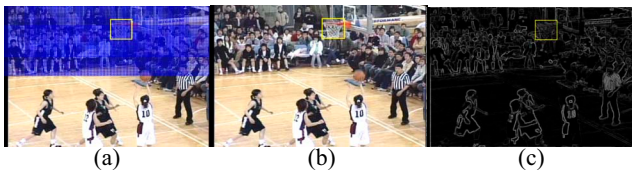

(a)   (b)   (c)
Figure 3. (a) the search range of the basket detection. (b) detected the basket. (c) the edge information of (b).

2.2. Relationship Feature extraction

Based on the information of the identified ball, shooter, and basket, we apply the SVM to find the relationships between them, such as the ball release, and the ball-ring contact etc.

a) Ball released Detection

Due to camera zooming, the appearance of the player changes very often. The different height and arm length of the players makes the timing estimation of the ball release a nontrivial problem. To identify whether the ball is being released, we compare the ball-shooter distance (h_1) with the product of shooter's height (h_2) as shown in Fig. 4.

Here, we find certain relationship between h_1 and h_2, (i.e., $r_1=h_1/h_2$), when the player is start to release the ball. Once $h_1> r_1h_2$, the shooter is releasing the ball, and the shot event may be occurring. Here, r_1 will be achieved in off-line training. Through the training sequence of about 70 video clips of ball-releasing, we find the average ratio $r_1 = 0.675$.
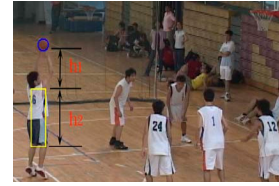

Figure 4. h_1 is ball-shooter distance, r_1 is ratio-1, h_2 is height of clothing of shooter.

b) Ball-ring contact Detection

We identify ball-ring contact by computing the ball-ring distance between the ball and the basket. We may locate the ball and basket by using SVT. When the bounding circle of ball and the bounding box of basket are very close, the ball-ring contact is occurs. As shown in Fig. 5, the distance between of the ball and the basket is calculated by computing: (1) the vertical distance between the upper bounding of the basket and the lower boundary of the ball as the ball-ring distance Y, and (2) the horizontal distance between the left/right boundary of the basket and the right/left boundary of the ball as the ball-ring distance X.

Figure 5. ball-ring distance

c) Basket-oriented ball moving direction Identification

After the ball-ring contact detection, the bouncing direction of the ball is usually unpredictable. We have to identify the moving direction of the bouncing ball relative to the basket. When the moving direction of the bouncing ball relative to the basket is downward, we may infer the video shot as a scoring shot. The results are illustrated in Fig. 6.

(a)   (b)
Figure 6. (a) the bouncing ball moving downward, (b) the bouncing ball moving sideward or upward

d) Ball inside net Verification

For a scoring shot, ball is enveloped by the basket net and retained inside for a couple of frames. If the bounding circle of ball is overlapped with the basket which is enclosed by a rectangle such as Fig. 7 (a) and (b), we find that the ball is either inside basket net or outside basket net. We distinguish the two cases by using luminance information and SVM. The SVM score is termed *in/outside score*. When the in/outside score is positive, it is a scoring shot, otherwise, it is not a scoring shot.
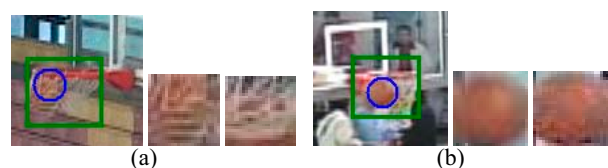

(a)   (b)
Figure 7. The ball-net in/outside identification. (a) scoring shot;(b)

non-scoring shot. The in/outside score respectively is 3.106 and -3.256 in (a) and (b).

2.3. Panning/Zooming Ratio

Similar to [9], we calculate the camera motion by analyzing the characteristics the vertical and horizontal projections of the m×n picture elements {p(i, j)}. First, we calculate the vertical projection and the horizontal projection as $f_x(i)=\Sigma_j p(i, j)$, and $f_y(j)=\Sigma_i p(i, j)$, where $i=1,...m$ and $j=1,...n$. Then, we divide the projection into small slices with N pixels width. For two consecutive frames a and b, we take a slice of frame a and slide it over frame b to find the best match. The sliding distance indicates the motion vector.

We calculate the *sum of absolute difference* (SAD) value for frames a and b with the displacements s_x and s_y in the vertical and horizontal directions to interpret the global horizontal motion of the video caused by camera panning or zooming. Once s_x and s_y are small enough, the camera is stationary else the camera motion is identified. The number of frames with motion is defined as panning/zooming ratio that is,

$$P/Z \ ratio = \frac{the \ number \ of \ frames \ with \ motion}{total \ number \ of \ frames}$$

In empirical, the P/Z ratio<0.3, then the camera is stationary. This information can be used to determine a free throw.

3. EVENT ANALYSIS USING BBN

Bayesian Belief Network (BBN) has been proved to be an effective statistical model for knowledge representation and inference. Similar to [4], we propose a semantics framework for the basketball video understanding. We apply BBN to identify the specific events of the basketball video and the corresponding semantics.

3.1. BBN for Shot event

As shown in Fig. 8, the shaded nodes represent the input evidences. When the system identifies *ball released* is true, it may determine the node of *ball flying period* for inferring category of the shot event. The *ball flying period* is a mid-level feature node which is influenced by three nodes: *ball released, ball-ring contact* and *frame count*. *Ball flying period* node represents the time period (in frame numbers) from the instance of *ball released* to *ball-ring contact*. When our system identifies the occurrence of *ball released* and *ball-ring contact*, *Frame count* is triggered which causes the *ball flying period* to change from one state to other state until the *ball-ring contact* occurs.
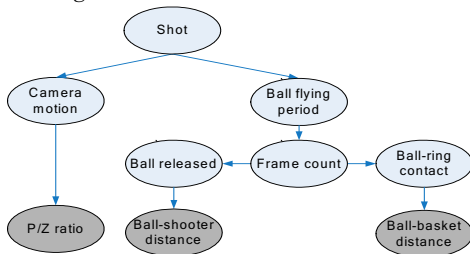

Figure 8. BBN structure of the shot event.

3.2. BBN for Score event

After the shooter release the ball, the occurrence of ball inside the basket net means the scoring shot probably happen. If a scoring shot happens, the ball moving direction is always downward, and the ball is always enveloped by basket net for a while.
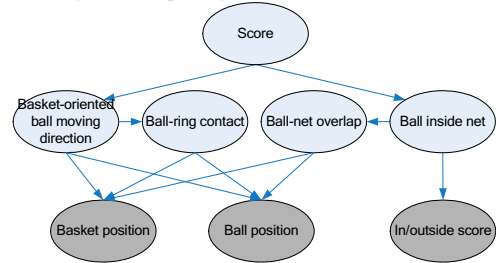

Figure 9. BBN structure of the scoring event.

The occurrence probability inference of *basket-oriented* b*all moving direction* node and *ball inside net* node for each frame are not activated by low-level evidence only. The hypothesis of all moving direction activated relies on the other hypothesis of *ball-ring contact*. Once *ball-ring contact* is activated, then the inference of the *basket-oriented* b*all moving direction* will proceed. Similarly, the inference process of *ball-inside-net* is activated only when b*all net overlapped* occurs. The hypothesis of b*all-inside-net* is not based on the b*all-net inside/outside score* only. Fig. 9 illustrates the scoring event detection.

3.3. Training Process

The training procedure can be divided into two phases. In the 1st training phase, we heuristically determine the causality between the root node and the hidden node. In the 2nd training phase, we find the conditional probability of the existing linkage between the evidence nodes and the hidden nodes.

Four events will be taken into account in *shot node* that is free throw, short shot, medium shot, and long shot. Since the four states are defined mutual exclusive. For example ,once *Ball inside net*=Y, we can count the number of times of the joint appearance of positive *In/outside score* and *Ball inside net*="Y" to determine the conditional probability as

$$P(\ In/outside \ score > 0 | \ Ball \ inside \ net = Y)$$
$$= \frac{P(\ In/outside \ score > 0 , Ball \ inside \ net = Y)}{P(Ball \ inside \ net = Y)}$$

3.4 Understanding Phase

Given the conditional probability and the prior probability for each node, we can interpret the video shots. The video shots are processed by several video analyzers such as shooter/ball/basket position using SVT, motion analyzer, in/outside score using SVM classifier. These analyzers extract the lowest-level features as the input evidence to BBN. We use the evidence propagation procedure from the low-level features to the relationship features (*e.g.*, ball flying period etc.) to identify the video shot events and scoring events.

4. EXPERIMENTAL RESULTS

Our experimental video data are selected from the video scene captured by the main camera. The video sequences of five basketball games of the NTHU-NCTU campus competition are segmented into many video shots [10] from which we select about 157 video shots from these scenes. There are five basketball matches consists of two female basketball matches and three male basketball matches. The precision rate and recall rate of testing for

each class are shown in Table 1. Besides, the node of the score event consists of two states, Y or N, indicating the scoring shot or non-scoring shot. The testing results are shown in Table 2. The ground truth video shots are assigned manually. We demonstrates the results by using the precision and the recall defined as follows,

$$Precision = \frac{number\ of\ correct\ detection}{number\ of\ correct\ detection + number\ of\ false\ alarm}$$

$$Recall = \frac{number\ of\ correct\ detection}{number\ of\ correct\ detection + number\ of\ miss}$$

In Figure 10, we represent a user interface of shot event and scoring event identification system. The system starts identifying the event once the ball is released, and stops after ball ring contact. Then, it will also identify whether the scoring event occurs. The short shots may be miss-identified as medium shots because the shooters are standing on the boundary. The ball flying period node can not be used differentiate their difference effectively. Especially, the bank shot usually causes a longer ball flying period, and the trajectory of the basketball will also complicate the information of ball flying period. Higher curvature trajectory indicates longer ball flying period. The accuracy of identifying the medium shots is the worst one among all. The medium shot may be miss-identified as a short shot, a long shot, or a free throw.

Table 1. Understanding results for category of shot events.

Video shot	Ground Truth	Correct	Miss	False Alarm	Recall Rate	Precision Rate
Short Shot	34	31	3	1	91.2%	96.9%
Medium Shot	20	15	5	4	75%	78.9%
Long Shot	15	13	2	4	86.7%	81.3%
Free Throw	18	16	2	3	88.9%	84.2%
Total	87	75	12	12	86.2%	86.2%

Table 2. Understanding results of score events.

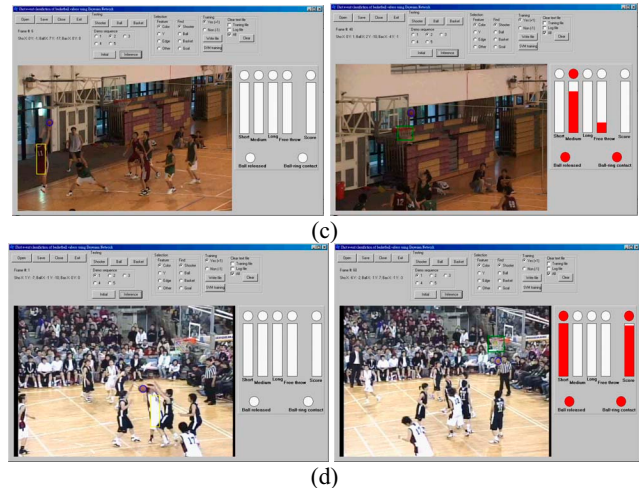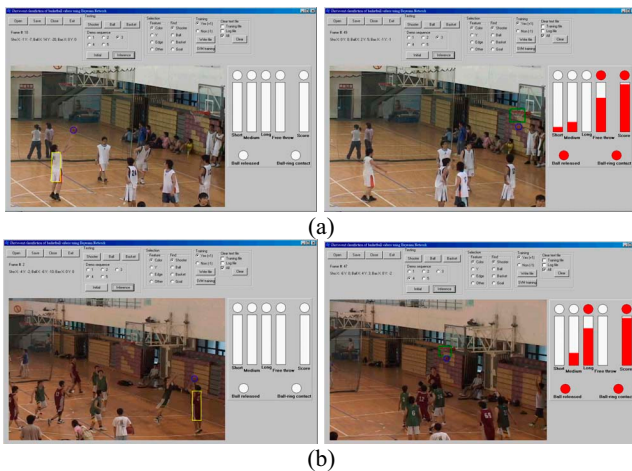| Video shot	Ground Truth	Correct	Miss	False Alarm	Recall Rate	Precision Rate
Score	35	31	4	3	88.6%	91.2%
Non-score	52	49	3	4	94.2%	92.5%
Total	87	80	7	7	92%	92%

(a)

(b)

(c)

(d)

Figure 10. Four different identified shot events: (a) Free throw; (b) Long shot; (c) Medium shot; (d) Short shot.

5. CONCLUSIONS

We have proposed a novel basketball video shot identification system based on BBN. The main contribution of this paper is to develop an inference system consisting of some linkages between unobservable concepts and observable concepts. We integrate the feature extraction with inferable BBN to fill the gap between the low-level visual domain and the high-level semantic classes. In the experiments, we have proved that our system can identify the shot events and scoring events effectively.

REFERENCES

[1] M. Xu, L.Y. Duan, C.S. Xu, M.S. Kankanhalli, and Q. Tian. "Event Detection in Basketball Video using Multiple Modalities," Proc. IEEE-PCM'03, Singapore, Dec. 2003.

[2] S. Nepal, U. Srinivasan, and G. Reynolds, "Automatic detection of goal segments in basketball videos," Proceedings of ACM Multimedia Conf. on Authoring Support, Oct. 2001.

[3] G. Xu, Y.-F. Ma, H.-J. Zhang, and S.Q. Yang. "A HMM based semantic analysis framework for sports game event detection," Proc. IEEE-ICIP'03, pp. 25-28, Sept. 2003.

[4] H. C. Shih and C. L. Huang, "*MSN*: Statistical Understanding of Broadcasted Sports Video Using Multi-level Semantic Network," IEEE Trans. on Broadcasting, pp. 449-459, Dec. 2005.

[5] C. Y. Chao, C. L. Huang, and H. C. Shih, "Semantic-based Highlight Extraction of Soccer Program Using DBN," Proc. IEEE-ICASSP 2005.

[6] X. Yu, Q. Tian, and K. W. Wan. "A novel ball detection framework for real soccer video," Proc. IEEE-ICME'03, pp. 265-268, 2003.

[7] S. Avidan, "Support Vector Tracking," IEEE Trans. on PAMI. Vol. 26, No 8, pp. 1064-1072, 2004.

[8] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.

[9] M. K. Kim, E. Kim, D. Shim, S. J, and G. Kim, "An Efficient Global Motion Characterization Methods for Image Processing Application," IEEE Trans, on Consumer Electronics, Vol. 43, No. 4, Nov. 1997.

[10] C. L. Huang and B. Y. Liao, "A robust scene-change detection method for video segmentation," IEEE Trans. on CSVT, Vol.11, No. 12, pp. 1281 -1288, Dec. 2001.