SEMANTIC UNDERSTANDING AND SHOT EVENT ANALYSIS FOR BASKETBALL VIDEO

Huang-Chia Shih(施皇嘉), Ching-Lun Chen (陳慶倫), and Chung-Lin Huang (黃仲陵)
Department of Electrical Engineering
National Tsing Hua University
HsinChu, Taiwan, R. O. C.

ABSTRACT

In this paper, we attempt to bridge the gap between low-level feature and high-level semantic event through the integration of image/video analysis algorithms with multi-level Bayesian Belief Network (BBN), and demonstrate how we can be effectively applied for fusing the evidences obtained from different video sources. Support Vector Tracking (SVT) is applied for ball/shooter/basket tracking. SVM classifier and camera motion analysis combined with low–level feature extract algorithms are applied to extract mid-level features from the video which treat as the input to the BBN. We have proposed a novel video shot classification system based on low-level features extraction. The semi-automatic semantic system is designated for the basketball game videos. Given the video shots of basketball game, the system can identify four categories of shot event such as short shot, medium shot, long shot, free throw, and the score event.

I. INTRODUCTION

Recently, a proliferation of digital media including image, audio, video, streaming video clips, panorama images and 3D graphics have been created. We need a flexible and scalable way to manage the mass media of which the digital video has been widely accepted as the most accessible media. Understanding the video content is necessary before querying and accessing the media based on the semantic items and key events.

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 5 temporal goal models which are constrained by the observation of crowd cheer, scoreboard display and change in direction. Nepal's work shows that the model-based method for 'goal' event detection is a feasible scheme.

More and more research topics with aspect to probabilistic approach have been proposed. For instance, a sport game event detection based on Hidden Markov Model (HMM) is 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. Huang *et al.* [5] proposed a Dynamic Bayesian Network (DBN) framework to extract the occurrence of highlights.

It is a nontrivial task to detect and track the ball from the sport video broadcasting. More keen than ever, researchers desire to obtain the position of the ball for each frame for soccer video broadcasting [6]. It can play a crucial role in analyzing video structure, ball possession, event detection, etc. Yu *et al.* [7] presents an improved trajectory-based algorithm for detecting and tracking the ball in soccer video broadcasting.

Here, we propose a novel video shot classification system based on the multi-level BBN. There is several root nodes in the proposed networks, the status of each node indicates the certainty of the specific event that occur in the videos. We propose a system for shot events classification in the basketball videos. The events may be the scenes of short shot, medium shot, long shot, free throw, and whether get points etc.

II. EXTRACTION OF DOMAIN-SPECIFIC FEATURE

For video understanding, it is important to extract the domain-specific semantic evidences, which in abstraction level would be a description of the object trajectory, identification of critical object, and extraction of semantic events.

In destination of tracking shooter, ball, and basket, we apply the Support Vector Tracking algorithm (SVT) [8]. We will briefly introduce SVT and describe the other features extraction framework based on different media components of the basketball game video. We detect five kinds of domain- specific features including ball released, ball-ring contact, basket orientated ball moving direction, ball inside basket net, and camera motion. These features treat as the evidences are essential for the BBNs.

1. Support Vector Tracking

SVT integrates the Support Vector Machine (SVM) classifier into an optical-flow-based tracker. Instead of minimizing an intensity difference function between successive frames, SVT maximizes the SVM classification score. In [8], they used edge information to distinguish between the vehicles and the non-vehicles as well as track a vehicle using intensity based on optical-flow. However, we are interested in tracking a particular class of objects such as human, ball and basket in basketball videos. We detect the basket using the edge information, and track the human/ball using the color histogram or luminance feature. We employ a coarse-to-fine processing in detection and tracking the objects.

SVM classification is given by SVM score function as

$$f(\mathbf{I}) = \sum_{j=1}^{N} y_j \alpha_j k(\mathbf{I}, \mathbf{x_j}) + b \qquad (1)$$

where $\mathbf{x_j}$ are the support vectors, y_j are their sign, α_j are their Lagrange multipliers, and bias b. $k(\mathbf{I}, \mathbf{x_j})$ is the kernel we choose to use and $\mathbf{I}$ is the image region we wish to test. If the above expression is positive, then the image region $\mathbf{I}$ is considered as a target object (such as ball, human or basket).We assume that the SVM score of $\mathbf{I_{final}}$ to be a local maximum in SVM score function, as follows,

$$\sum_{j=1}^{N} y_j \alpha_j k(\mathbf{I_{final}}, \mathbf{x_j}) + b = \max\{\mathbf{I} \mid f(\mathbf{I})\} \qquad (2)$$

where $\mathbf{I}$ are all possible (sub)images (in vector form) in the neighborhood of (sub)image $\mathbf{I_{final}}$. In our system, we use the linear kernel $k(\mathbf{x}, \mathbf{x_j}) = (\mathbf{x}^T \mathbf{x_j})$, i.e. the dot product of the test vector $\mathbf{x}$ with the support vector $\mathbf{x_j}$. In SVT, the image region to be tracked must be rescaled to the size of the support vectors. Once image region is rescaled to proper size, we have to look for the image region with the highest SVM score as the best position.

2. Color Histogram Quantization

To Track the ball and the shooter, we apply the SVT algorithm. The feature vectors for SVT are obtained by the color histogram of the bounding box of the ball and the shooter. The color histogram is a simple and efficient low-level feature. However, if it is calculated directly in a triple-dimension color space (e.g., RGB), both the storage space and computing time will be mass. Thus the color quantization is necessary for extraction of the image color features. We adopt a quantization algorithm based on subjective vision perception proposed in [9].

First, the color image is transformed from RGB to HSV, then, based on human vision perception for color, the triple-color components (H, S, and V) are quantized to non-uniform intervals: Hue, Saturation, and Value are quantized into 13, 5, and 3 bins respectively. Based on the quantization interval, the triple-color components are mapped into an one-dimension vector as,

$$C = HQ_sQ_v + SQ_v + V \qquad (3)$$

where Q_S and Q_V are the numbers of quantization levels for color components S and V respectively. In Eq. 3, we known that Q_S =5, Q_V =3, therefore Eq. 3 can be further described as:

$$C = 15 * H + 3 * S + V \qquad (4)$$

Therefore H, S, V which is originally represented by a vector, can be converted to a scale C with $0 \le C < 195$. To calculate the distribution of the quantized color component in scalar domain, we obtain the color histogram by using the following equation,

$$h[k] = \sum_{i=0}^{n-1}\sum_{j=0}^{m-1} f(i,j,k); \quad h_N[k] = \frac{h[k]}{mxn}$$

$$\text{where} \quad f(i,j,k) = \begin{Bmatrix} 1, \text{ if } C(i,j) = k \\ 0, \text{ otherwise} \end{Bmatrix} \qquad (5)$$

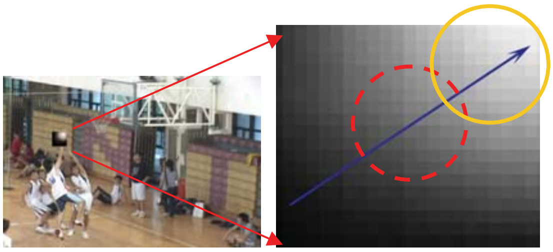
and, $h_N[k]$ is be normalized through image size, m and n are image width and height respectively. Final, we can extract the color histogram with 195 color bins.

3. Ball detection and tracking

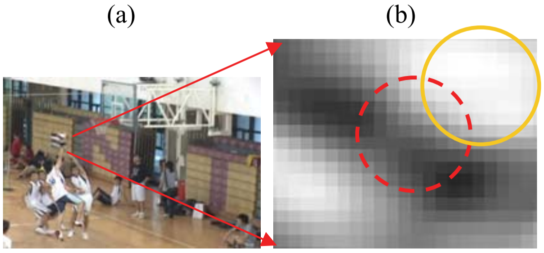
Ball detection is based on the coarse-to-fine approach. Coarse-processing searches for the possible location of the ball in a larger spatial area, whereas fine-processing is to exact the location of the shooter/ball within a smaller search area. The located position of the ball in the current frame will be used as the initial guess for the shooter identification and ball searching in the next frame.

In next frame we applied the pixel-wise ball tracking by checking the SVM score of neighboring candidates of the identified bounding circle. The ball tracking process in this frame is based on fine-processing. Fig. 1 shows SVM score distribution of ball tracking. SVT calculates the SVM scores using color histogram feature and luminance feature. Fig. 1(a) and 1(c) indicate the tracked ball from the initial position of previous frame (dashed bounding circle) to the best position of current frame (solid bounding circle).

The intensity of each small block in Fig. 1(b) and 1(d) indicate a SVM score. The block with lower intensity means lower SVM score and the location is measured in pixel unit. The arrowhead represents the tracking trajectory of the current searching from the initial position of previous frame to the best position of current frame with local maxima of SVM score. We search for the brightest pixel in which the location indicates the predicted position of the ball



(a) (b)



(c) (d)

Fig. 1 SVM score distribution. (a) and (b) show the SVM score distribution by using color histogram as input feature vector. (c) and (d) show the SVM score distribution by using gray value as input feature vector.

4. Shooter Identification

In basketball game video, the players in the two different teams always have dressed in two different colors. Here, we have trained two different SVM models for the shooters in the game. We pre-trained two distinguished SVM model by the color histogram of the shooter's clothes as the feature vector. Once the SVT locate the ball, we may search 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 (i.e. the block length may range from 50, 60,…, 140) and the aspect ratio equal to 0.3. Here we use SVM tracking to identify size of the shooter. The experimental result of shooter identification is shown in Fig. 2.

Fig. 2 Results of shooter identification. (the ball is located first as shown in Fig 1)

5. Shooter tracking

We use the color-based SVM to identify the coarse position of the shooter and then use the luminance-based SVM to identify its size. Different from shooter identification process, in shooter tracking process, we pre-trained the SVM model by using the color histogram of the shooter block as the feature vector. Once the location of the shooter is identified, we use the fine-processing to track the shooter by using color-based SVM tracking. We check the SVM scores of neighboring candidate of the shooter each pixel. The search range of the shooter tracking is 20x20 around the shooter located in the previous frame. The located position of the shooter in the current frame will be used as the initial guess for searching the shooter in the next frame. Fig. 3 shows the result of the occluded shooter who still can be tracked after five consecutive image frames.

Fig. 3 Results of the shooter tracking. (a) The shooter is occluded by other player, (b) After five image frames

6. Basket detection and tracking

We find that edge information of basket is obvious different from background region such as audience, player, and field etc. For the input image, we find the corresponding edge map, and then rescale the bounding box to normal size as the input feature for SVT. Due to the basket is frequent appear located in the upper half of the image frame. Therefore, we only check the candidate positions in the upper half of the image frame using coarse-processing (see Fig. 4(a)). We check the SVM scores of the block of candidate baskets at the upper half of the image frame for every five pixels to locate the best coarse position of the basket. However, due to camera panning or zooming, the position of basket is not identical in the subsequent frames. Therefore, we need to find the best position of basket by using fine-processing which searches for the location of current frame around the located the bounding box in the previous frame own the best SVM score. The tracking result is shows in Fig.4.

(a)

(b) (c)

Fig. 4(a) the search range of the basket detection. (b) detected the basket. (c) the edge information of (b).

7. Ball released Detection

Since the image of the player has different height and arm length, it makes the timing estimation when the ball is released more complicated. The taller player may probably also have longer arms. To detect whether the ball is released, we compare the ball-shooter distance (h_1) with the product of shooter's height and ratio1 (r_1). The ball-shooter distance in vertical direction (i.e., h_1) may be obtained by the distance between the ball tracking and the upper bounding of shooter torso.

Fig. 5 h_1 is ball-shooter distance, r_1 is ratio1, h_2 is height of clothing of shooter.

Here, we find certain relationship between h_1 and h_2, (i.e., $r_1=h_1/h_2$), when the player is about to release the ball. If ball-shooter distance is greater than the product of r_1 and h_2, the shooter is releasing the ball, and the shot event may be occurred. The relationship between h_1, h_2 and r_1 is illustrated in Fig 5. Here the ratio1 (i.e., r_1) may be achieved through a proper training. Through the training sequence of about 70 examples of ball-released, we find the average ratio $r_1 = 0.675$.

8. Ball-ring contact Detection

The ball-ring contact is important information for recognizing the shot event. We identify ball-ring contact by computing the ball-ring distance between the ball and the basket. We may obtain the locations of the ball and basket through ball tracking and basket tracking step. When the locations of the bounding circle of ball and the bounding box of basket are much closed, we may find the ball-ring contact. As shows in Fig. 6, the distance between of the ball and the basket is calculated by computing:

1) the upper bounding of the basket and the lower boundary of the ball as the ball-ring distance Y.

2) the left/right boundary of the basket and the right/left boundary of the ball as the ball-ring distance X.

Figure 6 ball-ring distance

(a) (b)

Fig. 7 (a) and (b) show the ball-ring contact in left side view and right side view respectively.

9. Basket-oriented ball 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. First, we acquire the ball position and basket position in image using ball tracking and basket detection. When the moving direction of the bouncing ball relative to the basket is downward, we may infer the video shot as a scoring shot. Fig. 8(a) illustrates the scoring shots in left side and right side respectively. If the moving direction of the bouncing ball is not downward, it may be moving upward or sideward which are shown in Fig. 8(b).

(a) the bouncing ball moving downward

(b) the bouncing ball moving sideward or upward.

Figure 8 The bouncing ball may move in various direction.

10. Ball inside net Verification

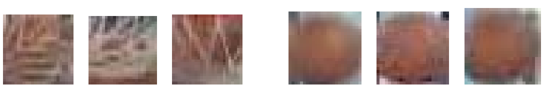
For a scoring shot, ball is enveloped by the basket net and retained for a couple of frames. If the bounding circle of ball is overlapped with the basket which is enclosed by a rectangle such as Fig. 9 (a) and (b), we find two cases:

Case 1: Ball is inside basket net.
Case 2: Ball is outside basket net.

We distinguished the two cases by using luminance information and SVM. Then, we accumulate the SVM scores in the frames in which the bounding circle of ball is overlapped with the bounding box of basket (illustrated in Fig. 9). We called the SVM score is *in/outside score*. When the in/outside score is positive, it is a scoring shot. In contrary, it is not a scoring shot. The SVM training data is illustrates in Figs. 10 (a) and (b). Fig. 10(a) shows the training images in which the ball is covered by basket net Fig. 9(a), indicated the scoring shot. On the other hand, the in/outside score is negative in Fig. 9(b), indicating a non-scoring shot.

(a) (b)

Fig. 9 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).



(a) (b)

Fig. 10 Training set of SVM (a) ball is inside the net (b) ball is outside the net

11. Panning/Zooming Ratio

In the basketball game, the camera always traces the ball, so that it keeps panning or zoom, and it dose not stop unless a particular event occurs such as free throw. The camera motion information consists of two cases：(1) static (2) panning or zoom. We apply the method mentioned in [10] to decide whether the camera is static or not. We calculate the motion of the two-dimensional m×n picture elements using two one-dimensional vectors. First, we calculate the vertical projection for each frame, and the horizontal projection described as follows,

$$f_x(i) = \frac{1}{n}\sum_{j=1}^{n} p(i,j) \qquad for \;\; i = 1,\ldots,m \tag{6}$$

$$f_y(j) = \frac{1}{m}\sum_{i=1}^{m} p(i,j) \qquad for \;\; j = 1,\ldots,n \tag{7}$$

where p(i,j) denotes the pixel values at location (i, j).

We divided the projection into small slices with N pixels width. Assume two consecutive frames A and B, and taking a slice of frame A and sliding it over frame B. To calculate the *sum of absolute difference* (SAD) value for a shifted of frame A, we define the SAD as,

$$SAD\;(n_0, s) = \sum_{i=n_0-N/2}^{n_0+N/2} |\, f_a(i) - f_b(i+s)\,| \tag{8}$$

where n_0 is the index of the center position of the slice from frame A, s is the displacement value. We find the displacement vector that makes the minimum SAD values.

To obtain the number of frames in a video shot with motion, therefore, we define the so-called panning/zooming ratio (*P/Z ratio*) as follows:

$$P/Z\;\; ratio \;=\; \frac{the\;\; number\;\; of\;\; frames\;\; with\;\; motion}{total\;\; number\;\; of\;\; frames} \tag{9}$$

If the P/Z ratio is less than 0.3, the camera is stationary. This ratio is a useful feature to determine whether a free throw occurred is (see Fig. 11).

(a) (b)

Fig. 11 (a) camera panning during the game. (b) camera static during free throw.

III. SHOT EVENT ANALYSIS USING BAYESIAN NETWORK

In this paper, we attempt to bridge this gap through the integration of image/video analysis algorithms with multi-level Bayesian Brief Network (BBN). Our application domain is the video shots classification of the sports game, and our system is designated for the basketball game videos. Given the video shots of basketball game, we can identify four different shot events such as short shot, medium shot, long shot, free throw, and the scoring event etc.

1. Bayesian Belief Network Modeling

The major objective of our approach is using the BBN to link the low-level media evidences to the high-level concepts. We generate a BBN model, which supports an inference of unobservable concepts based on their relevance with the observable evidences. Given evidences as the input, the statistical model-based classifiers and BBN may infer certain high-level concepts.

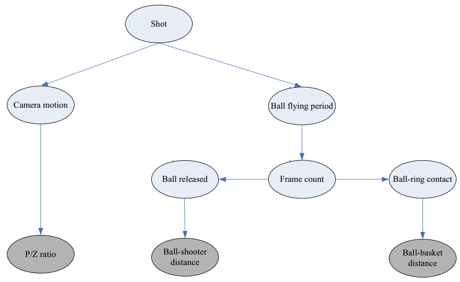


Figure 12 BBN structure of the shot event.

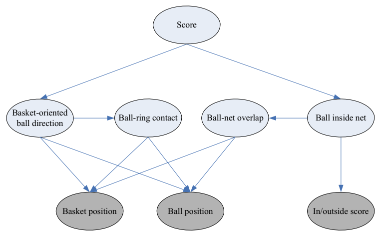


Figure 13 BBN structure of the score event.

Here, we propose two BBNs that are used to model the semantic feature of basketball game such as *short shot*, *medium shot*, *long shot*, *free throw*, and *scoring shot*. Each root node represents the category of video shot and is connected to several mid-level nodes. The linkage characteristics of the BBN are also manually determined, and the probabilities of these links can be obtained by training procedure. The root node of the shot event consists of four independent states such as short shot, medium shot, long shot and free throw, which are supported by camera motion and ball flying period. (see Fig. 12) But the root node of the score event only consists of two states such as “yes” and “no”, which supported by basket-oriented ball direction and ball inside net (see Fig.13).

1.1 Training Phase

Here, we apply a semi-automatic method to construct the BBN. In training process, the dependence between the nodes and the occurring possibility of each node in the network will be determined. The training procedure can be divided into two phases: event node v.s. mid-level nodes, and mid-level nodes v.s. domain-specific feature nodes.

1.2 Understanding Phase

Once the conditional probability and the prior probability for each node of BBN are known, we can utilize this model to understand and identify video shots. First, the video sequence is segmented into many video shots with shot event manually. Second, these 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 system.

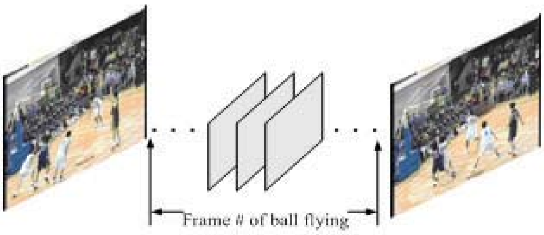
We use the evidence propagation procedure from the domain-specific feature layer to the mid-level semantic layer such as ball flying period, camera motion, basket-oriented ball direction and ball inside net. Finally, we may infer the video shot category of testing sequence from mid-level semantic information.

2. BBN for Shot Event

When the system identified the node of ball released occurring, we have to 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.

The flying period is divided into 10 independent states, from state1 which indicates the time period less than 2 image frames, to state10 which indicates the time period is between 35 and

45 image frames. When our system identify occurrence of ball released is true and ball-ring contact is false, the node of frame count is triggered which causes the ball flying period node to change from one state to other state until the ball-ring contact occurs. Fig. 14 illustrates the temporal relation of ball flying period and Fig. 15 shows the corresponding sub-network.



(a) (b)

Fig. 14 Ball flying period. (a) ball released; (b) ball-ring contact

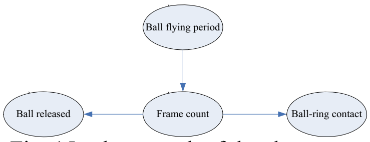


Fig. 15 sub-network of the shot event

The shot event is defined the distance between the basket and the shooter where he shoots the ball. The distance may be short distance, medium distance or long distance. Fig. 16 shows three zones where the shooter shot the ball. Short shot and medium shot are both 2-points shot, long shot is 3-points shot and free throw is 1-point shot.

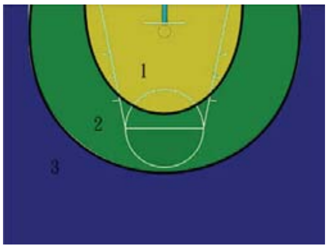


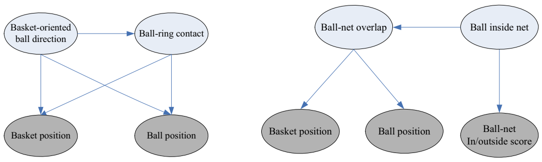
Figure 16 shooting zone; 1: short shot; 2: medium shot or free throw; 3: long shot.

In our experiments, BBN will determine the occurrence possibility of the mid-level nodes and the root node. For mid-level nodes, based on its posterior probability, if it exceeds 0.5, we consider the occurrence of the corresponding mid-level feature in the video shot is true. However, for root node of the shot event, based on its posterior probability, we only select the states of the node with largest probability to indicate the correctly classification class.

3. BBN for Score event

In Fig. 13, the shaded nodes applied to infer the score event are ball position, basket position, in/outside score. After the shooter released the ball, as well as 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 in the moment. Thus, the ball inside net and the ball moving direction are important evidences for score event inference.

The probabilities inferencing of the ball moving direction node and the ball inside net node for each frame are not activated by low-level evidence. Once the node of ball-ring contact is activated, then the inference of the ball moving direction will proceed. Similarly, the inference process of the node of ball-inside-net is activated only when the ball net overlapped occurs. The hypothesis of ball-inside-net is not based on the ball-net inside/outside score only. Figure 17 illustrates the two subnet of score event.



(a) (b)

Fig. 17 (a) the sub-net of basket-oriented ball moving direction inference; (b) the sub-net of ball inside net inference.

IV. EXPERIMENTAL RESULTS

We have proposed a novel video shot classification system using BBN for the basketball videos. The low-level features for the BBN inference included the object position/direction, SVM score, motion etc. Further, we can accomplish the goal of retrieve the video such as video shots semantic understanding and classification a lager number of the input video shots. In this section, we illustrate how to test the performance of our system, and show the experimental results in some aspects. We emphasized that the measurement of recognition rate was based on video shot unit, but the updated probability of BBN is based on frame unit.

1. Shot-based event identification

Our experimental video data are selected from the video scene captured by the main camera. First, the five basketball game video sequences of the NTHU-NCTU competition are segmented into many video shots [11] from which we select about 157 video shots from these scenes. The five basketball matches consists of two female basketball matches and three male basketball matches. For the lowest-level nodes, we apply the pre-processing and feature extraction to obtain shooter position, ball position and basket position in each frame.

$$Precision = \frac{number\ of\ correct\ detection}{number\ of\ correct\ detection + number\ of\ false\ alarm} \quad (10)$$

$$Recall = \frac{number\ of\ correct\ detection}{number\ of\ correct\ detection + number\ of\ miss} \quad (11)$$

The BBN-based video shot classification system has been developed to interpret of the basketball video by categorizing the video shots. The root node of shot event consists of four states, indicating four different categories of the shot events. The input video shot is categorized based on the posterior probability of the corresponding states which is the largest. The precision rate (Eq. 10) and recall rate (Eq. 11) has been calculated to measure the performance of shot event classification and score event detection shown in Table 1-2.

Table 1. results of shot event classification

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. the results of score event detection

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%

V. CONCLUSIONS

We have proposed a novel basketball video shot classification system based on Bayesian Belief Network. Given an input sequence, the image analyzer can be applied to collect and detect the low-level evidence, and the inference engine in the BBN can be applied to infer a high-level semantic 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.

VI. ACKNOWLEDGEMENTS

This work has been partially funded by Department of Industrial Technology (Industrial Technology Development Program) under project no. 94-EC-17-A-02-S1-032.

REFERENCE

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

[2] S. Nepal, U. Srinivasan, and G. Reynolds, "Automatic detection of goal segments in basketball videos," in Electronic 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," in Proc. IEEE-ICIP'03, pp. 25-28, September 2003.

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

[5] C. L. Huang, H. C. Shih, and C. Y. Chao, "Semantic Analysis of Sports Video using Dynamic Bayesian Network," to be appear in IEEE Transactions on Multimedia, 2005.

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

[7] X. Yu, C. S. Xu, H. W. Leong, Q. Tian, Q. Tang, and K. W. Wan "Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video," in Proc. of ACM MM'03, Berkeley, pp. 11-20, 2003.

[8] S. Avidan, "Support Vector Tracking," IEEE Trans. on Pattern Anal. Mach. Intel. Vol. 26, No 8, pp. 1064-1072, 2004.

[9] L. H. Cao, W. Liu, G. H. Li, "Research and implementation of an image retrieval algorithm based on multiple dominant colors," Journal of Computer Research and Development, Vol. 36, No. 1, pp. 96-100, 1999.

[10] 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.

[11] 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.