Event Detection and Semantic Identification Using Bayesian Belief Network

Maheshkumar H. Kolekar
Dept. of Computer Science,
University of Missouri, Columbia, MO, USA
mkolekar@gmail.com

K. Palaniappan
Dept. of Computer Science,
University of Missouri, Columbia, MO, USA
palaniappnk@missouri.edu

S. Sengupta
Dept. of Electronics and ECE,
Indian Institute of Technology, Kharagpur, INDIA
ssg@ece.iitkgp.ernet.in

G. Seetharaman
Air Force Research Lab,
RITB, 525 Brooks Road, Rome, NY, USA
Gunasekaran.seetharaman@rl.af.mil

Abstract

A probabilistic Bayesian belief network (BBN) based framework is proposed for semantic analysis and summarization of video using event detection. Our approach is customized for soccer but can be applied to other types of sports video sequences. We extract excitement clips from soccer sports video sequences that are comprised of multiple subclips corresponding to the events such as replay, field-view, goalkeeper, player, referee, spectator, players' gathering. The events are detected and classified using a hierarchical classification scheme. The BBN based on observed events is used to assign semantic concept-labels, such as goals, saves, and card to each excitement clip. The collection of labeled excitement clips provide a video summary for highlight browsing, video skimming, indexing and retrieval. The proposed scheme offers a general approach to automatic tagging large scale multimedia content with rich semantics. Our tests using soccer video shows that the proposed semantic identification framework is more efficient.

1. Introduction

With the rapid development in multimedia creation, storage, and compression technologies, large amounts of digital video is produced daily for applications such as news, sports, movies, video surveillance, etc. Searching and summarizing multimedia repositories to manage the content explosion has been motivating force over the past decade to develop more robust multimedia information mining technologies. One of the main research issues is the semantic gap between low-level video features and high-level semantic concepts. Similarity of low-level features is often not correlated with similarity at the user perceived semantic concept. Scenes or shots are associated together due to semantics rather than low-level features like color histograms or motion trajectories. Some examples of approaches to discover meaningful information at the event level include: Xu and Chang [14] have proposed video event recognition using kernel methods for news video sequences, Albanese *et. al.* [1] have presented probablisitic framework for human activity detection, and Zhou *et. al.* [15] have presented method for activity analysis for indoor video sequences.

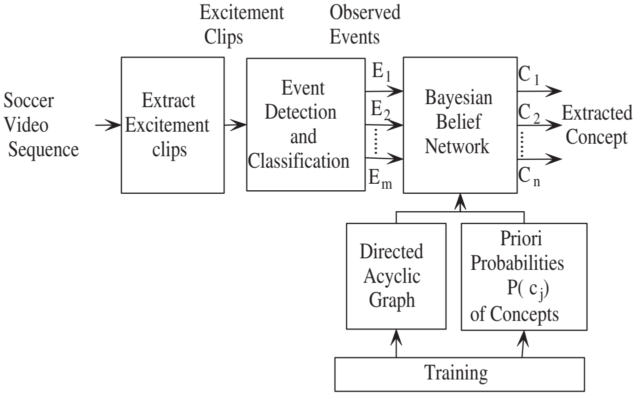

Figure 1. Proposed system diagram for concept extraction

Semantic indexing is an emerging research area that identifies events in audiovisual streams and facilitates semantic retrieval. Semantic indexing can be separated into two levels: isolated video event detection and semantics identification. Isolated events are not as intuitive for users as events with semantic context at the right granularity for retrieval; for example, replay event associated with a goal by one team, is more meaningful than identifying just the replay event.

2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops
978-1-4244-4441-0/09/$25.00 ©2009 IEEE

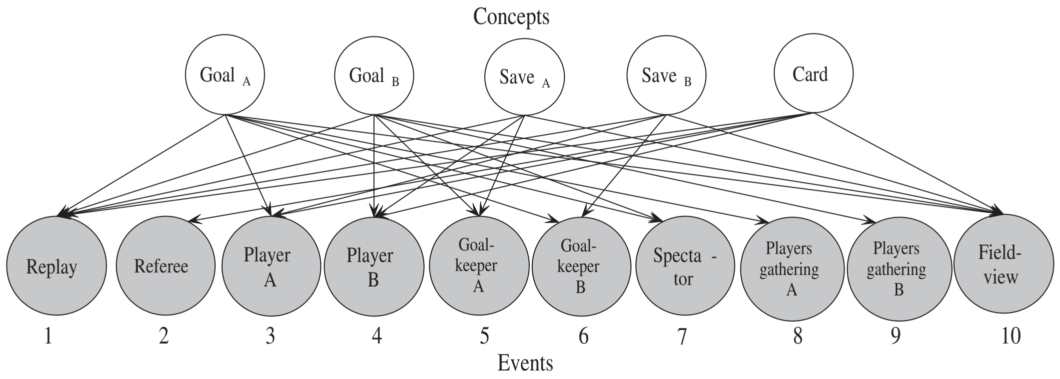

Figure 2. Directed Acyclic Graph for Bayesian belief network of soccer video sequence. Shaded nodes represent observed event nodes and unshaded nodes represent hidden nodes.

To reduce the gap between current techniques for semantic analysis and users expectation, the following approach is proposed. First, we detect interesting parts of sports video by exploiting domain knowledge and production rules. Second, we detect the events associated with those interesting parts by extracting low-level features. Finally, for modeling visual semantics, we propose a Bayesian belief network (BBN) based approach to fuse the information of visual events with the high-level semantic concepts, such as ”goal” in soccer video.

Bayesian networks [12] are a popular graphical model approach used in computer vision. Many multi-variate probabilistical and statistical inference problem can be formalized using the framework of Bayesian networks. A large number of research advances have been made to apply the Bayesian networks in various computer vision areas including video analysis, tracking, image segmentation, object recognition, etc. In [13], Sun *et. al.* uses Bayesian network for scoring event detection in soccer video based on using six different low-level features including gate, face, audio, texture, caption and text. Huang *et. al.* [2] developed the scheme for identifying special events in soccer games using dynamic Bayesian network.

As shown in Figure 1, excitement clips are extracted from soccer sports video using audio cues. Events are extracted using low-level feature based hierarchical classification tree. Each extracted event is annotated by appropriate label. The detected events within the clip are considered as evidences for the Bayesian network. Based on the posteriori probability of all the concepts given evidences, the concept-label is selected for the particular excitement clip. Each selected excitement clip will have one concept-label and several event-labels. For soccer video, we use 10 labels for the annotation of events and 5 labels for annotation of concepts in this paper. The main contribution and novelty of this paper is that we use a probabilistic BBN for extracting semantically meaningful concepts in soccer video using low-level feature-based evidences. Our proposed approach uses retrieval-friendly semantic concept labels.

Table 1. Noisy-OR CPD for node C with two parent A and B

A	B	$P(C = off)$	$P(C = on)$
F	F	1.0	0.0
T	F	$q(A)$	$1 - q(A)$
F	T	$q(B)$	$1 - q(B)$
T	T	$q(A)q(B)$	$1 - q(A)q(B)$

2. Bayesian Belief Network

A BBN is a directed acyclic graph (DAG) where the causal relationship between multiple variables are represented with conditional probabilities. BBN defines the probabilistic independency structure by a directed graph (V, E), whose nodes V represent the random variables $X = \{x_1, x_2, ..., x_N\}$ and edges E represent the conditional independent probabilities between the nodes. The joint probability distribution has the form $P(x_1, x_2, ..., x_N) = \prod_{i=1}^{N} P(x_i/Par(x_i))$, where $Par(x_i)$ denotes the set of parent nodes of the random variable x_i. The nodes outside $Par(x_i)$ is conditionally independent of x_i. The inference can be performed using various algorithms such as expectation maximization, variational algorithms [3], belief propagation, Markov Chain Monte Carlo, particle filter, etc. In this paper, we have customized the application for soccer video sequence whose DAG is shown in Figure 2 and used variational algorithm for inference.

2.1. Noisy-OR Distribution

A Conditional Probability Distributions (CPD) defines $P(x_i|Par(x_i))$, where x_i is the i^{th} node, and $Par(x_i)$ are the parents of node i. There are many ways to represent this distribution, which depend in part on whether x_i and $Par(x_i)$ are discrete, continuous, or a combination. In our case, since both are discrete, we use Noisy-OR distribution.

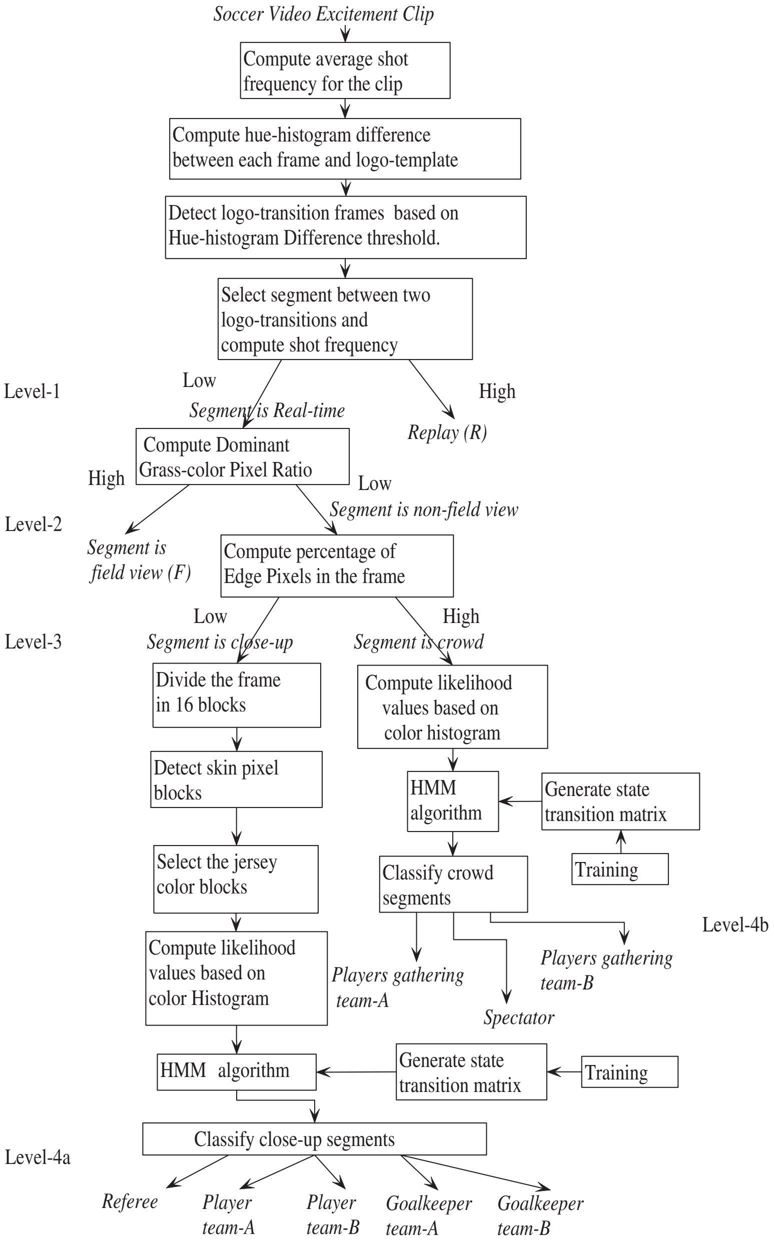
Soccer Video Excitement Clip
Compute average shot frequency for the clip
Compute hue-histogram difference between each frame and logo-template
Detect logo-transition frames based on Hue-histogram Difference threshold.
Select segment between two logo-transitions and compute shot frequency
Level-1
Low
Segment is Real-time
High
Replay (R)
Compute Dominant Grass-color Pixel Ratio
High
Low
Segment is non-field view
Level-2
Segment is field view (F)
Compute percentage of Edge Pixels in the frame
Low
Segment is close-up
High
Segment is crowd
Level-3
Divide the frame in 16 blocks
Detect skin pixel blocks
Select the jersey color blocks
Compute likelihood values based on color Histogram
HMM algorithm
Generate state transition matrix
Training
Compute likelihood values based on color histogram
HMM algorithm
Generate state transition matrix
Training
Classify crowd segments
Level-4b
Players gathering team-A
Spectator
Players gathering team-B
Classify close-up segments
Level-4a
Referee
Player team-A
Player team-B
Goalkeeper team-A
Goalkeeper team-B

Figure 3. Proposed system diagram for event annotation

A noisy-OR node is like a regular logical OR gate except that sometimes the effects of parents that are on get inhibited. Let the probability that parent i gets inhibited be $q(i)$. Then a node, C, with two parents, A and B, has the CPD shown in Table 1, where we use F and T to represent off and on. Thus, we see that the causes get inhibited independently. It is common to associate a ”leak” node with a noisy-OR CPD, which is like a parent that is always on. This can account for all other unmodelled causes which might turn the node on.

3. Event Detection and Classification

The fundamental problem associated with the sports video analysis is the large volume of data that we have to deal with. In every game, there are moments of excitement, with relatively dull periods in between. Excitements are always accompanied by significant audio content resulting from spectators cheering, increase in the audio level of the commentators voice etc. Based on this observation, we have used two popular audio content analysis techniques-short-time audio energy and zero-crossing rate for extract-

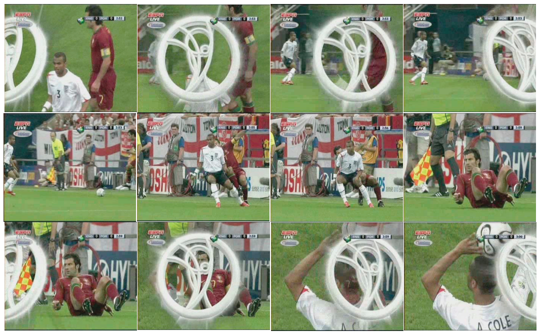

Figure 4. Row 1 and 3 show representative frames of logo-transition. Row-2 shows representative frames of replay segment.

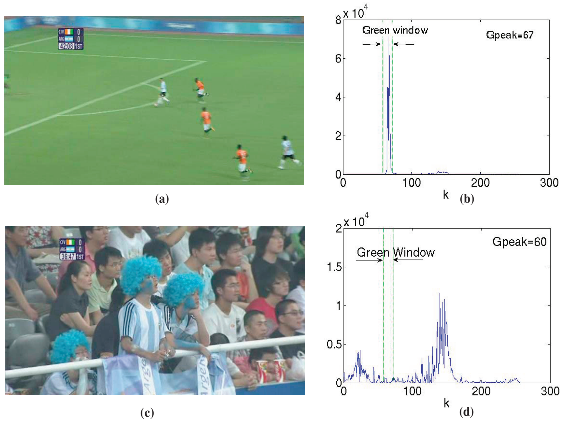

Figure 5. (a) Field view Image, (b) Hue-Histogram of field view image, (c) Non-field view Image, (d) Hue-Histogram of non-field view image

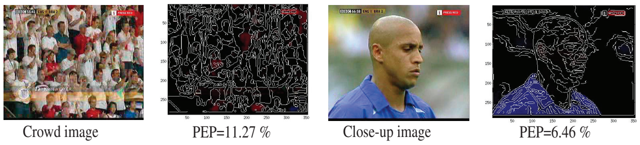

Figure 6. Crowd image has $PEP > 8\%$, and close-up image has $PEP < 8\%$.

ing excitement clip in [11].

Figure 3 shows the system diagram for event detection and classification of excitement clips. We have used algorithms proposed by us in [8],[7], [5] for soccer video sequences. At level-1, we classify the excitement clip into real-time and replay segments. Generally, replays are broadcasted with logo-transition indicating the start and end of the replay as shown in Figure 4. Based on this observation, we have proposed the logo-based replay detection algorithm in [6].

At level-2, we are using dominant green-color pixel ratio ($DGPR$) to classify the real-time clips into field view and non-field view. By testing several field view frames, we observed that the green color occur between bin 58 to 62 of the 256-bin hue histogram as shown in Figure 5. We define $DGPR$ as the ratio of the green color pixels to the total number of pixels in the frame. At level-3, we have to classify the frames of non-field views. Generally crowd images will have more edge density as shown in Figure 6. Hence, we are classifying the non-field views broadly into close-up and crowd classes. We have used the percentage of edge pixels (PEP) based algorithm to classify the frame into crowd or close-up.

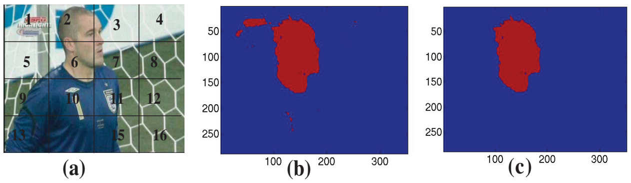

Figure 7. (a) Image of goalkeeper (b) Image showing skin detection, (c) Connected component image

At level-4a, we classify the close-up frames into referee, player team-A, player team-B, goalkeeper team-A, goalkeeper team-B using players jersey color information. As shown in Figure 7, we divide the frame into 16 blocks and computer the skin percentage of each block. Based on skin pixels,we select the blocks containing face location of the player. The blocks immediately below the face block is selected as jersey color block for the particular player. We compute the likelihood function based on hue histogram of the jersey block of the player. Using HMM algorithm [9], the jersey color block is classified. At level-4b, we classify crowd image into players gathering team-A, players gathering team-B, and spectator. We compute likelihood function based on the hue histogram of the crowd frame. By using HMM algorithm, we will classify the frame into the known classes.

We have defined the events as scenes in the video with some semantic meaning attached to it. The events for soccer excitement clips are replay, referee, player of team-A, player of team-B, goalkeeper of team-A, goalkeeper of team-B, spectator, players gathering of team-A, players gathering of team-B, and field view.

4. Semantic Analysis Using BBN

4.1. Soccer BBN

The major objective of our approach is to link low-level events to the high level concepts using Soccer BBN. For each excitement clip, we will extract the events using hierarchical classification scheme. We use Soccer BBN to computer posteriori probability of the concepts based on the observed events. The label of the concept-class which has larger posteriori probability is assigned to the excitement clip. By using 40 excitement clips from each concept-class for training, we have generated the bipartite graph as shown in Figure 2.

Let c and e are binary vectors referring to presence/absence state of concepts and the positive/negative state of events respectively. The prior probabilities of the concepts $P(c_j)$, were obtained from the training data. The joint probability of concepts and events are computed as:

$$P(c,e) = P(c|e)P(e) = \prod_i P(e_i|c) \prod_j P(c_j) \quad (1)$$

where the conditional probabilities $P(e_i|c)$ are represented by the 'noisy-OR model'.

$$P(e_i = 0|c) = P(e_i = 0|L) \prod_{j \in \pi_i} P(e_i = 0|c_j) \quad (2)$$

$$= (1 - q_{i0}) \prod_{j \in \pi_i} (1 - q_{ij})^{c_j} \quad (3)$$

$$= exp^{-\theta_{i0} - \sum_{j \in \pi_i} \theta_{ij} c_j} \quad (4)$$

where π_i is the set of concepts that are parents of the events e_i in the SBN graph, $q_{ij} = P(e_i = 1|c_j = 1)$ is the probability that the concept c_j, if present, could alone cause the event to have a positive outcome, and $q_{i0} = P(e_i = 1|L)$ is the 'leak' probability, i.e., the probability that the event is caused by means other than the concepts included in the SBN model. In the Eq. 4, we reparameterize the noisy-OR probability model using an exponentiated notation. In this notation, the model parameters are given by $\theta_{ij} = -log(1 - q_{ij})$.

4.2. Soccer BBN Inference

The inference mechanism using Noisy-OR model is shown in Figure 8. The inference involves computing the posterior marginal probabilities of the concepts given a set of observed positive ($e_i = 1$) and negative ($e_i = 0$) events. Note that the unobserved events have no effect on the posterior probabilities for the concepts. Let e_i^+ corresponds to the event $e_i = 1$, and e_i^- refers to $e_i = 0$ (positive and negative events respectively). Thus the posterior probabilities of interest are $P(c_j|e^+, e^-)$, where e^+ and e^- are the vectors of positive and negative events.

The negative events e^- can be incorporated into the posterior probability in linear time in the number of associated concepts and in the number of negative events. As we discuss below, this can be seen from the fact that the probability of a negative finding is the exponential of an expression that is linear in the c_j. The positive findings, on the other hand, are more problematic. In the worst case the exact calculation of posterior probabilities is exponentially costly in the number of positive findings.

Let us consider the inference calculations in more detail. To find the posterior probability $P(c|e^+, e^-)$, we first consider the evidence from negative events, i.e., we compute $P(c|e^-)$. This is just $P(e^-|c)P(c)$ with normalization. Since both $P(e^-|c)P(c)$ and $P(c)$ factorize over the concepts, the posterior $P(c|e^-)$ must factorize as well. The normalization of $P(e^-|c)P(c)$ therefore reduces to independent normalizations over each concepts and can be carried out in time linear in the number of concepts (or negative events).

Now, let us compute $P(c_j|e^+)$, the posterior marginal probability based on the positive events. Formally, obtaining such a posterior involves marginalizing $P(e^+|c)P(c)$ across the remaining concepts:

$$P(c_j|e^+) \alpha \sum_{c \neq c_j} P(e^+|c)P(c) \quad (5)$$

where, the summation is over all the possible configurations of the concept variables other than c_j. In the SBN model $P(e^+|c)P(c)$ has the form:

$$P(e^+|c)P(c) = \prod_i P(e_i^+|c) \prod_j P(c_j) \quad (6)$$

$$= \prod_i (1 - exp^{-\theta_{io} - \sum_j \theta_{ij} c_j}) \prod_j P(c_j) \quad (7)$$

To perform the summation in Eq. 5 over the concepts, we would have to multiply out the terms $1 - exp^{\{\cdot\}}$ corresponding to the conditional probabilities for each positive event. The number of resulting terms would be exponential in the number of positive events and this calculation is not feasible. Hence, we use variational method to simplify the complicated joint distribution function.

4.3. Variational Method

The variational transformation [4] involves replacing an exact conditional probability of a event with a lower bound and an upper bound.

$$P(e_i^+|c, q|i) \leq P(e_i^+|c) \leq P(e_i^+|c, \xi_i) \quad (8)$$

where, ξ_i is variational parameter for event e_i, and $q|i$ is variational distribution for event e_i.
Once the variational parameters are optimized, the resulting variational distribution can be exploited as an inference engine for calculating approximations to posteriori probabilities.

4.4. Concept Classification

The concept-label for the excitement clip is selected based on the concept class which has maximum posteriori probability. It is given by:

$$Concept\ \ Label = argmax_j P(c_j|e) \quad (9)$$

for $j = 1, 2, 3, 4, 5$.

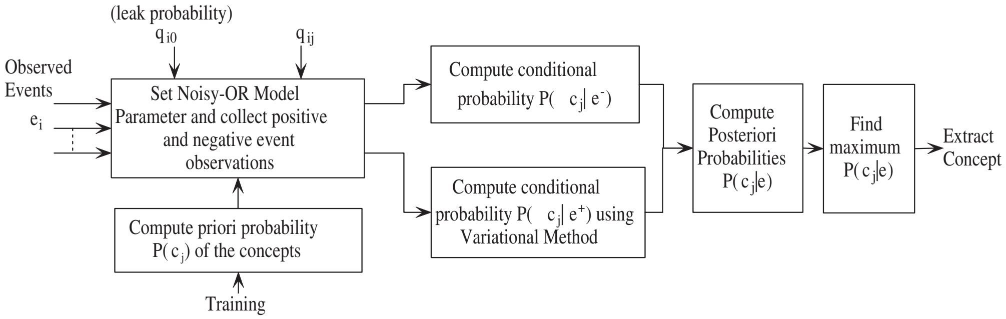
(leak probability)
q i0
q ij
Observed Events
e i
Set Noisy-OR Model Parameter and collect positive and negative event observations
Compute priori probability P(c j) of the concepts
Training
Compute conditional probability P(c j| e-)
Compute conditional probability P(c j| e+) using Variational Method
Compute Posteriori Probabilities P(c j|e)
Find maximum P(c j|e)
Extract Concept

Figure 8. Inference mechanism for extracting concept

Table 2. Soccer Video Sequences used for testing

Name of the Match	Duration	A vs B	Date
Olympic 2008	96 min	Argentina vs Ivory	07/08/2008
Olympic 2008	96 min	Brazil vs Belgium	07/08/2008
FIFA World Cup 2006	94 min	Germany vs Argentina	30/06/2006
FIFA World Cup 2006	95 min	England vs Portugal	01/07/2006
FIFA World Cup 2006	96 min	Brazil vs France	01/07/2006
Scottish Cup 2002	97 min	Celtic vs Rangers	04/05/2002
Champions League 2002	99 min	Real Madrid vs B. Leverkusen	15/05/2002
FIFA World Cup 2002	98 min	Germany vs Korea	25/06/2002
FIFA World Cup 2002	90 min	England vs Brazil	21/06/2002
FIFA World Cup 2002	90 min	Germany vs Brazil	30/06/2002

5. Experimental Results

We have used 45 hours and 16 hours of live-recordings of soccer video for training and testing purpose respectively. Table 2 shows the details of the video sequences used for testing. Training video data is used to construct DAG and compute prior probabilities of each concept. For measuring the performance of annotation, we use following parameters:

$$Recall = \frac{N_c}{N_c + N_m} \quad and \quad Precision = \frac{N_c}{N_c + N_f}$$

Where, N_c, N_m, N_f represents the number of clips correctly detected, missed and false positive, respectively.

The performance of event annotation is shown in Table 3. We observed 85.65 % recall and 87.12 % precision for replays. Replays generally occur at the end of the excitement clips. If audio is low during the last wipe of the replay, it will not be extracted as a part of the excitement clip, and hence there will be possibility of missing replays. At level-2, we observed 97.97 % recall and 94.60 % precision because of the high discrimination power of $DGPR$ ratios. At level-3, we observed 91.72 % recall and 92.63 % precision for close-up detection. We have trained our classifier for classifying the close-up frames into five classes. For close-up classification, we observed the average 86.38 % and 86.22 % recall and precision respectively. For crowd classification, we observed the average 85.24 % and 85.71 % recall and precision respectively.

The overall performance of the semantic concept annotation is shown in the Table 4. Table 5 shows the posteriori probabilities of 12 excitement clips of test video sequences. Figure 9 shows the representative event frames for the first five clips of Table 5. For $Goal_A$, $Goal_B$, and $Card$ concept clips, we observed posteriori probabilities equal to 1. For $Save_A$ concept clips, we observed posteriori porbabilities slightly less than 1, since concept class $Save_A$ is slightly confused with $Goal_B$. They have R, P_B, K_A, F are commonly observed events. Similarly $Save_B$ and $Goal_A$ have R, P_A, K_B, F as common events. Because of this overlapping, we observed posteriori probability slightly less than 1.

Table 3. Performance of event annotation

Level	Event Class	N_c	N_m	N_f	Recall (%)	Precision (%)
1	Replay	203	34	30	85.65	87.12
	Real-time	258	36	39	87.76	86.87
2	Field-view	193	4	11	97.97	94.60
	Non-field-view	298	11	4	96.44	98.68
3	Close-up	465	42	37	91.72	92.63
	Crowd	153	16	21	90.53	87.93
4a	Player-A	128	21	20	85.90	86.49
	Player-B	149	28	24	84.18	86.13
	Goalkeeper-A	42	8	9	84.00	82.35
	Goalkeeper-B	38	6	7	86.36	84.44
	Referee	44	4	4	91.67	91.67
4b	Players Gathering-A	28	4	4	87.5	87.5
	Players Gathering-B	31	6	5	83.78	86.11
	Spectator	76	14	15	84.44	83.52

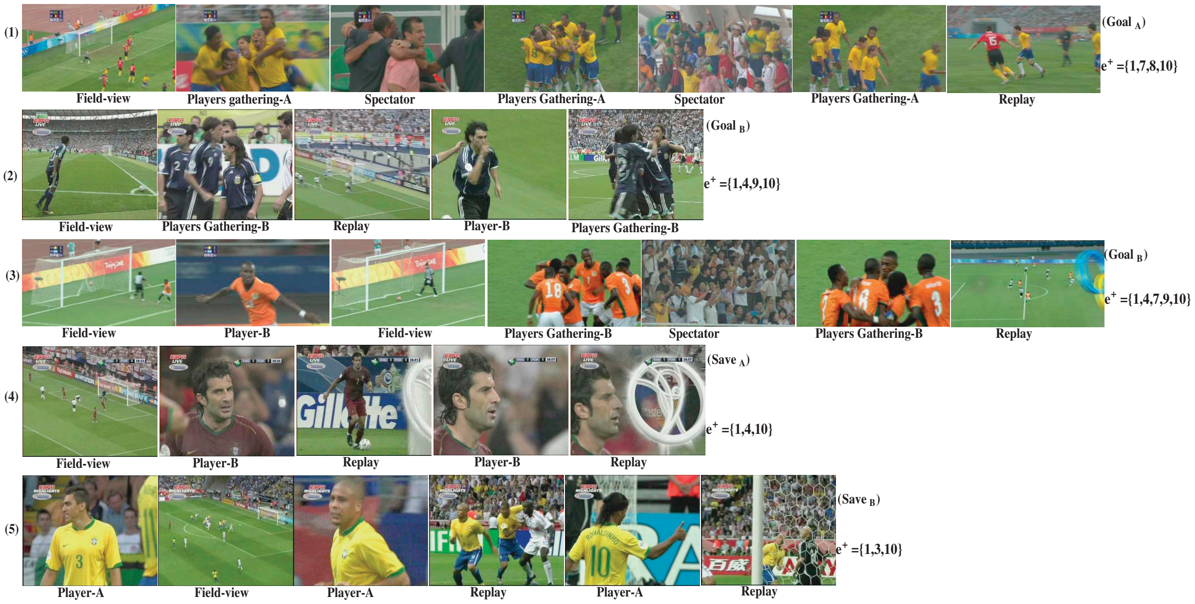

Figure 9. Video event sequence for the first five clips of Table 5.

Table 4. Performance of concept annotation

Concept Class	N_c	N_m	N_f	Recall (%)	Precision (%)
$Goal_A$	13	1	2	92.86	86.67
$Goal_B$	14	2	2	87.50	87.50
$Save_A$	54	8	9	87.10	85.71
$Save_B$	40	6	7	86.95	85.11
$Card$	42	4	4	91.30	89.36

6. Conclusion

In this paper, we have presented a Bayesian belief network based framework for extracting and identifying excitement clips with semantic concept-labels. The proposed framework creates probabilistic links between a set of low-level features extracted directly from sports video clips and high-level semantic concepts. The low-level features used for event detection and classification are field color, edge density, skin tone, player's jersey color, etc. Each excitement clip will have several event-labels and one semantically meaningful concept-label. Our results demonstrate that our proposed concept-annotation technique is effective showing an average recall 89 % and precision 87 % for soccer video with goal, save and card concept-labels. The proposed scheme facilitates the tasks such as highlight generation [10], indexing and retrieval. We intend to evaluate additional experiments involving additional concepts and large datasets to further evaluate the utility of BBN for soccer video summarization. Our proposed scheme has scope for improvement and extension

Table 5. Semantic Concept Extraction Based on Posteriori Probabilities

Clip No.	Positive Events (e^+)	Negative Events (e^-)	$P(c_1\mid e)$	$P(c_2\mid e)$	$P(c_3\mid e)$	$P(c_4\mid e)$	$P(c_5\mid e)$	Extracted Concept	Actual Concept
1	1,7,8,10	2,3,4,5,6,9	1.00	0.00	0.00	0.00	0.00	$Goal_A$	$Goal_A$
2	1,4,9,10	2,3,5,6,7,8	0.00	1.00	0.00	0.00	0.00	$Goal_B$	$Goal_B$
3	1,4,7,9,10	2,3,5,6,8	0.00	1.00	0.00	0.00	0.00	$Goal_B$	$Goal_B$
4	1,4,10	2,3,5,6,7,8,9	0.00	0.06	0.94	0.00	0.00	$Save_A$	$Save_A$
5	1,3,10	2,4,5,6,7,8,9	0.02	0.00	0.00	0.97	0.01	$Save_B$	$Save_B$
6	1,3,7,8,10	2,4,5,6,9	1.00	0.00	0.00	0.00	0.00	$Goal_A$	$Goal_A$
7	1,3,6,7,8,10	2,4,5,9	1.00	0.00	0.00	0.00	0.00	$Goal_A$	$Goal_A$
8	1,4,5,7,9,10	2,3,6,8	0.00	1.00	0.00	0.00	0.00	$Goal_B$	$Goal_B$
9	1,4,5,10	2,3,6,7,8,9	0.00	0.02	0.98	0.00	0.00	$Save_A$	$Save_A$
10	1,3,6,10	2,3,4,5,7,8,9	0.06	0.00	0.00	0.94	0.00	$Save_B$	$Save_B$
11	1,2,3,4,10	5,6,7,8,9	0.00	0.00	0.00	0.00	1.00	$Card$	$Card$
12	1,2,4,10	3,5,6,7,8,9	0.00	0.00	0.00	0.00	1.00	$Card$	$Card$

including: (1) considering more high level semantic concepts such as foul, corner kick, (2) investigating dynamic Bayesian network to explore the temporal information in soccer video, (3) extending the approach to surveillance and news video sequences.

Acknowledgement

Research was sponsored in part by the Leonard Wood Institute (LWI-181223) in cooperation with the U.S. Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-07-2-0062. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Leonard Wood Institute, the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

References

[1] M. Albanese, R. Chellappa, V. Moscato, A. Picariello, V. S. Subrahmanian, P. Turaga, and O. Udrea. A constrained probabilistic petri net framework for human activity detection in video. *IEEE Trans. on Multimedia*, 10(6):982–996, 2008.

[2] C. L. Huang, H. C. Shih, and C. Y. Chao. Semantic analysis of soccer video using dynamic bayesian network. *IEEE Trans. Broadcasting*, 8(4):749–760, 2006.

[3] T. S. Jaakkola and M. I. Jordan. Bayesian parameter estimation via variational methods. *Springer Journal on Statistics and Computing*, 10:25–37, 2000.

[4] M. I. Jordan. Graphical models. *Statistical Science (Special Issue on Bayesian Statistics)*, 19:140–155, 2004.

[5] M. H. Kolekar and K. Palaniappan. A hierarchical framework for semantic scene classification in soccer sports video. *IEEE Region 10 Conference (TENCON)*, pages 1–6, 2008.

[6] M. H. Kolekar, K. Palaniappan, and S. Sengupta. A novel framework for semantic annotation of soccer sports video sequences. *IET Int. Conf. on Visual Media Production*, pages 1–9, 2008.

[7] M. H. Kolekar, K. Palaniappan, and S. Sengupta. Semantic event detection and classification in cricket video sequences. *IEEE Indian Conf. Computer Vision, Graphics and Image Processing*, pages 382–389, 2008.

[8] M. H. Kolekar, K. Palaniappan, S. Sengupta, and G. Seetharaman. Semantic concept mining based on hierarchical event detection for soccer video indexing. *Int. Journal on Multimedia*, 4(5):298–312, 2009.

[9] M. H. Kolekar and S. Sengupta. Hidden markov model based video indexing with discrete cosine transform as a likelihood function. *IEEE INDICON Conference*, pages 157–159, 2004.

[10] M. H. Kolekar and S. Sengupta. Event-importance based customized and automatic cricket highlight generation. *IEEE Int. Conf. Multimedia Expo*, pages 1617–1620, 2006.

[11] M. H. Kolekar and S. Sengupta. A hierarchical framework for generic sports video classification. *in Lecture Notes on Computer Science (LNCS), Springer-Verlag Berlin Heidelberg*, 3852:633–642, Jan 2006.

[12] S. K. Kopparapu and U. B. Desai. Bayesian approach to image interpretation. *Kluwer Academic Publisher*, 616, 2001.

[13] X. Sun, G. Jin, M. Huang, and G. Xu. Bayesian network based soccer video event detection and retrieval. *Multispectral Image Processing and Pattern Recognition*, 2003.

[14] D. Xu and S.-F. Chang. Video event recognition using kernel methods with multilevel temporal alignment. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(11):1985–1997, 2008.

[15] Z. Zhou, X. Chen, Y. C. Chung, Z. He, T. X. Han, and J. M. Keller. Activity analysis, summarization, and visualization for indoor human activity monitoring. *IEEE Trans. on Circuits and Systems for Video Technology*, 18(11):1489–1498, 2008.