

A Query Description Model Based on Basic Semantic Unit Composite Petri-Nets for Soccer Video Analysis

Songyang Lao¹ Alan F. Smeaton, Gareth J. F. Jones, Hyowon Lee

Centre for Digital Video Processing, Dublin City University

Glasnevin, Dublin 9, Ireland

+353 -1 -7007535, 7005262, 7005559, 7008565

{lsongyang, asmeaton, gjones, hlee}@computing.dcu.ie

ABSTRACT

Digital video networks are making available increasing amounts of sports video data. The volume of material on offer means that sports fans often rely on prepared summaries of game highlights to follow the progress of their favourite teams. A significant application area for automated video analysis technology is the generation of personalized highlights of sports events. One of the most popular sports around world is soccer. A soccer game is composed of a range of significant events, such as goal scoring, fouls, and substitutions. Automatically detecting these events in a soccer video can enable users to interactively design their own highlights programmes. From an analysis of broadcast soccer video, we propose a query description model based on Basic Semantic Unit Composite Petri-Nets (BSUCPN) to automatically detect significant events within soccer video. Firstly we define a Basic Semantic Unit (BSU) set for soccer videos based on identifiable feature elements within a soccer video, Secondly we design Composite Petri-Net (CPN) models for semantic queries and use these to describe BSUCPNs for semantic events in soccer videos. A particular strength of this approach is that users are able to design their own semantic event queries based on BSUCPNs to search interactively within soccer videos. Experimental results based on recorded soccer broadcasts are used to illustrate the potential of this approach.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *query formulation, retrieval models.*

I.2.10 [Artificial Intelligence]: Vision and Scene Understanding – *video analysis.*

H.2.3 [Database Management]: Languages – *Query languages.*

General Terms

Management, Design, Experimentation

Keywords

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MIR '04, October 15–16, 2004, New York, New York, USA.

Copyright 2004 ACM 1-58113-940-3/04/0010...\$5.00.

Basic Semantic Unit, Composite Petri-Nets, Soccer Video Analysis, Query Description Model.

1. INTRODUCTION

The development of high-speed broadband networks and digital video technology (including storage, compression and processing) is giving viewers the opportunity to select from among ever increasing amounts of video data. One of the areas of greatest expansion in video content is sports broadcasting. In many different sports ever greater numbers of events are broadcast live in their entirety. It is often not possible for even the most avid sports fan to watch more than a small fraction of the available coverage of complete events. Further, for many sports much of the time during an event is often not significant to the progression of the game or its outcome. Broadcasters often prepare manual “highlights” broadcasts summarizing the significant events. However, these are not always available and the material included is selected by a single editor. This situation is inflexible with respect to individual viewers who may want a longer or shorter summary or to focus on certain event types; and the need for manual editing means generation of summaries is often not cost effective. Digital video processing technologies give the opportunity to develop systems to interactively select and search for personalized highlights of sporting events.

One of the most popular sports across the world is soccer. Matches are now shared and viewed internationally via live satellite broadcasting. Soccer fans often follow the progress of many teams and players, but to do this they frequently have to rely on highlights of the most interesting or significant events supplied by the broadcasters. Video technology can potentially provide new ways to view soccer videos which are more interactive and personal, rather than to view passively pre-edited highlights (when they are actually available). For example, spectators typically pay most attention to shots on goal, attacks, corners, fouls, substitutions and other soccer semantics [1]. Usually they are less interested in open play where the players are passing the ball around the pitch. In other words, it is important for video processing and retrieval researchers to develop new ways for users to search for the above semantic events within soccer videos. To date work in this area has focused on query-by-

¹ Prof. Lao is a faculty member of National University of Defense Technology and now as a scholar is working along with CDVP at DCU. He can be contacted at School of Information System & Management, National University of Defense Technology, Changsha, China, 410073, +0086-731-4826061, laosongyang@vip.sina.com.

text. The user enters the word “goal”, then the system searches for video clips including the word goal in the soundtrack or possibly in manually added metadata. This clearly relies on either a well annotated soundtrack or a costly manual labeling of semantic events. The huge number of soccer matches now routinely broadcast and recorded means that it is unrealistic to expect this level of hand annotation for all matches. Hence automatic detection of semantic events that capture the essential contents of a match is becoming more and more important. Related prior work towards automatic event detection in sports videos is described in [2], [3], [4], [5]. Although this existing work addresses the issue of the overload in manual work, the event detection algorithms are embedded in systems and cannot easily be redefined. This means that users cannot adapt the event types detected or the system refined to the different editorial rules used by different broadcasting corporations.

In this paper we introduce a new video query model. For this model we first define semantic units to describe the contents of soccer videos. These Basic Semantic Unit (BSUs) are combined in Basic Semantic Unit Composite Petri-Nets (BSUCPNs) which describe the spatio-temporal relations between the units associated with significant events in a game of soccer. Petri-Nets are a graphical and mathematical modeling tool [6] which have been used successfully in video meta-data modeling and query processing [7], supporting content-based queries in digital on-demand video services [8], modeling the synchronization and spatial aspects of a multimedia presentations [9], but have not previously been applied to model soccer query descriptions. The meta-data modeling in [7] proposed the use of Hierarchical Petri-nets (HPNs) for multi-level representation of video data. Each level consists of an augmented Petri-Net structure. Although it showed that this representation was able to capture video data semantics, from very low level semantics such as scene change and object movements, to higher level semantics, involving high-level textual description of events, the model is too complex to apply in soccer video, especially when relying on the efficiency of object detection. The BSUCPN approach proposed in this paper overcomes these problems giving a less complex model in which existing Petri-Nets for semantic events can easily be modified by the user or new ones added.

This paper begins by analyzing soccer video characteristics in order to identify suitable BSUs, these are then used in Composite Petri-Net models to design a description model of soccer semantic queries, and then describes example query BSUCPNs including goal, attack, corner, foul and substitution. Finally we illustrate the validity of this model using experiments on recorded soccer videos.

2. SOCCER VIDEO CHARACTERISTICS

Sports can be classified into a number of genres including ball games, athletics track and field, aquatic, ice-sports, and fight sports. It is our observation that more than half of sports videos are some form of ball game. Although there are many notable characteristics in ball games in general, here we concentrate on the specific details of soccer. These include:

- Video production is characterized by the use of a limited number of cameras at fixed positions. For example, in

Figure 1, a limited number of cameras are distributed on three sides around the field of play (the “pitch”).

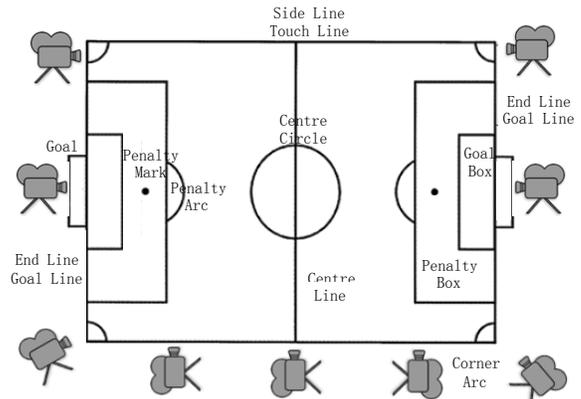


Figure 1. Cameras positions around a soccer pitch

- The techniques of shooting and playback (such as camera motion & change, focus zoom, slow motion etc.) are tied closely to the semantics of the game. For example, when a goal is scored a consistent and predictable sequence of shots is observed: from “Goal Area” to “Player Close-Up”, “Out-of-Field”, “Slow Motion Replay”, and another player “Close-Up”, and finally returning to “Infield Long” view (see Figure 2).

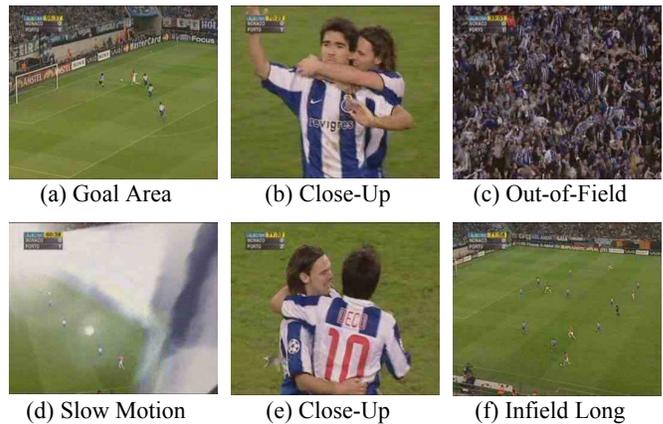


Figure 2. Soccer video clips playing order of Goal

- The shot, object and audio classes of ball games are different from heterogeneous video sources such as news video and movies. In soccer video we observe just three distinct shot classes “Long view” (including Infield and Out-of-Field), “Infield Medium”, and “Close-Up” (see Figures 5, 8, 9).
- Within this limited number of shot classes there are many variations, e.g. camera focus, the area of playing field in view, objects, and so on. These shot classes and their contents are then available for query description of soccer videos (see Figures 7, 10).

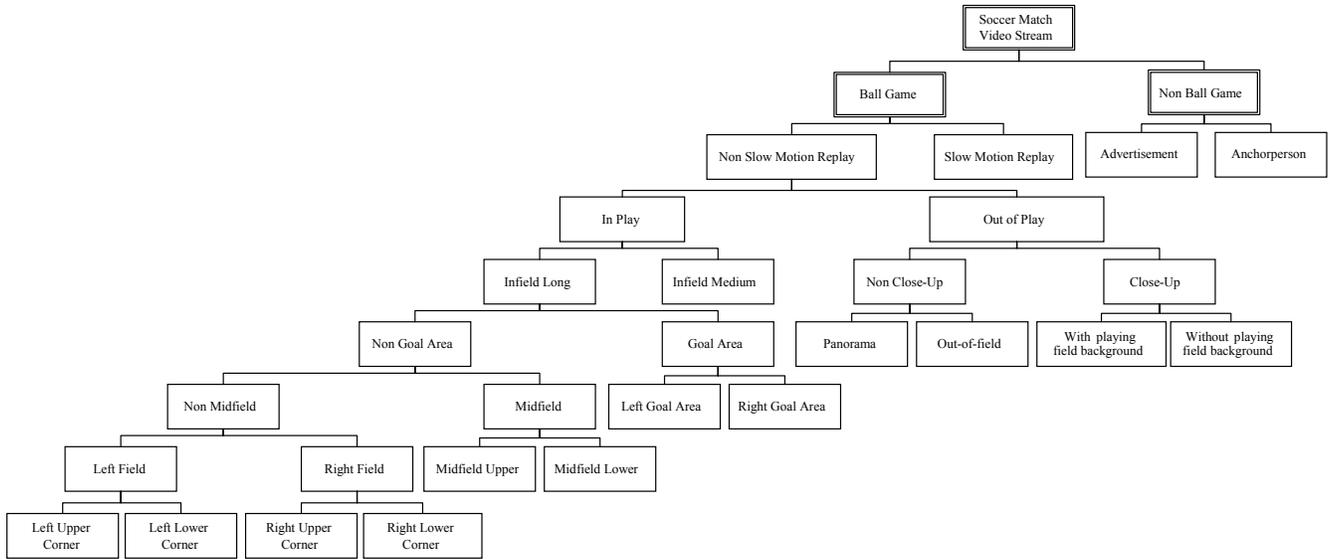


Figure 3. Hierarchy bintree of shot BSUs

The limited range of shot types and their contents can be used to define a set of BSUs describing simple and compound semantics of events in a soccer match. In summary, firstly a fixed set of BSUs can be defined for soccer videos, and secondly these BSUs can be used to describe both simple and compound semantics of soccer videos. The next section analyses the available BSUs for soccer event description.

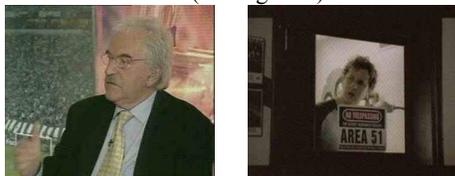
3. SOCCER VIDEO BSUs

Soccer videos are composed of a regular set of units which are used by editors in a reliable periodical manner which can help users to understand soccer semantics. Every BSU within a soccer video is a video shot or object, or an audio event which has a reliable semantic interpretation. Soccer video BSUs can be classified as: shot BSUs, object BSUs and audio BSUs. In this section we outline the characteristics of these BSU types.

3.1 Shot BSUs

Using shot level BSUs we can define a hierarchy bintree tree to describe soccer shot BSUs (see Figure 3). In this tree the root nodes are aggregative semantics of the subtrees and leaf nodes are shot BSUs. The leaf nodes consist of the soccer shot BSU set. Various classes of node in the subtrees can be defined as follows:

BSUs of Non Ball Game For soccer video BSUs of “Non Ball Game” shots include interview shots in the soccer live broadcast (for example studio discussion between an anchorperson and guests), and advertisements (see Figure 4).

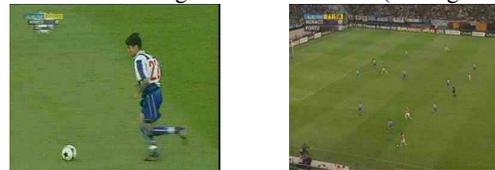


(a) Anchorperson (b) Advertisement
Figure 4. BSUs of Non Ball Game

BSUs of Ball Game BSUs of “Ball Game” can be divided into BSUs of “Non Slow Motion Play” and “Slow Motion Replay”.

BSUs of Non Slow Motion Play BSUs of “Non Slow Motion Play” can be divided into BSUs of “In-Play” and “Out-of-Play”.

BSUs of In-Play During a soccer match when the ball is in play the camera view mainly focuses on the field with an “Infield Medium” or “Infield Long” view of the action (see Figure 5).



(a) Infield Medium (b) Infield Long

Figure 5. BSUs of In Play

Based on the defined areas of the playing area (the “pitch”), the “Infield Long” shot can be divided into four classes and eight areas (see Figure 6):

- **Goal Area:** “Goal Area” includes “Left Goal Area” and “Right Goal Area” (see Figure 7(a), 7(b)). The meaning of left and right is the position and motion of the main cameras relative to the playing field (see Figure 1).
- **Midfield:** “Midfield” is composed of “Midfield Upper” and “Midfield Lower” (see Figure 7(c), 7(d)). The upper and lower areas are tracked by the moving position of the main cameras.
- **Left Field:** See Figure 7(e), 7(f), this consists of “Left Upper Corner” and “Left Lower Corner”. The definitions of left, upper & lower are the same as above.
- **Right Field:** See Figure 7(g), 7(h), this includes “Right Upper Corner” and “Right Lower Corner”. The definitions of right, upper & lower are the same as above.

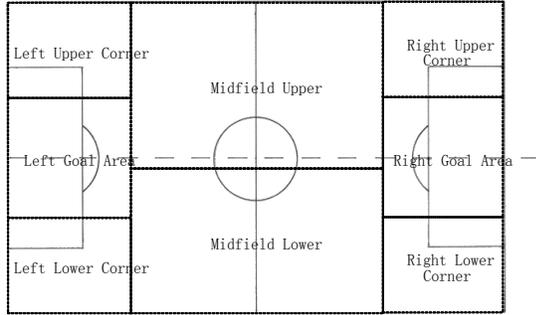


Figure 6. Eight areas of a soccer field

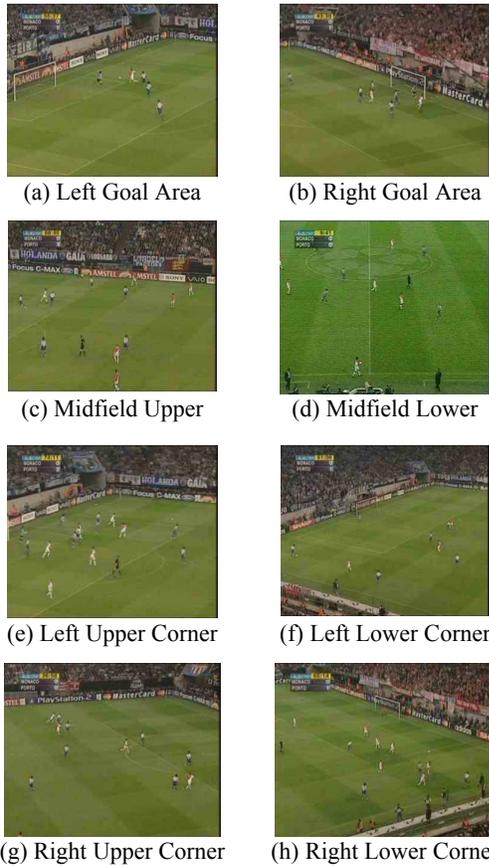


Figure 7. Eight areas of soccer field

BSUs of Out-of-Play BSUs of “Out-of-Play” include BSUs of “Close-Up” and “Non Close-Up”. Soccer is a fast moving game so close up shots always represent a break in play, for example after a goal or foul, before a kick-off, a corner, a goal kick, a free kick, or during substitution of players. The main objects in close-up shots are players and referees. Another kind of close-up is “Out-of-Field” focusing on spectators or coaches. For shot detection based on extended features not objects, BSUs of “Close-Up” can be classified “Close-Up with Playing Field Background” (Figure 8(a)) and “Close-Up without Playing Field Background” (Figure 8(b)). BSUs of “Non Close-Up” consist of two kinds shot: “Panorama” (Figure 9(a)) and “Out-of-Field” (Figure 9(b)).



(a) Close-Up With Playing Field Background

(b) Close-Up Without Playing Field Background

Figure 8. BSUs of Close-Up



(a) Panorama

(b) Out-of-Field

Figure 9. BSUs of Non Close-Up

3.2 Object BSUs

Only a limited number of object types are observed in soccer videos. These objects include: ball, player, referee (and assistant referees), coach, goal, sideline, corner arc (and other field markings), and on screen captions.

Automatic object detection is a much more difficult task than shot detection. For this reason in this paper we only select four types objects: caption, goal, corner arc and referee (see Figure 10). These can be detected reliably in soccer video [10], [11] and are thus available to be described in query semantics. In order to further increase the reliability of object recognition we introduce the following constraints:

- Caption: only detect captions in the lower part of the screen.
- Goal: only detect in shots of “Infield Long” and “Slow Motion Replay” BSU.
- Corner Arc: only detect in shots of “Infield Medium” BSU.
- Referee: only detect in shots of “Close-Up” BSU.



(a) Caption

(b) Goal



(c) Corner Arc

(d) Referee

Figure 10. Object BSUs

We illustrate later that these object BSUs can be used effectively in the description of significant events in soccer video.

3.3 Audio BSUs

Audio BSUs are useful for soccer semantic description. In general, in a soccer match there are three kinds of audio: pure speech, speech with crowd noise and pure crowd noise. After audio processing, soccer audio can be separated into two tracks: one track for high speech and low crowd noise, and one for high crowd noise with low or absent speech. In each track there are two BSUs: “High Energy” and “Low Energy” (relative to a bound range).

4. DEFINITION OF BSUCPN

A Petri-Net [6] is a graphical modeling tool. It consists of *places*, *transitions*, and *arcs* that connect them. *Input arcs* connect places with transitions, while *output arcs* start at a transition and end at a place. Places can contain *tokens*. The current state of the modeled system (the *marking*) is given by the number (and type if the tokens are distinguishable) of tokens in each place. Transitions are active components which model possible activities (the transition *fires*), thus changing the state of the system (the marking of the Petri-Net). Transitions are only allowed to fire if they are *enabled*, which means that all the preconditions for the activity must be fulfilled (there are enough tokens available in the input places). When the transition fires, it removes tokens from its input places and adds some at all of its output places. Based on the definition of BSUs in the previous section, we define a Basic Semantic Unit Compose Petri-Net (BSUCPN) model.

4.1 BSUCPN

Formally, a BSUCPN is defined as follows:

Definition 1 A BSUCPN is a 7-tuple

$$C_{BSUCPN} = \{T, P, A, Op, Du, Bsu, Ma\}$$

where

$T = (t_1, t_2, \dots, t_n)$ is a finite set of places with $n \geq 0$,

$P = (p_1, p_2, \dots, p_m)$ is a finite set of transitions with $m \geq 0$ and $P \cap T = \emptyset$,

$A: \{T \times P\} \cup \{P \times T\} \rightarrow I$, $I = \{1, 2, \dots\}$ is a set of arcs from places to transitions and from transitions to places,

$Op: T \rightarrow \{and, or, following^i, before^i, synchronization\}$ is a mapping from transition to its operators set where $i = \{1, 2, \dots\}$

$Du: P \rightarrow (Operate, Time)$ is a mapping from place to its duration where $Operate = \{\leq, =, \geq\}$, $Time \in R$

$Bsu: P \rightarrow \{null, A_{Bsu}, VO_{Bsu}, VS_{Bsu}\}$ is a mapping from place to its BSU set

$Ma: P \rightarrow I'$, $I' = \{0, 1, 2, \dots\}$ is a mapping from place to its marking

4.2 Logic Operator Set

For the BSUCPN model we define three classes of operator: and/or logic, sequence logic and synchronization logic.

- The and/or logic: the “and” logic of a transition is defined such that the transition is only allowed to fire if all input arcs are enabled. In other words it can remove the token from its input places and add some at all of its output places.

The fire precondition “or” logic is that when any input arc is enabled, the transition is fired.

- The sequence logic: video is different from text since there are temporal relations between BSUs, so we define before and following (n) as *tokens* to describe the sequence of shot BSUs.
- The synchronization logic: we use the synchronization logic to define the synchronization relationship of different kinds of BSU (such as among shot BSUs, object BSUs and audio BSUs). There are 13 synchronization relationships between two BSUs: equal, before, meet, during, overlap, start, finish and the six corresponding opposite relationships.

4.3 BSU Set

The BSU set includes shot BSU set, object BSU set and audio BSU sets as defined in Section 3, and also a “Null” BSU which is introduced in order to describe sequences and synchronization relationships more easily. In addition, we extend root nodes into the shot BSU set to simplify the query description model. The BSU Set is defined as follows:

Definition 2 A BSU set consists of 3-subsets and “Null”

$$BSU = \{Null, A_{BSU}, VO_{BSU}, VS_{BSU}\}$$

where

$A_{BSU} = \{hes, les, hecn, lecn\}$ is the audio BSU set, where *hes*, *les*, *hecn* and *lecn* are represented the element of “High Energy Speech”, “Low Energy Speech”, “High Energy Crowd Noise” and “Low Energy Crowd Noise”.

$VO_{BSU} = \{caption, referee, goal, corner_arc\}$ is the object BSU set which include four object elements.

$VS_{BSU} = \{LNVS_{BSU}, RNVS_{BSU}\}$ is the shot BSU set, where

$VS_{BSU} = \{luc, llc, ruc, rlc, mu, ml, lga, rga, pnr, of, im, cb, cnb, sm, adv, anc\}$ is a subset of shot BSU which include leaf nodes in bintree, where *luc*, *llc*, *ruc*, *rlc*, *mu*, *ml*, *lga*, *rga*, *pnr*, *of*, *im*, *cb*, *cnb*, *sm*, *adv*, *anc* represent the element of “Left Upper Corner”, “Left Lower Corner”, “Right Upper Corner”, “Right Lower Corner”, “Midfield Upper”, “Midfield Lower”, “Left Goal Area”, “Right Goal Area”, “Panorama”, “Out-of-Field”, “Infield Medium”, “Close-Up with Field Background”, “Close-Up without Field Background”, “Slow Motion”, “Advertisement” and “Anchorperson”.

$VS_{BSU} = \{lf, rf, nm, m, nga, ga, il, nc, c, ip, op\}$ is a subset of shot BSU which include root nodes in bintree, where *lf*, *rf*, *nm*, *m*, *nga*, *ga*, *il*, *nc*, *c*, *ip*, *op* represent the element of “Left Field”, “Right Field”, “Non Midfield”, “Midfield”, “Non Goal Area”, “Goal Area”, “Infield Long”, “Non Close-Up”, “Close-Up”, “In Play” and “Out of Play”.

Null is an element which is a short circuit of two transitions or used to describe sequences and synchronization relationships.

4.4 Duration

Duration is defined as the bound period range of a BSU instantiation as follows:

Definition 3 The duration of BSU is a 2-tuple

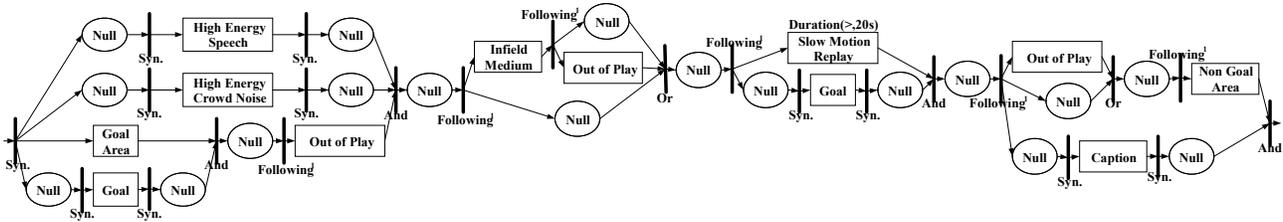


Figure 11. BSUCPN description of a “goal”

$Du(\text{Operator}, \text{Time})$ where $\text{Operator} = \{\leq, =, \geq\}$, $\text{Time} \in R$

The duration of a BSU is an important parameter in the query model. For example, in the semantic of substitution, the duration of shot BSU “Close-Up” is very long.

5. SOCCER SEMANTIC DESCRIPTION BASED ON BSUCPN MODEL

Four matches of soccer video (three European UEFA Champions League and one U.K. FA Cup) were used to design BSUCPN models for 5 soccer semantic events. The events described are: goal, foul, substitution, corner and open play attack. A particular strength of this approach is that the user can modify or define new BSUCPNs to describe soccer semantic events based on their knowledge. For example, the user could define a different “goal” BSUCPN from the one used here.

5.1 Goal

The BSUCPN of a “goal” is shown in Figure 11. At first there is a “goal area” shot, and then a long time “Out-of-Play or Out-of-Field” shot with “High Energy Speech” and “High Energy Crowd Noise”. After this there is a long-playing “Slow Motion Replay” (more than 30seconds) in which there is a “goal” object. Then the shot switches to “Out-of-Play or Out-of-Field” or “Non Goal Area” in which the object of “Caption” is presented. According to the semantic description of a “goal”, we can easily get the semantic of “left goal” and “right goal”.

5.2 Foul

The BSUCPN of a “foul” is described in Figure 12. A “Non Goal Area” is followed by a “Slow Motion Replay” or a “Close-Up” with the object of “Referee” or “Caption”.

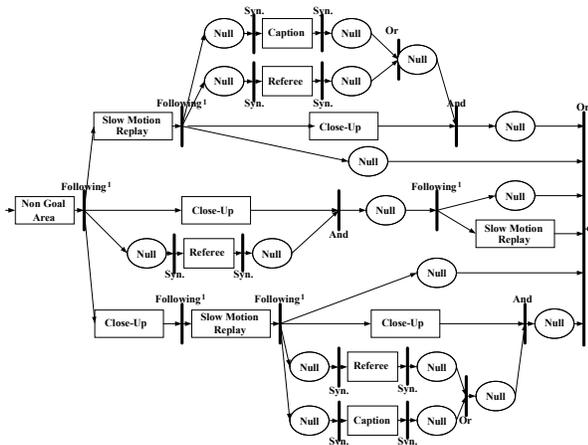


Figure 12. BSUCPN description of a “foul”

5.3 Substitution

The BSUCPN of a “substitution” is described in Figure 13. This comprises a long-playing “Close-Up” (more than 30 seconds) with the object of “Caption”.

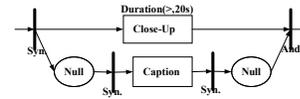


Figure 13. BSUCPN description of a “substitution”

5.4 Corner

The BSUCPN of a “corner” appears in Figure 14. After an “Infield Medium” shot with the object of “Corner Arc”, there is a “Goal Area” shot. According to the semantics of “corner”, we can easily get the semantics of “left lower corner”, “left upper corner”, “right lower corner” and “right upper corner”.

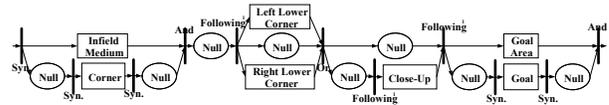


Figure 14. BSUCPN description of a “corner”

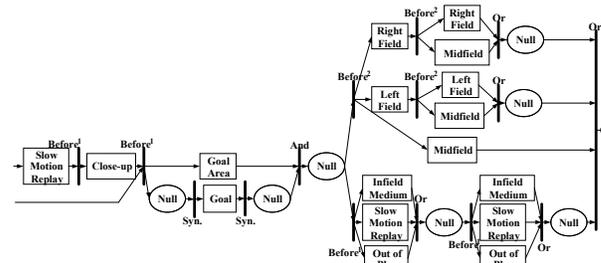


Figure 15. BSUCPN description of an “attack”

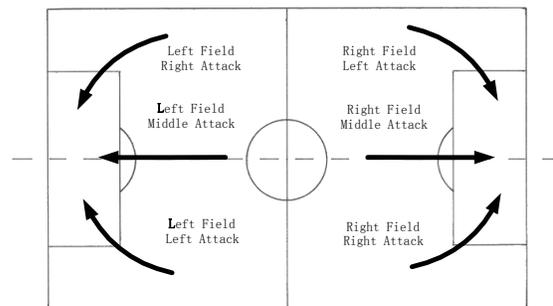


Figure 16. Six attack routes

5.5 Attack

The BSUCPN of an “attack” is shown in Figure 15. Before “Goal Area” there is “Left Field” or “Right Field” or “Midfield”, otherwise there is “Infield Medium” or “Out of Play” or “Slow Motion Replay”. The former indicates an attack in open play (including left, right and middle route attack), the latter represents a place kick attack (including corner kick, goal kick, penalty kick, free kick, etc.). According to the semantics of “attack”, we can easily get the semantics of “left goal area place kick”, “right goal area place kick”, “left field left route attack”, “left field right route attack”, “left field middle route attack”, “right field left route attack”, “right field right route attack” and “right field middle route attack”. The attack route of the latter six semantics is shown in Figure 16. The interpretation of left, right and middle route is made from the perspective of the attacking player.

6. EXPERIMENT AND RESULTS

6.1 Video Source of Experiments

In order to demonstrate our approach to identifying semantic events in soccer video we conducted a set of preliminary set experiments. These were carried out using a number of soccer match recordings captured from 4:2:2 YUV PAL tapes, which have a 375x288 resolution for each frame and a 50Hz field rate and save as MPEG1 files. Details are shown in Table 1.

Table 1. Video source of experiments

ID	Name of Match	Broadcast	A VS. B	Date	Duration
1	UEFA Champions League Final	ITV	FC Porto Vs. AS Monaco FC	26-May-2004	1:48:32
2	FA Cup Final	BBC1	Manchester Utd Vs. Millwall	22-May-2004	1:47:20
3	UEFA Champions League	Network2	Chelsea Vs. AS Monaco FC	5-May-2004	1:47:38
4	UEFA Champions League	Network2	FC Porto Vs. RC Deportivo La	4-May-2004	1:48:10

The soccer videos are from a range of broadcasters (ITV, BBC1 and Network2), and are taken from two competitions (European UEFA Champions League and the U.K. FA Cup), and thus capture a variety of editorial practices associated with different broadcasters and competitions. The BSUCPN models were developed initially using broadcast 1 with some subsequent minor adjustments based on the other three.

6.2 Query Process

We set up a table for each BSU type. Thus there are three tables: shot BSU table, object BSU table and audio BSU table. The data structure of each table is shown in Figure 17.

In the four matches all BSUs are written into a separate BSU database for automatic or manual annotation. The experiments described here used the manually annotated BSU database. This was used in order to eliminate effects of errors in automatic annotation, since in this paper we focus on demonstrating the validity of the BSUCPN model.

The query process is shown in Figure 18. The system first transforms the user query described by the BSUCPN model into SQL sentences (Figure 19 is an example of the “Substitution” SQL generated from the BSUCPN model), and then uses an SQL search engine to search in the three BSU tables. Lastly based on to the results of the BSU list, the system returns the relevant soccer video clips to the user. The query process using BSUCPNs is thus an exact match, similar to other video search systems.

```

BSU Structure{
  Integer BSUID;
  Integer VideoID; //Define BSU belong to which Video file
  Integer BSUNmae; //Name of BSU
  Byte BSUType; //Type of BSU, 0: shot BSU, 1:object BSU, 2: audio BSU
  Time BSUStartTime; //Start time of BSU
  Time BSUEndTime; //End time of BSU
  Time Duration;
  Integer GroupNo;
}

```

Figure 17. Data structure of BSU

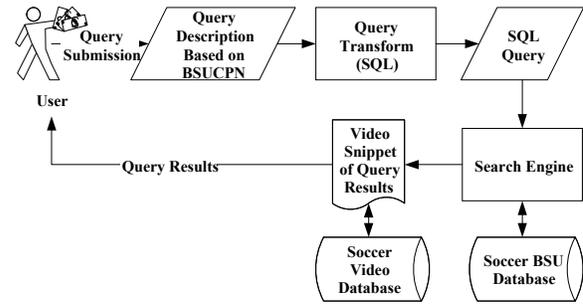


Figure 18. BSUCPN query process

```

//Select Shots of Close-Up
Q1=Select BSUID, VideoID, BSUStartTime, BSUEndTime, Duration, GroupNo
  From ShotBSU
  Where BSUName=cb or BSUName=cnb
//Set GroupNo of every substitution video clip
Select COUNT(*) "Count" FROM Q1
i=0
GroupNo=1
Do
  UPDATE Q1 SET Q1[i].GroupNo = GroupNo
  If Q1[i]. BSUEndTime <> Q1[i+1]. BSUStartTime Then GroupNo++
  i++
While (i=Count)
//Get the Start Time & End Time of every substitution video clip
Q2=Select Min(Q1. BSUStartTime), Max(Q1. BSUEndTime)
  From Q1 Group By Q1.VideoID, Q1.GroupNo
  Having Sum(Duration)>20

```

Figure 19. Program of SQL

6.3 Experimental Results

Table 2 shows the time and percentage of each shot BSU type in the four matches. It can be seen that about 60% of the time belongs to “In-Play”, and that half of it is taken up by open play. In other words, the semantic events defined in the previous section occupy 20% of the duration of a game.

Table 3 shows “Precision” and “Recall” for detection of the semantic events. “Actual Num” is the actual number of events in whole matches; “True Num” is the number of detected correct

matches, and “False Num” is the number of false matches. From Table 3, it can be seen that most results are higher than 80%, except “Substitution”, and that the recall results are generally better than those for precision.

Table 2. The duration and BSU events percentage

ID	Shot BSU Name (Leaf Nodes)	Percentage of Duration	Percentage of BSU Events
1	Anchorperson	5.12%	0.31%
2	Advertisement	12.21%	0.42%
3	Slow Motion Replay	8.01%	3.74%
4	Infield Medium	8.60%	14.97%
5	Panorama	0.84%	0.52%
6	Out-of-Field	3.46%	2.60%
7	Close-Up with playing field background	5.76%	7.59%
8	Close-Up without playing field background	15.26%	15.07%
9	Left Goal Area	2.07%	4.47%
10	Right Goal Area	0.53%	1.66%
11	Midfield Upper	17.34%	15.49%
12	Midfield Lower	9.07%	11.33%
13	Left Upper Corner	2.52%	6.03%
14	Left Lower Corner	2.69%	4.89%
15	Right Upper Corner	2.34%	4.47%
16	Right Lower Corner	4.19%	6.44%

Table 3. Precision and recall for five soccer semantics

Semantic	Actual Num	True Num	False Num	Precision (%)	Recall (%)
Goal	10	10	2	83	100
Foul	95	88	11	82	93
Substitution	16	11	5	52	69
Corner	40	36	0	90	90
Attack	215	215	10	96	100
Average				81	90

Table 4. Reasons of false matched & missed

Semantic	Reasons of False	Reasons of lost
Goal	Replay of shot	None
Foul	Replay previously shooting or goal	Foul in goal area (included in an attack)
Substitution	Player is injured or before start of game or disputed foul	Duration is too short
Corner	None	No corner arc object
Attack	Pass back to goalie or goal kick	None

Table 4 lists the reasons for false matches and missed events. From this we can see that sometimes it is difficult to distinguish a goal and replay of a shot on goal, substitution and injured player or disputed foul, an attack and passing back to goalkeeper or goal kick.

Based on the above experimental results, we believe that this approach to searching in sports video has considerable potential. We are currently conducting a more thorough experimental investigation using a larger set of independent videos.

7. CONCLUSIONS AND DISCUSSIONS

In this paper, based on analyzing of soccer video characteristics, we proposed a query description model for soccer video using Basic Semantic Unit Composite Petri-Nets (BSUCPN). The effectiveness of this model was demonstrated through a preliminary experiment.

Our method can be utilized for other ball games, and even for other sports video which demonstrate the same characteristics as ball game video. Future work will explore interface design which is very important to enable rapid development of new queries. Based on this work, we plan to design and implement a sports video query and browsing system based on our Fischlár digital video retrieval [12] system.

8. REFERENCES

- [1] FIFA, *Laws of the Game*, <http://www.fifa.com/en/game/laws>.
- [2] D. Yow, B. L. Yeo, M. Yeung, et al, Analysis and Presentation of Soccer Highlights from Digital Video. *In Proceedings of Asian Conference of Computer Vision*, 1995.
- [3] Y. Gong, L. T. Sin, C. H. Chuan, et al. Automatic Parsing of TV Soccer Programs. *In IEEE International Conference on Multimedia Computing and Systems*, Washington D.C, 1995.
- [4] S. Nepal, U. Srinivasan, G. Reynolds. Automatic Detection of 'Goal' Segments in Basketball Videos. *In Proceedings of ACM Multimedia '01*, Canada, 2001.
- [5] G. Sudhir, J. C. M. Lee, A. K. Jain. Automatic Classification of Tennis Video for High-Level Content-Based Retrieval. *In IEEE International Workshop on Content-based Access of Image and Video Database*, Bombay, India, Jan, 1998.
- [6] Peterson, James Lyle. Petri Net: Theory and the Modeling of Systems. Englewood Cliffs, N.J: Prentice-Hall, 1981.J.Y.
- [7] W. Al-Khatib, A. Ghafoor, An Approach for Video Meta-Data Modeling and Query Processing. *In Proceedings of the seventh ACM international conference on Multimedia*, Orlando, Florida, USA, October, 1999, 215 - 224.
- [8] T. D. C. Little, G. Ahanger, R. J. Folz, et al, A Digital On-Demand Video Service Supporting Content-Based Queries. *In Proceedings of the first ACM international conference on Multimedia*, Anaheim, California, USA, September, 1993.
- [9] A.F. Smeaton and A Gregan, Distributed Multimedia QoS Parameters from Presentation Modelling by Coloured Petri Nets. *Lecture Notes in Computer Science; Multimedia, Hypermedia, and Virtual Reality: Models, Systems, and Applications; 1st International Conference, MHVR'94*, Moscow, Russia, September, 1994, 47 – 60.
- [10] J.Y. Chen, Y.H. Li, S.Y. Lao, et al, *Detection of Scoring Event in Soccer Video for Highlight Generation*. Technical Report, National University of Defense Technology, 2004.
- [11] J.Y.Chen, Y.H. Li, L.D. Wu, et al, A Unified Framework for Semantic Content Analysis in Sports Video. *In Proceedings of the 2nd International Conference on Information Technology for Application*, Haerbin , P.R. China, Jan, 2004.
- [12] A.F. Smeaton, et al., The Fischlár-News-Stories System: Personalised Access to an Archive of TV News. *RIA0 2004*, Avignon, France, 26-28, April, 2004.