



COVER SHEET

This is the author-version of article published as:

Tjondronegoro, Dian and Chen, Yi-Ping Phoebe and Pham, Binh (2006) Extensible Detection and Indexing of Highlight Events in Broadcasted Sports Video. In Estivill-Castro, Vladimir and Dobbie, Gill, Eds. *Proceedings Australasian Conference on Computer Science*, Hobart. Australia.

Accessed from <http://eprints.qut.edu.au>

Copyright Australian Computer Society 2006

Extensible Detection and Indexing of Highlight Events in Broadcasted Sports Video

Dian W. Tjondronegoro¹, Yi-Ping Phoebe Chen², Binh Pham³

¹ School of Information Systems, Queensland University of Technology, Brisbane, Australia

² School of Information Technology, Deakin University, Melbourne, Australia

³ Faculty of Information Technology, Queensland University of Technology, Brisbane, Australia

dian@qut.edu.au, phoebe@deakin.edu.au, b.pham@qut.edu.au

Abstract

Content-based indexing is fundamental to support and sustain the ongoing growth of broadcasted sports video. The main challenge is to design extensible frameworks to detect and index highlight events. This paper presents: 1) A statistical-driven event detection approach that utilizes a minimum amount of manual knowledge and is based on a universal scope-of-detection and audio-visual features; 2) A semi-schema-based indexing that combines the benefits of schema-based modeling to ensure that the video indexes are valid at all time without manual checking, and schema-less modeling to allow several passes of instantiation in which additional elements can be declared. To demonstrate the performance of the events detection, a large dataset of sport videos with a total of around 15 hours including soccer, basketball and Australian football is used.

Keywords: Extensible sports video indexing, multi-modal event detection

1 Introduction

Sports video indexing approaches can be categorised based on low-level (perceptual) *features* and high-level *semantic* annotation (Djeraba, 2002). There are some elements beyond perceptual level (known as the *semantic gaps*) which can make feature based-indexing tedious and inaccurate. For example, users cannot always describe the visual characteristics of certain objects they want to view for each query. In contrast, the main benefit of *semantic-based indexing* is the ability to support more intuitive queries. However, semantic annotation is generally time-consuming, and often incomplete due to the limitations of manual supervision and the currently available techniques for automatic semantic extraction. Therefore, video should be indexed using semantic that can be extracted automatically with minimal human intervention. Events-based indexing can be noted as the most suitable indexing technique for sport videos as sport highlights on TV, magazine or internet are commonly described using a set of events, particularly the important or exciting ones.

As there is yet a complete solution that can extract all events automatically, we need to design frameworks that support extensible detection and indexing of (highlight) events. Extensibility is emphasized as the algorithms developed for automatic extraction of features and semantic in sports video need to be extended gradually while improving the performance. As a result of more extractable contents, the indexing scheme needs to support continuous updates. The first and second section of this paper addresses each of these issues respectively. Following this, the experimental results that use a large dataset are reported before we close with some conclusions and future work.

2 Extensible Events Detection

It has become a well-known theory that sports events can be detected based on the occurrences of specific audio and visual features which can be extracted automatically. To date, there are two main approaches to fuse audio-visual features. One alternative, called *machine-learning* approach, uses probabilistic models to automatically capture the unique patterns of audio visual feature-measurements in specific (highlight) events. For example, Hidden Markov Model (HMM) can be trained to capture the transitions of *still*, *standing*, *walking*, *throwing*, *jumping-down* and *running-down* states during athletic sports' events (Wu et al., 2002). The main benefit of using such approach is the potential robustness, thanks to the modest usage of domain-specific knowledge which is only needed to select the best features set to describe each event. However, one of the most challenging requirements for constructing reliable models is to use features that can be detected flawlessly during training due to the absence of manual supervision. Moreover, adding a new feature into a particular model will require re-training of the whole model. Thus, it is generally difficult to build extensible models that allow gradual development or improvement in the feature extraction algorithms. To tackle this limitation, our statistical-driven models are constructed based on the characteristics of each feature. Any addition of a new feature will only result on the updates of the rules that were associated with that feature.

Another alternative for audio-visual fusion is to use manual heuristic rules. For example, the temporal gaps between specific features during basketball goal have a predictable pattern that can be perceived manually (Nepal et al., 2001). The main benefit of this approach is the absence of comprehensive training for each highlight and

the computations are relatively less complex. However, this method usually relies on manual observations to construct the detection models for different events. Even though the numbers of domains and events of interest are limited and the amount of efforts is affordable, we primarily aim to reduce the subjectivity and limitation of manual decisions.

These two approaches still have two major drawbacks, namely, 1) the lack of a definitive solution for the scope of highlight detection such as where to start and finish the extraction. For example, Ekin et al (Ekin and Tekalp, 2003b) detect goals by examining the video-frames between the global shot that causes the goal and the global shot that shows the restart of the game. However, this template scope was not used to detect other events. On the other hand, Han et al (Han et al., 2003) used a static temporal-segment of 30-40 sec (empirical) for soccer highlights detection. 2) The lack of a universal set of features for detecting different highlights and across different sports. Features that best describe a highlight are selected using domain knowledge. For instance, whistle in soccer is only used to detect foul and offside, while excitement and goal-area are used to identify goal attempt (Duan et al., 2003).

In order to solve the first drawback, some approaches (Xu et al., 1998, Li and Ibrahim Sezan, 2001) have claimed that highlights are mainly contained in a *play* scene. However, based on a user study as reported in our earlier paper (Tjondronegoro et al., 2004b), we have found that most users need to watch the whole play and break to understand fully an event. For example, when a *whistle* is blown during a *play* in soccer video, we would expect that something has happened. During the break, the *close-up views* of the players, a *replay scene*, and/or the *text display* will confirm whether it was a *foul* or *offside*. Consequently, it is expected that automated semantic analysis should also need to use both *play* and *break* segments to detect highlights. As for the second drawback, we aim to reduce the amount of manual choice of features set. For instance, it is quite intuitive to decide that the most effective event-dependent features to describe a *soccer foul* are *whistle*, followed by *referee appearance*. However, we were able to identify some additional characteristics of *foul* that could be easily missed by manual observation such as shorter *duration* (compared to shoot) and less *excitement* (compared to foul), based on statistical features that will be discussed in section 2.2.

2.1 Play-Break as Standard Scope of Events

Most broadcasted sport videos use transitions of typical shot types to emphasize story boundaries while aiding important contents with additional items. For example, a long global shot is normally used to describe an attacking play that could end with scoring of a goal. After a goal is scored, zoom-in and close-up shots will be dominantly used to capture players and supporters celebration during the break. Subsequently, some slow-motion replay shots and artificial texts are usually inserted to add some additional contents to the goal highlight. Based on this example, it should be clear that play-break sequences

should be effective containers for a semantic content since they contain all the required details. Using this assumption, we should be able to extract all the phenomenal features from play-break that can be utilized for highlights detection. Thus, as shown in Figure 1, the scoping of highlight (event) detection should be from the last play-shot until the last break shot.

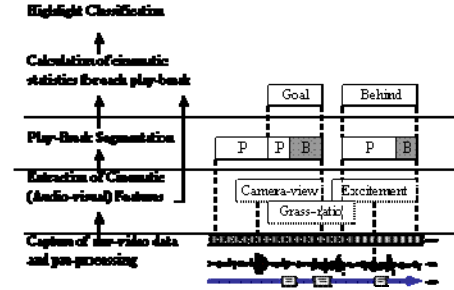


Figure 1. Extracting Events from Play-Break.

Analysis of *camera-views* transition in a sports video has been used successfully for play-break segmentation (Ekin and Tekalp, 2003a). We have extended this approach by adding *replay-based correction* to improve the performance. Figure 2 shows how a replay scene (R) can fix the boundaries of play-break sequences – which are formed by a sequential play scene (P) and break scene (B). Please note that “s” indicates start while “e” indicates end. For example, R.s is short for the start of replay scene.

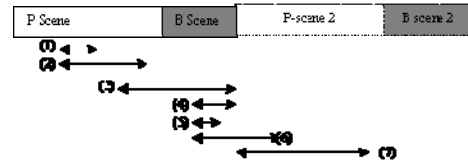


Figure 2. Locations of Replays in Play-breaks.

Based on these scenarios, an algorithm to perform replay-scene based play-break segmentation has been developed. This algorithm aims to: 1) fix the inaccurate boundaries of play-break sequences due to shorter breaks; 2) locate missing sequences due to missed breaks; and 3) avoid false sequences due to falsely detected play which is followed by a break.

Algorithm to fix play-break boundaries, based on replay scene locations

- | |
|---|
| <p>If $(A.s > B.s) \ \& \ (A.e < B.e)$
 <i>A strict_during B</i>
 If $(A.s > B.s \ \& \ A.e \leq B.e)$ OR $(A.s \geq B.s \ \& \ A.e < B.e)$
 <i>A during B</i>
 If $A.e = B.e$
 <i>A meets B</i></p> <ol style="list-style-type: none"> (1) If $[R \ \textit{strict_during} \ P] \ \& \ [(R.e - P.e) \geq \textit{dur_thres}]$
 $B.s = R.s; B.e = R.e;$ Create a new sequence where $[P_2.s = R.e+1] \ \& \ [P_2.e = P.e]$ (2) If $[R \ \textit{strict_during} \ P] \ \& \ [(R.e - P.e) \leq \textit{dur_thres}]$
 $P.e = R.e; B.s = R.e+1$ (3) If $[R \ \textit{meets} \ B] \ \& \ [R.s < P.e]$
 $P.e = R.s$ (4-5) If $[R \ \textit{during} \ B] \ \& \ [R \ \textit{meets} \ B]$) OR (If $[R \ \textit{strict_during} \ B]$)
 No processing required (6) If $[R \ \textit{during} \ B] \ \& \ [(R.e - P_2.s) \geq \textit{dur_thres}]$
 $B.e = R.e;$ Amend the neighbor sequence: $[P_2.s = R.e+1]$ (7) If $[R \ \textit{during} \ P_2] \ \& \ [(R.e - P_2.s) \geq \textit{dur_thres}]$
 Attach sequence 2 to sequence 1 (i.e. combine seq 1 and seq 2 into one sequence) |
|---|

It is important to note that some broadcasters insert some advertisements (*ads*) in-between or during the replay. To obtain the correct length of the total break, the total length of the ads has to be taken into account.

2.2 Statistical-Driven Events Detection

As most of the current cinematic-heuristics for highlight detection are heavily based on manual discoveries and domain-specific rules, we aim to minimize the amount of manual supervision in discovering the phenomenal features that exist in each of the different highlights. Moreover, in developing the rules for highlight detection, we should use as little domain knowledge as possible to make the framework more flexible for other sports with minimum adjustments. For this purpose, we have conducted a semi-supervised training from different broadcasters and different matches for each highlight to determine the characteristics of play-break sequences containing different highlights and no highlights. It is semi-supervised training as we manually classify the specific highlight that each play-break sequence contains. Moreover, the automatically detected play-break boundaries and mid-level features locations within each play-break such as excitement are manually checked to ensure the accuracy of training.

During training, statistics of each highlight are calculated with the following parameters (the examples are based on AFL video):

- SqD = duration of currently-observed play-break sequence. For example, we can predict that a sequence that contains a goal will be much longer than a sequence with no highlight.
- BrR = duration of break / SqD . Rather than measuring the length of a break to determine a highlight, the ratio of break segment within a sequence is more robust and descriptive. For example, we can distinguish goal from behind based on the fact that goal has a higher break ratio than behind due to a longer goal celebration and slow motion replay.
- PIR = duration of play scene / SqD . We find that most non-highlight sequences have the highest play ratio since they usually contain very short break.
- RpD = duration of (slow-motion) replay scene in the sequence. This measurement implicitly represents the number of replay shots which is generally hard to be determined due to many camera changes during a slow motion replay.
- $ExcR$ = duration of excitement / SqD . Typically, a goal consists of a very high excitement ratio whereas a non-highlight usually contains no excitement.
- NgR = duration of the frames containing goal-area/duration of play-break sequence. A high ratio of near goal area during a play potentially indicate goal.
- CuR = length of close-up views that includes crowd, stadium, and advertisements within the sequence / SqD . We find that the ratio of close-up views used in a sequence can predict the type of highlight. For example, goal and behind highlights generally has a

higher close-up views due to focusing on just one player such as the shooter and goal celebration.

The statistical data of the universal feature sets within each highlight after a training that uses 20 samples is presented in Table 1. Based on the trained statistics, we have constructed a novel set of ‘*statistical-driven*’ heuristics to detect soccer, AFL, and basketball highlights. We do not need to use any domain-specific knowledge, thereby making the approach less-subjective and robust when applied for similar sports. As each feature can be considered independently, more features can be introduced without the necessity to make major changes in the highlight classification rules. Moreover, our model does not need to be re-trained as a whole, thereby promoting extensibility. Hence, our approach will reap the full benefit when larger set of features are to be developed/improved gradually.

Highlight classification is performed as:

$$[HgtClass] = \text{Classify_Highlight}(D, NgR, ExcR, CuR, PIR, RpR)$$

where, $HgtClass$ is the highlight class most likely contained by the sequence, while D , NgR , and so on are the statistical parameters described earlier. This equation will be performed according to the sport genre.

In order to classify which highlight is contained in a sequence, the algorithm uses some *measurements*. For example, in soccer, G , S , F , and Non are the highlight-score for goal, shoot, foul and non-highlight respectively. Each of these measurements is incremented by 1 point when certain rules are met. Thus, users should be able to intuitively decide the most-likely highlight of each sequence based on the highest score. However, to reduce users’ workload, we can apply some post-processing to automate/assist their decision.

Feature	Soccer	AFL	Basketball
	G=Goal, S=Shoot, F=Foul, N=Non (avg; max; min)	G=Goal, B=Behind, M=Mark, T=Tackle, N=Non (avg; max; min)	G=Goal, F=Foul, FT=Free throw, T=Timeout (avg; max; min)
Duration (D)	Gd (73; 104; 43) Sd (36; 73; 10) Fd (38; 72; 14) Nd (24; 40; 5)	Gd (72; 120; 40) Bd (31; 53; 7) Md (26; 65; 8) Td (25; 63; 10) Nd (20; 42; 8)	Gd (24; 51.6; 9.6) Fd (28.8; 60; 12) FTd (20.4; 30; 11) Td (124.8; 255; 25)
Play Ratio (PIR)	Gp (0.30; 0.46; 0.07) Sp (0.57; 0.87; 0.15) Fp (0.64; 0.97; 0.08) Np (0.73; 0.91; 0.47)	Gp (0.17; 0.33; 0.06) Bp (0.38; 0.92; 0.10) Mp (0.62; 0.86; 0.26) Tp (0.55; 0.83; 0.08) Np (0.52; 0.81; 0.17)	Gp (0.71; 0.94; 0.27) Fp (0.48; 0.72; 0.13) FTp (0.50; 0.81; 0.23) Tp (0.12; 0.24; 0.05)
Near Goal (NgR)	Gn (0.47; 1; 0.13) Sn (0.55; 0.93; 0) Fn (0.23; 0.81; 0) Nn (0.17; 0.1; 0)	Gn (0.13; 0.43; 0.02) Bn (0.10; 0.39; 0.02) Mn (0.02; 0.23; 0) Tn (0.01; 0.05; 0) Nn (0.01; 0.08; 0)	Gn (0.49; 0.92; 0.04) Fn (0.43; 0.93; 0) FTn (0.55; 1; 0.05) Tn (0.34; 0.85; 0)
Excitement (ExcR)	Ge (0.45; 0.83; 0.10) Se (0.35; 0.79; 0) Fe (0.20; 0.50; 0) Ne (0.2; 0.6; 0)	Ge (0.29; 0.54; 0) Be (0.38; 0.86; 0) Me (0.32; 0.91; 0) Te (0.22; 0.59; 0) Ne (0.30; 0.75; 0)	Ge (0.41; 0.82; 0.05) Fe (0.34; 0.78; 0) FTe (0.44; 0.90; 0) Te (0.24; 0.43; 0.05)
Close-up (CuR)	Ge (0.26; 0.51; 0.08) Sc (0.23; 0.74; 0) Fc (0.12; 0.29; 0) Nc (0.2; 0.6; 0)	Ge (0.35; 0.86; 0) Bc (0.35; 0.76; 0) Mc (0.28; 0.56; 0) Tc (0.18; 0.44; 0) Nc (0.29; 0.69; 0)	Ge (0.11; 0.3; 0) Fc (0.27; 0.69; 0) FTc (0.26; 0.68; 0) Tc (0.49; 0.78; 0.16) Nc (0.2; 0.63; 0)
Replay (RpD)	Gr (25; 34; 20) Sr (6; 16; 0) Fr (6; 23; 0) Nr (0; 0; 0)	Gr (9; 23; 0) Br (6; 40; 0) Mr (1; 14; 0) Tr (4; 14; 0) Nr (0; 0; 0)	Gr (0; 0; 0) Fr (4.8; 13; 0) FTr (0; 0; 0) Tr (16; 40; 0)

Table 1. Statistics of Soccer, AFL, and Basketball Highlights.

The essence of highlight classification is on comparing the value of each input parameter against the typical statistical characteristics: min, avg, and max which are

denoted as a *stat*. The following algorithm describes the calculation that can be applied to any sport (using soccer as an example).

Common event classification algorithm

```
Let Det_Soccer_Region(val) = Region(val, statG, statS, statF, statN)

Perform
region1..n = Det_Soccer_Region(D), (NgR), (ExcR), (CuR), (PIR), (RpR)
For region1 to regionn
Increment the corresponding highlight score //G, Sh, F, Non in this case
```

where,

$$\text{Region}(val, stat_1, stat_2, \dots, stat_n) = \begin{cases} 1, & \text{if } (AvgD_1 \leq MinAvgD) \& (TD_1 \leq MinTD) \\ 2, & \text{if } (AvgD_2 \leq MinAvgD) \& (TD_2 \leq MinTD) \\ \dots \\ n, & \text{if } (AvgD_n \leq MinAvgD) \& (TD_n \leq MinTD) \end{cases}$$

$$\begin{aligned} stat_n &= \{avg_n, min_n, max_n\}, AvgD_n = |val - stat_n^{avg}|, \\ TD_n &= |val - stat_n^{max}| + |val - stat_n^{min}|, \\ MinAvgD &= \min(AvgD_1, AvgD_2, \dots, AvgD_n), MinTD = \min(TD_1, TD_2, \dots, TD_n). \end{aligned}$$

It is to be noted that in $\text{Det_soccer_region}(val), stat_x$ matches the value input. Therefore, when val is NgR , then $stat_G = \{Gn_avg, Gn_max, Gn_min\}$ is used according to the statistics-table.

In addition to the common algorithm, we can improve the accuracy of the event classification for a particular sport based on its statistical phenomena. This concept is described in the rest of this section.

2.2.1 Events Classification in Soccer

When play ratio, sequence duration and near goal ratio fall within the statistics of goal or shoot, it is likely that the sequence contains goal or shoot. Otherwise, we will usually find a foul or non-highlight. However, shoot often has similar characteristics with foul. In order to differentiate *goal* from *shoot*, and *shoot/foul* from *non-highlight*, we apply some statistical features:

- *Goal vs. Shoot*: Compared to shoot, goal has longer duration, more replays and more excitement. However, goal has shorter play scene due to the dominance of break during celebration.
- *Shoot, Foul, vs. Non-highlight (None)*: None does not contain any replay whereas foul contains longer replay than shoot in average. Foul has the lowest close-up ratio as compared to shoot and none. None has the shortest duration as compared to shoot and foul. None contains the least excitement as compared to shoot and foul, whereas foul has less excitement than shoot.

Based on these findings, the following algorithm is developed.

Specific algorithm to classify highlight events in soccer

```
Perform region1..3 = Det_Soccer_Region (PIR), (D), (NgR)
accordingly
If all region1, 2 and 3 = 1 or 2
//Most likely to be goal or shoot
Increment G and Sh
```

```
Perform region4..7 = Det_Soccer_Region(ExcR), (RpD), (PIR), (D)
For region4 to region7
If current region = 1, Increment G
Else if current region = 2, increment Sh
Else
//Most likely to be foul, shoot, or non
Increment F, Sh, Non
Perform region4..7 = Det_Soccer_Region (CuR), (ExcR), (D), (RpD)
For region4 to region7
If current region = 2, increment Sh
Else if current region = 3, Increment F
Else if current region = 4, increment Non
```

It should be noted that the more compact representation of this algorithm is presented in Figure 3, where $\{val\}$ is the convention of $\text{region}_{1..N} = \text{Det_Soccer_Region}(val_1), (val_2), \dots, (val_N)$. Thus, squares denote the statistics that need to be checked, whereas the non-boxed texts are the associated highlight point(s) that will be incremented based on the outputs of each region. This representation is used for describing other sports.

2.2.2 Events Classification in AFL

In AFL, a *goal* is scored when the ball is kicked completely over the *goal-line* by a player of the attacking team without being touched by any other player. A *behind* is scored when the football touches or passes over the goal post after being touched by another player, or the football passes completely over the behind-line. A *mark* is taken if a player catches or takes control of the football within the playing surface after it has been kicked by another player a distance of at least 15 meters and the ball has not touched the ground or been touched by another player. A *tackle* is when the attacking player is being forced to stop from moving because being held (tackled) by a player from the defensive team. Based on these definitions, it should be clear that goal is the hardest event to achieve. Thus, it will be celebrated longest and given greatest emphasis will be given by the broadcaster. Consequently, behind, mark and tackle can be listed in the order of its importance (i.e. behind is more interesting than mark).

Figure 4 shows the highlight classification rules for AFL. Let G, B, M, T, Non be the highlight-score for *goal, behind, mark, tackle* and *non-highlight* respectively. Thus, for AFL event detection:

$$\text{Det_AFL_Region}(val) = \text{Region}(val, stat_G, stat_B, stat_M, stat_T, stat_N)$$

The algorithm firstly checks that if current PIR belongs to $stat_G$ (i.e. output = 1) and NgR is greater than the minimum of the typical value for goal and behind, then the sequence is most likely to contain either goal or behind. This is followed by comparing: $ExcR, RpD$, and PIR values: the outputs determine which score is incremented from G or B .

Else (if PIR does not belong to $stat_G$), it is more likely to contain mark, tackle, or none. This is followed by comparing: D, CuR, PIR , and RpR values: the outputs determine which score is incremented from M, T , or N .

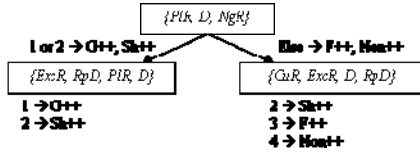


Figure 3. Highlight Classification Rules for Soccer

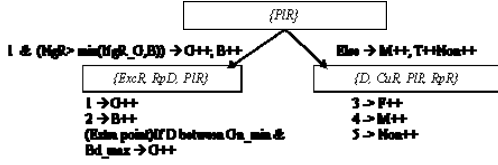


Figure 4. Highlight Classification Rules for AFL.

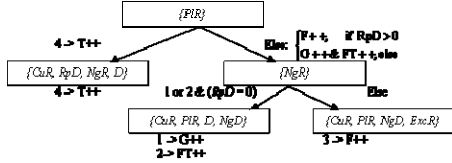


Figure 5. Highlight Classification Rules for Basketball

2.2.3 Events Classification in Basketball

Compared to soccer and AFL, goals in basketball are not celebrated and do not need a special resume such as kick off. Therefore, it is noted that the rules applied to soccer and AFL cannot be used directly for basketball goals.

Figure 5 shows the highlight classification rules for basketball. Let G , FT , F , T be the highlight-score for goal, free-throw, foul, and timeout respectively. Thus, for basketball event detection, let:

$$\text{Det_Basketball_Region}(val) = \text{Region}(val, \text{stat}_G, \text{stat}_{FT}, \text{stat}_F, \text{stat}_T)$$

The algorithm firstly checks if current PIR belongs to stat_T (i.e. output = 4), then the sequence is most likely to contain timeout. This is followed by comparing: Cur , RpD , NgR , and D values: each time that the output of comparison is equal to 4, T is further incremented.

Else (if current PIR does not belong to stat_T), it is more likely to contain goal, free-throw, or foul (if $RpD > 0$). This is followed by checking:

If NgR belongs to region stat_G or stat_{FT} (i.e. output = 1 or 2), then the comparison is based on the values of: CuR , PIR , D , and NgD : the outputs determine which score is incremented from G or FT .

Else, (if NgR does not belong to region stat_G or stat_{FT}), then the comparison is based on the values of: CuR , PIR , NgD , and $ExcR$: each time that the output of comparison is equal to 3, F is further incremented.

3 Extensible Indexing

For the indexing of events, OO modeling is recognized for its ability to support complex data definitions. We have identified two main alternatives in using O-O for modelling data based on the models presented in AVIS (Adali et al., 1996) and OVID (Oomoto and Tanaka, 1997), namely, schema-based and schema-less,. A schema-based model (Adali et al., 1996) can be

composed of three types of *entities* (i.e. index-able items) in a video database, namely, 1) *video objects*, which capture entities that present in the video frames, 2) *activity types*, which is the subject of a frame sequence, and 3) *event*, which is the instantiation of an activity type. Thus, their model has allowed users to query the location of the occurrence of their desired object or events. The main benefit of using a schema-based model is its capability to support easy updates due to the strict components that have to be followed exactly for each entity. However, the main limitation is the difficulty to include new description during instantiation of video models due to the static schema; therefore, the model is not extensible.

In contrast, schema-less modeling (Oomoto and Tanaka, 1997) is designed based on the fact that each video interval can be regarded as a video object, in which the attributes can be objects, events, or other video objects. Thus, the content of a video object is more dynamic. Moreover, they also proposed dynamic calculation of inheritance, overlap, merge and projection of intervals to satisfy user queries. However, there are two main problems of schema-less modelling. First, query difficulties arise as users/developers must inspect the attribute definition of each object to develop a query because each object has its own attribute structure. Second, the total dependency on users or applications for supervising the instantiation of video objects occurs due to the fact that a schema is not present.

In order to combine the strengths of schema-based and schema-less modelling, this section demonstrates the utilization of XML to design and construct a semi-schema based video model. Schema-based matching ensures that the video indexes are valid during data operations such as insertion, thereby minimizing the need of manual checking. However, the model is also semi-schema based as it allows additional declared elements in the instantiated objects as compared to its schema definition. Moreover, not all elements in an object need to be instantiated at one time as video content extraction often requires several passes due to the complexity and lengthy processing; thereby supporting an extensible modeling scheme. In addition to the strength of OO modeling, the video model also attempts to benefit from relational modeling scheme. In particular, the utilization of *referential integrity* (Connolly and Begg., 2002) allows an object to include elements which are referenced from the existing objects within the database. The main purpose is to reduce objects being added within another object(s), thereby avoiding complex hierarchies and potential redundancies. Hence, in overall, the proposed video model supports object-relational modeling approach while adopting semi-schema based index construction and maintenance.

The sport video indexing is designed using two main abstraction classes, namely, *segment* and *event*. Each segment is instantiated with a unique key of segment Id into either: video-, visual-, or audio-segment. A segment, as shown in Figure 7, can be instantiated as video-, audio- or visual-segment which are extracted from a raw video track when mid-level features (e.g. whistle and

excitement) can be detected. An event can be instantiated into generic (e.g. interesting event), domain-specific (e.g. soccer goal), or further-tactical (e.g. soccer free kick) semantics. Events and segments are chosen as they can provide an effective description for many sport games. For example, most users will benefit from watching soccer goals as the most celebrated and exciting event. Segments are used as the text-alternative annotations to describe the goal. As shown in Figure 6, the last near-goal segment in a play-break sequence containing goal describe *how* the goal was scored. Face and text displays can inform *who* scored the goal (i.e. the actor of the event) and the updated score. Replay scene shows the goal from different angles to *further emphasize* the details of how the goal is scored. In most cases, when the replay scene is associated with excitement, the content is more important. Excitement during the last play shot in a goal is usually associated with descriptive narration about the goal. In fact, we (human) often can hear a goal without actually seeing it.

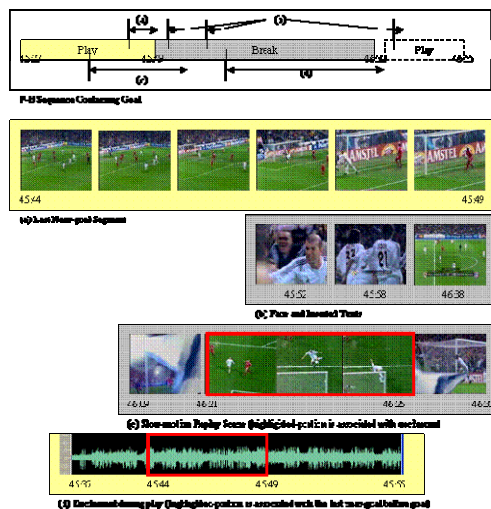


Figure 6. Goal Event with Segment-Based Annotations.

We have utilized some of the main benefits from using XML to store and index the extracted information from sport videos:

- XML is extensible by allowing additional information without affecting others. This is important to support gradual developments of feature extraction techniques that can add extractable segments and events.
- XML is internally descriptive and can be displayed in various ways. This is important to allow users browsing the XML data directly, while search results can also be returned as XML that can provide direct link(s) to the video location.
- XML fully supports semi-structured aspects that match video database characteristics: 1) Object can be described using attributes (properties), other objects (i.e. nested object), or heterogeneous elements (i.e. *any* element). Instantiated objects from the same class may not have the same number of attributes as not all attributes are compulsory, depending on the min and max occurs. 2) XML

supports two types of relationships: nesting and referencing. However, to reduce redundancy, we have used referencing instead of nested object class.

We have used *XML Schema* to define and construct the XML-based video schema as it has replaced *DTD* as the most descriptive language. Due to its expressive power, XML schema has also been used as the basis of MPEG-7 DDL (Data Definition Language) and XQuery data model. Therefore, we should be able to easily leverage our proposed model to support MPEG-7 standard multimedia descriptions and XQuery implementation. For a more compact representation of XML schema, this section will demonstrate the use of *ORA-SS (Object-Relationship-Attribute notation for Semi-Structured data)* (Dobbie et al., 2000) to design the video model as shown in Figure 7 to Figure 9 (that is located on the last page). ORA-SS notation is chosen for its ability to represent most of XML schema's features. It is to be noted that our diagrams extend the ORA-SS notation by demonstrating a more complex sample which integrate inheritance diagram with schema diagram. We have also introduced two additional notations: 1) *italic texts* indicate abstract object, 2) ∇ (in Figure 9) indicates repeated object to avoid complex/crossing lines.

The followings describe the overall video indexing model. As shown in Figure 9, a *sport video (SV)* is a type of video segment which consists of SV components, overall summary, and hierarchical summary. *SV components* are composed of: 1) *segment collection* which stores a flat-list of audio, visual and audio segments that can be extracted from the sport video, 2) *syntactic relation collection* which stores all the syntactic relations such as 'composed of' and 'starts after' between one source segments and one or more destination segments, and 3) *semantic relation collection* which records all the semantic relations such as 'is actor of' and 'appears in' between one source segment or semantic object and one or more destination segments or semantic objects. *Overall summary* describes the sport video game as a whole; it includes where (stadium), when (date time), who (teams that compete), final result, and match statistics. Match statistics can be stored as XML tags or a visual frame such as text displays that depicts the number of goals, shots, fouls, red/yellow cards, and counter attacks in a soccer game. *Hierarchical summary* is composed of *comprehensive summary* and *highlight events (HE)* summary. *Comprehensive* summary describes sport video in terms of play-break sequences which are the main story decomposition unit in most of sport videos. For example, an attacking attempt during a play is stopped when there is a goal or foul. Each play-break can contain zero or one (key) event and can be decomposed into one or more play and break shots. Each play or break can be described by text-alternative annotations, including face, replay and excitement which are referenced (segments) from segment collection. On the other hand, *HE* summary organizes highlight events into common summary theme such as soccer goals and basketball free throws.

Each time sport video is instantiated, it will be specialized into the classified genre, such as soccer video, basketball

video and AFL video. Therefore, a *soccer video* will inherit all components of (general) sport video while providing extra attributes such as *sport category* and some extra components. In particular, for each type of sport video, we can extract *domain specific events* such as soccer goal. Each domain event can be described using specific roles such as goal scorer. It should be noted that goal scorer will reference to a player that is defined elsewhere in order to avoid nested components. Similarly, domain events are referenced by hierarchical summaries. Finally, a *sport video database* is composed of one or more classified *sport videos*, and one *semantic object collection*. Semantic object collection defines the details of all the semantic objects that appear in the sport videos. For example, player can be instantiated into soccer player which is described by the specific attributes of a soccer player such as squad number, and preferred position.

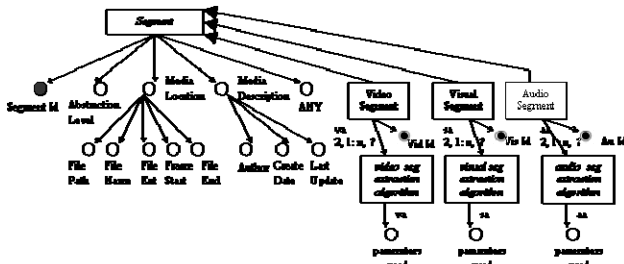


Figure 7. Extensible Indexing Scheme (1).

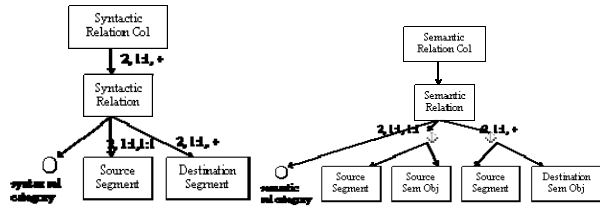


Figure 8. Extensible Indexing Scheme (2).

It is to be noted that in order to achieve a faster gradual index construction, all segments should be able to be extracted incrementally in the same level, without concerning about the hierarchy. For example, assuming that *PBI* contains *PI*, *P2*, and *BI*, the system should be able to add *BI* without necessarily attaching it to *PBI*. This allows the system to easily add *PI* and *P2* at later time. Therefore, hierarchy structures should be stored separately as a *hierarchical view* or processed dynamically when required by users for browsing.

Using the proposed video model, we have demonstrated a sport video indexing scheme that supports:

- *Extensible video indexes* that allow gradual extraction of segments and events without affecting the others. For example, we can introduce more segments and events incrementally without affecting the existing ones. Similarly, more semantic objects, such as stadium and referee, can be introduced at a later stage when many sport videos share the same stadium and referee.
- *Object-Relationship modeling scheme*. In particular, we have demonstrated that inheritance and referencing are important features in video database

modeling. Inheritance enables us to reuse existing parent components while refining them with more specific items. Referencing enables us to store video components into a flat list which can be referenced by hierarchical structures to avoid redundancies.

- *Semi-schema based modeling scheme*. As shown in Figure 7, we allow users/applications to add *ANY* additional elements (or attributes) into a segment description as long as the element has been declared somewhere else in the proposed schema, or other schema within a particular scope. In fact, we may attach *ANY* into other elements in our data model to allow more flexibility as users often know better what they want to describe than developers. However, we aim gradually modifying the schema with new components, especially when the extra information provided by users can be used to enrich the current video model.

4 Experimental Results

Performance results for mid-level features extraction (that are required during training and evaluation) including view classification, near-goal, and excitement, have been presented in our previous papers (Tjondronegoro et al., 2004a). For AFL and basketball videos, we only need to ensure that the adaptive thresholds are effective for each video sample. For this purpose, we compare the truth and the automatic results of features detection on each video for duration of 5-10 minutes. We then select the best empirical thresholds that can be applied to all videos within the same domain. Missing and/or false detections on individual mid-level features detection have less significant impacts on the highlights classification as the models depend on the fusion of all features. For example, soccer goal will still be detectable even if the near goal ratio and excitement is not detected perfectly. Nevertheless, the more accurate mid-level features can be extracted, the highlight points will be more accurately calculated. Hence, during experiment we have set a minimum value that highlight point should reach to be trusted. For all sport videos, we have successfully applied a minimum of 3 points for all highlights which means that at least 3 mid-level features can be detected. In almost all cases, highlights can be detected with a 6 to 7 point minimum threshold.

Table 2 will describe the video samples used during experiment. For each sport, we have used videos from different competitions, broadcasters and/or stage of tournament. The purpose is, for example, final match is expected to contain more excitement than a group match while exhibition will show many replay scenes to display players' skills. Our experiment was conducted using MATLAB 6.5 with image processing toolbox. The videos are captured directly from a TV tuner and compressed into 'mpg' format which can be read into MATLAB image matrixes.

Sample Group (Broadcaster)	Videos "team1-teams2_period-[duration]"
Soccer: UEFA Champions League Group Stage Matches (SBS)	ManchesterUtd-Deportivo1,2-[9:51, 19:50] Madrid-Milan1,2[9:55,9:52]
Soccer: UEFA Champions league (SBS)	Juventus-Madrid1,2:[19:45,9:50] Milan-Internazionale1,2:[9:40,5:53]

Elimination Rounds	Milan-Depor1,2-[51:15,49:36] (S1) Madrid-BayernMunich1,2-[59:41,59:00] (S2) Depor-Porto-[50:01,59:30] (S3)
Soccer: FIFA World cup Final (Nine)	Brazil-Germany [9:29,19:46]
Soccer: International Exhibition (SBS)	Aussie-SthAfrica1,2-[48:31,47:50] (S4)
Soccer: FIFA 100 th Anniversary Exhibition (SBS)	Brazil-France1,2-[31:36,37:39] (S5)
AFL League Matches (Nine)	COL-GEEL_2-[28:39] (A3) StK-HAW_3-[19:33] (A4) Rich-StK_4-[25:20] (A5)
AFL League Matches (Ten)	COL-HAW_2-[28:15] (A1) ESS-BL_2-[35:28] (A2) BL-ADEL_1,2-[35:33,18:00] (A6)
AFL League Final rounds (Ten)	Port-Geel_3,4-[30:37,29:00] (A7)
Basketball: Athens 2004 Olympics (Seven)	Women: AusBrazil_1,2,3-[19:50,19:41,4:20] (B1) Women: Russia-USA_3-[19:58] (B2) Men: Australia-USA_1,2-[29:51,6:15] (B3)
Basketball: Athens 2004 Olympics (SBS)	Men: USA-Angola_2,3-[22:25,15:01] (B4) Women: Australia-USA_1,2-[24:04-11:11] (B5)

Table 2. Sample Video Data.

4.1 Performance of Play-Break Segmentation

Play-break scoping plays a significant role to ensure that we can extract all of the features that usually exist in each highlight. Moreover, the statistics (especially play-/break-dominance) will be affected when the play-break sequences are detected perfectly. Table 3 to Table 5 depicts the performance of the play-break segmentation algorithm on soccer, AFL and basketball videos, respectively. It is to be noted that that RC = Replay-based (P-B sequence) Correction, PD = perfectly detected, D = detected, M = missed detection, F = false detection, Tr = Total number in Truth, Det = Total Detected, RR = Recall Rate, PR = Precision Rate, and PD_{decr} = perfectly detected decrease rate if RC is not used; $Tru = PD + D + M$, $Det = PD + D + F$, $RR = (PD + D + M) / Tru * 100\%$, $PR = (PD + D) / Det * 100\%$, and $PD_{Decr} = (PD - D) / PD * 100\%$. The results demonstrate that RC is generally useful to improve the play-break segmentation performance. It is due to the fact that many (if not most) replay scenes, especially soccer and AFL use global (i.e. play) shots. This is shown by all PD_{decr} , RR , and PR as RC always improves all of these performance statistics. In particular, the RR and PR for soccer 1-1 with RC are 100% each but they are reduced to below 50% without RC . In soccer 1-1 without RC , the PD dropped from 49 to 12 (i.e. 75% worse) whereas M increases from 0 to 25 and F increases from 0 to 5. This is due to the fact that soccer1 video contains many replay scenes which are played abruptly during a play, thereby causing a too-long play scene and missing a break. However, based on the statistics shown in Table 5, RC for basketball may not be as important as that of soccer and AFL. It is because basketball's replay scene uses more break shots such as zoom-in and close-up, as compared to soccer and basketball.

4.2 Performance of Soccer Events Detection

Based on Table 6 and Table 7, most soccer highlights can be distinguished from non-highlights with high recall and precision. As there are normally not many goal highlights in a soccer match, it would be ideal to have a high RR

over a reasonable PR ; 5 out of 7 goals are correctly detected from the 5 sample videos while 2 shoots and 1 non-highlight are classified as goals. The shoot segments detected as goals very exciting and nearly result in goal. On the other hand, the non-highlight detected as a goal also consist of a long duration and replay scenes and excited commentaries due to a fight between players. The foul detection is also effective as the RR is 81% and most of the misdetections are either detected as shoot or non which have the closest characteristics. However, the PR is considerably low since some shoots and non-highlights are detected as foul. An alternative solution is to use whistle existence for foul detection, but we still need to achieve a really accurate whistle detection that can overcome the high-level of noise in most of sport domains. Only 46 out of 266 non-highlight sequences were incorrectly detected as highlights. These additional highlights will still be presented to the viewers as there are generally not many significant events during a soccer video. In fact, most of these false highlights can still be interesting for some viewers as they often consist of long excitement, near-goal duration and replay scene.

4.3 Performance Basketball Events Detection

Highlights detection in basketball is slightly harder than soccer and AFL due to the fact that: 1) goals are generally not celebrated as much as soccer and AFL, 2) non-highlights are often detected as goal and vice versa. Fortunately, non-highlights mainly just include ball out play which hardly happen in basketball matches. Thus, we have decided to exclude non-highlight detection and replace it with timeout detection which can be regarded as non-highlights for most viewers. However, for some sport fans, timeouts may still be interesting to show the players and coaches for each team and some replay scenes. In addition to these problems, sequences containing fouls are sometimes inseparable from the resulting free throws. For such cases, the fouls are often detected as goal due to the high amount of excitement and long near-goal. However, fouls which are detected as goals can actually be avoided by applying a higher minimum highlight point for goal but at the expense of missing some goal segments. For our experiment, we did not use this option as we want to use a universal threshold for all highlights.

Based on Table 8 and Table 9, basketball goal detection achieves high RR and reasonable PR . This is due to the fact that goals generally have very unique characteristics as compared to foul and free throw. Timeouts can be detected very accurately (high RR and PR) due to their very long and many replay scenes. Moreover, most broadcasters will play some in-between advertisements when a timeout is longer than 2 minutes, thereby increasing the close-up ratio. Free throw is also detected very well due to the fact that free throw is mainly played in near-goal position; that is, the camera focuses on capturing the player with the ball to shoot. However, it is generally distinguishable from goal based on: less excitement, higher near goal, and more close-up shott; that is, goal scorer is often just shown with zoom-in views to keep the game flowing. However, the system only detected 28 out of 54 foul events. This problem is

caused by the fact that after foul, basketball videos often abruptly switches to a replay scene which is followed by time-out or free-throw. This can be fixed with the introduction of additional knowledge such as whistle-detection.

4.4 Performance of AFL Events Detection

As shown in Table 10 and Table 11, the overall performance of the AFL highlights detection is found to yield promising results. All 37 goals from the 7 videos were correctly detected. Although the RR of behind detection seems to be low, most of the miss-detections are actually detected as goal. Moreover, behind is still a sub-type of goal except that it has lower point awarded. The slightly lower performance for detection of mark and tackle detection is caused by the fact that our system does not include whistle feature which is predominantly used during these events. Based on the experimental results, mark is the hardest to be detected and needs additional knowledge. In Table 11, *PR* and *RR* for behind is N/A as 1 behind was detected as goal while Mark = N/A because 5 marks were detected as goal.

5 Conclusion and Future Work

We have proposed an extensible approach for detecting events in sports video. The use of play-break scoping for all highlights have enabled us to obtain statistical-phenomena of the features contained in each highlight. Since the rules for highlight classification are driven by the statistics, none or low amount of domain-specific knowledge is required. Therefore, the proposed algorithms should be more robust for different sports, especially, field-ball goal oriented games. Based on the experimental results, play-break sequences are proven to be effective containers for detecting highlights. Thus, play-breaks need to be perfectly segmented and we have shown that replay-correction improves the performance. We have also proposed a segment-event based video data model which is designed using semi-schema-based and object-relationship modeling schemes. The schema is developed into XML schema with ORA-SS notation. The proposed schema is extensible as it supports incremental development of algorithms for feature-semantic extraction. Moreover, the schema does not need to be complete at one time while allowing users to add additional elements. We have also emphasized the usage of referencing relationship to avoid redundant data. Referencing also allows the system to add segments and events to achieve more straightforward and faster data insertions. In order to further verify and improve the robustness of the proposed algorithms for events detection we have incorporated more sport genre such as volleyball, tennis and gymnastics, into the existing dataset. The extracted information will allow the system to construct a larger sample of video database data which consequently would verify the benefits from using the proposed video indexing model.

Video	Soccer Play-break detection								
	PD	D	M	F	Tru	Det	RR	PR	PD decr
S1-1 (RC)	49	0	0	0	49	49	100.00	100.00	
S1-1	12	12	25	5	49	54	48.98	44.44	75.51
S1-2(RC)	53	0	0	1	53	54	100.00	98.15	

S1-2	36	10	7	1	53	54	86.79	85.19	32.08
S2-1(RC)	54	1	1	12	56	68	98.21	80.88	
S2-1	53	2	1	12	56	68	98.21	80.88	1.85
S2-2(RC)	58	1	0	7	59	66	100.00	89.39	
S2-2	55	4	0	7	59	66	100.00	89.39	5.17
S3-1 (RC)	49	0	0	4	49	53	100.00	92.45	
S3-1	45	4	0	5	49	54	100.00	90.74	8.16
S3-2 (RC)	69	0	0	3	69	72	100.00	95.83	
S3-2	65	4	0	5	69	74	100.00	93.24	5.80
S4-1(RC)	49	0	0	9	49	58	100.00	84.48	
S4-1	40	8	1	13	49	62	97.96	77.42	18.37
S4-2(RC)	47	0	0	9	47	56	100.00	83.93	
S4-2	36	11	0	12	47	59	100.00	79.66	23.40
S5 (RC)	48	0	0	0	48	48	100.00	100.00	
S5	24	16	8	1	48	49	83.33	81.63	50.00

Table 3. Play-Break Detection in Soccer Videos.

Video	AFL Play-break detection								
	PD	D	M	F	Tru	Det	RR	PR	PD decr
A1 (RC)	34	0	0	5	34	39	100.00	87.18	
A1	29	5	0	8	34	42	100.00	80.95	14.71
A2 (RC)	21	6	0	8	27	35	100.00	77.14	
A2	16	10	1	5	27	32	96.30	81.25	23.81
A3 (RC)	20	3	0	4	23	27	100.00	85.19	
A3	17	6	0	6	23	29	100.00	79.31	15.00
A4 (RC)	29	0	0	1	29	30	100.00	96.67	
A4	21	6	2	2	29	31	93.10	87.10	27.59
A5 (RC)	34	0	0	1	34	35	100.00	97.14	
A5	23	4	7	3	34	37	79.41	72.97	32.35
A6 (RC)	50	2	0	3	52	55	100.00	94.55	
A6	36	10	6	7	52	59	88.46	77.97	28.00
A7 (RC)	41	10	4	4	55	59	92.73	86.44	
A7	39	12	4	6	55	61	92.73	83.61	4.88

Table 4. Play-Break Detection Results in AFL Videos.

Video	Basketball Play-break detection								
	PD	D	M	F	Tru	Det	RR	PR	PD decr
B1 (RC)	32	6	2	3	40	43	95.00	88.37	
B1	31	7	2	4	40	44	95.00	86.36	3.13
B2 (RC)	19	2	0	2	21	23	100.00	91.30	
B2	18	3	0	3	21	24	100.00	87.50	5.26
B3 (RC)	39	3	0	1	42	43	100.00	97.67	
B3	38	4	0	2	42	44	100.00	95.45	2.56
B4 (RC)	26	5	2	2	33	35	93.94	88.57	
B4	25	6	0	3	31	34	100.00	91.18	3.85
B5 (RC)	39	0	1	1	40	41	97.50	95.12	
B5	25	13	2	5	40	45	95.00	84.44	35.90

Table 5. Play-Break Detection Results in Basketball.

Ground truth	Highlight classification of 5 videos				
	Goal	Shoot	Foul	Non	Truth
Goal	5	0	2	0	7
Shoot	2	66	32	12	112
Foul	0	13	91	13	117
Non	1	11	34	220	266
Detected	8	90	159	245	

Table 6. Events Detection Results in Soccer Videos.

	S1		S2		S3		S4		S5		Average	
	RR	PR	RR	PR	RR	PR	RR	PR	RR	PR	RR	PR
Goal	60	100.0	100	50.0	N/A	N/A	100	33.3	N/A	N/A	86.7	61.1
Shoot	39.4	76.5	64.0	84.2	80.0	66.7	78.9	71.4	40.0	66.7	60.5	73.1
Foul	85.2	53.5	68.0	53.1	71.4	78.9	88.9	38.1	92.9	52.0	81.3	55.1
Non	86.5	82.1	86.3	88.5	90.5	90.5	75.8	100.0	60.0	80.0	79.8	88.2

Table 7. Distribution of Soccer Events Detection

Ground truth	Highlight classification of 5 basketball videos				
	Goal	Free throw	Foul	Timeout	Truth
Goal	56	0	0	2	58
Free throw	4	14	0	0	18
Foul	21	2	28	3	54
Timeout	0	0	0	13	13
Total Detected	81	16	28	18	

Table 8. Basketball Events Detection Results

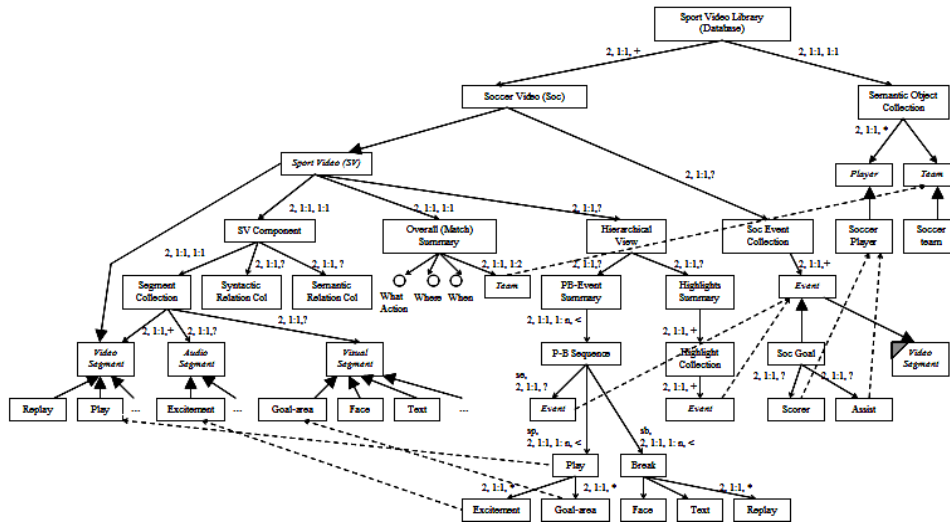


Figure 9. Extensible Indexing Scheme (3).

	B1		B2		B3		B4		B5		Average	
	RR	PR	RR	PR	RR	PR	RR	PR	RR	PR	RR	PR
Goal	100	72.2	75	50.0	95	70.4	100	72.2	100	66.7	94	66.3
Free throw	100.0	66.7	100.0	75.0	80.0	100.0	50.0	100.0	66.7	100.0	79.33	88.3
Foul	64.7	100.0	50.0	100.0	30.8	100.0	37.5	100.0	75.0	100.0	51.59	100.0
Timeout	100.0	100.0	100.0	50.0	100.0	40.0	100.0	66.7	100.0	100.0	100	71.3

Table 9. Distribution of Basketball Events Detection

Ground truth	Highlight classification of 7 videos						Truth
	Goal	Behind	Mark	Tackle	Non		
Goal	37	0	0	0	0		37
Behind	11	12	7	0	2		32
Mark	15	1	35	8	5		64
Tackle	4	0	9	20	2		35
Non	4	4	11	3	33		55
Detected	71	17	62	31	42		

Table 10. Events Detection Results in AFL Videos

	A1		A2		A3		A4		A5		A6		A7		AVG	
	RR	PR	RR	PR	RR	PR	RR	PR	RR	PR	RR	PR	RR	PR	RR	PR
Goal	100.0	44.4	100.0	52.9	100.0	57.1	100.0	33.3	100.0	50.0	100.0	63.6	100.0	53.3	100.0	50.7
Behind	50.0	100.0	N/A	N/A	33.3	33.3	33.3	66.7	50.0	100.0	33.3	100.0	33.3	50.0	38.9	75.0
Mark	50.0	60.0	N/A	N/A	60.0	60.0	77.8	77.8	60.0	42.9	66.7	42.1	47.1	80.0	60.3	60.5
Tackle	80.0	66.7	100.0	75.0	25.0	100.0	100.0	100.0	12.5	33.3	85.7	66.7	50.0	50.0	64.7	70.2
Non	77.8	100.0	33.3	100.0	50.0	50.0	66.7	71.4	71.4	46.2	100.0	75.0	69.2	57.7	79.6	

Table 11. Distribution of AFL Events Detection

6 References

Adali, S., Candan, K. S., Chen, S.-S., Erol, K. and Subrahmanian, V. S. (1996) 'The Advanced Video Information System: Data Structures and Query Processing' *Multimedia Systems*, **4**, 172-186.

Connolly, T. M. and Begg., C. E. (2002) *Database systems : a practical approach to design, implementation, and management*, Addison-Wesley, Harlow [England] ; [New York].

Djeraba, C. (2002) 'Content-based multimedia indexing and retrieval' *Multimedia, IEEE*, **9**, 18-22.

Dobbie, G., Xiaoying, W., Ling, T. W. and Lee, M. L. (2000) In *Technical Report Department of Computer Science*, National University of Singapore.

Duan, L.-Y., Xu, M., Chua, T.-S., Qi, T. and Xu, C.-S. (2003) In *ACM MM2004* ACM, Berkeley, USA, pp. 33-44.

Ekin, A. and Tekalp, A. M. (2003a) In *International Conference on Multimedia and Expo 2003 (ICME03)*, Vol. 1 IEEE, pp. 6-9 July 2003.

Ekin, A. and Tekalp, M. (2003b) 'Automatic Soccer Video Analysis and Summarization' *IEEE Transaction on Image Processing*, **12**, 796-807.

Han, M., Hua, W., Chen, T. and Gong, Y. (2003) In *Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*, Vol. 2, pp. 950-954.

Li, B. and Ibrahim Sezan, M. (2001) In *Content-Based Access of Image and Video Libraries, 2001. (CBAIVL 2001). IEEE Workshop on Practical, Sharp Labs. of America, Camas, WA, USA*, pp. 132-138.

Nepal, S., Srinivasan, U. and Reynolds, G. (2001) In *ACM International Conference on Multimedia* ACM, Ottawa; Canada, pp. 261-269.

Oomoto, E. and Tanaka, K. (1997) In *The Handbook of Multimedia Information Management* (Ed, William I. Grosky, R. J. a. R. M.) Prentice Hall, Upper Saddle River, NJ, pp. 405 - 448.

Tjondronegoro, D., Chen, Y.-P. P. and Pham, B. (2004a) 'Integrating Highlights to Play-break Sequences for More Complete Sport Video Summarization' *IEEE Multimedia*, **Oct-Dec 2004**, 22-37.

Tjondronegoro, D., Chen, Y.-P. P. and Pham, B. (2004b) In *The 6th International ACM Multimedia Information Retrieval Workshop* ACM Press, New York, USA, pp. 267-274.

Wu, C., Ma, Y.-F., Zhang, H.-J. and Zhong, Y.-Z. (2002) In *Multimedia and Expo, 2002. Proceedings. 2002 IEEE International Conference on*, Vol. 1, pp. 805-808.

Xu, P., Xie, L. and Chang, S.-F. (1998) In *IEEE International Conference on Multimedia and Expo* IEEE, Tokyo, Japan.