Towards Universal and Statistical-Driven Heuristics for Automatic Classification of Sports Video Events

Dian Tjondronegoro[1] Yi-Ping Phoebe Chen[2]
School of Information Systems, Queensland University of Technology, Brisbane, Australia[1]
School of Information Technology, Deakin University, Melbourne, Australia[2]
dian@qut.edu.au *phoebe@deakin.edu.au*

Abstract

Researchers worldwide have been actively seeking for the most robust and powerful solutions to detect and classify key events (or highlights) in various sports domains. Most approaches have employed manual heuristics that model the typical pattern of audio-visual features within particular sport events. To avoid manual observation and knowledge, machine-learning can be used as an alternative approach. To bridge the gaps between these two alternatives, an attempt is made to integrate statistics into heuristic models during highlight detection in our investigation. The models can be designed with a modest amount of domain-knowledge, making them less subjective and more robust for different sports. We have also successfully used a universal scope of detection and a standard set of features that can be applied for different sports that include soccer, basketball and Australian football. An experiment on a large dataset of sport videos, with a total of around 15 hours, has demonstrated the effectiveness and robustness of our algorithms.

1. Introduction

Automatic content analysis is an essential requirement for constructing an effective sports video summary. It has become a well-known theory that the high-level semantics in sport video 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, which are detected based on color, texture and global-motion measurements [8]. 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 probabilistic 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 new feature will only result on 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 [6]. 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 also have two major drawbacks, namely,

- 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 [3] 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 [4] used a static temporal-segment of 30-40 sec (empirical) for soccer highlights detection.
- 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 [1].

1-4244-0028-7/06/$20.00 ©2006 IEEE

To solve the first drawback, some approaches have claimed that highlights are mainly contained in a *play* scene [5, 9]. However, based on a user study, 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, based on statistical features that will be discussed in section 3, we were able to identify some additional characteristics of foul that could be easily missed by manual observation such as shorter duration as compared to shoot and less excitement as compared to goal.

The focus of this paper is to present a statistical-driven framework for automatic highlight classification that is based on a universal scope-of-detection and a standard set of audio-visual features. The effectiveness and robustness of this framework has been tested with a large dataset of soccer (around 7 hours), basketball (3 hours) and Australian Football (4.5 hours). At this stage, our algorithms have successfully detected and classified soccer highlights, including *goal, shoot* (goal attempt), and *foul*, and detecting *non-highlights*. With very minor changes, the system can also distinguish *goal, behind, mark, tackle,* and *non-highlight* in Australian Football (AFL), and *goal, free throw, foul* and *timeout* in basketball. Soccer and basketball are chosen as the case domain since they have a world-wide audience with many different national leagues and international competitions. AFL is selected as one of largest sectors in Australia's sport and recreation industry, attracting more than 14 million people to watch an average of 10 hour per week live-broadcasted matches Moreover, there is yet any significant work presented for this domain.

2. Utilizing Play-Break as a Definitive and Universal Detection Scope

A play is when the game is still flowing, such as when the ball is being played in soccer and basketball. A break is when the game is stopped or paused due to specific reasons, such as when a foul or a goal happens. 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.

Given that: a) the start of a play sequence is marked by the first frame of a long global shot (e.g. > 5 sec) and b) the start of a break sequence is marked by the first frame of a long medium shot, (slow-motion) replay shot, or zoom shot of medium length; it should be clear that play-break sequences should be effective containers for a semantic content since they contain all the required details. Moreover, most events are contained within a play-break sequence. Using this assumption, we should be able to extract all the phenomenal features from play-break that can be utilized for highlights detection.

As shown in Figure 1, the scoping of event detection should be from the last play-shot until the last break shot. However, more play shots can be included for viewing, depending on how much detail on the play that users prefer, thereby reducing the subjectivity level rather than selecting particular frames. It is important to note that if the scope of play and break for detection is changed, we need to re-calculate the statistics.

Benefits of using play-break to serve as a definitive scope for the start and end of features observation:

- It becomes possible to use comparative measurements (e.g. break ratio) which are more robust and flexible as compared to definitive measurements such as length of break.
- We can potentially design a more standard benchmarking of different highlight detection approaches. For example, we cannot literally compare two approaches. If one uses play-break segment while the other one uses play-break-play segment, or a static empirical based.
- We can reduce the level of subjectivity during manual observations for ground-truth. For example, we should not simply conclude that an artificial text always appear after/during a goal highlight as text can be used during the break segment and/or the first play segment after the break segment. We should therefore take a precaution to include a text when it is too far from the highlight itself (e.g. two or three play segments after the highlight) as it can belong to another highlight (or no highlight at all).

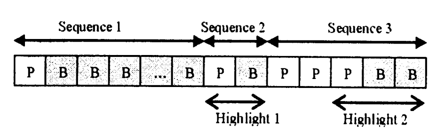

Figure 1. Play-break Scoping for Highlight Detection

Analysis of *camera-views* transition in a sports video has been used successfully for play-break segmentation (such as in [2]). We have extended this approach by adding *replay-based correction* to improve the performance. Replay detection is very important to locate additional breaks which are often recognized as play shots (i.e. replay shot often use global view). Replay scenes should be regarded as part of a break since they contain *non real-time* match contents. Based on the experimental results which has been reported in [7], replay-based correction on play-break segmentation can fix a large number of imperfect sequences due to shorter breaks, locate missing sequences due to missed breaks, and avoid false sequences due to falsely detected play which is followed by a break.

Figure 2 illustrates the main processing required for our semantic analysis scheme. First, play-break sequences are segmented using the outputs from view classification and replay detection. Second, in order to classify the highlight contained in each sequence, statistics of the mid-level features are calculated and compared to the trained statistics using specific heuristic rules. Finally, for each (classified) highlight, some text-alternative annotation can be extracted to construct the summary. It should be noted that dashed boxes represent processes that are only used during training. In particular, dominant-hue index training is usually required for new video while training of statistics is required for new highlights.

3. Semi-supervised Discovery of Heuristics

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 on 20 samples 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 as we manually classify the specific highlight that each play-break sequence (for training) 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. It should be noted that a separate training should be performed for non-highlight to find its distinctive characteristics.

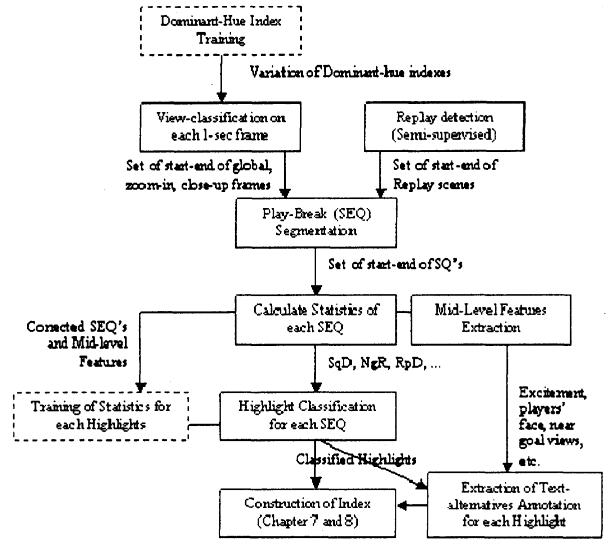

Figure 2. Processing Steps of Highlight Classification

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 higher break ratio than behind due to a longer goal celebration and slow motion replay.
- *PlR* = 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 slow motion 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, goal consists of a very high excitement ratio whereas non-highlight usually contain 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. Advertisements after a goal will be detected as close-up or no grass.

This set of features is selected as they are generally effective for describing sport events, in particular, soccer, AFL, basketball and any sports with similar characteristics. However, whistle occurrence is not used even though it is very useful for many sports; it is due to the fact that whistles are hardly audible and often falsely detected from whistle blown by audience. Similarly, inserted texts occurrence is not used as their location within a sequence is not predictable. For example, caption for a goal is usually displayed in the next play shot after goal celebration while caption for a shot is usually displayed during the break.

Table 1 shows the training data based on an AFL match in terms of the locations of play-break sequences that make up the video and the mid-level features contained within each sequence. In this table, the highlighted segments are used for training purposes while others are used for detection experiment. Using this type of data, the statistical parameters of each highlight for each sport genre can be calculated. Table 2 is an example of the training data used for AFL goal event. After the mid-level features based parameters are calculated for each sample, the statistical characteristics are then derived as minimum, maximum, and average values. The statistical data of the universal feature sets within each highlight are presented in Table 3.

4. Constructing Statistical-Driven Heuristics

Based on the trained statistics, we have constructed a novel set of '*statistical-driven*' rules 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 without the necessity to make major changes in the highlight classification rules, can be introduced. 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.

Play-break location (duration)	Near Goal	Exc ratio 1.2 -> 1.5	Close-up ratio	Play-break ratio	Slomo
1:07-1:32 (26) Out of play	30-32	7-9, 18-22 (8)		7-27 (21)	
1:33-1:54 (22) Tackle			42 (1)	33-41 (9)	
2:07-2:24 (18) mark		13-21 (9)		7-18 (12)	
4:00-4:38 (39) behind (goal)	27-29, 37-38	11-33 (23)	36-39 (4)	0-35 (36)	
5:33-5:45 (13) Out of play		34-45 (12)	43-46 (4)	33-42 (10)	
5:46-6:06 (21) Mark (someone else injured)		47-59 (13)	55-59 (14)	46-54 (9)	58-06 (9)

Table 1. Example of Training Data

Sample	Video	play ratio	duration(2min)	excitement	break ratio	Replay Duration (/40s)	near goal ratio	Close-up ratio
1	AFL_ColHaw2	0.30	0.96	0.32	0.70	0	0.14	0.30
2	AFL_ColHaw2	0.20	0.43	0.22	0.80	0.28	0.08	0.53
3	AFL_ColHaw2	0.14	0.35	0.40	0.86	0.00	0.43	0.86
4	AFL_ColHaw2	0.14	0.48	0.44	0.86	0.00	0.02	0.66
5	BnsLion2	0.06	0.91	0.39	0.94	0.58	0.18	0.06
6	BnsLion2	0.15	0.33	0.10	0.85	0.43	0.18	0.05
7	BnsLion2	0.16	0.37	0.32	0.84	0.43	0.27	0.00
8	Col_Gel2	0.14	0.61	0.33	0.86	0.30	0.04	0.53
9	StK-HAW3	0.18	0.78	0.18	0.82	0.30	0.02	0.13
10	StK-HAW3	0.15	0.66	0.23	0.85	0.00	0.05	0.77
11	Col-Haw2	0.19	0.46	0.00	0.81	0.00	0.07	0.25
12	Bns-Lion2	0.30	0.80	0.42	0.70	0.18	0.04	0.00
13	Bns-Lion2	0.12	0.57	0.12	0.88	0.53	0.34	0.03
14	Bns-Lion2	0.08	0.71	0.39	0.92	0.35	0.21	0.35
15	Bns-Lion2	0.17	0.38	0.28	0.83	0.00	0.09	0.59
16	Bns-Lion2	0.23	0.33	0.54	0.77	0.49	0.18	0.00
17	Col-Gel2	0.13	0.65	0.35	0.87	0.00	0.08	0.50
18	Col-Gel2	0.07	0.63	0.34	0.93	0.03	0.03	0.33
19	Col-Gel2	0.13	1.00	0.23	0.88	0.35	0.06	0.54
20	Rich-Stk4	0.33	0.51	0.15	0.67	0.00	0.08	0.38
21	Rich-Stk4	0.27	0.31	0.41	0.73	0.59	0.03	0.08
22	Rich-Stk4	0.31	0.58	0.84	0.69	0.36	0.01	0.26
	MIN	0.06	0.33	0.00	0.67	0.00	0.02	0.00
	AVG	0.17	0.60	0.29	0.83	0.21	0.13	0.35
	MAX	0.33	1.00	0.54	0.94	0.58	0.43	0.86

Table 2. Training Samples Used for AFL Goal Event

Feature	Soccer G=Goal, S=Shoot, F=Foul, N=Non_(avg; max; min)	AFL G=Goal, B=Behind, M=Mark, T=Tackle, N=Non_(avg; max; min)	Basketball 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 (PlR)	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)	Gc_(0.26; 0.51; 0.08) Sc_(0.23; 0.74; 0) Fc_(0.12; 0.29; 0) Nc_(0.2; 0.6; 0)	Gc_(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)	Gc_(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 3. Statistics of Soccer, AFL, and Basketball Highlights after 20 Samples Training

4.1. Event Classification Algorithm

Highlight classification is performed as:

[HgtClass] = Classify_ Highlight *(D,NgR,Exc R,CuR,PlR, 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.

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).

```
Let Det_Soccer_Region(val) = Region(val, stat_G, stat_S, stat_F, stat_N)

Perform
region_1..n = Det_Soccer_Region(D),(NgR),(ExcR),(CuR),(PIR),(RpR)
For region_1 to region_n
Increment the corresponding highlight score //G, Sh, F, Non in this case
```

where,

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

$\text{stat}_n = \{\text{avg}_n, \text{min}_n, \text{max}_n\}$, $AvgD_n = \left|val - stat_n^{avg}\right|$,

$TD_n = \left|val - stat_n^{max}\right| + \left|val - stat_n^{min}\right|$,

$MinAvgD = \min(AvgD_1, AvgD_2, \ldots AvgD_n)$, $MinTD = \min(TD_1, TD_2, \ldots TD_n)$.

It is to be noted that in 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.

4.2. 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, to classify highlight events in soccer

```
Perform region_1..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 region_4..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 region_4..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 $\text{region}_{1..N} = \text{Det\_Soccer\_Region}(\text{val}_1), (\text{val}_2) \ldots (val_N)$. 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.

4.3. 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: $\text{Det\_AFL\_Region}(val) = \text{Region}(val, \text{stat}_G, stat_B, stat_M, stat_T, stat_N)$

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*.

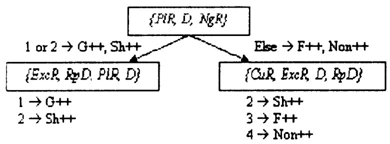

Figure 3. Highlight Classification Rules for Soccer

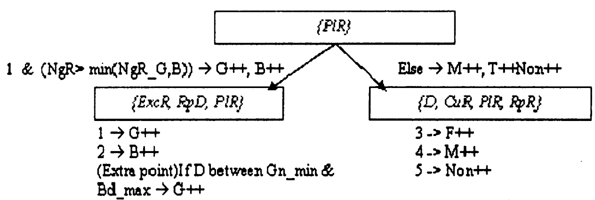

Figure 4. Highlight Classification Rules for AFL.

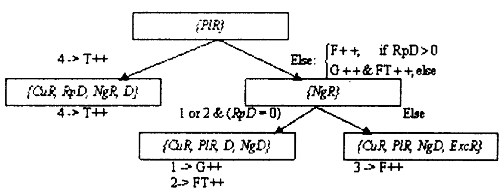

Figure 5. Highlight Classification Rules for Basketball

4.4. 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: $\text{Det\_Basketball\_Region}(val) = \text{Region}(val.\text{stat}_G, stat_{FT}, stat_F, stat_T)$

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.

5. Experiment Results

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.

For highlights classification, we manually developed the ground truth for each sequence with the highlight contained. In order to measure the performance of highlights classification, Recall (*RR*) and Precision Rate (*PR*) are not sufficiently accurate and expressive. The main reason is that we need to see precisely where the miss- and false-detections are. Therefore, we have provided the *RR*, *PR* and the actual detections results.

5.1. Performance of Soccer Events Detection

Based on Table 6 and Figure 5a, most soccer highlights can be distinguished from non-highlights with high recall and precision. It is to be noted that that *D* = detected, *M* = missed detection, *F* = false detection, *Tr* = Total number in Truth, *Det* = Total Detected, *RR* = Recall Rate, and *PR*= Precision Rate; $Tru = PD+D+M$, $Det = PD+D+F$, $RR = (PD+D+M)/Tru * 100\%$, $PR = (PD+D)/Det * 100\%$.

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.

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) Elimination Rounds	*Juventus-Madrid1,2:[19:45,9:50]* *Milan-Internazionale1,2:[9:40,5:53]* 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 100th 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 4. Details of Sample Data for Experiments

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.

5.2. Performance of 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 7 and Figure 5b, 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.

5.3. Performance of AFL Events Detection

Based on the information from Table 8 and Figure 5c 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. It should also be noted that in Table 11, *PR* and *RR* for behind is N/A because 1 behind was detected as goal while Mark = N/A because 5 marks were detected as goal.

References

[1] Duan, L.-Y., Xu, M., Chua, T.-S., Qi, T. and Xu, C.-S., A mid-level representation framework for semantic sports video analysis. in *ACM MM2004*, (Berkeley, USA, 2003), ACM, 33-44.

[2] Ekin, A. and Tekalp, A.M., Generic play-break event detection for summarization and hierarchical sports video analysis. in *International Conference on Mulmedia and Expo 2003 (ICME03)*, (2003), IEEE, 6-9 July 2003.

[3] Ekin, A. and Tekalp, M. Automatic Soccer Video Analysis and Summarization. *IEEE Transaction on Image Processing, 12* (7). 796-807.

[4] Han, M., Hua, W., Chen, T. and Gong, Y., Feature design in soccer video indexing. 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*, (2003), 950-954.
[5] Li, B. and Ibrahim Sezan, M., Event detection and summarization in sports video. in *Content-Based Access of Image and Video Libraries, 2001. (CBAIVL 2001). IEEE Workshop on*, (Sharp Labs. of America, Camas, WA, USA, 2001), Practical, 132-138.
[6] Nepal, S., Srinivasan, U. and Reynolds, G., Automatic detection of 'Goal' segments in basketball videos. in *ACM International Conference on Multimedia*, (Ottawa; Canada, 2001), ACM, 261-269.
[7] Tjondronegoro, D., Chen, Y.-P.P. and Pham, B., A Statistical-driven Approach for Automatic Classification of Events in AFL Video Highlights. in *The 28th Australasian Computer Science Conference*, (Newcastle, Australia, 2005), ACS, 209-218.
[8] Wu, C., Ma, Y.-F., Zhang, H.-J. and Zhong, Y.-Z., Events recognition by semantic inference for sports video. in *Multimedia and Expo, 2002. Proceedings. 2002 IEEE International Conference on*, (2002), 805-808.
[9] Xu, P., Xie, L. and Chang, S.-F., Algorithms and System for Segmentation and Structure Analysis in Soccer Video. in *IEEE International Conference on Multimedia and Expo*, (Tokyo, Japan,, 1998), IEEE.

Ground truth	Highlight classification of 5 soccer videos				
	Goal	Shoot	Foul	Non	Total 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
Total Detected	8	90	159	245	

Table 6. Highlight Classification Performance in Soccer Videos

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 7. Highlight Classification Performance in Basketball Videos

Ground truth	Highlight classification of 7 AFL videos					
	Goal	Behind	Mark	Tackle	Non	Total Truth
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
Total Detected	71	17	62	31	42	

Table 8. Highlight Classification Performance in AFL Videos

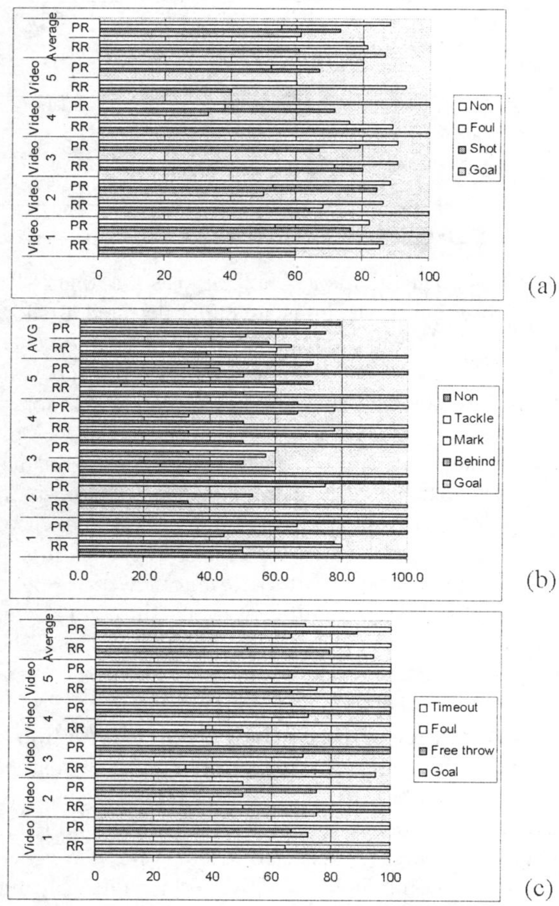

Figure 6. Distribution of Highlight Classification Performance in a) Soccer, b) AFL and c) Basketball