A Statistical-driven Approach for Automatic Classification of Events in AFL Video Highlights

Dian Tjondronegoro[1 2 3]   **Yi-Ping Phoebe Chen**[1]   **Binh Pham**[3]

School of Information Technology, Deakin University [1]
School of Information Systems, Queensland University of Technology [2]
Centre for Information Technology Innovation, Queensland University of Technology [3]

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

Abstract

Due to the repetitive and lengthy nature, automatic content-based s*ummarization* is essential to extract a more compact and interesting representation of sport video. State-of-the art approaches have confirmed that high-level semantic in sport video can be detected based on the occurrences of specific audio and visual features (also known as cinematic). However, most of them still rely heavily on manual investigation to construct the algorithms for highlight detection. Thus, the primary aim of this paper is to demonstrate how the statistics of cinematic features within play-break sequences can be used to less-subjectively construct highlight classification rules. To verify the effectiveness of our algorithms, we will present some experimental results using six AFL (Australian Football League) matches from different broadcasters. At this stage, we have successfully classified each play-break sequence into: *goal, behind, mark, tackle*, and *non-highlight*. These events are chosen since they are commonly used for broadcasted AFL highlights. The proposed algorithms have also been tested successfully with soccer video.

Keywords: Sports video summarisation, semantic analysis, self-consumable highlights, algorithms, AFL.

1   Introduction

For more than a decade, researchers around the world have proposed many techniques for automatic content extraction which take full advantage of the fact that sports videos have typical and predictable temporal structure, recurrent events, consistent features and a fixed number of camera views. It has become a well-known theory that high-level semantic in sport video can be detected based on the occurrences of specific audio and visual features. Another alternative that is object-motion based offers high-level analysis, but this approach is in general computationally expensive. On the other hand, cinematic features offer a good trade-off between computational requirements and the detectable semantics.

For example, a goal in soccer is scored when the ball passes the goal line inside of the goal-mouth. While object-based features, such as ball-tracking are capable of detecting such semantic, specific features like slow-motion replay, excitement, and text display should be able to detect it more efficiently or at least helping to narrow down the scope of the analysis. For example, Nepal et al (Nepal et al., 2001) proposed some *temporal models* to describe the temporal gaps of specific features in basketball goals which include crowd cheer, scoreboard, and change of direction. However, the scope of the detection (i.e. the start and end of observation) was not definitive. Similarly, maximum-entropy based models have been used to combine low-level and mid-level features for detecting soccer highlights, such as m*otionless-regions* for locating the 'human-wall' during a free-kick or corner kick (Han et al., 2003). Yet again a static temporal-segment of 30-40 sec (empirical) was used as the scope of "contextual information".

To achieve a more definitive scope of highlight detection, some approaches have claimed that highlights are mainly contained in a play scene, see, for example, (Xu et al., 1998). However, based on a user study reported in (Tjondronegoro et al., 2004b), we have identified that most users need to watch the whole play and break to fully understand 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 and/or a replay scene will confirm whether it was a *foul* or *offside*. Thus, it is expected that automated semantic analysis should also need to use both play and break segments to detect highlights since a play-break sequence should contain all the necessary features required.

Using this approach, Ekin et al (Ekin and Tekalp, 2003b) has recently defined a cinematic template for soccer goal events detection. This template examines the video-frames between the global shot that causes the goal and the global shot that shows the restart of the game. Firstly, the duration of a break must be between 30 and 120 seconds. Secondly, at least one close-up shot and one slow motion replay shot must be found. Finally, relative position of replay shot should be after the close-up shot. However, this template scope was not used to detect other events, such as *yellow/red cards, penalties and free-kicks, shot/saves, penalties and free-kicks* which are based on the occurrence of referee shot and goal area respectively. Similarly, Duan et al (Duan et al., 2003) introduced a *mid-level representation layer* to separate sports specific knowledge and rules from the low-level and mid-level

feature extraction; thus making it less domain-specific. However, their event detection is still too domain specific since each event has different cinematic templates. For example, corner kick is detected when whistle is detected in the last two shots in the break segment and there are some goal-area views and (player) zoom-in views within the break segment. Moreover, the detection method for goal and foul/offside depends only on one feature (i.e. long excitement and long break for goal while whistle existence is used for foul/offside).

Based on these related work, we have outlined three main limitations from previous work on utilising cinematic features. First of all, they have not used a 'uniform' set of measurements to classify different highlights. For example, referee is only used for foul detection and not applicable for other highlights (Ekin and Tekalp, 2003b). Secondly, the templates are mostly based on manual observation which is very subjective and cumbersome (i.e. human's attention is very limited). Finally, there is yet a definitive suggestion on selecting the scope of features extraction (i.e. where to start and finish the extraction). To overcome these limitations, we have developed: 1) Novel algorithms for AFL play-break segmentation which is to be used as the definitive scope of *self-consumable* highlight detection (self-consumable means that the highlight segment can be watched as it is without referring to what happens before and after); 2) A novel statistical-driven template for highlight classification in AFL using a uniform set of audio-visual features. AFL is chosen as the primary domain due to the fact that there is yet any significant work presented this domain (as far as our knowledge). Moreover, AFL is one of the largest sectors in Australia's sport and recreation industry attracting more than 14 million people to watch all levels of the game across diversed communities. It is evident by the fact that AFL games are broadcasted live in Australia for (a total of) more than 10 hours per week (from Friday to Sunday), therefore increase the necessity for summarization.

The rest of this paper is structured as follows. In Section 2, we will present the overall framework for sports video summary extraction. In Section 3 and 4, we will present the algorithms for play-break and highlight classification respectively. Section 5 will be dedicated for experimental results while Section 6 will provide conclusion and future work

2   Summarization Framework

Play-break and *highlights* have been widely accepted as the semantically-meaningful segments for sport videos (Xu et al., 1998, Rui et al., 2000, Yu, 2003, Ekin and Tekalp, 2003b). 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. A highlight or key event represents an interesting (semantically important) portion of the game, such as goal and foul in soccer. Thus, we should integrate highlights to play-break to achieve a more complete summary (Tjondronegoro et al., 2004a). A Play-break sequence depicts a particular event which can be

classified into specific events (or highlights). The play segment describes the cause while break segment describes the outcome. Play segment provides the description of the event and can be annotated by specific key frames and/or audio and/or statistical diagrams. Break segment usually depicts the actors (or players) who were involved in the event. Thus, break segment can be annotated by the frames which contain face(s) region.

Using a simple browsing structure (described in Figure 1), users can choose to browse a sport video either by play-break sequences (like CD audio tracks), or collection of highlights (based on the category, such as goal, foul, etc). When a particular collection is selected, users can select the particular highlight segment. Each highlight segment will consist of play and break shots. On the other hand, if users prefer to browse by sequences, they can check whether the sequence contain a highlight or not. Users can watch the entire sequence or watch the highlight only (for shorter version).   The graphical interface of our video browser is depicted in Figure 3.
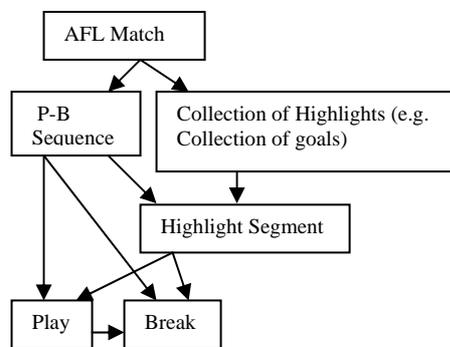

Figure 1: Browsing Structure

In order to construct this browsing structure, the summarization framework starts with cinematic features which can be directly extracted from the raw video data. For example, by applying experimental thresholds and domain knowledge, we can classify frames based on its camera-view (i.e. global, zoom-in and close up) from grass-ratio. Based on camera-views classification, play-break sequences are segmented. In the end, the statistical characteristics of each play-break will be used to classify the sequence into one of AFL highlights, including *goal, behind, mark, tackle and non-highlight.* This process is described in Figure 2, while the next two sections will describe play-break segmentation and highlight classification in more details.
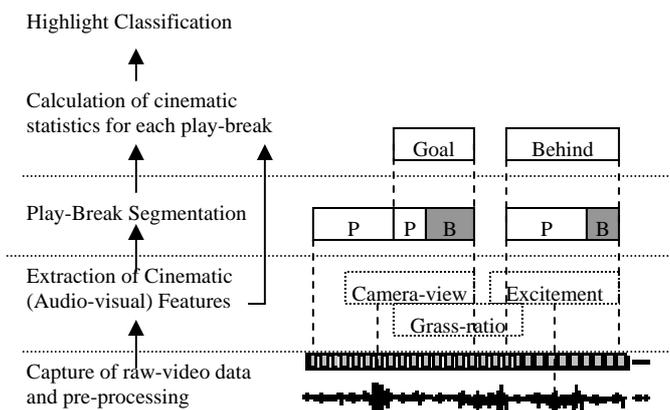

Figure 2: Summarization Processing Framework

Figure 3: Graphical User Interface for Browsing

3 Play-Break Segmentation

As shown in Figure 4, broadcasted AFL videos use transitions of typical shot types (i.e. global, zoom-in, and close-up) in order to emphasize story boundaries of the match. For example, a long global shot (interleaved shortly by other shots) is usually used to describe attacking play which could end because of a goal. After a goal is scored, zoom-in and close-up shots will be dominantly used to capture players and supporters celebration. 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 are the effective self-consumable containers for a semantic content since they contain all the required details. In particular, since most highlights lead to a break, we only use the last play-shot when there is a long sequence of play shots. However, users can choose to include more play shots, depending on how much detail on the play they want. Thus, we are reducing the subjectivity level of highlight's scope (e.g. compared to the case where users select particular frames). This concept is described in Figure 5.
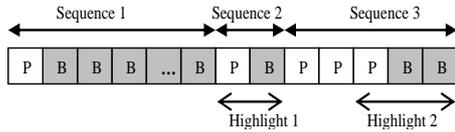

Figure 5: Scope of Highlights in Play-break Sequences

For play-break segmentation, view classification is firstly performed for each frame (with 1 second gap). We can use g*rass (or dominant color)-ratio*, which measures the amount of grass pixels in a frame, to classify the main shots in soccer video (Ekin and Tekalp, 2003a, Xu et al., 1998). Global shots contain the highest grass-ratio, while zoom-in contains less and close-up contains the lowest (or none). Thus, close-up shots generalize other shots,

such as crowd, substitutes, or coach close-ups which contain no grass. The first challenge is to decide the grass-hue index which is typically different from one stadium to another. A simple yet effective approach is to take random, equally-spread frame samples for an unsupervised training. Since global and zoom-in shots are most dominant, the peak from the total hue-histogram of these random frames will indicate the grass-hue. For our experiment, we take 20 random frames within a window of 5 minutes length. We also checked that the grass-hue value is within 0.15-0.25 since the initial segment of a video may contain non-match scenes. This process is repeated 10 times to calculate 10 variations of grass-hue indexes (i.e. $G_1, G_{2,\cdots} G_{10}$).

Grass Ratio (GR) is calculated on each frame as:

$$GR = P_G / P \tag{1}$$

where, P_G is the number of pixels which belong to grass-hue and P is the total pixels in a frame. Since there are 10 grass-hue indexes, the final GR is obtained from $\max(GR_1, GR_2, \cdots, GR_{10})$.
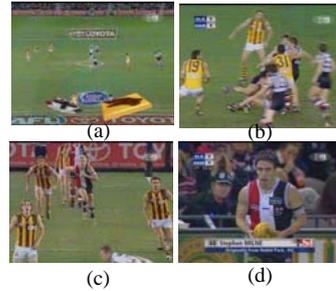

Figure 4: Camera-views in AFL video; a) Global, b) and c) Zoom-in, d) Close-up

We then need a set of thresholds which can distinguish the grass-ratio for the different shot types. For our AFL experiment we applied thres1=0.04 to 0.06 and thres2=0.2*thres1. Using these two thresholds, each frame can be classified into global, zoom-in, or close-up based on this rule:

$$FrameType \begin{cases} global, & GR \geq thres1 \\ zoom-in, & thres1 \geq GR \geq thres2 \\ close-up, & GR \leq thres2 \end{cases} \tag{2}$$

Duration of camera-views have been used successfully for play-break segmentation (Ekin and Tekalp, 2003a). We have applied a similar approach to design the algorithms for play-break segmentation which have been effectively for soccer video (Tjondronegoro et al., 2004b). The algorithm is described by Figure 6.
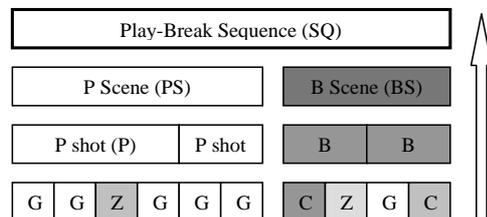

Figure 6: Play-break Segmentation

This play-break algorithm is generally applicable for AFL. Nevertheless, in order to achieve higher performance, we accommodated two main differences between AFL and soccer: Firstly, in AFL, there are many global shots which show a large portion of crowd due to the way that AFL is played (i.e. one player kick the ball high, and the other player of the same team can make a catch). Secondly, AFL uses more zoom-in shots during play than in soccer. To overcome the first problem, we applied a crop for all frames during *camera-view classification*. To accommodate the second difference, we applied *AFL_thres*.

Play-break segmentation Algorithms:

If there is yet an array of grass-hue indexes for the current video, perform semi-supervised training.

Run view-classification on each 1-sec gap frames [output: arrays for {G, Z, C}.fs and fe]

// Determines the start of Play and Break

Loop in G.fs array

  If abs(current G.fs - G.fe) > Pthres, add the G.fs to P.fs array

Loop in Z.fs array

  If abs(current Z.fs - Z.fe) > Bthres1, add the Z.fs to B.fs array

  Else if abs(current Z.fs - Z.fe) > AFL_thres, add the Z.fs to P.fs array

Loop in C.fs array

  If abs(current C.fs - C.fe) > Bthres2, add the C.fs to B.fs array

// Determines the end of Play and Break.

Sort both P.fs and B.fs arrays.

Concatenate P.fs and B.fs array into SEQ.

Sort SEQ ascending.

Loop in SEQ

  If current SEQ is element of P.fs array, add (next SEQ - fr) to P.fe array

  Else, add (next SEQ - fr) to B.fe array

Where: *G*=Global-, *Z*=Zoom-in-, *C*=Close-up- frames, *fs*, *fe, fr* = frame start, frame end, frame rate (e.g. 25 for PAL). For our AFL experiment, we applied: *P_thres* = 5*fr*, *AFL_thres* = 3*fr*, *B_thres1*=5*fr*, *B_thres2* = 2*fr*.

During camera-view classification, we applied a uniform cropping for all frames:

Frame' = imcrop(Frame,[1 (sz(:,1)/3) sz(:,2) sz(:,1)]);

Where: *Frame'* is the cropped frame and *Frame* is original. *Imcrop* performs an image cropping to a specified set of rectangle coordinates (specified as XMIN=*1*, YMIN=s*z(:,1)/3*, WIDTH *sz(:,2)*, HEIGHT= *sz(:,1)*). *Sz* is the size of frame; thus sz(:,1) is the height-size and sz(:,2) is the width-size.

The results of this cropping, as shown in Figure 8, have normalised the grass-ratio for global-view which has a large portion of crowd, as well as close-up with a large grass portion. The assumption is that, the playing ground (grass) will always be in the bottom half of the frame.

In most cases, there should be a long break after a goal is scored due to goal celebrations, and a wait for the players

to get into their formation. Thus, most broadcasters play a long replay scene(s) after the celebration to emphasize the goal and to keep viewers' attention while waiting for the play to be resumed. However, as described in Figure 8, some broadcasters insert some advertisements (ads) in-between the replay (e.g. after the first scene), or straight after the celebration. To obtain the correct length of the total break, we should not remove the ads or at least take into account the total length of the ads. Using the same method as camera-view classification, which is based on grass-ratio, we can detect frames that belong to ads (i.e. ads are classified as close-up since there is no grass). There is also a subtle increase of audio-volume in ads and a short total-silence before entering the ads. Moreover, the audio characteristics of ads and live sport match are very different (Han et al., 1998).

Figure 7: Results of Frame-cropping
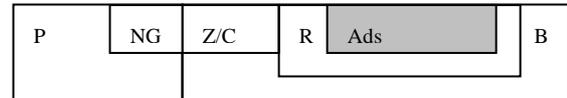
P=Play, NG=NearGoal, Z/C=Zoom-in/Close-up, R=Replay, Ads=Advertisements, B=Break.

Figure 8: Ads in between a Play-break Sequence

Benefits of using play-break as a definitive scope for the start and end of features observation (to detect highlight):

- It becomes possible to use comparative measurements (e.g. break ratio) which are more robust and flexible compared to definitive measurements (e.g. 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 is using play-break segment only while the other one is using play-break-play segment (Ekin and Tekalp, 2003b) or a static, empirical based, segment (Han et al., 2003)
- 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 since text can be 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) since it can belong to another highlight (or no highlight at all).

In addition to the play-break segmentation algorithm, replay detection is very important to locate additional breaks which are often recognized as play shots (i.e. replay shot often use global view). Moreover, to calculate the duration of a replay scene (for highlight detection), we need to identify the start and end of the replay scene. Replay scene is generally structured with editing effects at the beginning and end of one or more combinations of: slow motion shot, normal replay shot and still (paused) shot (Pan et al., 2001). We have investigated a wide variety of logo used by different broadcasters to mark the boundary of replay scene in soccer, AFL, rugby, and basketball. Based on the investigation, we constructed a generic and robust *logo-model*. Logo is meant to be contrast from the background and is usually animated within 10-20 frames with a general pattern of: "smallest– biggest (take up 40-50 percent of the whole frame – smallest". The main benefit of our approach, compared to the *color-based logo* model (Pan et al., 2002), is that we do not need to perform training for different broadcasters. Moreover, our logo template should comply to the examples of logo and the pattern used in (Pan et al., 2002) and (Babaguchi et al., 2000) which are depicted in Figure 9.
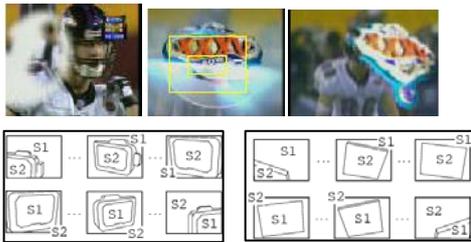

Figure 9: Typical Logo-pattern Model

Figure 10: AFL Replay Logo

Figure 11: Various Logo in Soccer and Rugby

Based on some experiments with different sports and broadcasters, this *logo-model* has been effective and robust for AFL logo (Figure 10) as well as various logos from soccer and rugby (Figure 11). Nevertheless, it has been noted that some frames from advertisement are primarily white, thus may be falsely detected as a logo (in our algorithm). To avoid this, we checked if the neighbouring frames contain grass. Moreover, some broadcasters do not use logo to emphasize replay scene. In this case, frame repetition (or frame rate) can be used to detect slow-motion shot (Ekin and Tekalp, 2003b, Pan et al., 2001). This approach relies on the fact that frequent and strong fluctuations in frame difference are generated by shot-repetition/drop (depending on the camera used during recording). In addition, some logos are not contrast from the rest of the frame. In such case, we need to use colour-based logo model (Pan et al., 2002).
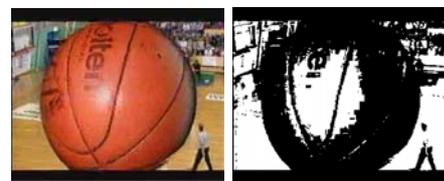
Figure 12: Non-contrast Replay Logo

Thus, to detect each replay scene, the following algorithms were developed.

Logo-model based Replay Scene Detection Algorithm

Find the frame with very large contrast object:

> Convert current frame into a stretched black and white (binary)

> Calculate the ratio of white pixels (*Pw*), large contrast object is found when the ratio is nearest to 0.5

Set the value of this *Pw* as the (current) largest ratio

Set the location of the frame as the *Middle* of transition

Check neighboring frames to find the Start and End of transition:

Keep on calculating the ratio of the previous frames while *Pw >= 0.25 & Pw < largest ratio*

Set the last previous frame-index as the *Start* of transition

Apply the same method to find the *End* of transition

(Note: 0.5 and 0.25 are empirical thresholds which can be adjusted to 0.6 and 0.15 respectively)

Post processing

If [abs (Start - Middle) <= 10 frames] & [abs (End - middle) <= 10 frames] set slo-mo location = *End*

Pair-up Slo-mo location with distance <= 20 sec

To remove false detections, remove slomo locations which do not have any pair.

Set each pair's locations as start and end of slow motion replay scene

To find large contrast object, we can use the following MATLAB 6.5 functions:

```
Frame' = rgb2gray(frame);
Frame'' = imadjust(frame',stretchlim(frame'),[]);
Frame''' = roicolor(frame'',128,255);
[Hcounts,Hi] = imhist(Frame''');
Contrast ratio = Hcounts(2)/sum(Hcounts);
```

In order to obtain the final play-break sequences, Figure 13 below shows the various scenarios on 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. Thus, R.s is short for the start of replay scene. Scenarios 6 and 7 fix the neighbouring play-break sequence (i.e Seq1.e = Seq2.s OR [Seq2.e - Seq1.e] < short_dur). The scenarios are described as follows (bullet points are the algorithms for the outcome):

(1) [R *strict_during* P] & [(R.e – P.e) >= dur_thres] : locate additional breaks (from play shots)
- 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) [R *strict_during* P] & [(R.e – P.e) <= dur_thres]
- P.e = R.e
- B.s = R.e+1

(3) [R *meets* B] & [R.s < P.e]
- P.e = R.s

(4) [R *during* B] & [R *meets* B]
- Do nothing

(5) [R *strict_during* B]
- Do nothing

(6) [R *during* B] & [(R.e – P_2.s) >= dur_thres]
- B.e = R.e
- Amend the neighbor sequence: [P_2.s = R.e+1]

(7) (Pan et al.) & [(R.e – P_2.s) >= dur_thres]
- Attach sequence 2 to sequence 1 (i.e. combine seq 1 and seq 2 into 1 sequence)

Where:

If A *strict_during* B, (A.s > B.s) & (A.e < B.e)

If A *during* B, (A.s > B.s & A.e <= B.e) OR (A.s >= B.s & A.e < B.e)

If A *meets* B, A.e = B.e

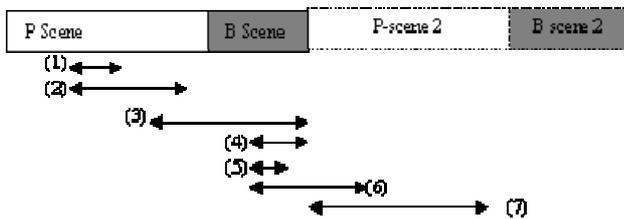dur_thres can be set to 2-4 seconds

Figure 13: Locations of Replays in Play-breaks

4 Automatic Highlights Classification

Highlights are generically the interesting events that may capture user attentions. Thus, most broadcasters will distinguish them by inserting slow-motion replay scene(s), artificial text display. For most sports, highlights can be detected based on specific audio-visual characteristics, such as excitement, whistle, and goal-area. While generic key events are good for casual video skimming, domain-specific (or classified) highlights will support more useful browsing and query applications. For example, users may prefer to watch all the goals only. Thus, we will present in this paper some statistical-driven algorithms for highlight classification applied for AFL videos (please note that this approach has been applied successfully for soccer videos). In particular, we have successfully classified a sequence into: *goal, behind* (regardless they are during normal play or free kicks)*, mark, tackle, and non-highlight.*

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 metres 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 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).

Unlike most of previous work which rely on manual investigation and knowledge to construct the highlight detection algorithms, 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 very little adjustments. For this purpose, we have conducted a semi-supervised training on 20 samples from 5 matches for each highlight in order to determine the characteristics of play-break sequences containing different highlights and no highlights. It is semi-supervised since we manually classified the specific highlight that each play-break sequence (for training) contains. It should be noted that a separate training should be performed for non-highlight to find its distinctive characteristics (rather than just applying a threshold).

Based on the training data, we have produced the statistics of each highlight (depicted in last page - Figure 16) using the following variables:

- *SqD* = duration of (currently-observed) play-break sequence. We can predict that a sequence in which a a goal can be found will be much longer than a sequence with no highlight .

- *BR* = duration of break / *SqD*. Rather than measuring the length of a break to determine a highlight (like in (Ekin and Tekalp, 2003b)), 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.

- *PR* = duration of play scene / *SqD*. We found that non-highlight sequence has the highest play ratio since it contains very little break.

- *SlD* = 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.

- *ER* = duration of excitement / *SqD*. Typically, goal consists a very high excitement ratio while 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 play potentially indicate goal or behind.

- *CR* = length of close-up views within the sequence / *SqD*. We found 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 (i.e. the shooter) and goal celebration. Advertisements after a goal will be detected as close-up (i.e. no grass).

Based on the trained statistics, we have constructed a novel '*statistical-driven*' cinematic template for AFL highlights as shown in Figure 6. Thus, when we add more training, these values need to be updated.
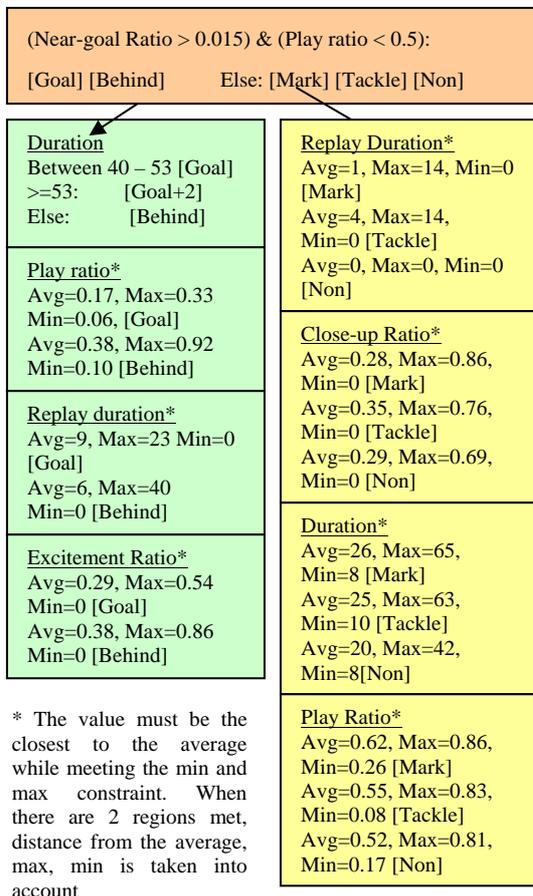
This highlight-classification template was designed primarily based on the statistics since we did not need to use any domain-specific knowledge and thus less-subjective. In most cases, when a near goal is detected and break ratio is more dominant that play, it is likely that the sequence contains goal or behind. Otherwise, it is more likely that we will find a mark, tackle or non-highlight. Thus, we need to further distinguish goal from behind, and then mark/tackle/non:

- *Goal vs. Behind:* Compared to behind, goal has longer duration, less replay and excitement (due to advertisement in-between).

- *Mark vs. Tackle vs. Non:* Non does not contain any replay, while tackle in average contains longer replay than mark. Non has the lowest close-up ratio compared to mark and tackle. Non has the shortest duration compared to mark and tackle.
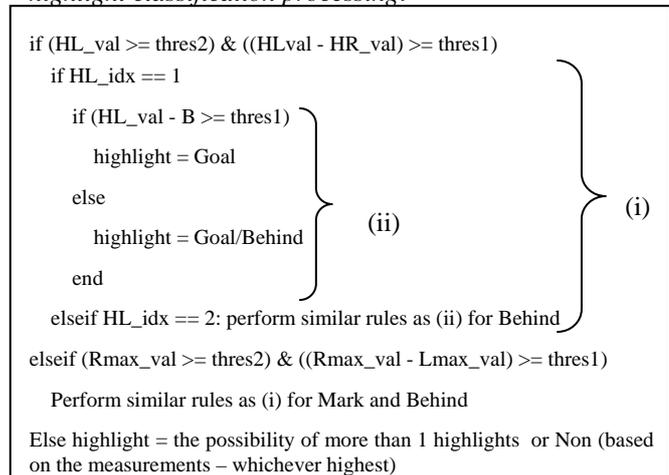
In order to classify which highlight is contained in a sequence, we used some *measurements*: *G,B,M,T* and *Non*, where *G* is the possibility point that the sequence contains a goal and *B,M,T, Non* is the possibility for behind, mark, tackle, and non-highlight respectively. Each of these measurements is incremented by 1 point when certain rules are met (as indicated in the diagram). It should be evident that the maximum possible points for goal/behind and mark/tackle/non should be equal (i.e. 50:50 chances). The only bonus point is when goal/behind is more likely and the duration is >=53 which is the maximum possible duration for behind and the minimum duration for goal. In addition, we applied some post-calculations by performing (* in Figure 14) for each statistics on: duration, play ratio, near-goal, excitement, close-up ratio, and replay duration.

Based on these measurements, the following variables for highlight classification are calculated

$$[HL\_val, HL\_idx] = max(G,B)$$

$$[HR\_val, HR\_idx] = max(M,T)$$

where, *HL_val* is the maximum value of (*G,B*). Thus, *HL_idx* is the index of *HL_val*. For example if the maximum value is B, HL_idx will be equal to 2. The same concept is applied for *HR_val* and *HR_idx*. Using these variables, the followings describe the rules for *highlight classification processing*:

```
if (HL_val >= thres2) & ((HLval - HR_val) >= thres1)
   if HL_idx == 1
      if (HL_val - B >= thres1)
         highlight = Goal
      else
         highlight = Goal/Behind
      end                                    (ii)     (i)
   elseif HL_idx == 2: perform similar rules as (ii) for Behind
elseif (Rmax_val >= thres2) & ((Rmax_val - Lmax_val) >= thres1)
   Perform similar rules as (i) for Mark and Behind
Else highlight = the possibility of more than 1 highlights  or Non (based
on the measurements – whichever highest)
```

(Near-goal Ratio > 0.015) & (Play ratio < 0.5):

[Goal] [Behind]   Else: [Mark] [Tackle] [Non]

Duration
Between 40 – 53 [Goal]
>=53:   [Goal+2]
Else: [Behind]

Play ratio*
Avg=0.17, Max=0.33
Min=0.06, [Goal]
Avg=0.38, Max=0.92
Min=0.10 [Behind]

Replay duration*
Avg=9, Max=23 Min=0
[Goal]
Avg=6, Max=40
Min=0 [Behind]

Excitement Ratio*
Avg=0.29, Max=0.54
Min=0 [Goal]
Avg=0.38, Max=0.86
Min=0 [Behind]

Replay Duration*
Avg=1, Max=14, Min=0
[Mark]
Avg=4, Max=14,
Min=0 [Tackle]
Avg=0, Max=0, Min=0
[Non]

Close-up Ratio*
Avg=0.28, Max=0.86,
Min=0 [Mark]
Avg=0.35, Max=0.76,
Min=0 [Tackle]
Avg=0.29, Max=0.69,
Min=0 [Non]

Duration*
Avg=26, Max=65,
Min=8 [Mark]
Avg=25, Max=63,
Min=10 [Tackle]
Avg=20, Max=42,
Min=8[Non]

Play Ratio*
Avg=0.62, Max=0.86,
Min=0.26 [Mark]
Avg=0.55, Max=0.83,
Min=0.08 [Tackle]
Avg=0.52, Max=0.81,
Min=0.17 [Non]

* The value must be the closest to the average while meeting the min and max constraint. When there are 2 regions met, distance from the average, max, min is taken into account

Figure 14: Statistical Template for AFL Highlight Classification

Consequently, *thres2* is the minimum points for a measurement to be accurate, while *thres1* is the minimum difference between measurements (i.e. how significant is the confident). For the experiment, we have set thres2 >= 4 (while 3 is still considered as a low chance if no measurement is above 3) and thres1 = 2.

For extraction of cinematic features, such as excitement, and near-goal, readers can find the algorithms and thresholds used in (Tjondronegoro et al., 2004b, Tjondronegoro et al., 2004a). The only adjustment we made for AFL is the goal-area detection. In AFL, goal and behind posts can be detected as vertical (usually parallel) lines, as shown in Figure 15. These lines are detected as strong peaks in the *Hough* transform (compared to a threshold) which is calculated from the gradient image of a frame. A gradient image can be produced either by *Canny* or *Sobel* transform. The more goal lines we can detect in a frame, the higher probability that the frame shows goal-area.

Figure 15: Goal- and Behind- Posts in AFL

5   Experimental Results

In this section, we will only focus on discussing the performance of the algorithms for highlight classification. Readers can find comprehensive reports on the performance of cinematic features extraction in our earlier papers, such as (Tjondronegoro et al., 2004b).

During experiment, we have used 5 AFL matches from *channel 9* for training and highlight classification. Video 6 was recorded from *channel 10* to show that our algorithms are robust for different broadcasters. The algorithms were implemented using MATLAB 6.5 with standard image processing toolbox.

In order to measure the performance of highlights classification, Recall and Precision rates are not so accurate and expressive. The main reason is: we need to see precisely where the miss- and false- detections are. Moreover, we should realise that when goal is detected as behind, it is not as bad as when it is detected as mark/tackle. Likewise, when mark detected as tackle or non-highlight it is not as bad as when it is detected as goal/behind. Hence, the following tables (Table 1-6) will present the results of highlight classification for each AFL video (each video contains 1 whole quarter – without any editing considerations). In these tables, highlighted numbers signify correct detections. In addition to these tables, however, we have provided the recall and precision rates in Table 8. Moreover, to show the robustness of our statistical-driven approach, we have applied the same method to soccer successfully. The results for soccer are depicted in Table 7 and 9.

Please note that Table 8 was derived from Table 1-6. In particular, Recall Rate (RR) is calculated as: (correct detection / total truth) * 100%, while Precision Rate (PR) is calculated as: (correct detection / total detected) * 100%. Based on Table 8, it is clear that our highlight classification is most accurate for *goal*, *tackle*, and *mark* in its respective order. Although the RR and PR are relatively low for behind detection, Table 1-6 should show that most *behinds* are detected as *goal*. Moreover, low RR for non-highlight detection (caused by miss detections) can be considered less significant since it means users will have additional highlights.

6   Conclusion and Future Work

Current approaches in sport video highlights classification have not used the definitive scope of detection and uniform set of measurements for different highlights (and sport genres). We have demonstrated in this paper that play-break can be used as an effective scope of highlights detection since they contain all the necessary details and features. We have also used a uniform set of measurements for all types of highlights which were also used for soccer.

Ground truth	Highlight classification of video 1 (Col-HAW3)					
	Goal	Behind	Mark	Tackle	Non	Truth
Goal	4	0	0	0	0	4
Behind	3	3	0	0	0	6
Mark	2	0	3	1	0	6
Tackle	0	0	1	4	0	5
Non	0	0	1	1	7	9
Detected	9	3	5	6	7	

Table 1: AFL Highlights Classification Results 1

Ground truth	Highlight classification of video 2 (BL-ESS)					
	Goal	Behind	Mark	Tackle	Non	Truth
Goal	9	0	0	0	0	9
Behind	1	0	0	0	0	1
Mark	5	0	0	0	0	5
Tackle	0	0	0	3	0	3
Non	2	1	0	1	2	6
Detected	17	1	0	4	2	

Table 2: AFL Highlights Classification Results 2

In order to avoid manual and subjective- based rules for highlight detection, we have proposed a novel approach that is based on the phenomenal statistics of features for each play-break containing different highlights. These statistics have been used to construct an effective template for AFL highlights classification with little domain-specific knowledge. Thus, we should be able to apply the same approach for other sport genres, such as soccer. Based on our experiment in AFL domain, we have used almost all the algorithms that we used for soccer in our earlier work. Thus, the algorithms presented in this paper should be robust for many other sports (at least the ones that have similar characteristics to AFL and soccer). For future work, we aim to experiment with the robustness of the proposed approach for team-based sports such as basket-ball, net ball, rugby, and hockey.

7   References

Babaguchi, N., Kawai, Y., Yasugi, Y. and Kitahashi, T. (2000): Linking Live and Replay Scenes in Broadcasted Sports Video. *Proc. ACM Workshop on Multimedia*, Los Angeles, California, United States, 205-208, ACM Press.

Duan, L.-Y., Xu, M., Chua, T.-S., Qi, T. and Xu, C.-S. (2003): A Mid-level Representation Framework for Semantic Sports Video Analysis. *Proc. ACM International Conference on Multimedia*, Berkeley, USA, 33-44, ACM Press.

Ekin, A. and Tekalp, A. M. (2003a): Generic play-break event detection for summarization and hierarchical sports video analysis. *Proc. International Conference on Mulmedia and Expo,* **1**:6-9, IEEE.

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

Han, K.-P., Park, Y.-S., Jeon, S.-G., Lee, G.-C. and Ha, Y.-H. (1998): Genre classification system of TV sound signals based on a spectrogram analysis. *IEEE Transactions on Consumer Electronics* **44**:33-42.

Han, M., Hua, W., Chen, T. and Gong, Y. (2003): Feature design in soccer video indexing. *Proc. International Conference on Communications and Signal Processing*, **2**:950-954, IEEE.

Nepal, S., Srinivasan, U. and Reynolds, G. (2001): Automatic detection of 'Goal' segments in basketball videos. *Proc. ACM International Conference on Multimedia,* Ottawa; Canada, 261-269, ACM Press.

Pan, H., Li, B. and Sezan, M. I. (2002): Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions. *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing,* **4**:3385-3388, IEEE.

Pan, H., van Beek, P. and Sezan, M. I. (2001): Detection of slow-motion replay segments in sports video for highlights generation. *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing,* Salt Lake City, USA, **3**:1649-1652, IEEE.

Rui, Y., Gupta, A. and Acero, A. (2000): Automatically extracting highlights for TV Baseball programs. *Proc. ACM International Conference on Multimedia,* Marina del Rey, California, USA, 105-115, ACM Press.

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-Dec2004**:22-37.

Tjondronegoro, D., Chen, Y.-P. P. and Pham, B. (2004b): The Power of Play-Break for Automatic Detection and Browsing of Self-consumable Sport Video Highlights. *Proc. ACM Workshop on Multimedia Information Retrieval,* New York, USA, 267-274, ACM Press.

Xu, P., Xie, L. and Chang, S.-F. (1998): Algorithms and System for Segmentation and Structure Analysis in Soccer Video. *Proc. IEEE International Conference on Multimedia and Expo*, Tokyo, Japan, IEEE.

Yu, X. (2003): Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. *Proc. ACM International Conference on Multimedia,* Berkeley, USA, 11-20, ACM Press.

Ground truth	Highlight classification of video 3 (Col-Gel2)					
	Goal	Behind	Mark	Tackle	Non	Truth
Goal	**4**	0	0	0	0	4
Behind	1	**1**	0	0	1	3
Mark	0	1	**3**	0	1	5
Tackle	1	0	2	**1**	0	4
Non	1	1	0	0	**2**	4
Detected	7	3	5	1	4	

Table 3: AFL Highlights Classification Results 3

Ground truth	Highlight classification of video 4 (StK-HAW3)					
	Goal	Behind	Mark	Tackle	Non	Truth
Goal	**2**	0	0	0	0	2
Behind	3	**2**	1	0	0	6
Mark	1	0	**7**	0	1	9
Tackle	0	0	0	**2**	0	2
Non	0	1	1	0	**2**	4
Detected	6	3	9	2	3	

Table 4: AFL Highlights Classification Results 4

Ground truth	Highlight classification of video 5 (Rich-StK4)					
	Goal	Behind	Mark	Tackle	Non	Truth
Goal	**3**	0	0	0	0	3
Behind	0	**2**	2	0	0	4
Mark	1	0	**6**	2	1	10
Tackle	1	0	5	**1**	1	8
Non	1	0	1	0	**5**	7
Detected	6	2	14	3	7	

Table 5: AFL Highlights Classification Results 5

Ground truth	Highlight classification of video 6 (BL-ADEL)					
	Goal	Behind	Mark	Tackle	Non	Truth
Goal	**7**	0	0	0	0	7
Behind	2	**3**	4	0	0	9
Mark	2	0	**8**	2	0	12
Tackle	0	0	1	**6**	0	7
Non	0	0	6	1	**6**	13
Detected	11	3	19	9	6	

Table 6: AFL Highlights Classification Results 6

Ground truth	Highlight classification of 4 full-match videos				
	Goal	Shot	Foul	Non	Truth
Goal	**3**	4	0	0	7
Shot	7	**91**	16	3	117
Foul	6	22	**57**	16	101
Non	2	10	32	**183**	227
Detected	18	127	105	202	

Table 7: Soccer Highlights Classification Results from 4 Full Matches
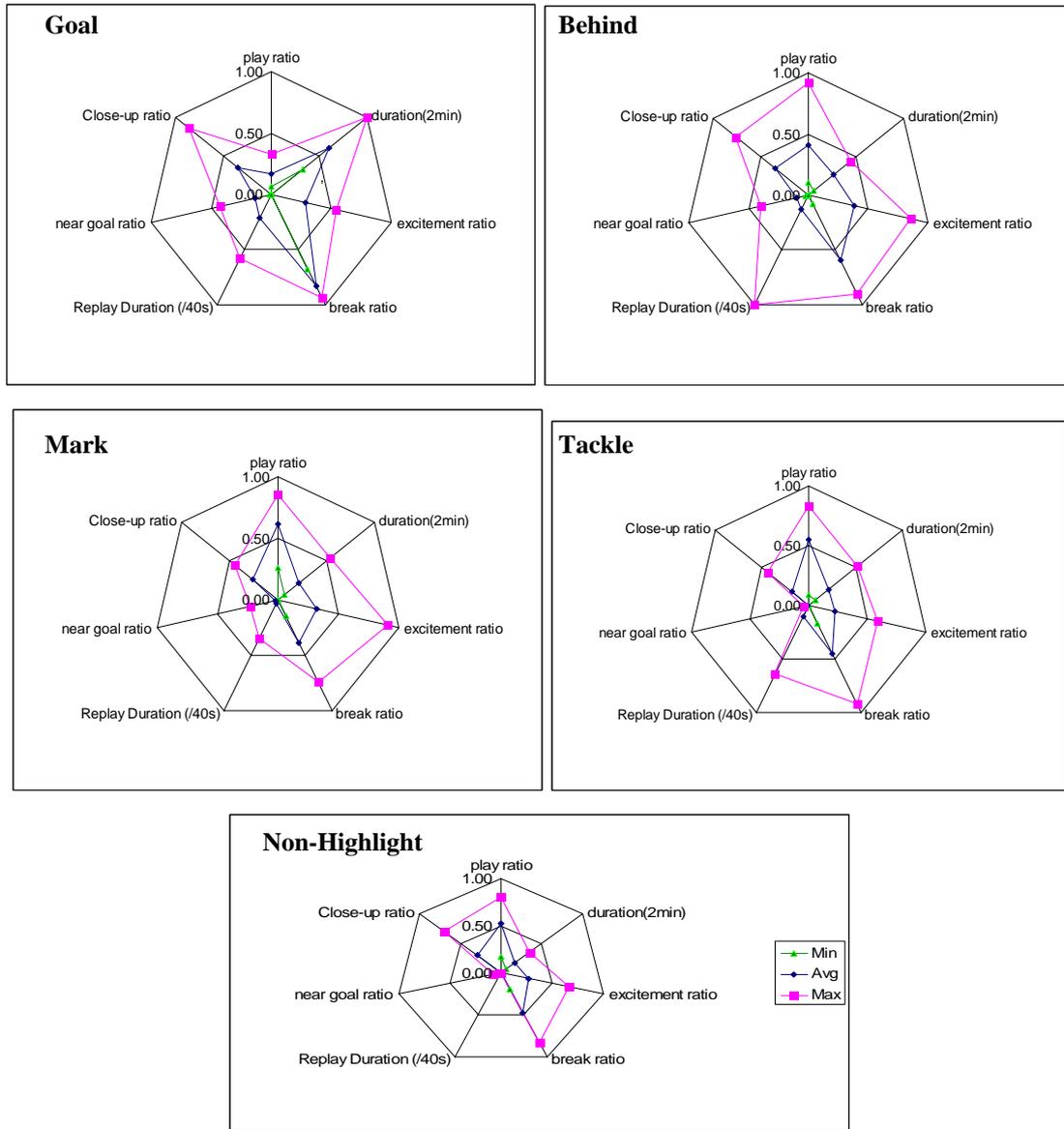
Figure 16: Statistics of Highlights After 20 Samples Training

	Video 1		Video 2		Video 3		Video 4		Video 5		Video 6		AVG	
	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	50.2
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	40.0	80.0
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	62.9	56.5
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	67.2	73.6
Non	77.8	100.0	33.3	100.0	50.0	50.0	50.0	66.7	71.4	71.4	46.2	100.0	54.8	81.3

Table 8: Recall (RR) and Precision Rates (PR) in AFL Highlights Classification Results

	Video 1		Video 2		Video 3		Video 4		Average	
	RR	PR	RR	PR	RR	PR	RR	PR	RR	PR
Goal	40	66.7	N/A	N/A	N/A	N/A	100	25.0	70	45.8
Shot	88.2	63.8	73.5	69.4	80.0	76.2	69.0	87.0	77.7	74.1
Foul	28.6	85.7	55.6	42.9	65.9	90.0	75.0	27.3	56.2	61.5
Non	97.1	89.2	71.9	86.8	91.5	90.0	71.4	96.2	83	90.5

Table 9: Recall (RR) and Precision Rates (PR) in AFL Highlights Classification Results