USING CAMERA MOTION TO IDENTIFY TYPES OF AMERICAN FOOTBALL PLAYS

Mihai Lazarescu, Svetha Venkatesh

School of Computing, Curtin University of Technology,
GPO Box U1987, Perth 6001, W.A.
Email: lazaresc@cs.curtin.edu.au

ABSTRACT

This paper presents a method that uses camera motion parameters to recognise 7 types of American Football plays. The approach is based on the motion information extracted from the video and it can identify short and long pass plays, short and long running plays, quaterback sacks, punt plays and kickoff plays. This method has the advantage that it is fast and it does not require player or ball tracking. The system was trained and tested using 782 plays and the results show that the system has an overall classification accuracy of 68%.

1. INTRODUCTION

The automatic indexing and retrieval of American Football plays is a very difficult task for which there are currently no tools available. The difficulty comes from both the complexity of American Football itself as well as the image processing involved for an accurate analysis of the plays.

American Football is a highly structured game and it involves set plays based on a set of predefined player formations. The player formations can be recognised with good accuracy. The plays, however, are extremely difficult to identify because it requires either the accurate tracking of the ball or the accurate tracking of the player that has the ball. As this is a contact sport where players tackle each other, it is very difficult to track the player or the ball consistently because of occlusion.

Based on the hypothesis that each play has a unique signature in terms of motion of the camera tracking the play, we formulate a method to recognise the major types of American Football plays. Our approach is based only on the camera motion parameters and has the advantages that it is fast and it avoids detailed low level image segmentation. Furthermore, it is more robust as it does not depend on the reliable tracking of the players or the ball.

The paper is organised as follows: in Section 2 we cover previous work done in interpreting American Football and in extracting camera motion parameters. Section 3 describes our approach to identify American Football plays using camera motion while Section 4 briefly covers the supervised learning algorithm used to train the system. The results are described in Section 5 with the conclusions presented in Section 6.

2. PREVIOUS WORK

Motion parameters have been used in the past since they provide a simple, fast and accurate way to search multimedia databases for specific shots (for example a shot of a landscape is likely to involve a significant amount of pan, whilst a shot of an aerobatic sequence is likely to contain roll).

Previous work done towards the recognition of American Football plays has generally involved methods that combine clues from several sources such as video, audio and text scripts.

A method for tracking players in American Football greyscale images that makes use of context knowledge is described by Intille [2]. In his work, Intille proposes the use of "closed-world" analysis to incorporate contextual knowledge into low-level tracking. Though the tracking method in general produces good results, it is still likely to fail in situations where the player models are imprecise, spatial resolution is low or when the object tracked is very close in appearance to some nearby feature. Intile and Bobick [3] also used belief networks to model American Football plays. In [3] they present an approach that uses temporal graphs to represent complex multi-agent actions. These temporal graphs could be used to recognise American Football plays from noisy data.

A different approach to process American Football video was used by [4] and [5]. The approach by [4] uses the closed caption of the games to extract highlights from American Football games. The method involves a search for predefined keywords which are used to label segments of video for later retrieval. The work by [5] uses a similar approach but it involves a more detailed analysis of textual and audio clues from commentary of the game. The clues extracted from text and audio are combined to generate a high level description of the video which is then used to retrieve the requested video segments

3. DEFINING AMERICAN FOOTBALL OFFENSIVE PLAYS USING CAMERA MOTION PARAMETERS

Each one of the 7 major types of American Football offensive plays were defined using the pan and tilt camera motion paramters of the video shot containing the play. The major assumption of our work is that video analysed is obtained from cameras that provide pictures from the main stands. The pan and tilt parameters were extracted for each frame in the video containing the play.

To extract the camera motion parameters from a sequence of images we use a method developed by Srinivasan *et al* [1]. The method can qualitatively estimate camera pan, tilt, zoom, roll, and horizontal and vertical tracking (for a detailed description of the technique see [1]).

The raw camera parameter values were first filtered to remove noise and then were used to derive a symbolic description of the plays. The filtering process involved the analysis of the camera motion parameter values and the removal of values that were not consistent with the overall motion of the camera.

Each play was defined in terms of the following five features:

Number of stages in the camera motion: This feature indicates the number of times the camera changes direction during the play. The reason for extracting this feature is to determine when the main action of the play occurs. Typically an offensive play has three stages. In the first stage, the ball is given to the quarterback and the quarterback moves back from the scrummage line while looking for a teammate to pass or give the ball. In the second stage of the play, the quarterback passes the ball or gives the ball to one of his teammates who in turn attempts to take the ball as far forward as possible. In the third stage, the player carrying the ball is either tackled, runs out of the playing area or completes the play by scoring a touchdown. Therefore, the camera pan motion is generally as follows (assuming that movement to the left is movement forward): stage 1 - pan to the right, stages 2 and 3 pan to the left. An example is shown in Figure 1. While this feature is in most cases used to determine when the main action of the play occurs (such as the quarterback passing the ball), the number of stages in the play can also be used to quickly identify two major types of plays: kickoffs and quarterback sacks. In the case of the kickoff there are only two stages in the play: the kickoff stage and the kickoff return stage. The quarteback sack play has only one stage when the quarterback moves back from the line of scrummage. The symbolic values for this feature are: *one,two,three, four.*

Net pan movement of the camera: The net movement was computed by computing the difference between the pan movement occuring during the first and second stages of

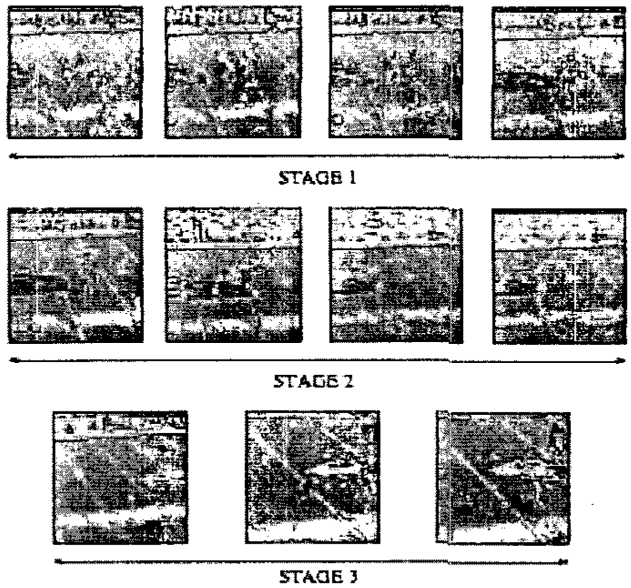

Fig. 1. Stages in a typical NFL play (permission from Channel 9, Australia).

the plays. Essentially, this feature represents the difference between the movement of the camera to cover the backward motion of the quarterback at the start of the play and the movement of the camera to cover the forward motion required to track the forward movement during a passing action or a ball carry action. The net pan movement indicates whether the type of play is long or short. We defined short plays as those that involve the offensive team gaining a small number of yards (up to 10 yards). The long plays involve the offensive team gaining a large number of yards. The symbolic values for this feature along with net pan movement are shown in Table 1.

Net Pan Movement Value Range	Symbolic Value
0-50	Small
50-100	Medium
≥ 100	Large

Table 1. Symbolic values classifying the net pan movement

Speed Value Range	Symbolic Value
0-3	Slow
3-5	Medium
≥ 5	Fast

Table 2. Symbolic values classifying the speed of the pan movement

Speed of the pan movement: This feature was used to determine whether a pass play or a running play occured as the two types of plays typically have different speeds. In the case of the running play, the quarterback gives the ball to a player who then attempts to carry it forward. For such a play, the camera motion which tracks the player's movement is generally consistent and the speed of the movement is slow. In the case of the pass play the camera tends to move rapidly for a brief period of time as the speed of the ball when it is passed is significantly faster than that of a player running with the ball. The symbolic values for this feature are shown in Table 2.

Tilt motion during the play action: Once the set of frames containing the play action had been delimited, we analysed the tilt motion of the camera to determine whether a significant amount of movement up or down occured during the play. The symbolic values for this feature are shown in Table 3.

Tilt Value Range	Symbolic Value
Large Positive Values	Up
Large Negative Values	Down
Close to Zero	Centered

Table 3. Symbolic values classifying the tilt movement

Angle Value Range	Symbolic Value
0 - 45	SmallR
45 - 90	LargeR
-45 - 0	SmallL
-90 - -45	LargeL

Table 4. Symbolic values for the Camera Angle

Camera angle: The purpose of this feature is to help refine the definition of the short and long running plays once the overall tilt motion has been determined. Running plays involve the movement of the player carrying the ball through the gaps that may occur in the defence line of the opposing team. Two approaches are used. The first approach involves the offensive team attempting to create a gap in the middle of the opposing team's defence line (see Figure 2). The second approach involves the player carrying the ball going around the extremities of the opposing team's defence line (the player runs to the left or right of the defensive line as shown in Figure 3). This feature was used to help determine whether the player attempted to run through the middle of the defensive line or whether the player attempted to go around the defensive line. The symbolic values for this feature are shown in Table 4.

Once the system had analysed the video and generated a symbolic description of the play, the definitions were used to train and classify the plays.

4. RESULTS

For the classification and learning process we used the incremental learning algorithm ILF [6]. ILF learns by creating new concepts that are added to the concept hierarchy or by updating the existing concepts. Each concept in the hierarchy has a *age* value associated with it that indicates the number of times that concept has been observed by the system. To keep the concept hierachy consistent with the data observed, ILF applies a forgetting mechanism that uses the age value to prune the conceptual hierarchy of noise or irrelevant information.

The learning process has two steps. In the first step the symbolic definitions of the plays are analysed by ILF to determine whether the knowledge base contains a definition similar to the one observed. If such a defition exists, then depending on the level of similarity it may be reinforced. If no match is found, then ILF creates a new concept to represent the play. The matching process involves computing a similarity score that takes into account both the number of matching features as well as the age associated with each feature value.

For the testing and classification process we used 10 hours of American Football footage which covered 3 complete games as well as several highlights from 20 other games. We extracted a total of 782 plays from the American Football footage. The length of each shot covering the play varied between 70 and 420 frames. The resolution of the video frames was 320x200 and the capture rate was 25 frames per second. The plays extracted covered all 7 types of play to recognise but the distribution of the plays varied depending on the play type. The most frequent plays were the short running plays and the passing plays, while the quarterback sacks and the long running plays had a very low frequency (generally the quarterback releases the ball before he gets tackled and the running players rarely manage to gain a large number of yards). To train the system we used a total of 440 plays while the remaining number of plays were used to test the system. The overall results are shown in Table 5. In many of the cases where the short and long plays were missclassified, the action involved the quarterback faking passing movements (which affected the camera motion). The results show that the system has an overall accuracy of 68%. The quarterback sack and the kickoff are classified with better accuracy mainly because of the fact that the plays are very distinctive when compared to the other five plays (smaller number of stages). The system was also able to distinguish between the short and long plays with reasonable accuracy. There were 243 short and long plays in the test and the confusion matrix is shown in Table 6.

We also attempted to obtain a more detailed classification of the short running plays using the tilt motion of the camera. We therefore attempted to determine whether the

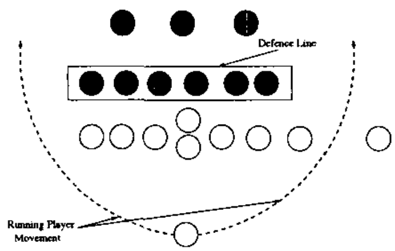

Fig. 2. Running play type 1.

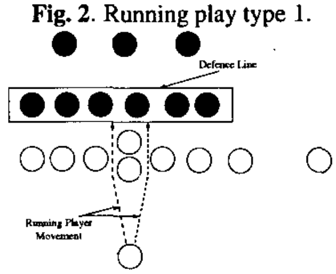

Fig. 3. Running play type 2.

Play Type	No. Test Instances	TP	FP	FN	Accuracy%
Short Pass	76	49	18	1	64
Long Pass	29	20	12	5	68
Short Run	112	71	40	2	63
Long Run	26	17	5	2	65
QB. Sack	21	17	3	2	80
Punt	36	23	4	15	63
KickOff	32	25	4	6	78
Total	332	222	86	24	68

Table 5. Classification results for the American Football plays test data.

Play Type	Short Pass	Long Pass	Short Run	Long Run
Short Pass	49	10	17	2
Long Pass	5	20	1	3
Short Run	18	3	71	5
Long Run	5	8	4	17

Table 6. The confusion matrix for the short and long plays

system could distinguish between running plays through the middle of the defence line and running plays around the defence line. Of the 99 test plays used, 78 were plays involving runs through the middle of the defence line while 21 plays involved runs around the defence line. The system was able to recognise the running plays around the defence line with an accuracy of 62%, while the running plays through the middle of the defence line were classified with an accuracy of 76%.

5. CONCLUSIONS

In this paper we describe a method based on camera motion parameters to classify several types of American Football plays. The approach is simple, fast and avoids complex image segmentation. It also has the advantage that it could be applied to different camera positions. However, the method has the disadvantage that it does not provide a detailed recognition of the American Football plays. It can be used to determine whether the ball is passed or carried but it cannot provide enough detail to have a specific recognition of a certain play. A total of 782 plays were used for the training and classification process and the results show that the system correctly classifies 68% of the plays. The method described in this paper can be applied to other sports such as cricket and baseball.

6. REFERENCES

[1] Srinivasan M., S. Venkatesh, and R. Hosie, "Qualitative estimation of camera motion parameters from video sequences," *Pattern Recognition*, vol. 30, pp. 593–606, 1997.

[2] Stephen S. Intille, "Real-time closed-world tracking," in *International Conference on Computer Vision and Pattern Recognition*. June 1997, pp. 697–703, IEEE Computer Society Press.

[3] Stephen S. Intille and Aaron F. Bobick, "A framework for recognizing multi-agent action from visual evidence," in *Proceedings of AAAI99*, April 1999.

[4] Noboru Babaguchi and Ramesh Jain, "Event detection from continuous media," in *Proceedings of ICPR98*, August 1998, pp. 1209–1212.

[5] Naoko Nitta, Noboru Babaguchi, and Tadahiro Kitahashi, "Story based representation for broadcasted sports video," in *Proceedings of ICME2002*, August 2002, pp. 813–817.

[6] Mihai Lazarescu, Svetha Venkatesh, and Geoff West, "Incremental learning with forgetting (i.l.f.)," in *Proceedings of ICML-99 Workshop on Machine Learning in Computer Vision*, June 1999.