Video Summarization Based on User Interaction

Dan R. Olsen Jr.
Brigham Young University
olsen@cs.byu.edu

Brandon Moon
Pixelture, Inc
brandon@pixelture.com

ABSTRACT

It is frequently the case that viewers want to watch sports in less time than a game takes to play. Computing a summarization of a game depends upon information about each play. In contrast with previous video summarization work this paper describes how the interactive behavior of prior viewers can be used to compute which plays are most interesting. This passive feedback is examined as a possible source for the degree of interest in a play. In addition, a mechanism is described for briefly showing the game action that was removed to create the summary. This helps to preserve the continuity of the game in the viewer's mind.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *video.*

General Terms

Design, Human Factors

Keywords

Video, interactive television, sports

1. INTRODUCTION

Recent advances in internet video technology are the precursors of a massive shift in television delivery. Throughout the world, television is the dominant form of mass communication and entertainment. However, the underlying technology is shifting from channels defined by RF spectrum to IP-based internet delivery. There are two major trends in the current shift. The first is the democratization of video as seen on YouTube and the second is the viewer's ability to watch what they want whenever they want rather than conforming their lives to a television schedule. These trends are both extremely important but there is another, less obvious, capability that is waiting to bloom.

Modern internet video protocols such as MOVE Networks, Flash Server and Microsoft Smooth HD all have the property that a viewer can switch to any point in a video stream in a very few seconds. The painful "video buffering" step is rapidly disappearing from internet television content. This ability to shift video without time delays opens up numerous possibilities for interactive video experiences. The viewer is no longer confined to a single linear experience as defined by broadcast directors. The viewer can skip to any point and watch what they like, however they like it.

EuroITV'11, June 29–July 1, 2011, Lisbon, Portugal.

This interactive power to skip anywhere in a video and watch anything at the user's discretion does not necessarily lead to a quality user experience. Experiences such as stories or games unfold over time. Random jumping can be a jarring and distasteful experience. This new interactive video freedom must have structure to support the viewer and address the entertainment orientation of the family room rather than the task orientation of the desktop. This opens a large number of research questions around the kinds of interactive structure that will be effective. This paper is one such structure.

1.1 Interactive football

In this paper we focus on interactive video experiences for American football. We chose football because 1) sports are an important and highly lucrative segment of television, 2) many sports, by their nature are highly structured and 3) the authors like American football. Previously published work on interactive football allows viewers to skip from play to play, conduct their own instant replays including switching camera angles under viewer control. In-home tests of this technology were highly successful [5].

In this paper we address the problem of watching a game in less time than the original. Though we specifically addressed American football the need is much broader. People frequently have much less time to spend on an otherwise fixed length segment of video entertainment. An American football game can take three or more hours to play. The amount of viewer time available depends greatly on the individual, their personal interest in the sport and the level of interest they find in a particular game.

The most common form of game summarization is the highlight reel shown on the nightly news or other sports broadcasts. Though informative, such clips are only 2-5 minutes in length to cover a 3 hour event. There is no sense of the story of the game and no opportunity to spend more than 5 minutes but less than 180 minutes. This paper shows how summaries of various lengths can be produced such that viewer can get a sense of the game rather than just a snippet. We also show how viewers can control how much they watch and when.

To watch less of something there must be a decision on what should be removed. We, and others, have based this decision on a degree of interest function (DOI). When this function is applied to information about a particular play it returns a value as to how interesting this play might be. The design of this DOI function is critical to any video summarization technique. We offer some new approaches to such a function and measure their value against viewer opinion. When a segment is removed from a video there is a loss of what happened in between. This can sometimes be quite disorienting for the viewer. We will show one technique for filling the missing gaps and describe viewer response to our gap-filling approach. Lastly we will show how viewers can step out of the summary into full interactive viewing of selected portions of the game. When their interest is piqued they can interactively choose to spend more time on a particular segment before returning to the summary. It is our contention that variable, adaptive summaries with interactive viewer control are superior to traditional fixed summaries required by traditional broadcast schedules.

We will first provide an overview of prior efforts in this area, followed by a discussion of the Time Warp Football (TWF) [5] project. TWF is the foundation for this work and an understanding of its features is important for this implementation. We will then look at how we produce a game summarization. In particular we discuss our *degree of interest* (DOI) function and how we developed it. Our particular contribution is in the use of other viewer's interactive behavior to inform this function. Next we will present our animation solution for preserving the continuity of the game even when plays have been skipped to shorten the time. Lastly we discuss how interactivity can augment the value of a game summarization. We will also discuss our evaluations with potential viewers.

2. Prior Work

Much of the work in this paper is related to video summarization[10]. One of the most popular approaches is to automatically process the video to determine plays that are of the most interest. This includes looking for replays, identifying shots that involve a soccer goal [9], identifying scores in baseball [3] or similarly in football [2]. The problem we have with these techniques is that they are relatively imprecise and miss many interesting plays. They are also inaccurate as to the start and the end times of a play.

A related approach is to use the audio level to indicate interesting plays[8]. This is a rough form of audience response. The louder the noise, the more interested the crowd must be. In American football this can be deceptive because frequently the home crowd makes noise to confuse the opposing team. This also does not help to place the start and end of a play because crowd noise tends to lag the actual events. Another approach is to add closed captioning information to the video and audio streams [1] to derive features that can be used for summarization.

Some approaches use human intervention to identify where the plays are located. Babaguchi, et al. [2] use computer vision to parse out the game clock and then match with internet game statistics to identify where events happened in the game. They also classify each shot into various kinds. Takahasi, et al. [7] uses annotations embedded into MPEG-7 with information about each baseball play. One very simple technique is to simply discard every part of video where no play is taking place [4]. In American football this eliminates a lot of time.

In virtually all of these techniques there is some *degree of interest* (DOI) function that is computed from some set of *features* that are derived from the video source or other inputs. This degree of interest function is used to predict which plays or camera shots would be of the most interest to viewers.

A separate, but relevant, work is Newstream [6], where the goal is to detect when two pieces of media that cover the same topic provide additional information or just repetition. This is not the same as summarization but their approach is enlightening. They use the interactive behavior of previous viewers to inform their similarity function. When the same person views two pieces of media known to be on the same topic and views the first in depth and skips out of the second immediately, it is inferred that there is too much overlap between the two. In our DOI function, we similarly use prior interactive behavior.

The differences in the various papers generally are in the kinds of features about the video or the game that they extract and the way that they create the DOI function from those features. In this work we mirror this approach. However, we add three new things. 1) We use people rather than algorithms to identify the plays and to provide game stat features for our DOI function. 2) We use the interactive behavior of actual viewers as features to the DOI function. 3) We use animation to stitch together the plays we present into a full story of the game rather than merely a highlight reel.

3. Time Warp Football

Our summarization system is built upon the framework of Time Warp Football[5] (TWF). TWF provides viewers with the interactive ability to skip plays, instantly replay a down (a single play in American football), play a down from any camera angle and get live game stats. The full TWF interactive experience is not important here, but the underlying architecture is. As the game progresses all of the camera feeds (usually 5-6) are digitized to a server. In addition, there is an annotation tool that allows staff in the broadcast truck to create game information that parallels these video streams. It is this annotation information that enables the TWF interactivity and also drives our summarization system.

The key pieces of annotation information are the time for play start and play end. These time markers give us the basic structure for our summarization. Though we are using football as our driving example in this paper, a similar approach works for basketball, volleyball, baseball, curling and a variety of other sports. The summarization approach we describe here will work in similar fashion to all of these other sports.

In addition to the timing markers for start and end of play, each play is annotated with what happened during that play. These annotations include yards gained/lost, position on the field, what down is it, any scoring that occurred, fumbles, interceptions, sacks and other events of interest to fans. These information items about a play provide some of the features to our DOI function.

When a fan watches a game using TWF they can engage in a variety of interactive behaviors including skipping ahead, replay and change camera angle. Each time the fan selects one of these options a message goes from the player software to the server to select the desired video fragment to be played next. We can then log these server requests and tie them to the particular play involved. This provides us with information about the interactive behavior of each fan for each play. This interactive behavior information forms the second set of inputs to our summarization technique. Our hypothesis is that larger numbers of "replay" and "new camera angle" interactions are indicative of plays that will be of high interest to future viewers and can be used to select plays to form part of a summary.

4. Game Summarization

The simplest form of summarization is to remove all segments that are not part of an actual play[4]. In American football that reduces a game from 3 hours to 30-40 minutes. To get shorter summaries we need to remove some plays and keep only the ones of highest interest.

From TWF we now have the following information:

- Time segments for each play in the game
- Data about what happened during each play
- Data about interactive behavior of fans who have watched the game previous to our generating a summary.

Based on the play data and the interactive data we generate a Degree of Interest (DOI) function that can be applied to each play. Our summarization algorithm is:

- Select a time *T* for the summary
- Sort all of the plays *P* by DOI(*P*) in descending order.
- Select the first *N* plays such that the sum of their times does not exceed *T*.

The *N* plays form our summary. The algorithm itself is not novel. Our key interest is in the computation of the DOI function.

4.1 DOI Standard

To select an appropriate DOI function we needed a standard against which to compare our candidate functions. One of the challenges of this work is that it is difficult to get college football networks to cooperate in providing all of the camera angles. For competitive reasons, coaching staffs are reluctant to provide their video resources. We were able to collect a main feed and two coach's feeds (high sideline and end zone) that were taken from a game that was played about five years previous. From this we produced our TWF demo for user testing. This game also forms the basis for our summarization experiments.

Figure 1 – Rating football plays

Taking this one game we prepared a web-based program (see figure 1) that displayed the main broadcast feed for the game. At the end of each play, viewers are asked to rate how interesting each play is on a scale of 1-10 with 10 being the most interesting. We had 12 college students rate all 120 plays in our example game. We used these ratings as a standard against which to compare our DOI functions. We do not claim that the data we collected is a definitive representation of fan interest. When TWF is deployed commercially, video software can be augmented to allow thousands of fans to rate hundreds of games from which a definitive DOI profile can be computed. Our interest is to explore how this process would work and understand better how appropriate DOI functions can be generated.

One possible summarization approach is to allow some fans access to a game for free or without advertising if they will rate every play. The actual fan ratings can then be used as a DOI function to drive generating summaries for later viewers. The problem with this approach is that raters can become apathetic. We do use fan ratings of games, but only as a mechanism for training our DOI function. In the long run it is the nature of a particular game and the interactive behavior of other people watching the game that drives our summarization algorithm.

One of the first questions to answer is whether there is any uniformity in the ratings generated by fans. The data we received varied in the way raters behaved. Some would rate most things high, some low. This variation in their basic rating style made the direct use of their data difficult. Because our goal is to select plays to include in a game summary, we counted how many times each play occurs in some rater's top 20 plays (see figure 2). This measure gets directly at what we are trying to do and resolves differences in how individuals rate plays.

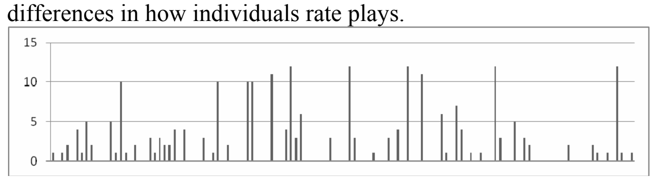

Figure 2 – Number of times each play is in top 20

There are 5 plays that appear in every rater's top 20 and 11 plays that appear in the top 20 of 10 or more raters. There are 64 of the 120 plays that appear in nobody's top 20. This means that 37% of the plays get mixed ratings. It should also be pointed out that all of the raters were from the same university as one of the teams. Looking at the data shows a bias of interest towards plays that are positive for one's own team. Full deployment of such ratings must take team bais into account when sampling. A key finding from this data is that only about half of the plays show consistency among the raters. This means that careful optimization of the DOI function to human ratings is not going to be possible. It will be a rough approximation at best. Once we get past really exiting plays the interest becomes rather mixed. This "confusion in the middle" is not novel to this problem.

4.2 Degree of Interest (DOI) function

Our approach to developing a DOI function is to collect a set of features about each play and then compute a linear least squares approximation to the average rating for that play. Because there is large range variation among the features we used the Z-score of each feature

((feature - mean) / stdDev).

The first set of features to consider are the statistics that can be derived directly from the game play. The statistics that we used were:

- Did the defensive team score? (interception or fumble run back for a touchdown)
- Touchdown
- Field goal
- Penalty yards
- Turnover (fumble or interception)
- Punt or kickoff
- First down
- Third Down Conversion (successful)
- Fourth down attempt
- Yards gained

For readers acquainted with American Football these are clearly the interesting events that can occur in a game. The list of such events will vary with other sports but sports in general have such statistics that are collected, discussed, memorized and generally talked about. Others have used such statistics for DOI functions. However, TWF gives us finer grained statistics than previous work. We have this advantage because TWF has humans directly encode all of this information rather than trying to automatically estimate it from the video after the fact.

The features of most interested in are the interactive behaviors exhibited when watching the game. Our assumption is that many fans will watch the game in close to live time (TWF interactions can cause fans to fall behind live time as they review various plays). When they make interactive requests during their viewing those requests can be logged at the web site that is serving the interactive video.

Collegiate football in the U.S. is very competitive and involves large amounts of money. As yet we have not been able to deploy a live game using our interactive technology. However, we do have a game for which we have the main broadcast feed as well as two additional camera angles. From this we built a prototype of TWF that we were able to deploy into individual homes. We recruited 11 homes from the local area offering pizza to anyone who wanted to come and watch a game. We then connected our technology to their own television in their home and their own internet connection. We pointed a video camera at the fans, gave them the controls and left the room until the game was over. In the process we logged all of their interactive behavior.

From the logs of this experiment we collected the following statistics for every play.

- Number of times an alternative camera angle was requested (indicator of interest because replay is being requested)
- Whether "Next play" was requested (indicator of disinterest because the viewer is purposely skipping commentary and replay)
- Whether "Previous play" was requested (indicator of interest because this implies a replay)
- Whether a replay of a particular play was requested (indicator of interest)

In a real commercial deployment these numbers would be readily available from the TWF server. We were interested in these features because they can be collected without explicit user action. They are a passive byproduct of normal user behavior. Unlike paid or volunteer raters, that can get bored or pursue their own agendas, this interactive data reflects actual sports fan behavior and interest.

4.3 Developing a DOI function

Once we have our features collected and a DOI standard from the human raters, we can develop a DOI function that can drive our summarization. For these experiments we chose a simple linear function for our DOI. We used the user ratings from the data we collected as the desired result and the z-scores of the features as the variables. Using linear regression we produced coefficients for the features which yielded a DOI function. We then compared the resulting DOI function's prediction of interest against the actual user ratings. Obviously more complex function learning could be applied but the amount of data that we had available did not justify such a fine analysis and was actually orthogonal to the questions we were interested in.

There were three feature sets for the DOI function that we wanted to test.

1. Game event features alone
2. Viewer interactivity features alone
3. All of the features together.

4.3.1 Game Events DOI

Figure 3 shows a comparison of the user ratings with the DOI function generated only from game event features (scoring, turnovers, yards gained. . .).

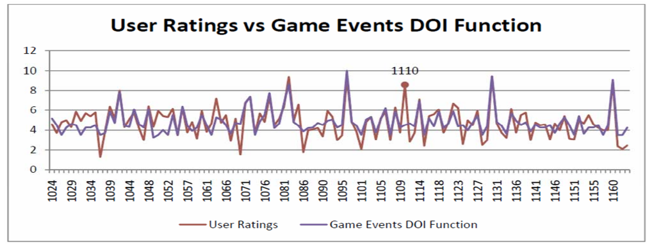

Figure 3 – Comparison of Game Events DOI to ratings

For the most part these two measures track each other quite closely. In comparing the user ratings with our game events DOI we achieved a 0.979 correlation with a mean squared error of 1.075. In particular these two measures agree strongly only the most interesting plays. However, there are exceptions. Figure 3 shows that the users rated play 1110 very high while the game events DOI rated it only average. Looking at the game video we see that the punter drops the ball, picks it up, runs to the right and at the last minute punts the ball to the opponent's two yard line (for those who are not fans of American football, that is a very exciting and improbable play.) The problem with this play is that the statistical result is rather normal and thus not very exciting. No points were scored, there was no turnover and modest yards were gained. In this particular case the game statistics are not telling the whole story.

4.3.2 Viewer Interactivity DOI

Figure 4 shows the same sequence of plays only this time the graph shows the number of times viewers replayed the down using a different camera angle. Note that our play number 1110 shows up very strongly. Other strong plays also are highlighted. It is also important to point out in this data that plays in the first part of the game (far left) show lots of viewer activity as well. We attribute this to new viewers exploring the technology for the first time rather than actual interest in the plays.

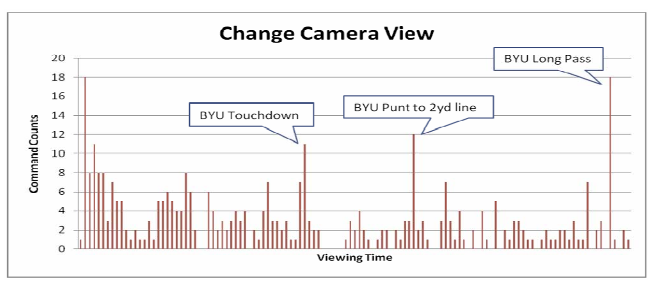

Figure 4 – Frequency of "replay with different camera"

Encouraged by the data in figure 4 we computed a new DOI function using the interactive behavior of viewers. For these features we used the data collected during the in-home trials of TWF.

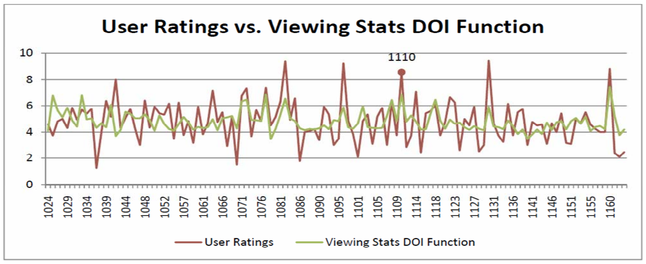

Figure 5 – Comparison of User Activity DOI to ratings

This DOI produces a correlation of 0.960 which is slightly worse than the game events DOI and a MSE of 1.998, which is also slightly worse. This is not a bad DOI function but does not yield the kinds of improvements that we had hoped. Notice that at the far left of figure 5 we see that the user ratings and the DOI are not well correlated. This confirms our belief that viewers were exploring the technology more than the game. We explored the hypothesis the fans were biased in favor of good plays for BYU. Separating out the BYU offense from Notre Dame offense we did not detect a difference in the ratings.

In looking at the game video we did find that some plays with high game significance did not produce increased interactive behavior. For example, if a running back scored from the 2 yard line without difficulty it was an important play but viewers did not feel a need to view it again. Viewer interactivity seems to be more a measure of unexpectedness and controversy than actual game interest.

4.3.3 Combined Features DOI

Our next experiment was to combine all of the features from both game events and viewer interactivity. The results are shown in figure 6. This produced a correlation of 0.981 and an MSE of 0.971 which is better than either of the feature groups separately. Note, however, that play 1110 is still not as prominent in the DOI as we would like.

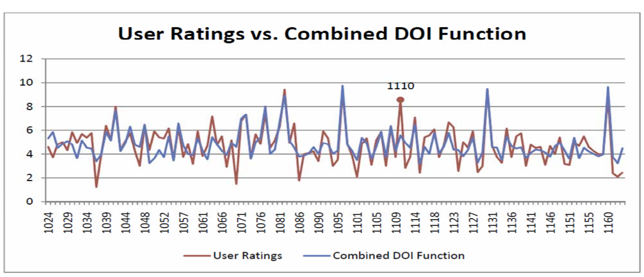

Figure 6 – Comparison of Combined DOI to ratings

4.3.4 Game summary validation

The correlation and MSE measures are good for indicating a DOI function that corresponds to user interest but are not actually the measure we want. The summarization algorithm we used is to compute a DOI for all plays and then pick the top N plays that fit within the requested time. The key measure is whether a particular DOI captures a top N that is similar to the top N from the viewer ratings.

To test our DOI against this criterion we picked 20 our number of plays (N) to show in the summary. We then averaged the user ratings for each play and picked the top 20. If our DOI was based on manual user ratings, this was the summary that would have been generated. Figure 7 shows a graph of how many times each of these plays appeared in some rater's top 20.

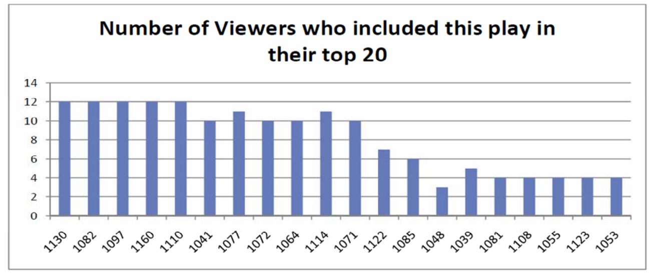

Figure 7 – Plays sorted by number of times rated in top 20

As is shown in figure 7 there are 5 plays that are in every rater's top 20 and 11 plays that are in the top 20 of 10 out of 12 raters. The agreement on top 20 plays drops off sharply from there. Obviously a good DOI will put those 11 plays in its top 20. Play 1055 which is a field goal by Notre Dame early in the game is not in these top 11 plays. Not all scoring events are particularly exciting.

We then used our combined DOI function (game stats plus interactive behavior) to select a top 20 plays. The DOI function's top 20 matches 16 of the manual ratings top 20. The combined DOI function captures 10 of the 11 plays where raters agree, as well as most of the remaining high rated plays. Our combined DOI disagrees with the raters on exactly the plays where they disagree with each other. Our combined DOI function clearly performs as well as human viewers in picking the top 20 plays that most viewers would be interested in.

We were concerned about play 1064 that the human raters put in the top eleven but our DOI function left out. It turns out that Notre Dame was tackled for a loss which is not a game statistic that we gathered. The DOI based on game statistics ranked this play as number 36 while the DOI based on viewer interactive behavior ranked it as number 25, just barely missing the top 20 by 4%.

4.3.5 Conclusions about DOI functions

The evaluations described above are not the last word on selecting a correct DOI function for football summaries. There are too few manual reviewers and too few games to get a good globally applicable function. However, the data does show that logging the interactive behavior of viewers and using those features as part of a DOI function will substantially improve the results. The viewer behavior alone, however, is not enough. The analysis clearly shows that many plays that are important to a game are not always replayed or interactively examined by viewers.

5. Continuity of the Story

A game is a story that unfolds over time. Like a story, dropping pieces out can create a discontinuity. In our preliminary TWF work we noticed this problem when multiple people are watching a game and only one is holding the controller. For those not holding the controller a jump to the next play or a request for replay with a change of camera angle can be very disorienting. In TWF we had to inject a banner that announced the yard-line, ball possession and down whenever someone invoked one of these controls.

We have similar problems in a game summary as the video jumps from play to play, leaving out those plays that are skipped. To resolve this problem we created an animated play summary technique that shows the ball moving on the field and visually shows forward and backward movement as well as punts, interceptions, fumbles and penalties. Each play's animation takes 0.5 seconds which is much faster than watching those plays in their entirety. This rate is adjustable but 0.5 seconds per play seemed to create a timely yet understandable representation for each play.

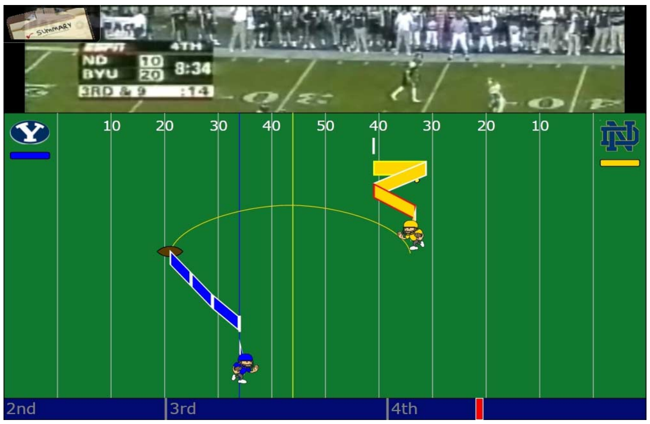

Figure 8 – Animation of skipped plays

The horizontal axis of this animation is the playing field. The vertical axis is time. Each successive play moves down in time. If the animation is long then the history simply scrolls off the top.

Figure 8 shows one team losing yardage, then gaining, then losing again and then punting. The opposing team then gains yardage on three successive plays. At this point a selected play from the game summary is shown.

This kind of animation is possible because of the game statistics that are included in the annotation mechanisms of Time Warp Football. We know how many yards were gained/lost. We know when penalties, punts or change of possession occurs. The transition animation is simply a user friendly presentation of that information.

The key question is whether these kinds of transitions that cover the removed plays will actually make the whole summary more interesting to viewers. We took 20 subjects and showed them the same game summary in two different ways. Half of them saw the summary without the transitions and then again with the transitions. The other half saw it with transitions and then without. At the end of each viewing we asked to agree or disagree with several statements about their experience.

When asked about understanding the story of the game they seemed to understand slightly more when the transitions are present (figure 9). When asked about what was missing from the game summary, they felt less was missing when the transitions were present (figure 10). When asked about ease of following the summary, again the transitions fared better (figure 11). These preferences for the transitions in head-to-head comparisons with the simple highlight reel show a consistent but not strong preference for the additional information.

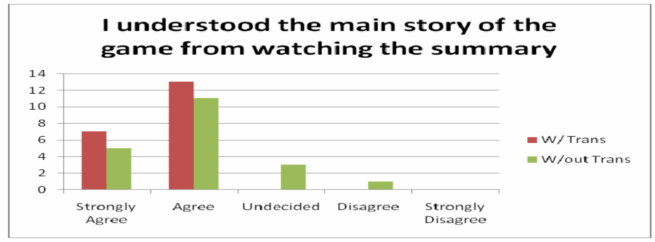

Figure 9

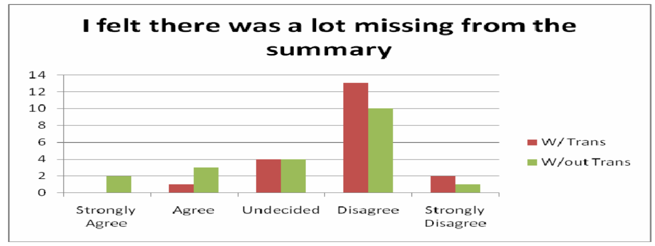

Figure 10

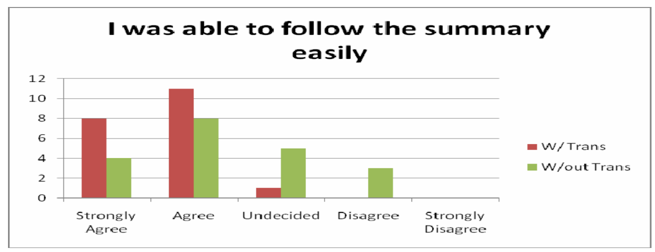

Figure 11

These results appear to favor inclusion of the transitions. However, when asked asked directly about usefulness the response was almost identical both with and without the transition animations (figure 12).

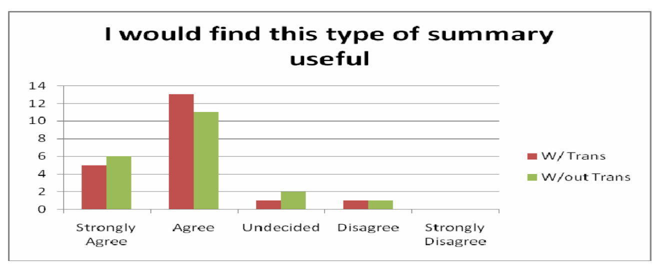

Figure 12

A series of more direct statements on viewer judgement of the transitions yields a stronger result. Viewers clearly agreed that the transitions helped (figure 13), were not distracting (figure 14) but also indicated that they had problems understanding what the transition animation actually meant (figure 15). This interpretation is supported both by their open ended comments "the animation was confusing but I soon figured it out" and also in review of the video tapes of their usage. There is early confusion about what the animation is doing. After a while they figure out what the animation means and then they understand it just fine.

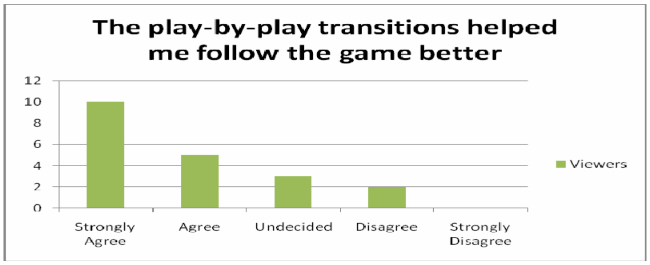

Figure 13

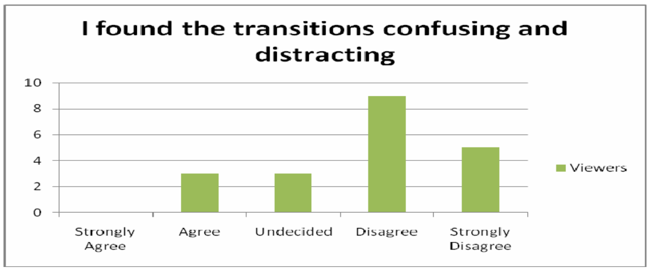

Figure 14

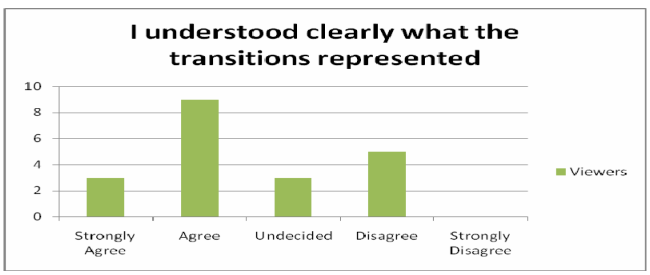

Figure 15

6. Interaction within the summary

One of the features that we wanted to test was the ability for viewers to drop out of a summary and into our more complete interactive football system [5]. We felt that this would be a distinct advantage because the summary might lead a viewer to a play of interest about which they would want to know more. The interactive football interface gives viewers access to replays, slow motion and multiple camera angles. We thought this might be of particular value from the transition animations where a play of interest was omitted but the viewer might decide to look at it anyway. This feature also has commercial value because it entices viewers to watch more than they originally intended and thus offer more advertising opportunities. As part of the interactive controls the viewer can step in and out of the summary by pressing the summary button as shown in figure 16. In addition if a viewer was intrigued by any play they could simply press "previous play" or any of the alternate camera angle controls to drop out of the summary into interactive mode.

Figure 16 – Reentering game summary

As part of our study of how viewers used the summary we told them about the interactive controls that would take them out of the summary and into full interactive viewing. We gave them 15 minutes to familiarize themselves with the various controls but we did not explain to them how to use the controls. We very much want the interaction to be intuitive without instruction.

As before, we logged their use of all of the user interface controls. The viewers dropped out of summary mode into full interaction an average of 6 times with a standard deviation of 4. When dropped out they stayed in full interaction mode an average of 41 seconds with a standard deviation of 22.5 seconds. Because plays average 10.7 seconds it is clear that specific plays are being watched several times which is a key advantage of the TWF system.

When asked, the viewers claimed to like the possibility of interactive exploration of particularly plays (figure 17). However, a significant minority viewers were confused by the feature (figure 18).

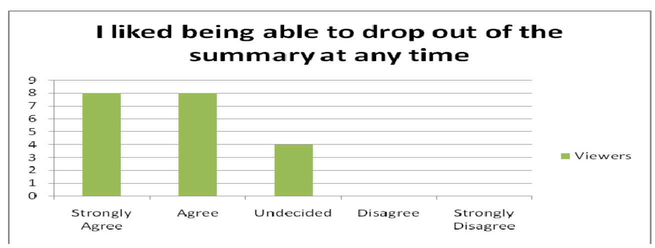

Figure 17

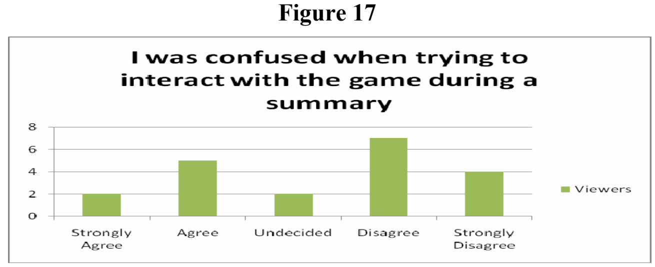

Figure 18

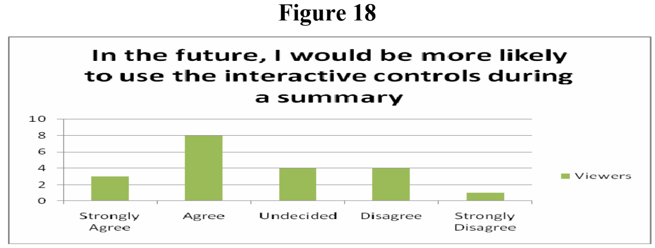

Figure 19

A majority of the viewers did claim that they would use the controls in the future (figure 19). However, support for the feature is mixed. Some viewers would prefer to just let the summary run and passively watch. In reviewing the comments and the video tapes we see viewers trying out the controls and getting stuck in the interactive player and not knowing how to get back to the summary. This is particularly true when entering interactive mode by previous play or camera angle selection. Having not used the summary button they were confused about how to get back to the summary.

We think the main problem is our attempt to repurpose the interactive football implementation without smoothly integrating it with the summary. The support for the feature in general (figure 19) leads us to believe that a better UI design for the feature would lead to something of real value to the viewers.

7. Conclusions

We had three key questions in this work:

1. Can logging the interactive behavior of previous viewers provide feature that will produce better DOI functions for summarization?
2. Will animated transitions to cover skipped plays help the understanding of the game as a whole?
3. Will the ability to drop into interactive mode when desired improve the summary viewer experience?

In comparing DOI functions against manual user ratings we see a definite improvement when interactive behavior is included. However, interactive behavior alone is not sufficient as a DOI function because not all interesting plays induce interactive actions. The inclusion of game statistics along with viewer behavior produces a very good summarization which speaks well to the value of the Time Warp Football as a basis for game summarization. There are three important issues that this work does not address. First there is no evidence of how this might transition to other sports experiences. Secondly, this work does not address the relative value of other features such as crowd noise or video motion to indicate interest. Lastly, the data is too limited to produce a really valid DOI function. We have only pointed the way to valuable directions to pursue.

This use of interactive behavior to indicate viewer interest can also be expanded beyond sports. These live logs of behavior might guide news teams to developing stories that have lots of viewer interest and need more material.

The viewer responses do indicate that the animated transitions add value to the summary experience. Any additional information that is collected as part of the TWF annotation process could improve this animation. Further work will be required to determine if the novelty of the animation wears off over time and becomes annoying or if experience with the animation actually enriches understanding of the game. A longer and broader study will be required to answer these questions.

The interactive behavior logs as well as the after experiment questionnaires do indicate that most viewers like the ability to step out of the summary and interactively explore particular plays in detail. This speaks well to the integration of our summary system with the larger interactive football system. This also speaks well to the commercial value of this feature. Viewers who initially claim to not have a lot of time can be enticed to interactively view more of the game than they originally intended. It would be an interesting further experiment to include the crowd noise track with the play summaries to induce viewers to explore something the crowd likes but that they might be missing. Clearly this use of interaction in the context of a summary needs a longer and broader study of its effectiveness.

8. References

[1] Babaguchi, N., Kawai, Y., et al. (2002). "Event based indexing of broadcasted sports video by intermodal collaboration." IEEE Transactions on Multimedia **4**(1): 68-75.

[2] Babaguchi, N., Kawai, Y., et al. (2004). "Personalized Abstraction of Broadcasted American Football Video by Highlight Selection." IEEE Transactions on Multimedia **6**(4): 575-586.

[3] Chu, W.-T. and Wu, J.-L. (2008). "Explicit semantic events detection and development of realistic applications for broadcasting baseball videos." Multimedia Tools and Applications **38**(1): 27-50.

[4] Logan, J. D., Durgin, S. A. and Burd, D. (2005). Methods and Apparatus for Recording and Replaying Sports Broadcasts. U. S. Government. United States, Gotuit Video, Inc. US Patent 2005/0005308 A1.

[5] Lynn, S. G., Olsen, D. R., and Partridge, B. G. (2009). Time Warp Football. European Conference on Interactive Television, ACM.

[6] Martin, R., and Holtzman, H., (2010) “Newstream: A Multi-Device, Cross-Medium, and Socially Aware Approach to News Content.” European Conference on Internet Television, EUROITV ’10, pp 83-90.

[7] Takahashi, Y., Nitta, N., et al. (2005). Automatic Video Summarization of Sports Videos Using Metadata. Advances in Multimedia Information Processing - PCM 2004**:** 272-280.

[8] Tjondronegoro, D., Chen, Y.-P. P., et al. (2003). Sports video summarization using highlights and play-breaks. Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval. Berkeley, California, ACM.

[9] Tong, X., Liu, Q., et al. (2005). Highlight ranking for sports video browsing. Proceedings of the 13th annual ACM international conference on Multimedia. Hilton, Singapore, ACM.

[10] Truong, B. T. and Venkatesh, S. (2007). "Video abstraction: A systematic review and classification." ACM Trans. Multimedia Comput. Commun. Appl. **3**(1): 3.