#EpicPlay: Crowd-sourcing Sports Video Highlights

Tang, Anthony; Boring, Sebastian

Publication date:
2012

Document version
Peer reviewed version

#EpicPlay: Crowd-sourcing Sports Video Highlights

Anthony Tang **Sebastian Boring**
Department of Computer Science, University of Calgary
2500 University Drive NW, Calgary, AB T2N 1N4, Canada
{tonyt, sebastian.boring}@ucalgary.ca

ABSTRACT
During a live sports event, many sports fans use social media as a part of their viewing experience, reporting on their thoughts on the event as it unfolds. In this work, we use this information stream to semantically annotate live broadcast sports games, using these annotations to select video highlights from the game. We demonstrate that this approach can be used to select highlights specific for fans of each team, and that these clips reflect the emotions of a fan during a game. Further, we describe how these clips differ from those seen on nightly sportscasts.

Author Keywords
Crowd-sourcing; Video summarization; Sports; Twitter; Microblogging; Broadcast sports; Video annotation.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors

INTRODUCTION
Live broadcast sports events (e.g. a football game) attract large audiences. However, truly interesting footage is only a small fraction of the entire broadcast. Sportscasts summarize these interesting events into highlight reels, though these are limited by: (1) time (a highlight reel is on average two minutes long), and (2) the inherent bias of the video editor. Most automated approaches to highlight reel generation rely heavily on the video itself, using computer vision to detect motion, field orientation, and score [1,5]. Others use statistics and play-by-play information published to the web by on-site statisticians to help annotate the video [9].

Watching broadcast sports can be a social experience—many viewers watch with other fans. Today, this social experience can transcend physical boundaries: for example, during the event, viewers might microblog on Twitter as part of their viewing practice. In this work, we make use of this Twitter data, using it to address the problem of sports video annotation and summarization. This approach provides two meaningful advantages combining play-by-play information with the video feed: (1) it recognizes events that are interesting or important to fans in the moment, and (2) it avoids the bias of a single video editor, relying instead on fans to inform us of when an important or interesting event has occurred.

Prior work uses peaks in the rate of incoming data to detect when interesting events occur in a live broadcast [2,3,4,7]. We build on this approach by also separating the incoming tweet stream by home/away teams to provide more nuanced interpretation of the stream. Using this approach, we can annotate the video with time spans that indicate "interesting snippets" of the video; this data can help us clip the video stream, producing different highlight reels to suit fans of each team. Thus, we use Twitter as an additional semantic layer to understand the video stream.

In our work, we evaluate this approach by using a dataset of 15 American football games, drawing on live national broadcasts from both the Canadian Football League (CFL) and National Football League (NFL). In comparing the highlight reels generated from our tool with those on broadcast news, we found that our approach is able to generate team-specific highlights that capture "in-the-moment" excitement well (i.e., in response to fan excitement). In contrast, clips from broadcast news better capture which plays were ultimately meaningful to the outcome of the game.

This work makes two contributions: (1) we demonstrate that crowdsourcing techniques can be applied to the domain of sports video summarization, and (2) with a deep probe into the sports problem domain, we uncover and highlight some interesting subtleties in both the data and requirements for crowd-sourced sports summarization.

RELATED WORK
To set the scene, we review existing approaches to sports video summarization, which widely focus on extracting information from the video data itself. We further describe a growing body of work that examines crowd-sourcing as an approach to assist in annotating and summarizing events.

Sports Video Summarization. Many researchers have focused on summarizing sports videos for two reasons: first, sports have underlying structure that can be exploited (compared to other types of video streams); second, sports broadcasts are limited to a small range of shot types, which eases the recognition burden. Early work in this space focused on computer vision approaches (e.g. [1,5,9]), which segment shot types (e.g. close, far, audience, etc.) before extracting semantic information from the video itself (e.g.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

scores from score boards). More recent work attempted to combine this data with secondary feeds to help annotation, segmentation, and event boundary detection (e.g. [9]). These data streams can be of textual nature (e.g., play-by-play data from statisticians published live online), or related to our approach, analyze the video's audio (e.g., higher volumes indicate crowd excitement, perhaps signaling important events) [5].

Social Event Summarization. There is a growing body of research (largely focused on Twitter[1]) demonstrating the utility of using crowd-sourced data to summarize and interpret events. Several tools exist to both gather and interpret the data from Twitter (e.g. [4]). This crowd-sourced data allows for identifying sports [3] or political events [7] that are temporarily interesting/relevant to the population. In addition, the actual text of tweets can be used to construct textual summaries of the event itself [2].

Applying crowd-sourcing to video summarization is a far less explored area. Wu et al. showed that it is possible to use explicit tasks to crowd-source video summarization [8]. Olsen et al. show that user interaction data from an interactive TV application can be mined to detect events and plays that are likely of interest to a sports fan [6]. Hannon et al. make use Twitter data in *PASSEV* [3], combining summaries of tweet frequency and user-specified search for terms in tweets to generate a highlight reel. Unfortunately, preference data for this system was mixed—potentially because these were *general* fans (of the sport) rather than fans of the teams involved in the games themselves. Our work builds on this work, applying algorithms originally designed for social event summarization [4] to the sports video domain. Here, we focus on understanding the types of video clips generated by the algorithm rather than evaluating preference data.

CHARACTERIZING AMERICAN FOOTBALL: CFL & NFL

American football is a team-based contact sport where the goal is to move the football into the other team's end zone. Points are scored for touchdowns (bringing the football into the end zone: 6 points) or field goals (kicking the football through a set of uprights: 3 points). Other plays can result in 1 and 2 points depending on the situation. In North America, two professional leagues play American football (with slight variations on some rules). In their respective countries, both leagues broadcast games live nationally, and command huge fanbases. To serve this fanbase, both leagues have web-based play-by-play trackers that report summaries of each play as they occur (i.e., "web-text").

For our interest in crowd-sourced sports video summarization, American football thus has a number of useful characteristics: (1) it is a highly structured game, with well-defined beginning and ends of plays (rather than being a game with continuous action, as in soccer or hockey); (2) gameplay actually occurs for only a small fraction of the duration of the entire event (e.g. teams play through four quarters of 15 minutes in a game, but broadcasts regularly last three hours), and (3) the fanbase for football teams actively use Twitter during live broadcasts of games.

Scoring plays are generally considered significant unless the game score is already decidedly lopsided. In games where the score is close, plays toward the end of the game often carry far more significance than those earlier in the game. Most importantly, however, is that web-text provides summaries of plays as they occur, but they do not always capture the significance of (or excitement generated by) each play to the game. Olsen et al. note that obvious plays (even if they are scoring plays) may not be considered as interesting as successful execution of trick plays (even if they are not scoring plays) [6]. Thus, web-text is inadequate in picking out these plays; we believe that Twitter, on the other hand, can help find such plays.

#EPICPLAY: CROWD-SOURCED SPORT HIGHLIGHTS

#EpicPlay visualizes four different data sources (Figure 1): video data from the live broadcast, time-stamped web-text from league websites, and two time-stamped Twitter streams (collected through Twitter's Search API). The two Twitter data streams are collected by determining the hashtags used most frequently by fans of each of the teams playing (e.g. BC Lions fans typically tweet with the hashtag `#BCLions`). This is not a clean separation—some fans use hashtags of both teams for two reasons: (1) greater visibility (in our sample, 10%-25% tweets had both hashtags); (2) fans of one team may complain about a "dirty play" involving the other team. However, the vast majority of tweets were about one team.

With this data, #EpicPlay computes two different sets of highlights using the TwitInfo algorithm [4] on the team-specific streams. We calculate local peaks, as we believe that a peak in activity reflects interest by fans of that team (or that something interesting happened to that team at that time). From the beginning of a peak we move one minute backward in the video stream, and select a one-minute video clip. This is based on the heuristic that tweets will likely appear after the end of the play (rather than before or during the play), and that plays typically last less than a minute, thus ensuring that the clip contains the play in question.

CHARACTERIZING CROWD-SOURCED VIDEO CLIPS

What constitutes a good set of highlights from a sports event is necessarily subjective, complicating the evaluation of the highlight selection process. As a quantitative benchmark, we compare #EpicPlay's selected highlights with those of sportscast news highlights (see Table 1); however, our primary interest is in understanding the *types* of highlights selected by these different processes.

Data Set & Highlight Selection

Data Set. For the period from Sept. 2 until Sept. 19, 2011, we collected the aforementioned data for 15 regular-season

[1] http://twitter.com/

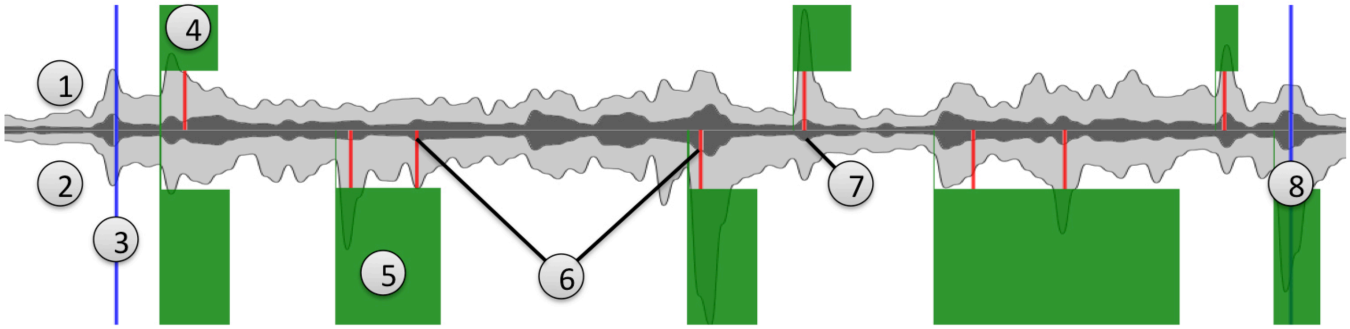

Figure 1. A visualization from #EpicPlay that shows the twitter activity during an NFL football game, along with our annotations: (1,2) tweets with the home/away team's hashtag; (3,8) the start and end of the game; (4,5) are peaks detected by the TwitInfo algorithm for home/away team's tweets; (6) the red ticks represent scoring events in the game, where ticks above the centre are home team scoring events, and ticks below are away team scores and (7) the darkened area represent tweets with both teams' hashtags.

American football games (7 NFL games, 9 CFL games). For comparison purposes, we also collected highlight reels from two nightly sports newscasts (two national sports networks) for each game. In our sample of ~680k tweets across these games, an average of about 10% of tweets occurred within a minute following a scoring play.

Highlight Selection. We used TwitInfo's peak-finding algorithm as we assumed that "interesting events" would be highlighted by sudden bursts of activity. For NFL games, we used 1-minute intervals for binning the tweets; for CFL games, where the volume of tweets significantly lower, we binned tweets into 2-minute intervals (cf. [4]). The TwitInfo algorithm's sensitivity is controlled by two parameters: (1) the difference in number of tweets in the current bin as compared to the last one classified by the mean deviation τ (#EpicPlay uses $\tau = 4$), and (2) the correction of the mean deviation over time α (we use $\alpha = 0.125$). As illustrated in Figure 1, different peaks found for home/away teams reflect the differential excitement by the fans but are not necessarily aligned with scoring plays—as expected.

Comparing #EpicPlay to Sportscasts

Table 1 shows the average calculated *precision* (percentage of retrieved items being relevant) and *recall* (percentage of relevant items being retrieved) for the games, with the highlights from sportscast considered as relevant items. From a purely information retrieval perspective, crowd-sourcing video highlights on a per-team basis is not comparable to nightly sportscast highlights.

Why does #EpicPlay exhibit such poor precision and recall scores? To address this, we performed a more qualitative analysis on the highlight reels themselves. When we examined the differences between sportscast highlights and our crowd-sourced highlights, we noted several differences: (1) sportscasters can only show a limited number of highlights (in our sample, each game was given typically either 45 seconds or 2 minutes). (2) Highlights from sportscasts are retrospective narratives, dictated by the game's *outcome*. Thus, they typically reflected a build-up toward the final score (although some plays were spectacular enough to be included regardless of the outcome).

In contrast, the crowd-sourced highlights typically reflected the fan emotion *in the moment*. In practice, this resulted in several different types of plays being selected by the crowd-source method, summarized in the following:

- Only scoring plays that are interesting were selected. That is, if a score is "expected" or considered "easy", the play was not considered interesting; in contrast, a score that changes the lead was often selected.
- Plays that required a high degree of skill, or resulted from extraordinary effort are often selected by the crowd. These were sometimes considered more interesting than a possible subsequent scoring play.
- Since the algorithm identifies spikes in activity, this also resulted in "lowlight" selection. Plays that resulted in fans becoming frustrated with their team were selected.
- Plays that resulted in a controversial penalty called by the referee were often selected, with fans voicing their displeasure through Twitter.
- Unusual things that occurred during the broadcast were often captured. In one game, for example, a player dribbled the football in soccer-style around the field.

Crowd-sourcing using Twitter does have some shortcomings. We noticed, for instance, that many fans would tweet at the beginning and ends of each quarter (and at the beginning and end of the game). These did not typically result in an interesting highlight. Similarly, we observed in at least

		Avg # tweets / min		*Median # clips selected*		*TwitInfo Home Clips*		*TwitInfo Away Clips*	
	Games	**Home**	**Away**	**Home**	**Away**	**Precision**	**Recall**	**Precision**	**Recall**
NFL	7	128	359	4	5	0.53	0.41	0.36	0.33
CFL	8	2.7	2.0	5	6	0.48	0.38	0.56	0.42

Table 1. Quantitative summary of our dataset.

one game some "re-tweet spam" being propagated (where fans that "re-tweeted" a tweet would be entered in a drawing for a prize). Fortunately, both these shortcomings are systematic in nature, suggesting that they can be programmatically eliminated if necessary.

Variations & Caveats

Variations on Highlight Selection. We considered two alternative methods for selecting highlights: (1) using the TwitInfo algorithm on aggregated tweets for both teams, and (2) selecting highlights based on scoring plays. The first variation selected highlights that were interesting to fans of either one or of both teams (as in Figure 1, many scoring plays result in Twitter activity with hashtags of only one team). The resulting set of highlights was perhaps more balanced, but sports fans in our lab suggested that they would be more interested in highlights selected by like-minded fans. Using solely scoring events for highlight selection missed other "interesting" events that occurred in the game (e.g., spectacular plays, controversial penalties, or other interesting events that did not result in a score).

Caveats. TwitInfo requires a meaningful rate of tweets to detect peaks reliably. For some CFL games, we found that fans did not tweet with sufficient frequency (at times, dipping as low as 1 or 2 per minute) for meaningful peaks to be found. In general, NFL games had considerably higher rates of tweets (usually more than one tweet per second), resulting in more meaningful peak detection.

DISCUSSION & CONCLUSIONS

Evaluating #EpicPlay. To better understand the quality of #EpicPlay's selected highlights, we considered two approaches: (1) some form of user evaluation, and (2) comparing highlights against a suitable benchmark. With due consideration to others' work in this space [3], we felt that highlight reel selection was necessarily extremely subjective. Casual fans (i.e. fans of a sport, but not of a particular team) may not feel affinity to highlights selected by #EpicPlay vs. those of a sportscast. Rather, we expected the strongest reaction from team-specific fans to highlights generated about those specific teams. These fans would be able to relate better to the emotional content or meaning of particular plays within the context of the specific games. However, given our dataset, recruiting such a select group of participants would have proved extremely challenging.

Exploiting "Normal Behaviour" for Crowd-sourcing. This work builds on a considerable body of research that explores crowd-sourcing. What sets this work apart is that we are actually mining data that results as a consequence of existing behaviors. That is, we do not ask viewers to tweet about their experiences; instead, they already do this as a part of their everyday sports-viewing activities. We simply use this data to perform a task that is challenging for computers to do on their own: *understand* the event. In this work, we have explored how tweet volume maps to some notion of interest/excitement in the game (albeit with caveats mentioned earlier); yet, this only uses a small piece of the rich information embedded in Twitter data. For instance, tweets might actually be used as a source for annotating highlights, or even provide commentary for a sports event (with a slight delay for the video feed). The utility of Twitter data is that it adds a semantic layer of interpretation atop the broadcast and web-text feeds. If used in concert with existing techniques, this layer would add considerable strength to work in this area.

We have also demonstrated what should be straightforward for most sports fans: fans of different teams actually see a sports event differently (i.e. through the lens of being a fan of a team). As a consequence, the plays fans are interested in differ depending on their team allegiance. This is also information that is reflected in the Twitter data.

In this work, we show that we can extract meaningful highlights for sports events using Twitter. The resulting set of clips does not closely match those of nightly sportscasts; instead they are more tied to the drama and emotion of the game as experienced by fans. As such, they are a reflection of the fan's excitement and interest during the game. We believe this information can be used as another data source for information retrieval: rich information being generated as a consequence of fans' everyday viewing behavior.

ACKNOWLEDGMENTS

This research is partially funded by the iCORE/NSERC/SMART Chair in Interactive Technologies, Alberta Innovates Technology Futures, NSERC, and SMART Technologies Inc.

REFERENCES

1. Babaguchi, N., Kawai, Y., Ogura, T., and Kitahashi, T. Personalized abstraction of broadcasted American football video by highlight selection. In *IEEE Multimedia 6* (4), (2004), 575-586.
2. Chakrabarti, D., and Punera, K. Event summarization using tweets. In *Proc ICWSM 2011*, (2011).
3. Hannon, J., McCarthy, K., Lynch, J., and Smyth, B. Personalized and automatic social summarization of events in video. In *Proc IUI 2011*, (2011), 335-338.
4. Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., and Miller, R. C. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proc CHI 2011*, (2011), 227-236.
5. Nepal, S., Srinivasan, U., and Reynold, G. Automatic detection of 'Goal' segments in basketball videos. In *Proc Multimedia '01*, (2001), 261-269.
6. Olsen, D. R., and Moon, B. Video summarization based on user interaction. In *Proc EuroITV 2011*, (2011), 115-122.
7. Shamma, D. A., Kennedy, L., and Churchill, E. F. Conversational shadows: describing live media events using short messages. In *Proc ICWSM 2010*, (2010).
8. Wu, S.-Y., Thawonmas, R., and Chen, K.-T. Video summarization via crowdsourcing. In *Ext. Abstracts of CHI 2011*, (2011), 1531-1536.
9. Xu, C., Wang, J., Wan, K., Li, Y., and Duan, L. Live sports event detection based on broadcast video and web-casting text. In *Proc Multimedia '06*, (2006), 221-230.