Title	Personalized and automatic social summarization of events in video
Author(s)	Hannon, John; McCarthy, Kevin; Lynch, James; Smyth, Barry
Publication Date	2011-02-13
Publication information	IUI '11 Proceedings of the 16th international conference on Intelligent user interfaces
Publisher	ACM
Link to publisher's version	http://dx.doi.org/10.1145/1943403.1943459
Item record/more information	http://hdl.handle.net/10197/2952
Publisher's statement	This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Proceedings of the 16th international conference on Intelligent user interfaces available at http://doi.acm.org/10.1145/1943403.1943459
DOI	http://dx.doi.org/10.1145/1943403.1943459

Personalized and Automatic Social Summarization of Events in Video

John Hannon[1], **Kevin McCarthy**[1], **James Lynch**[2], **Barry Smyth**[1]

[1]CLARITY,
Centre for Sensor Web Technologies,
School of Computer Science & Informatics,
University College Dublin, Ireland.
{firstname.lastname}@ucd.ie

[2]Amdocs Research Centre,
NovaUCD,
Belfield,
Dublin 4, Ireland.
james.lynch@amdocs.com

ABSTRACT

Social services like Twitter are increasingly used to provide a conversational backdrop to real-world events in real-time. Sporting events are a good example of this and this year, millions of users tweeted their comments as they watched the World Cup matches from around the world. In this paper, we look at using these time-stamped opinions as the basis for generating video highlights for these soccer matches. We introduce the PASSEV system and describe and evaluate two basic summarization approaches.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: Graphical user interfaces

General Terms

Algorithms, Experimentation

INTRODUCTION AND RELATED WORK

Soon to be gone are the days when viewers consumed video content through their TV sets on a schedule dictated by the major networks. Already, we are moving to a *pull* model of media consumption where users are increasingly in control of how and when they consume their favourite media. Indeed users are no longer passive consumers of media. The rise of the social and real-time web means that many of us play a more contributory role via user-generated content. For any given event there will be a whole host of associated content much of it created by users themselves.

A concrete example of this is the role of Twitter as a conversational backdrop to a variety of real-world events. Increasingly we are finding users who effectively participate in events, as they unfold, via their Twitter messages. Twitter and Twitter-like content has already been used successfully in a variety of applications such as the ranking of news stories [6], the profiling of user preferences [3], and even the recommendations of products [2]. In this paper, we consider how this type of conversational, user-generated content might be used to add value to more traditional event media, such as video. In particular, we use the recent soccer World Cup as a case-study for video summarization by using Twitter data as the basis for the summarization process. Whether it is referred to as football, soccer, fútbol, futebol or fußball, one full match of *The Beautiful Game* lasts for at least 90 minutes. When half time and pre and post match analysis are factored in, the interested spectator must give over nearly two hours of their lives to watch a full game. Even the most loyal fan, with their busy schedules, cannot afford to watch every game from start to finish. Therefore most soccer enthusiasts enjoy watching games in the condensed form of a highlight video. These summarization videos typically encompass the most crucial incidents or events of a match. Most important of these defining moments is the goals scored, however, other events such as red and yellow cards, penalties, goalmouth incidents and contentious decisions are all gathered together to form the summarization highlights. Highlight videos are typically edited together by analysts working for the television companies. This requires going through the whole game, deciding which events are most important and then cutting and pasting the selected segments into a coherent summarization video. The length of these videos can be anything from 5 to 20 minutes, but they are always compiled by an editor and are broadcasted in a "one-size-fits-all" format even though individual fans may have more niche interests.

The task of video summarization has conventionally been a job for professional software/hardware experts using high-tech frame and object detection algorithms. Much research has been carried out into improving those algorithms by using various different techniques from histogram intersections to hidden Markov models [1, 5]. These techniques require a huge post-processing effort and do not facilitate anything close to real-time summarization. Some research has also looked at less conventional ways of detecting these key moments within a video by monitoring a user's interaction with that video and building up a profile over time of how users generally watch a specific video [4, 8]. Again, this has the notable disadvantage of a cold start problem when new videos are added. In this work, we want to examine the automatic production of these highlight videos and we intend on us-

IUI'11, February 13–16, 2011, Palo Alto, California, USA.

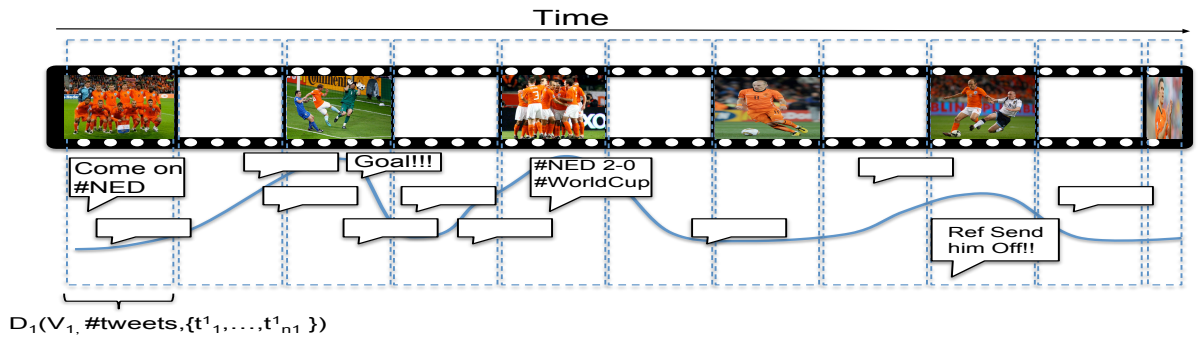

Figure 1. Temporal alignment of video sequence with Twitter stream.

ing content from the real-time web to accomplish this. We believe that the frequency and content of time-stamped Twitter messages, which can be reliably associated with a given event, have the potential to serve as a powerful form of summarization data. This paper serves as an initial case-study of one such summarization service.

PASSEV: PERSONALIZED HIGHLIGHT GENERATOR

PASSEV is the prototype system which we have developed to test different approaches to using real-time web data as the basis for event detection and summarization of video streams. As discussed in the next section, we will apply PASSEV to the summarization of World Cup video content, based on the tweets of viewers and spectators.

To begin with, PASSEV takes as input a video sequence and a collection of time-stamped tweets that are known to refer to the event captured by the video. Very briefly, these tweets are extracted from Twitter's streaming API[1] based on temporal alignment with World Cup matches and based on the presence of certain key terms and hashtags (e.g. #worldcup, team names, player names, etc). For the purpose of this short paper we will focus on two key components of the PASSEV system. First, we will describe how PASSEV segments and indexes the video stream using Twitter data. Second, we will describe the summarization process, based on this indexed content, exploring two different summarization techniques.

INDEXING VIDEOS AND TWEETS

The basic idea behind PASSEV is to slice a video sequence into a set of segments and then index these segments by the content of the tweets that are temporally aligned with the segments. Currently PASSEV is configured to use a fixed-duration time-slice so that, for example, a World Cup match video can be sliced into a set of 90 or so (depending on match length) video segments. In this way, each video V is sliced into a set of k clips V_1 - V_k. In turn each clip V_i is associated with a document D_i that is composed of this clip plus the set of terms that make up its corresponding tweets, plus a count of the number of tweets it is associated with ($\#tweets$); see Equations 1 and 2.

$$V = \{D_1, ..., D_k\} \quad (1)$$

[1]Twitter API http://apiwiki.twitter.com

$$D_i = (V_i, \#tweets, \{t_i^1, ..., t_n^1\}) \quad (2)$$

In this way each video clip can now be indexed by its tweet content. For this we use Lucene[2]. Essentially, each of these documents are indexed using Lucene's standard TF-IDF [7] term-weighting function. In this way, a term such as 'goal' which appears frequently in the tweets for a given document, but which is infrequent across the match as a whole, will receive a higher weight for that document, helping to distinguish this document with respect to this term.

The advantage of this document-centric approach to indexing, and the use of Lucene, is that we now have a flexible index that can be easily queried based on the interests of the user, which we can use during summarization.

GENERATING A SUMMARY

In this paper we explore two approaches to producing a summary from a given document index. Both, involve selecting a set of documents from the index, based on the duration of the required summary, and concatenating them temporally to produce the final highlight video.

Frequency-Based Summaries

Perhaps the most straightforward approach to producing a summary of a given duration D is to select documents on the basis of the frequency or *volume* of tweets during the time-period of the particular video clip (see Figure 2). As such, generating a frequency-based summary amounts to the selection of the top m documents that have the highest $\#tweet$ count, where m corresponds to the number of documents required for the given summary duration; thus if t is the fixed duration of each time-slice then $m = D/t$. The clips from these m documents are then concatenated in the correct temporal order to produce the final highlight reel.

The frequency of the incoming tweets about an event can tend to have a long tail, as tweets from users can be delayed, this delay can be attributed to anything from slow typing to the Twitter client application being used. To try to circumvent this we convert contiguous sequences of two or more documents into a single document by selecting that clip from the sequence that has the highest $\#tweet$ count to attempt to group the event into the build up, the event, the aftermath.

[2]Lucene available from http://lucene.apache.org

This ensures that the summary is not dominated by extended sequences of match-play which might have attracted a lot of Twitter traffic.

The simplicity of this frequency-based approach is countered by the one-size-fits-all nature of the end result. Clearly not every user is likely to be interested in the same type of summary. For example, an England fan might want to see more of England in that controversial England-v-Germany game, and certainly might want to avoid the Lampard goal that was disallowed; more of that later.

Content-Based Summaries

To solve the issue of one-size-fits-all summaries we provide a content-based summarization technique that is triggered by a user query; each PASSEV user can provide a set of terms that reflects the events they would like to see (e.g. goal, red card, penalty). This query q then provides a basis for selecting documents from the index, using Lucene's built-in retrieval functionality. This produces a set of *candidate clips* each of which is associated with tweet content that matches the user's query. As in the previous technique we prune contiguous sequences of documents by eliminating all but the one with the highest tweet count. And the final summary is composed of the concatenation of the clips from these documents, in the correct temporal order.

Frequency-Based Algorithm

I: Lucene Index, Duration: Highlight Length

```
1.   Define FrequencyRecommendations(Duration, I)
2.    Begin
3.    ordered-clips ← sort(I, tweets)
4.    selected-clips ← Cluster-clips(ordered-clips)
5.    return Top-k(ordered-clips, Duration)
6.    End
7.   End

8.   Define Cluster-clips(candidate-clips)
9.    Begin
10.   clusters ← contiguous clips
11.    For Each c in clusters
12.     clip ← identify clip with highest tweet count
13.     selected ← selected + clip
14.    End
15.    return selected
16.   End
17.  End

18.  Define Top-k (selected-clips, Duration)
19.   Begin
20.   While( index < Duration )
        selected-clips' ← selected-clips[index]
       index++;
21.   End
22.   return selected-clips'
23.   End
24.  End
```

Content-Based Algorithm

I: Lucene Index, Q: query, Duration: Highlight Length

```
1.   Define ContentRecommendations(Duration, Q, I)
2.    Begin
3.    candidate-clips ← SearchVideoIndex(Q,I)
4.    selected-clips ← Cluster-clips(candidate-clips)
5.    return Top-k(selected-clips, Duration)
6.    End
7.   End

8.   Define SearchVideoIndex(Q, I)
9.   return LuceneSearch(Q,I)
10.  End
```

Figure 2. Frequency and content-based summarization algorithms.

USER EVALUATION

In the previous section, we described two initial approaches to using Twitter data to summarize video content for sporting events: a one-size-fits-all frequency-based approach and a more personalized, content-based approach which allows the user to indicate their preferences for certain specific event types. In this section we describe the results of an initial user trial with a view to understanding the preference of users for each type of summary produced.

Setup

The trial is based on video footage recorded during this years World Cup 2010 and the Twitter messages generated for the matches. In total we collected 50m tweets from 6m individual users and aligned this content with the recorded match videos, based on the tweet time-stamps.

For the user study we recruited 13 male students. In terms of their soccer expertise/interest, 6 classified themselves as *novice*, while 7 classified themselves as *intermediate-expert*. To gauge their apathy towards football, we asked the participants if they had watched any of the games from the World Cup, 11 of the 13 users had watched at least some of the games from the World Cup. For the purpose of the study we limited the video content to 3 World Cup matches, and created frequency-based summaries of 3, 5, and 8 minutes in duration; for this we configured PASSEV to slice the video into 1-minute document segments. Each user was allowed to submit a set of query terms as the basis for the personalized summary. The basic test for the user was to compare a frequency-based summary of a given length to their personalized summary of the same length, and express a preference. In total we collected 24 preference pairs.

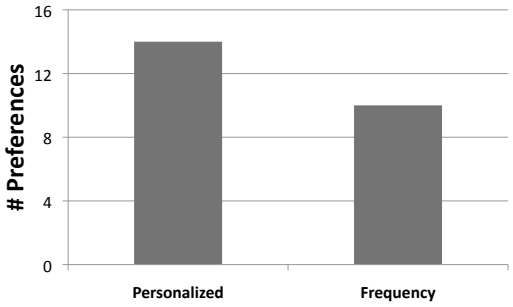

Figure 3. User preferences by summarization technique.

Overall, the test users expressed broad satisfaction with the summaries produced by both techniques which speaks to the potential for this form of summarization. Figure 3 compares the preferences expressed for the two summarization techniques and it is clear that there is a modest preference for the personalized summaries (14/24 preferences) compared to the one-size-fits-all frequency-based summaries (10/24). Figure 4 shows a break down of the preferences versus the selected highlight lengths. Overall users who selected 3 and 8 minute highlight lengths tended to prefer the summaries produced by the personalized approach, whereas, for the 5 minute games frequency-based summaries were preferred. The reason for this is interesting and reveals an important limitation of the personalized approach as it stands.

When we examined the user preferences in more detail we found that several users selected a 5 minute highlight video

of the England-Germany World Cup game. This was an especially controversial game in the World Cup as England had a crucial, early (and perfectly legitimate) goal mistakenly disallowed. This incident was included in the frequency-based summary since it correlated with a significant increase in tweet volume. However, it so happens that the users who requested a summary of this game did not tend to include tags that would help to identify this particular event. Clearly this is a limitation of the personalized approach since expecting users to choose the correct query terms is not guaranteed to be reliable. In the future we will explore the development of a conversational interface to provide direct feedback to users with a view to helping to recommend specific terms that may reflect important game events, if those terms are not present in the user's initial query. For example, the system might present a list of the most popular tweet terms for the match as hints to the user during query formation.

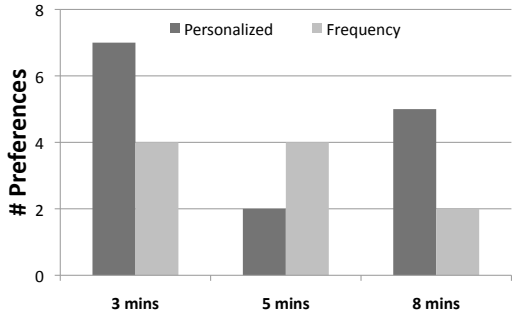

Figure 4. Summarization preference by highlight duration.

CONCLUSION

This paper represents work-in-progress. We propose that user-generated content on the real-time web can provide important insights into real-world events as they unfold. To test this we have examined video summarization by aligning time-stamped tweets with relevant video content, in this case World Cup matches. We have described how to use this real-time data as the basis for summarizing full-length matches, which translates sequences of video content into documents that can be indexed by their associated tweets. This facilitates an information retrieval approach to video summarization, and we describe two specific techniques. A simple frequency-based approach, which focuses on sequences within a game that attract higher than normal tweet volumes, provides a one-size-fits-all summarization approach. In contrast, a content-based approach, which selects video sequences whose associated tweets match the user's query, provides a more personalized summary. In addition by allowing users to control the duration of summarizations, users can now define how much time they want to spend watching highlights.

The results of a preliminary user trial are promising. Users were satisfied with the quality of the resulting summaries and expressed a preference for the personalized summaries. That being said, the trial also helped to highlight some limitations with the approach as it stands. For example, obviously enough, poor queries by the user limited the quality of the personalized summary because important events could be missed. Moreover, in retrospect, indexing the video into 1-minute segments was probably not the right level of granularity for the optimal summary quality and future work will certainly explore shorter segment lengths and the possible use of different segment lengths for different parts of the match. Also building upon our user evaluation, we plan to carry out a comparison of our system against commercial grade video analysis software to compare our event detection against that of a system that uses audio, motion and other metrics to detect events in video and this will become part of a large scale user study.

Acknowledgment

This work is supported by Science Foundation Ireland under grant 07/CE/I1147 and by Amdocs Inc.

REFERENCES

1. Mohamed Y. Eldib, Bassam S. Abou Zaid, Hossam M. Zawbaa, Mohamed El-Zahar, and Motaz El-Saban. Soccer video summarization using enhanced logo detection. In *Proceedings of the 16th IEEE international conference on Image processing*, ICIP'09, pages 4289–4292, Piscataway, NJ, USA, 2009. IEEE Press.
2. Sandra Garcia Esparza, Michael P. O'Mahony, and Barry Smyth. On the real-time web as a source of recommendation knowledge. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 305–308, New York, NY, USA, 2010. ACM.
3. John Hannon, Mike Bennett, and Barry Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 199–206, New York, NY, USA, 2010. ACM.
4. J Lanagan and A F Smeaton. SportsAnno: What Do You Think? In *RIAO'2007: Proceedings of the 8th conference on Information Retrieval and its Applications*, Pittsburgh, Pennsylvania, USA, 2007.
5. Baoxin Li and M. Ibrahim Sezan. Event detection and summarization in sports video. In *Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'01)*, CBAIVL '01, pages 132–, Washington, DC, USA, 2001. IEEE Computer Society.
6. Owen Phelan, Kevin McCarthy, and Barry Smyth. Using twitter to recommend real-time topical news. In *Proceedings of the third ACM conference on Recommender systems*, RecSys '09, pages 385–388, New York, NY, USA, 2009. ACM.
7. Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, NY, USA, 1986.
8. Bin Yu, Wei-Ying Ma, Klara Nahrstedt, and Hong-Jiang Zhang. Video summarization based on user log enhanced link analysis. In *Proceedings of the eleventh ACM international conference on Multimedia*, MULTIMEDIA '03, pages 382–391, New York, NY, USA, 2003. ACM.