VideoSkip: Event Detection in Social Web Videos with an Implicit User Heuristic

Chrysoula Gkonela, Konstantinos Chorianopoulos*

Department of Informatics, Ionian University, Corfu, Greece
+30-26610-87707
choko@ionio.gr
www.ionio.gr/~choko

Abstract. In this paper, we present a user-based event detection method for social web videos. Previous research in event detection has focused on content-based techniques, such as pattern recognition algorithms that attempt to understand the contents of a video. There are few user-centric approaches that have considered either search keywords, or external data such as comments, tags, and annotations. Moreover, some of the user-centric approaches imposed an extra effort to the users in order to capture required information. In this research, we are describing a method for the analysis of implicit users' interactions with a web video player, such as pause, play, and thirty-seconds skip or rewind. The results of our experiments indicated that even the simple user heuristic of local maxima might effectively detect the same video-events, as indicated manually. Notably, the proposed technique was more accurate in the detection of events that have a short duration, because those events motivated increased user interaction in video hot-spots. The findings of this research provide evidence that we might be able to infer semantics about a piece of unstructured data just from the way people actually use it.

Keywords: Video, Event detection, Semantics, Web, User-Based, Experiment.

Introduction

During the last years, the intense growth of the internet has given impetus to sharing of video material between users all over the world. Besides user generated content, movies, TV series, lectures, sports, news etc., are becoming available on-demand. One of the most successful platforms, that millions of users use on a daily basis, in order to browse all these videos, is YouTube (Cha et al. 2007). In addition to watching videos on the main YouTube video player, users can upload videos or perform other important tasks, such as commenting videos, replying with other videos, or just expressing their preference ("like"/"dislike"). Previous research has focused on the importance of the micro-blogs and other metadata information, in order to improve the semantic understanding of video content. Although there are various methods that collect and manipulate text-based

information (Aisopos et al 2011), the majority of them is usually burdensome for the users. Moreover, the percentage of users leaving a comment is too small according to the real number of viewers of a video (Mitra et al. 2011). In this research, we propose a collective intelligence method (Zhang et al. 2011) that leverages implicit user interactions for extracting useful information about a video.

We have developed a system and an experimental procedure that leverages seamless user interactions for extracting useful information about a video. In particular, we let the viewer browse the video and we store all the interactions with the player (e.g. play, pause) for future use. According to Shamma et al (2007) the more the emotive energy of a scene, the more the specific interval of the video containing that scene is used. Thus, the records of all interactions between the user and the video-player might be employed to infer the most important scenes of a video and to automatically generate thumbnails, or even implement a summarization feature.

Content- And User-Based Event Detection

Previous research has examined methods that provide summaries for videos. Girgensohn et al (2001) developed three visual interfaces and conducted user studies in order to identify whether a set of key-frames can describe a video and provide access to its relevant parts. They used piles based on visual similarity to organize the key-frames, calculated the importance of each key-frame according to its rarity and duration and examined the usability of three interfaces. They found that clustering is an effective way to manipulate large number of key-frames. Video Snapshot (Ma and Zhang 2005) is a pictorial video summary method that deals with algorithms that reveals the structure of the video elements, understand their content and detect their visual features. This technique reserves the primary elements of video and provides flexibility of content selection and flexibility of visualization in different shapes and sizes. Nevertheless, users cannot interact with those pictures.

Moreover, previous research has explored key-frames for video navigation. SmartSkip (Drucker et al. 2002) from Microsoft Research is an interface that, also, generates thumbnails. It uses the histogram of images in almost every 10 seconds of the video and looking at rapid overall changes in the color and

brightness. Similarly, Li et al (2000) developed an interface that generates shot boundaries using a detection algorithm that identifies transitions between shots. This is not the only feature on their development but it includes aids like a table of contents, a time compression (increase playback speed) and pause removal (detects and removes pauses and silence segments of the video) function.
Overall, in content-based information retrieval, researchers have employed automated techniques, which are depended to the video content. Thus, they use images' colors, shapes, textures, sounds, motions, events, objects or any other information that can be derived from the video signal itself. Some of them combine videos' metadata (Takahashi 2005) with picture, or sound editing and others provide affective annotation (Chen et al. 2008), or navigation aids (Kim et al. 2006). Even though they generate important content for the users they do not take into account their browsing behavior. Moreover, low-level features (e.g. colour, camera transitions) often fail to capture the high-level semantics (e.g. events, actors, objects) of the video content itself, yet such semantics are often what guide users, particularly, non-specialist users, when navigating (Crockford and Agius 2006). They suggest that a significant event could be a temporal relationship, a spatial relationship or an object that has a semantic aspect (Crockford and Agius 2006). Content-based systems have offered many practical advances to interactive video navigation (Doulamis and Doulamis 2004), but they have the drawback of being monolithic with respect to user preferences.
Besides the research interest in event detection, there have been also commercial systems that provide similar functionality. For example, Google Video provided thumbnails (Figure 1) to facilitate user's navigation in each video. A collection of still images is more preferred in many applications, because it is easy to display and deliver. Nevertheless, most of the techniques that extract thumbnails at regular time intervals or from each shot are not effective, because there might be too many shots in a video. In the case of Google Video there are so many thumbnails that a separate scroll bar was employed for navigating through them.

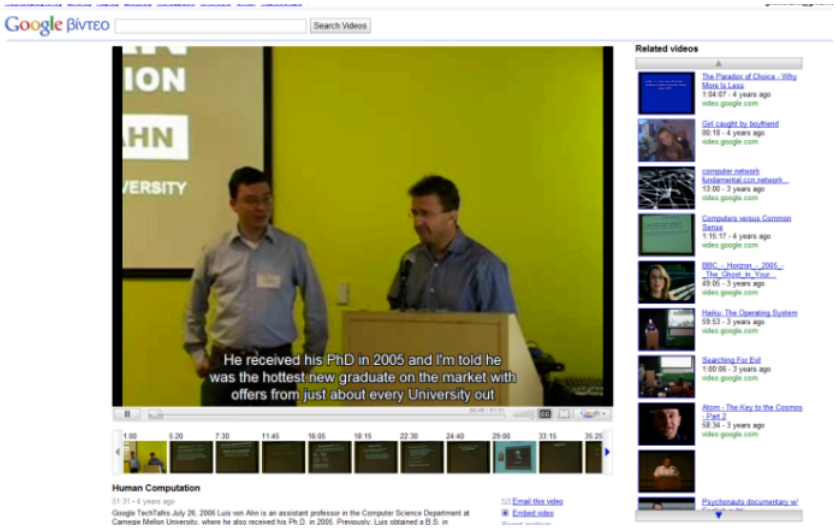

Figure 1: Google Video provides navigation between key-frames

Although many corporations and academic institutions are making lecture videos and seminars available online, there are few research efforts to understand and leverage user browsing behavior. Moreover, the techniques that use information such as the amount of motion, or the newness of visual context, are not applicable to informational presentations. He et al (1999) focused on informational talks accompanied by slides and extracted a summary removing portions of the content. They analyzed audio channel, speaker's channel, and users' actions watching the talk. Although they recorded user activity (e.g., play, pause, random seek) of user behavior they did not take advantage of them.

In addition to the video and audio content, researchers have considered the analysis implicit user actions. Shen et al. (2005) demonstrated a personalized search system that learned from previous users' queries. In the context of video, Yu et al [11] believe that viewers unintentionally leave footprints during their video-browsing process. They proposed ShotRank, a concept that measures the interestingness or importance of each video shot combining video content analysis and user log mining. They tried to generate meaningful video summaries that would assist future viewers, who in turn would render more meaningful browsing trails. Moreover, their work is influenced by the idea of the page-rank algorithm by Google. Therefore, they make the assumption that there is a shortest path in each video. They provide a list of video's shots and user has to choose among them which one he wants to watch. They try to generate video skims and hierarchical story trees. They use only play/stop/pause for each scene. They have developed a system on the assumption that users interact with video content by

means of browsing content using custom video browsing applications that record the user's interactions with the video content; whereas most users still view video content in more traditional ways that require no explicit interaction, such as viewing content on their home television sets.

Users' interactions are the basic element of our research. There is a need for detailed tracking of video browsing behavior. Syeda-Mahmood and Ponceleon (2001) developed a media player-based learning system called the Media Miner. They tracked video browsing behavior, modeled user's states transition with Hidden Markov Model and generated fast video previews to satisfy the "interestingness" constraint of them. MediaMiner featured the common play, pause and random seek into the video via a slider bar, fast/slow forward and fast/slow backward as well. Researchers tried to relate user activity to the user's browsing status (e.g., user is bored, or interested). Video understanding is part of a broader research effort (Doulamis and Gonzàlez 2010). Most notably, Ntalianis et al. (2010) proposed a method that combines implicit feedback information and visual concept models for semantically annotating images. Their method is applicable to video and could be further extended as soon as we have a better understanding of users' implicit interactions with the video player, which is the main topic of this research work.

System Design And Experimental Methodology

Event Detection Systems

Several applications have been developed by the researchers, in order to evaluate novel event detection methods. Macromedia Director, a multimedia application platform, was used to develop SmartSkip (Drucket et al. 2002). The system used re-encoded videos in QuickTime format. Similarly, Emoplayer (Chen et al. 2008) was running locally on a laptop and participants used a pointing device to interact with it. The system was developed with VC++ and DirectShow and the annotated video clips were stored in XML files. In the case of Li et al (2000) Microsoft Windows Media Player had been modified to develop the enhanced browser with its special features, because its default playback features are not sufficient for video navigation. Crockford and Agius (2006) designed a system as a wrapper around an ActiveX control of Windows Media Player. A video-recorder used to

collect video, at first, and then it was encoded in MPEG-1. The majority of previous systems runs locally, needs special modification on software, and at the same time on video clips. Another important procedural parameter is that subjects must be at a specific place where the experiment was conducted. Besides stand-alone applications, there are web-based systems. Hotstream (Hjelsvold et al. 2001) employed Java 2 Enterprise Edition (J2EE) to develop a multi-tier web based architecture system (web-tier, middle-tier, backend database-tier, streaming platform). HotStream delivers the personalized video content from streaming video servers. Shamma et al (2007) created different web-based platforms where the user can watch, browse, select and annotate video material.

VideoSkip System Design

The VideoSkip player provides the main functionality of a typical VCR device (Crockford and Agius 2006). We decided to use buttons which remind the main playing /browsing controls of VCR devices because they are familiar to users. ReplayTV system and TiVo provide the ability to replay segments, or to jump forward in different speeds. In this way, we have modified the classic forward and backward buttons to Goforward and Gobackward. The first one goes backward 30 seconds and its main purpose is to replay the last viewed seconds of the video, while the Goforward button jumps forward 30 seconds and its main purpose is to skip insignificant video segments. The thirty-second step is an average time-step used in previous research and commercial work due to the fact that it is the average duration of commercials. Next to the player's button, the current time of the video is shown followed by the total time of the video in seconds. We chose not to use a seek thumb because we wanted to avoid random guesses as it would have been is difficult to analyze users' interactions. Li et al (2000) observed that when seek thumb is used heavily, users have to make many attempts to find the desirable section of the video and thus causing significant delays.

Figure 2: VideoSkip player with simple navigation buttons

In this research, we have selected to develop an experimental system with the Google App Engine and the YouTube API. VideoSkip (Leftheriotis et al. 2010) is a web video player we developed to gather interactions of the users while they watch a video. Based on these interactions, representative thumbnails of the video are generated. Users of VideoSkip should have a Google account in order to sign in and watch the uploaded videos. Thus, users' interactions are recorded and stored in Google's database alongside with their Gmail addresses. Google App Engine's database, the Datastore, is used to store users' interactions. Each time a user signs in the web video player application, a new record is created. Whenever a button is pressed, an abbreviation of the button's name and the time it occurred are added to the Text variable. The time is stored within a second's accuracy.

User Heuristic for Event Detection

We consider that every video is associated with an array of k cells, where k is the number of the duration of the video in seconds. The user activity heuristic consists of three distinct stages. In the first stage, every cell is initialized to the number of users who have watched the video. We use this initial value to avoid extremely negative values, to increase viewing value of the whole video and to provide a balance for random interactions.

In the second stage, we increase by two the value for each cell that has been played by the user. Moreover, every interaction means something for the event detection scheme. Each time user presses the Gobackward button, the cells' values matching the last thirty seconds of the video, are incremented by two

again. On the other hand, each time user presses the Goforward button, the cells' values matching the next thirty seconds of the video, are decreased by two. We experimented with different values for interactions before we ended-up to these on table 1. For example we used for play, goforward, gobackward and pause «+1» or for play/pause «+1» and for goforwrd/gobackward «+2». We choose this combination that made our results distinguishable, without being complex.

Table 1: User activity heuristic provides a simple mapping between user action and value

User action	Play	GoForward (30 sec)	GoBackward (30 sec)	Pause
Heuristic	+2	-2	+2	+2

In the third stage, we take into account the highest values of the array and at the same time the number of values (interactions) that are gathered in a specific cell area (i.e., the surface size). Moreover, we defined a distance threshold of thirty seconds between the selected thumbnails in order to avoid having consecutive cells as a result. These specific scenes can be used as proposed thumbnails and improve users' browsing experience. Each proposed thumbnail begins at the first second of the selected area.

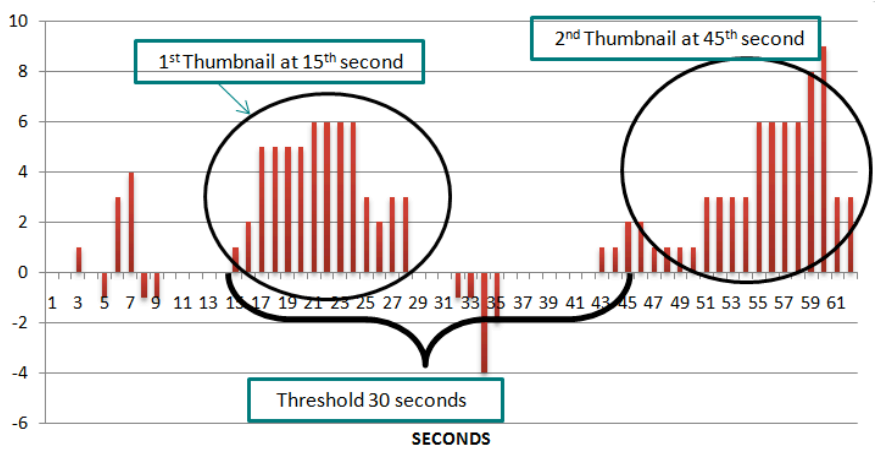

Figure 3 User activity graph with heuristic rules

In summary, we have employed a simple heuristic for the purpose of validating the system design (user centric approach for event detection). In this research, we focused on the design and implementation of a flexible system. In further research, we plan to elaborate in alternative user-centric algorithms, as well as to combine them with content-based ones.

Experimental Methodology

Materials

In this research, we are exploring a method for event detection, so we have elaborated on the selection of the suitable video content. We selected videos that are as much visually unstructured as possible, because content-based algorithms have already been successful with those videos that have visually structured scene changes. Another key factor is the length of a video. In general, YouTube allows video uploading up to 10 minutes. Although there are videos that exceed that limit, we decided not to use them, because it would be tiresome for the majority of users. Indeed, some early pilot user-tests have revealed that user attention is reduced after they have watched more than 3-4 videos of 10min each.
Narrative and entertainment have been the most popular category. According to He et al (1999) entertainment content is more likely to be watched in a leisurely manner and costs so much to produce that it is reasonable to have a human produce previews and summaries. Moreover, we decided to use lecture and how-to videos, because users are actively watching them to retrieve information about a specific topic. Documentary videos could be categorized as video or audio-centric, lectures have an audio-centric content and cooking videos have more video-centric features.
The documentary video features a segment of a television program called «Protagonists».[1] The selected segment refers to the use of internet by young people. The lecture video is a paper presentation from a local workshop.[2] Presentation's topic is «The acceptance of free laptops, that have been given to secondary education students ». Finally, the how-to video is a segment of a cooking TV show for a shouffle-cake.[3] Each one lasts ten minutes and is available on YouTube.

[1] http://www.youtube.com/watch?v=GOQfIXxbjlE

[2] http://www.youtube.com/watch?v=Z09ythJT9Wk

[3] http://www.youtube.com/watch?v=LzkYvtqlT5I

Measurement

This measuring process is based on the assumption of Yu et al (2003) that there are segments of a video clip that are commonly interesting to most users, and users might browse the respective parts of the video clip in searching for answers to some interesting questions. When enough user data is available, user behavior will exhibit similar patterns even if they are not explicitly asked to answer questions.

In order to experimentally replicate user activity we developed a questionnaire that corresponds to several segments of the video. We selected each scene in order to combine audio and video factors. Thus, each question corresponds to a visual and/or a structural cue that could be used as hints to find the answer. Furthermore there were some irrelevant questions in order to check that the users are searching for the answers and do not guess the correct ones. We took into account audio channel, speaker's channel, end-users' actions watching the talk to reveal the significant portions and video channel to select the thumbnails. We used Google Docs to create online forms for users' questionnaires and we integrated these forms in our user interface as it is presented in Figure 3.

Figure 3: Screenshot of VideoSkip with the questionnaire

Procedure

The goal of the user experiment is to collect activity data from the users, as well as to establish a flexible experimental procedure that can be replicated and validated by other researchers. There are several suggested approaches to the evaluation of interactive information retrieval systems (Kelly 2009). Instead of mining real usage data, we have designed a controlled experiment, because it provides a clean set of data that might be easier to analyze. The experiment took

place in a lab with Internet connection, general-purpose computers and headphones. Twenty-three university students (18-35 years old, 13 female and 10 male) spent approximately ten minutes to watch each video (buttons were muted). All students had been attending the Human-Computer Interaction courses at the Department of Informatics at a post- or under-graduate level and received course credit in the respective courses. Next, there was a time restriction of five minutes, in order to motivate the users to actively browse through the video and answer the respective questions. We informed the users that the purpose of the study was to measure their performance in finding the answers to the questions within time constraints. After a basic understanding between the user behavior data and the key-frame detection is established, further research could progress to larger scale studies, or even to field studies and data mining of large data-sets.

Table Summary of users' characteristics

Users Characteristics	
Number	23
Age	18-35
Gender	13 Female, 10 Male
Occupation	Studying informatics
Motivation	Course credit

Before the experimental procedure we introduced participants to the user interface of the video player and the questionnaire. The experimental session for each video consisted of two parts. Initially, the users had to watch a video and afterwards to answer the respective questionnaire. They could not see the questions from the beginning and the video player's buttons were hidden during the first part. Buttons were available for use in the second part and participants could use them to browse video and search for the answer. Figure 3 portrays the second part of the experimental procedure. Furthermore, there was a time restriction of five minutes in this part, in order to motivate the users to actively browse through the video. The procedure was repeated in a random sequence for each video, in order to minimize possible learning effects.

Results

The quantitative results confirmed our observations during the user experiment and indicated that twenty-three users produced more than one thousands interactions for three videos. Overall, the most popular button was the skip forward thirty seconds (skip30), because the users were under time pressure to retrieve the required information in order to answer the questions. Moreover, we noticed that there are more "play" than "pause" buttons, because sometimes the video would not start immediately so the users pressed play multiple times. We also observed that the interaction counts are identical between the different types of videos, because user behavior was rather motivated by the controlled conditions of the experiment than their own preferences about the content types. Nevertheless, we expect that in a natural setting (e.g., data mining of video logs from a web site) there will be variability depending on the type of the video content. In the following table we are presenting the counts of user interactions for each one of the buttons and then we analyze the user activity signals that we generated with a simple heuristic.

Table The most popular button is the skip forward because users had to retrieve information from a ten-minute video, within five minutes.

Video Interactions counts	Play	Pause	Skip30	Rewind30	Total
Documentary	38	36	271	85	430
Lecture	47	39	239	81	406
Cooking	32	20	302	68	422

In the case of the documentary video, the VideoSkip user activity graph matches most of the events (Figure 4). We found that the user activity graph is more accurate for those events that are short in duration. This is reasonable, because shorter events require higher attention by the user and thus they create more interaction with the player. On the other hand, longer events that repeat the same information are more difficult to register by the user-based heuristic, because the user interactions are spread over their duration without creating a hot-spot.

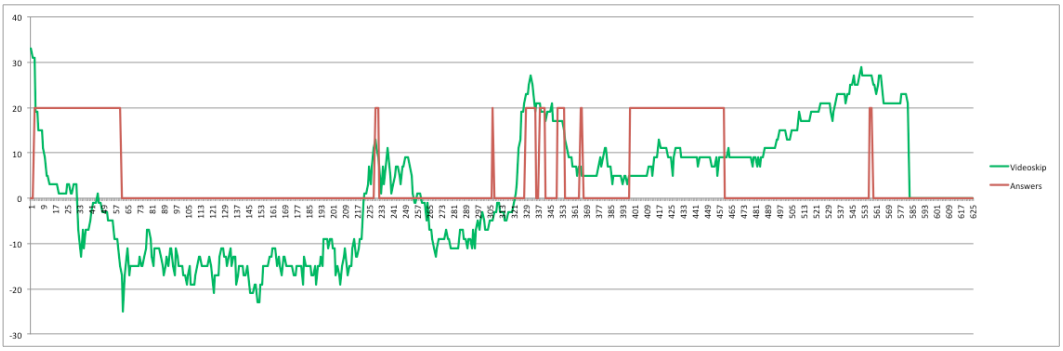

Figure 4: User activity graph for the documentary video (430 user interactions).

In the case of the lecture video (Figure 5), the VideoSkip user activity graph matches all but one (06:30) of the manually indicated events. Nevertheless, VideoSkip also indicated one event that did not correspond to any of the questions. It (00:24) is at the beginning of the presentation where the main topic is being presented. This element was not included to the questionnaire, but, it seems that the users actively browsed there in order to comprehend the content of the rest of the video.

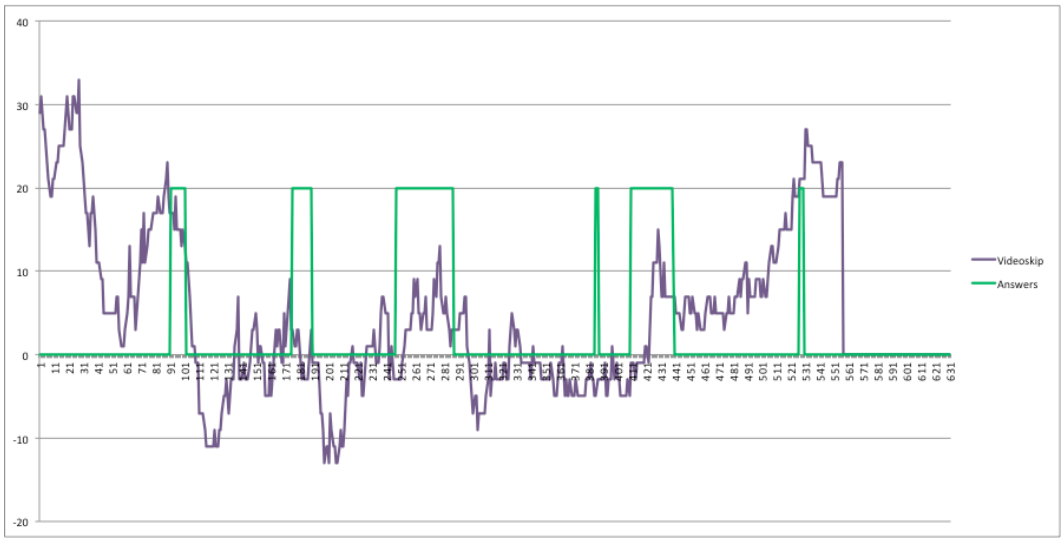

Figure 5: User activity graph for the lecture video (406 user interactions)

In case of cooking video, the VideoSkip user activity heuristic matches accurately most of the events (Figure 6). It is notable that despite the close distance of two events (08:00) we can distinguish between two local maxima of the user activity graph surface.

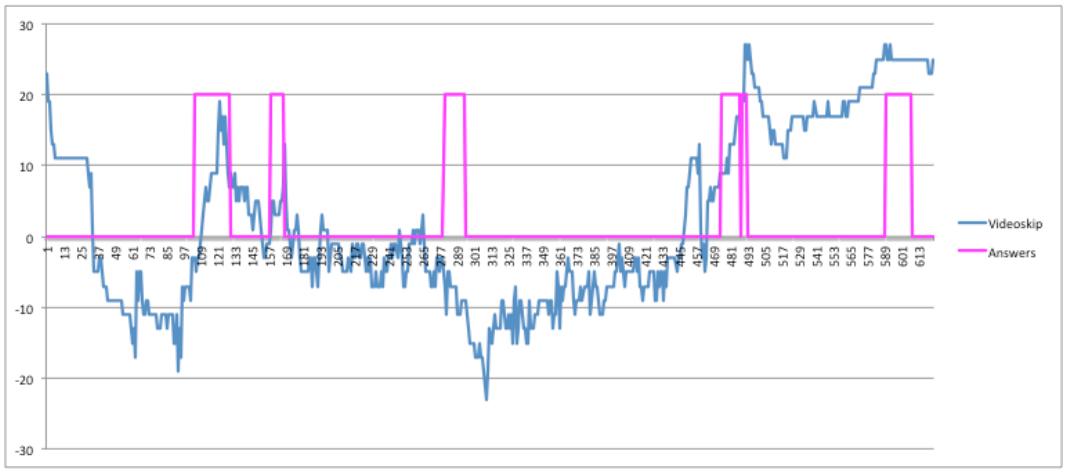

Figure 6: User Activity graph for the cooking video (422 user interactions)

In this work, we focused on simple analysis of the data set with an implicit user heurist. In further research, we plan to employ signal processing techniques, in order to better comprehend the attributes of the user activity graph.

Discussion

In terms of the experimental methodology, we have developed a flexible and lightweight process for the evaluation of user activity heuristics. He et al (1999) asked the users to rate video summaries in terms of clarity, conciseness, and coherence. Their users viewed the summaries in the traditional way but the basic interactions were not used. On the other hand VideoSkip exploits buttons' interactions to improve future viewers' navigation. Syeda-Mahmood and Ponceleon (2001) modeled viewers' interactions with VCR buttons to generate a video preview. Their approach has revealed that previous users' experiences can help future users in their browsing. Actually they propose a scene path that may detect something significant. We differentiated by using fewer player buttons and by using the interaction with each one as a meaningful event, instead of the transitional states. Additionally we compare the automated results with key frames selected manually. On their results the number of automatically detected interesting segments was higher than the actual number of such segments, but they included the correct segments.

In comparison to previous research, the proposed user activity heuristic is more maleable, because researchers can make various combinations and give different meaning to them. Yu et al (2003) experimental process used some questions to help mimic user interests and focus user behavior. Their algorithm should work

with any video as long as it contains some commonly attractive content. VideoSkip has been developed with the same assumption. On the other hand, they have implemented a system with a custom video browsing applications. Peng et al. (2011) have examined the physiological behavior (eye and head movement) of video users, in order to identify interesting key-frames, but this approach is not practical because it assumes that a video camera should be available and turned-on in the home environment. In contrast, the majority of users browses web video in more traditional ways that require no extra interactions or extra equipment, besides play, pause and seek, which are the main controls of VideoSkip.

Although the proposed user-based technique has several advantages, the present study has some limitations. We consider that both the type and the number of videos is very limited at least in comparison to the data-sets employed in established content-based research, but the granule of interest in user-based research is the number of user interactions. Therefore, we are concerned mostly with the number of users and in particular with the number of user interactions, which was in the scale of hundreds. We suggest that future studies employ a wider selection of video genres, and most importantly larger number of users and interactions with video.

Conclusion and Further Research

In this paper, we have experimentally validated a simple user activity heuristic for detecting events in a growing number of web videos, such as a lecture, and a how-to. Our approach has been successful, but there might be several steps required before a fully-working and practical application is in place. For example, the prototype system collects user activity data, but does not automatically detect the respective events. We have selected not to hard-code the algorithm in order to have flexibility in choosing different variations of the user activity heuristic. Moreover, we have selected to compute the heuristic with regard to a simple set of user actions (play/pause, fw30, rw30), because these are the simplest user actions with video. Nevertheless, there might exist further patterns of user interests in other actions, such as random seek and thumbnail navigation.

Another direction for further research would be to perform data mining on a large-scale web-video database. Indeed, social video interactions on web sites are very suitable for applying community intelligence techniques (Zhang et al. 2011) and

there are several projects exploring social networks on the internet (e.g., EU SOCIOS project). Nevertheless, we found that the experimental approach is more flexible than data mining for the development phase of the system. The design of our system (YouTube Chromeless player, Google App Engine) facilitates the incremental scaling of data collection from many users and videos, but we have selected to start with a controlled experiment. In contrast to data mining in large data-sets, a controlled experiment has the benefit of keeping a clean set of data that does not need several steps of filtering, before it becomes usable for any kind of simple user heuristic. In particular, the incremental and experimental approach is very suitable for user-centric information retrieval, because it is feasible to connect user behavior with the respective data-logs. Nevertheless, further research should consider: 1) a large scale deployment of users and videos, 2) the meaning of alternative user activity (e.g., random-seek, direct selection of thumbnail), and 3) a system that automatically generates and displays the selected thumbnails. Moreover, we have considered several types of video content, but we decided to focus on lectures and how-to videos, because they represent a growing and useful amount of content on the web. Moreover, user-centric techniques might be more suitable for lecture and how-to videos that do not have very structured scene changes like the narrative videos. Further research should consider a comparative approach between content-based and user-centric algorithms across different types of web videos. In addition, we expect that in contrast to content-based techniques, the user-centric design has the benefits of continuously adapting to evolving users' preferences, as well as providing additional opportunities for the personalization of content to different user groups. Therefore, user-based techniques for event detection in video provide the groundwork for future work in user modeling and user adapted interaction.

Finally, further research and practice should build upon the availability of user-centric and content-based approaches in order to provide the right mix in hybrid algorithms, which might represent the best of the two worlds. Although the proposed technique provides a simple and effective pointer to the most important moments in a video, it is not sufficient to understand the contents of that moment. Therefore, we propose that future research applies content-based methods to audio and video signals, as well as text-oriented (e.g., comments, tags) user-based

methods to social web videos, in order to paint the complete picture of the semantics in video hot-spots.

Acknowledgements. We are thankful to Ioannis Leftheriotis for his work on early system prototypes and to the participants of the user study.

References

Fotis Aisopos, George Papadakis, and Theodora Varvarigou. 2011. Sentiment analysis of social media content using N-Gram graphs. In Proceedings of the 3rd ACM SIGMM international workshop on Social media (WSM '11). ACM, New York, NY, USA, 9-14.

S. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y., And Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," Proceedings of the 7th ACM SIGCOMM Conference on internet Measurement, San Diego, California, USA: ACM, New York, NY, 2007, pp. 1-14.

L. Chen, G. Chen, C. Xu, J. March, and S. Benford, "EmoPlayer: A media player for video clips with affective annotations," Interacting with Computers, 2008, pp. 17-28.

C. Crockford and H. Agius, "An empirical investigation into user navigation of digital video using the VCR-like control set," International Journal of Human-Computer Studies, 2006, pp. 340-355.

Doulamis, A. D., & Doulamis, N. D. (2004). Optimal content-based video decomposition for interactive video navigation. IEEE Transactions on Circuits and Systems for Video Technology, 14(6), 757-775.

Anastasios D. Doulamis, Jordi Gonzàlez: First ACM international workshop on analysis and retrieval of tracked events and motion in imagery streams (ARTEMI 2010). ACM Multimedia 2010: 1749-1750

S. Drucker, A. Glatzer, and S.D. Mar, C, "SmartSkip: consumer level browsing and skipping of digital video content," In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves, Minneapolis, Minnesota, USA: 2002, pp. 219-226.

A. Girgensohn, J. Boreczky, and L. Wilcox, "Keyframe-based user interfaces for digital video," Computer, vol. 34, 2001, pp. 61-67.

L. He, E. Sanocki, A. Gupta, and J. Grudin, "Auto-summarization of audio-video presentations," Proceedings of the seventh ACM international conference on Multimedia (Part 1) - MULTIMEDIA '99, New York, New York, USA: ACM Press, 1999, pp. 489-498.

R. Hjelsvold, S. Vdaygiri, and Y. Léauté, "Web-based personalization and management of interactive video," Proceedings of the tenth international conference on World Wide Web - WWW '01, 2001, pp. 129-139.

D. Kelly (2009) Methods for Evaluating Interactive Information Retrieval Systems with Users, Foundations and Trends in Information Retrieval: Vol. 3: No 1—2, pp 1-224.

J. Kim, H. Kim, and K. Park, "Towards optimal navigation through video content on interactive TV," Interacting with Computers, vol. 18, Jul. 2006, pp. 723-746.

A. Money and H. Agius, “Video summarisation: A conceptual framework and survey of the state of the art,” Journal of Visual Communication and Image Representation, vol. 19, Feb. 2008, pp. 121-143.

I. Leftheriotis, C. Gkonela, and K. Chorianopoulos, “Efficient Video Indexing on the Web: A System that Leverages User Interactions with a Video Player,” In Proceedings of the 2nd International Conference on User-Centric Media (UCMEDIA 2010).

F.C. Li, A. Gupta, E. Sanocki, L.-wei He, and Y. Rui, “Browsing digital video,” Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’00, vol. 2, 2000, pp. 169-176.

Y.-fei Ma and H.-jiang Zhang, “Video Snapshot: A Bird View of Video Sequence,” 11th International Multimedia Modelling Conference, IEEE, 2005, pp. 94-101.

Siddharth Mitra, Mayank Agrawal, Amit Yadav, Niklas Carlsson, Derek Eager, and Anirban Mahanti. 2011. Characterizing Web-Based Video Sharing Workloads. ACM Trans. Web 5, 2, Article 8 (May 2011)

Klimis S. Ntalianis, Anastasios D. Doulamis, Nicolas Tsapatsoulis, and Nikolaos Doulamis. 2010. Human action annotation, modeling and analysis based on implicit user interaction. Multimedia Tools Appl. 50, 1 (October 2010), 199-225.

Peng, W.-T., Chu, W.-T., Chang, C.-H., Chou, C.-N., Huang, W.-J., Chang, W.-Y., and Hung, Y.-P. (2011). Editing by viewing: Automatic home video summarization by viewing behavior analysis. Multimedia, IEEE Transactions on, 13(3):539-550.

D.A. Shamma, R. Shaw, P.L. Shafton, and Y. Liu, “Watch what I watch: using community activity to understand content,” Proceedings of the international workshop on Workshop on multimedia information retrieval - MIR ’07, N. ACM, New York, ed., Augsburg, Bavaria, Germany: ACM Press, 2007, p. 275.

Xuehua Shen, Bin Tan, and ChengXiang Zhai. 2005. Implicit user modeling for personalized search. In Proceedings of the 14th ACM international conference on Information and knowledge management (CIKM '05). ACM, New York, NY, USA, 824-831.

SOCIOS EU project: http://www.sociosproject.eu/

T. Syeda-Mahmood and D. Ponceleon, “Learning video browsing behavior and its application in the generation of video previews,” Proceedings of the ninth ACM international conference on Multimedia - MULTIMEDIA ’01, New York, New York, USA: ACM Press, 2001, p. 119.

Y. Takahashi, N. Nitta, and N. Babaguchi, “Video summarization for large sports video archives,” In Proceedings of the 13th Annual ACM international Conference on Multimedia, Hilton, Singapore,: ACM, New York, 2005, pp. 820-828.

B. Yu, W.-Y. Ma, K. Nahrstedt, and H.-J. Zhang, “Video summarization based on user log enhanced link analysis,” Proceedings of the eleventh ACM international conference on Multimedia - MULTIMEDIA ’03, New York, New York, USA: ACM Press, 2003, p. 382.

Zhang, D., Guo, B., Yu, Z., Jul. 2011. The emergence of social and community intelligence. Computer 44 (7), 21-28.