Hindawi Publishing Corporation
Advances in Multimedia
Volume 2014, Article ID 712589, 9 pages
http://dx.doi.org/10.1155/2014/712589

Research Article

Video Pulses: User-Based Modeling of Interesting Video Segments

Markos Avlonitis and Konstantinos Chorianopoulos

Ionian University, 49100 Corfu, Greece

Correspondence should be addressed to Konstantinos Chorianopoulos; choko@ionio.gr

Received 11 September 2013; Accepted 29 November 2013; Published 12 January 2014

Academic Editor: Deepu Rajan

We present a user-based method that detects regions of interest within a video in order to provide video skims and video summaries. Previous research in video retrieval has focused on content-based techniques, such as pattern recognition algorithms that attempt to understand the low-level features of a video. We are proposing a pulse modeling method, which makes sense of a web video by analyzing users' *Replay* interactions with the video player. In particular, we have modeled the user information seeking behavior as a time series and the semantic regions as a discrete pulse of fixed width. Then, we have calculated the correlation coefficient between the dynamically detected pulses at the local maximums of the user activity signal and the pulse of reference. We have found that users' *Replay* activity significantly matches the important segments in information-rich and visually complex videos, such as lecture, how-to, and documentary. The proposed signal processing of user activity is complementary to previous work in content-based video retrieval and provides an additional user-based dimension for modeling the semantics of a social video on the web.

1. Introduction

The web has become a very popular medium for sharing and watching video content [1]. Moreover, many organizations and academic institutions are making lecture videos and seminars available online. Previous work on video retrieval has investigated the content of the video and has contributed a standard set of procedures, tools, and data-sets for comparing the performance of video retrieval algorithms (e.g., TRECVID), but they have not considered the interactive behavior of the users as an integral part of the video retrieval process. In addition to watching and browsing video content on the web, people also perform other "social metadata" tasks, such as sharing, commenting videos, replying to other videos, or just expressing their preference/rating. User-based research has explored the association between commenting and microblogs, primarily tweets, or other text-based and explicitly user-generated content. Although there are various established information retrieval methods that collect and manipulate text, they could be considered burdensome for the users, in the context of video watching. In many cases, there is a lack of comment density when compared to the number of viewers of a video. There are a few research efforts to understand user-based video retrieval without the use of social metadata.

In our research, we have developed a method that utilizes more so implicit user interactions for extracting useful information about a video. Our goal is to analyze the aggregated user interactions with the video using a stochastic pulse modeling process.

2. Related Work

Content semantics is an important concept that facilitates the retrieval of information from rich, yet complex, content, such as video. Semantic research in multimedia details two broad categories of approaches: content-based and user-based. Content-based methods extract meaning by analyzing the video itself (e.g., scene change, sound, and closed captioning). Alternatively, user-based methods extract meaning by analysis of the user activity on the video. Of these user-based actions, there are two subcategories; they can be explicit, like comments, annotations, and ratings, or implicit, such as play/pause events or seeking/scrubbing behavior [2]. One such set of experiments involves associating or finding a video's table of contents. Just like a book or a web site with many pages has a user navigation metaphor based on an index or a table of contents, a video needs structure to facilitate user through numerous scenes. Video table of contents is perceived by people to have high value for finding information, yet are seldom used for navigation when one is available [3]. Scenes are generally provided to the user with a set of thumbnails, which are called key-frames, if they are fixed pictures, or skims, if they are short videos [4]. A collection of still images has become popular in many applications, because it is easy to display and delivers a set of images, which stand as a table of contents for a video.

Besides the research interest in scene extraction, there have been also commercial systems that provide similar functionality. Many commercially available online players and devices, such as YouTube (Figure 1), provide thumbnails to facilitate user's navigation in each video. Nevertheless, most of the techniques that extract thumbnails at regular time intervals or from each shot are inefficient, because there might be too many shots in a video. In the case of Google YouTube, there is a very large number of thumbnails, which depending on the length of the video might be captured every second (for a three-minute video) or every five seconds (for an hour of video). Therefore, the selection of the thumbnails is actually completely random and stands for neither the content nor the semantics of the content.

2.1. Content-Based Semantics. Content-based information retrieval uses automated techniques to analyze actual video content. It uses images' colors, shapes, textures, sounds, motions, events, objects, or any other information that can be derived from only the video itself. Some techniques have combined the videos' metadata [5] with picture [6] or sound [7], while other researchers provide affective annotation [8, 9] or navigation aids [10]. Even though content-based techniques have begun to emphasize the importance of user perception, they do not take into account people's actual browsing and sharing behavior. Moreover, low-level features (e.g., color and camera transitions) often fail to capture the high-level semantics (e.g., events, actors, and objects) of the video content itself, yet such semantics are often what guide users, particularly nonspecialist users, when navigating [9] within or between videos [10].

According to Money and Agius [11], another classification for video summarization takes into account information from the videos during its production stage; this is called internal summarization as seen in SmartSkip by Drucker et al. [6]. Likewise, external summarization analyzes exterior information during any stage of the video lifecycle; however, most external summarization techniques ignore user activity with the video. Other approaches focus on personalization with the user. Hjelsvold et al. [12] employed hotspots and hyperlinks to match the content to the user profile. Although their framework is based on users' preferences, it requires extra user effort in order to build a profile. Overall, since it is very difficult to detect scenes and extract meaning from videos, previous research has attempted to model video in terms of better-understood concepts, such as text and images [13].

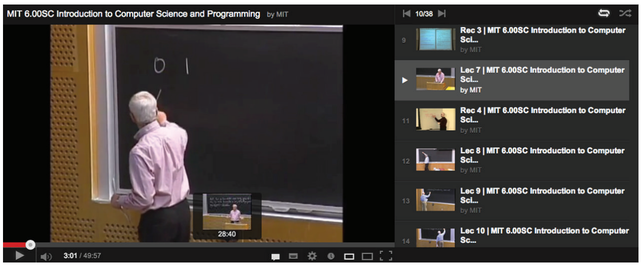

MIT 6.00SC Introduction to Computer Science and Programming by MIT
10/38
9
Rec 3 | MIT 6.00SC Introduction to Computer Sci...
by MIT
Lec 7 | MIT 6.00SC Introduction to Computer Sci...
by MIT
11
Rec 4 | MIT 6.00SC Introduction to Computer Sci...
by MIT
12
Lec 8 | MIT 6.00SC Introduction to Computer Sci...
by MIT
13
Lec 9 | MIT 6.00SC Introduction to Computer Sci...
by MIT
14
Lec 10 | MIT 6.00SC Introduction to Computer Sc...
28:40
3:01 / 49:57

Figure 1: Google YouTube provides several thumbnails for each video. Moreover, a thumbnail is used to represent related videos on the right. The selection (as well as the number) of these thumbnails is important for effective user navigation.

To evaluate methods for understanding video content, researchers and practitioners have been cooperating for more than a decade on a large-scale video library and tools for analyzing the content of video. The TRECVID (TREC Video Retrieval Evaluation: http://trecvid.nist.gov/) workshop series provides a standard-set of videos, tools, and benchmarks, which facilitate the incremental improvement of sense making for videos [14].

In summary, content-based techniques facilitate the discovery of a specific scene, the comprehension of a video in a limited time, and the navigation in multiple videos simultaneously. Again, here the video content is analyzed rather than the metadata associated with people or how people manipulated and consumed the video. Finally, content-based techniques are not applicable to some types of web video, such as lecture and how-to instruction, with a visually flat structure, or are semantically complex respectively.

2.2. User-Based Semantics. In comparison to the more so legacy content-based techniques, there are fewer works on user-based analysis of information retrieval for video content. One explanation for this imbalance is not the importance of content-based, but it is the relatively newer interest in the social web, the sharing, and the use of videos online. Nevertheless, there is a growing body of research and interest on user-based retrieval of video.

User interaction with video has been a basic element of multimedia research for many years. Yu et al. [15] suggested 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. Their work, influenced by the PageRank and centrality metrics, assumes there exists a short path in each video. Similarly, Syeda-Mahmood and Ponceleon [16] suggested that user interaction with video is a Markov-model chain of affect-based probability, and they developed a media player-based learning system called the MediaMiner. 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. They modeled implicit user activity according to the user's sentiment (e.g., user is bored, or interested) nowadays is not the main motivation for watching video content. For example, there is a growing number of lecture and how-to videos, which are being watched for their informational value.

Finally, social video interactions on web sites are very suitable for applying community intelligence techniques. In the seminal user-based approach to web video, Shaw and Davis [17] proposed that video representation might be better modeled after the actual use made by the users. Notably, Yew and Shamma [2] have recognized the importance of scrubs (fast forward and rewind), but they have only included counts in their classifier and not the actual timing of the scrub events. Thus, we propose to leverage implicit user activity (e.g., pause/play, seek/scrub), in order to dynamically identify video segments of interest.

In summary, as more media is posted and viewed in online contexts, we assert the importance of analyzing the implicit behavior of consumption along with the traditional video signal and contemporary social metadata.

3. Methodology

We employed an open data-set [18], which has been created in the context of a controlled user experiment (23 users, approximately 400 user interactions within each video), in order to ensure well-defined user-based semantics and noise-free user activity data. Previous work has highlighted the evidence of correlation between the local maximum of user activity and the regions of interest [19], but it has not provided a statistical measure of this correlation, which is the focus of this work. Next, we developed a user activity model for analyzing user interactions as a time-based signal. Since there are no similar works in user activity modeling of implicit user interactions within web video, we have developed a pulse modeling process, which is straightforward to replicate for the same set videos or different ones.

In the initialization phase, we consider that every video is associated with four distinct time series of length equal to the video duration in seconds. Each series corresponds to the four distinct buttons of *Play/Pause*, *Skip*, and *Replay*.

It is our aim to construct a general formalism to treat the statistical properties of the aforementioned discrete signals as well as correlation properties between them. We have adapted established techniques from similar signal processing domains such as material science and seismology (see, e.g., [20] and references therein). Let us consider N user interactions and denote with $\mathbf{r}$ the position vectors of those actions in the time domain. The type of the button pushed is labeled by m. The discrete system of user's actions can be formally characterized by discrete densities as follows:

$$\rho^{m}(\mathbf{r}) = \sum_{j}^{N} \delta^{m}\left(\mathbf{r} - \mathbf{r}_{\mathbf{j}}\right), \tag{1}$$

which is actually a counter of the series of pulses (here modeling the users' actions) of definite width the centers of which are determined by the position vectors $\mathbf{r}$ in time. The complete knowledge of the user's actions system is attributed to the fourth-dimension density function $\rho(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) dv_1 dv_2 dv_3 dv_4$ interpreted as being the joint probability to find the first button action in a time volume element dv_1 at $\mathbf{r}_1$, the second button action in a time volume element dv_2 at $\mathbf{r}_2$, the third button action in a time volume element dv_3 at $\mathbf{r}_3$, and the fourth button action in a time volume element dv_4 at $\mathbf{r}_4$. One possible way to take into account time correlation between the different bottom user actions is to assume that

$$\begin{aligned} \rho(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \rho^{1}(\mathbf{r}_1)\, \rho^{2}(\mathbf{r}_2)\, \rho^{3}(\mathbf{r}_3) \\ \times \rho^{4}(\mathbf{r}_4)\left(1 - d(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4)\right), \end{aligned} \tag{2}$$

where $d(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4)$ corresponds to the correlation function in a homogeneous system and which in a first approximation can be considered of higher order. To this end, in the rest of the paper, we assume the simplest case of uncorrelated button actions. On the other hand pair correlation functions between pulse signals may be treated as usual with the well-known Pearson correlation coefficient.

Initially, the user activity signal is created as follows: each time user presses the *Replay* (*Skip*) button; the moments matching the replayed (skipped) segment of the video are incremented by one. We assume that the user replays a video either because there is something interesting or because there is something difficult to understand, while the user skips a video because there is nothing of interest. In this way, an experimental time series is constructed for each button and for each video—a depiction of users' activity patterns over time. In order to extract pattern characteristics for each time series, that is, scenes with high user activity, the following methodology, consistes of four distinct stages (see Table 1), was used.

In the first stage, we use simple procedure in order to average out user activity noise (Figure 2). In the context of probability theory the noise removal can be treated with the notion of the moving average [21]: from a time series $s^{\exp}(t)$ a new smoother time series $s_{T}^{\exp}(t)$ may be obtained as

$$s_{T}^{\exp}(t) = \frac{1}{T} \int_{t-T/2}^{t+T/2} s^{\exp}\left(t'\right) dt', \tag{3}$$

where T denotes the averaging "window" in time. The larger the averaging window T is, the smoother the signal will be. Schematically the procedure is depicted in Figure 2. The procedure of noise removal of the experimentally recording signal $s^{\exp}(t)$ is of crucial importance for the following reasons: first, in order to reveal trends of the corresponding signals (regions of high user activity) and second in order to estimate local maxima for the second stage, as explained in the next paragraph. It must be noted that the optimum size of the averaging window T is completely defined from the variability of the initial signal. Indeed, T should be large enough in order to average out random fluctuations of the user's activities and small enough in order to reveal, and not disturb, the bell-like localized shape of the user's signal which in turn will demonstrate the area of high user activity.

Table 1: Overview of the user activity modeling and analysis.

Stage	User activity signal processing
1	Smoothness procedure
2	Pulse construction at local maximums
3	Construction of approximated reference pulses
4	Determination of correlation between pulse signals

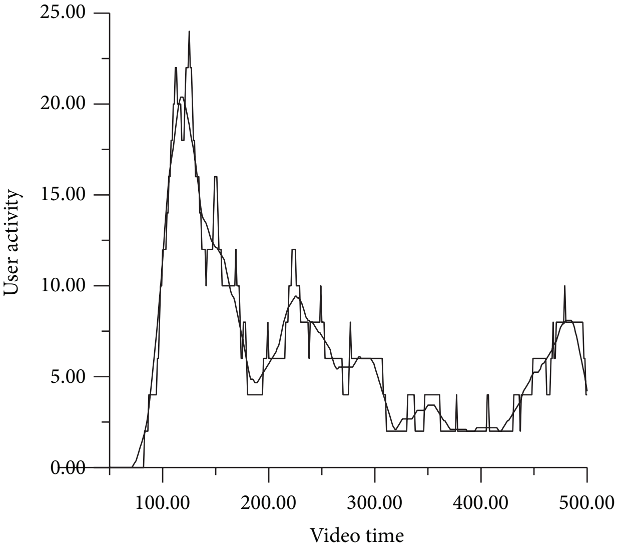


Figure 2: The user activity signal is approximated with a smooth signal. The y-axis is the measured user activity while the x-axis is the relative video time in seconds. The same notation is used throughout the paper.

In the second stage, we construct a pulse series from the above constructed user activity smooth signal (Figure 3). The pulse signal is to be compared with the corresponding pulse signal, which models the regions of interest of each video as explained in the third stage. The idea to construct a pulse signal from a time series is not new and several methods may be found in the literature (see, e.g., in [22] and references therein). At the basis of all those formalisms is the need to construct an analytical signal that models local areas of a given signal with significant value in contrast with the rest of the domain where almost zero values are encountered. Instead of pulses, other functional, for example, Gaussian-like, could be also used. In our analysis, the shape of the localized functional had no effect and as a result we kept the pulses since the following analysis was easier. Here, in order to construct the pulse signal the exact location of the pulses is defined by means of the generalized local maxima of the experimental smooth signal (Figure 4). By the term generalized local maxima, we mention the center of the corresponding bell-like area of the average signal, since the nature of our signal may cause more than one peak at the top of the bell. Although the height of the pulse does not affect our results, the width of the pulse D is a parameter that must be treated carefully. In particular, the variability of the average signal determines the order of the pulse width D. Here, we propose that the pulse width should be equal to the average half of the widths of the bell-like regions of the signals (see Figure 3(b)). In the context of our controlled experiment, this is a safe assumption, but it requires further elaboration in different experimental setups or in the field (e.g., data-mining of real video usage data). Moreover, we are providing a more detailed analysis of the interplay between the parameters in Section 5.

In the third stage we construct the corresponding pulse signal $s_{kf}(t)$ which models the regions of interest of each video (Figure 3). For compatibility reasons and without loss of generality the shape of the pulses (width and high) is the same as for $s_{kf}^{\exp}(t)$. On the other hand, the exact locations of the pulses are defined as the center of the corresponding regions of interest as defined in the data-set.

It is our aim to examine whether the two signals (user activity and reference pulses) are correlated, for example, whether the patterns revealed from the user's activity are correlated with objective regions of interest of each video. In order to check this hypothesis the cross-correlation coefficient was used which estimates the degree to which two series are correlated (e.g., [21]). The values of the correlation coefficient range from −1 to 1. Perfect uncorrelated time series has zero correlation coefficient, while positive or negative correlations may be scored as follows (we refer to absolute values): from 0.1 to 0.3 low correlation, from 0.3 to 0.5 medium correlation, and from 0.5 to 1 strong correlation. It is noted that the determination of the cross-correlation coefficient as well as the proposed signal process methodology was carried out via simple codes developed with standard math libraries in the C programming environment.

4. Results

We have focused on the analysis of the video seeking behavior, such as *Replay* and *Skip* the previously described smoothening procedure. An exploratory analysis with time series probabilistic tools verified what is visually depicted in the case of Video A, which is a lecture video (Figure 4). While the *Replay* signal has a quite regular pattern with a small number of regions with high user's activity, the *Skip* signal is characterized by a large number of merely random and abnormal local maxima of user's activity. We have also considered the use of the *Play/Pause* buttons, but there were few interactions. In the following, we present the results of the *Replay* signal analysis for four videos.

The analysis of the user activity signal was based on an exploration of several alternative averaging window sizes. The results of the pulse modeling methodology are depicted in Tables 2 to 5. The smoothed signals are plotted with the solid black curve. The pulse signals were extracted from the corresponding local maxima that are depicted with the red discontinued pulse signal while the pulse signals that model the regions of interest of each video are depicted with the blue solid pulse. Although the correlation of the constructed pulse signals for each video is visually evident in the graphs (figures embedded in Tables 2 to 5), the cross-correlation coefficient was used in order to establish the respective quantitative

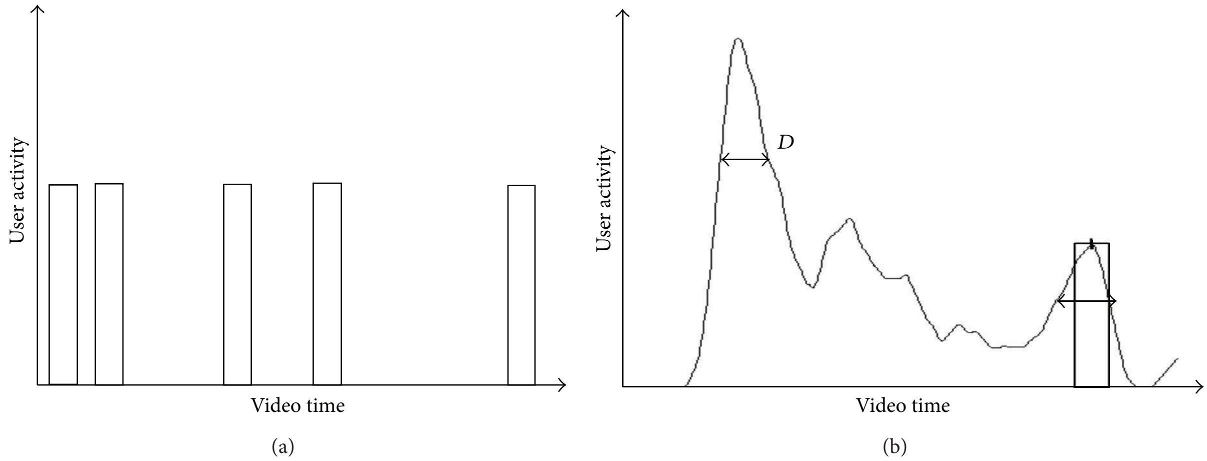


Figure 3: The pulse of reference (a), which is based on manually selected video scenes, is compared to the experimental pulse (b), which is created at the local maximum of the (smooth) user activity signal. The optimum pulse width D is also depicted schematically.

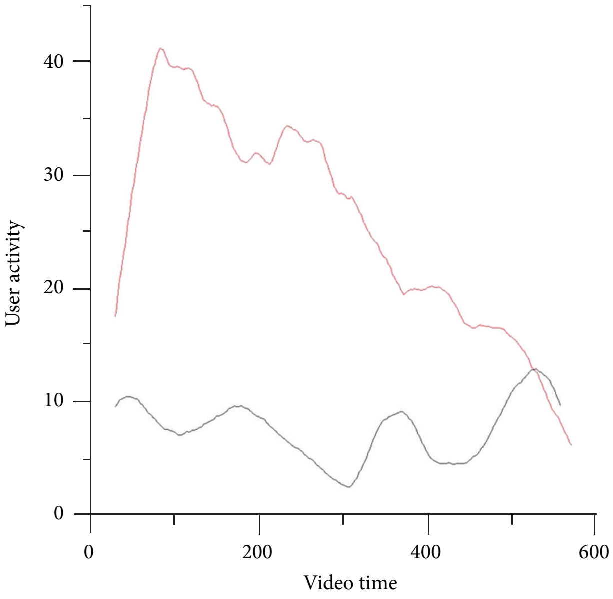


Figure 4: The *Replay* signal (blue, at the bottom) was compared to the *Skip* (red, at the top), in order to understand which one is closer to the semantics of the video. The higher values of the *Skip* signal stand for the popularity of the respective user activity.

measures. Indeed, the cross-correlation coefficients that we estimated were 0.67, 0.58, 0.76, and 0.62 correspondingly, indicating strong correlation between the two signals (reference and user signal). The pulse modeling process has identified the majority of the manually selected video scenes with high accuracy, but a few scenes were still not detected. In Tables 2, 3, 4, and 5, the video scenes (S1,. . .,S5) detected by the algorithm (user activity pulse modeling) are compared to the reference video scenes.

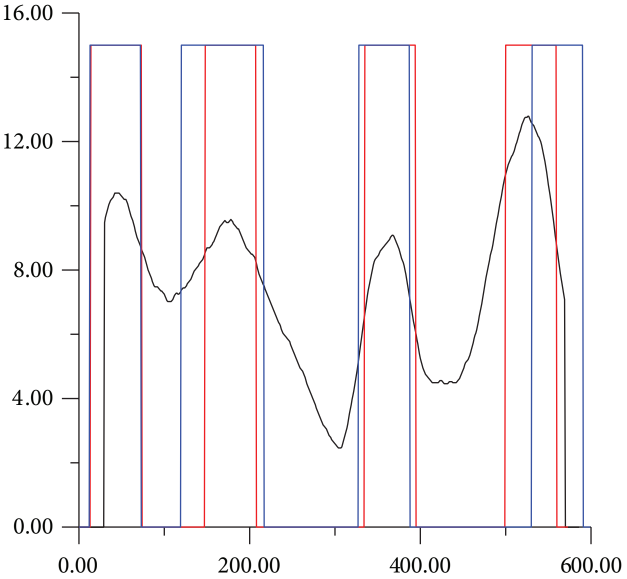
The most important parameter in the analysis of the user activity signal is the averaging window T and the relationship it has with the (1) skipping step, (2) video duration, and (3) number of user interactions. It must be noted that several values of the averaging window T were checked and the empirical relation $T \approx D$ was found, as the optimal one since it removes the underline signal noise without affecting user's activity characteristics. It is notable that if the skipping step was not fixed (e.g., random seek with a progress bar), then the analysis of the user activity signal would have required a dynamic size of the averaging window T, which would have made the process much more complex. In summary, the above results demonstrate the efficacy of this approach and provide a small set of parameters (video browsing actions, averaging window duration T, and pulse width D) that need to be further explored, as it is discussed next.

5. Discussion

In this work, we focused on an application for detecting important video segments, because it plays several roles in understanding video semantics. In particular, the important segments provide an additional navigation mechanism and an abstract of the video, either thumbnails or skims. The idea to interpret user's actions as a sum of discrete pulses as was mentioned before is borrowed from other fields, for example, material science [23]. Actually what is common is the existence of different populations (here different types of buttons) of discrete nature (discrete user's actions) and their patterning or morphogenesis in the corresponding space (here patterning of user's actions within the video duration). Note that since populations are discrete in nature the corresponding emerged patterns are also discrete thus resulting in theoretical models by means of pulses of definite width.

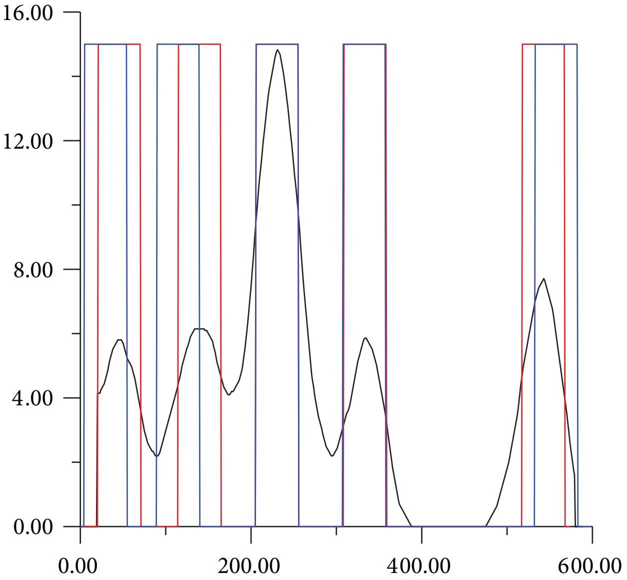
The determination of the optimum averaging window as well as the corresponding width pulse is of crucial importance and the analysis shows that these are dynamic-like variables meaning that their values require a careful balance between video and user activity attributes. On the one hand, a lengthy video might require a wider averaging window, in order to

Table 2: Video A is a lecture video (http://www.youtube.com/watch?v=8LebAtvulIY). The pulse width D is 60 seconds and the smoothing window T is 60 seconds. The pulse modeling is reported with respect to the center of each pulse.



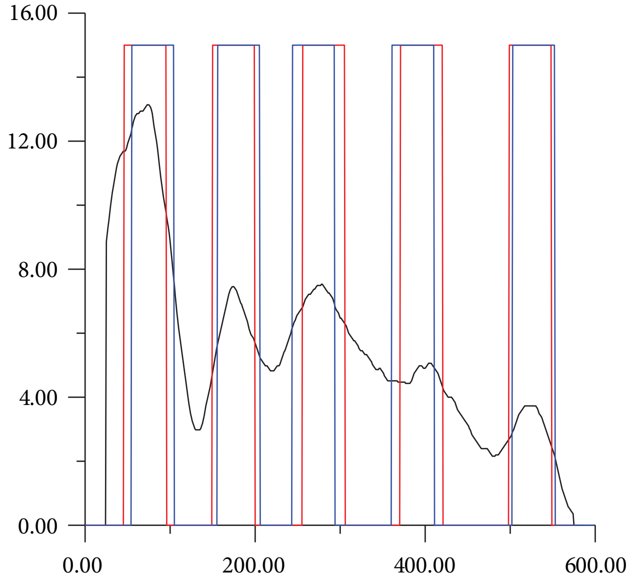
A	Reference pulse center (sec)	User activity pulse center (sec)
S1	43	44
S2	150	
S3	187	178
S4	358	365
S5	561	530

Table 3: Video B is a documentary video (http://www.youtube.com/watch?v=tSV2kAfkp5A). The pulse width D is 50 seconds and the smoothing window T is 40 seconds. The pulse modeling is reported with respect to the center of each pulse.



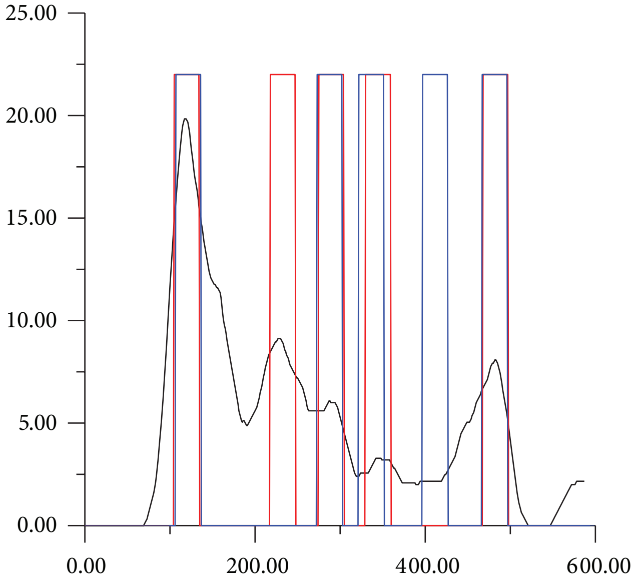
B	Reference pulse center (sec)	User activity pulse center (sec)
S1	30	46
S2	115	140
S3	231	231
S4	333	334
S5	558	543

Table 4: Video C is a lecture video (http://www.youtube.com/watch?v=Z09ythJT9Wk). The pulse width D is 50 seconds and the smoothing window T is 50 seconds. The pulse modeling is reported with respect to the center of each pulse.



C	Reference pulse center (sec)	User activity pulse center (sec)
S1	80	71
S2	181	175
S3	269	281
S4	386	396
S5	528	524

Table 5: Video D is a cooking (how-to) video (http://www.youtube.com/watch?v=LzkYvtqlT5I). The pulse width D is 30 seconds and the smoothing window T is 25 seconds. The pulse modeling is reported with respect to the center of each pulse.



D	Reference pulse center (sec)	User activity pulse center (sec)
S1	122	120
S2	288	233
S3	337	290
S4	412	345
S5	482	483

limit the possible number of detected scenes. For example, a typical one-hour lecture with many users would have produced too many local maximums, which could be filtered with a wider averaging window (e.g., ten minutes). On the other hand, the larger the number and the variability of the users' activity signal, the smaller the averaging window. Indeed, if a dense users' activity is recording (during the video time), then a small averaging window must be used in order to catch this dense activity, while a larger averaging window may result to a mutual overlapping of two different regions of interest. Further research should also explore these basic signal attributes (smoothing window T and pulse width D) in the context of other real systems. In this way, our knowledge about the user activity signal attributes could complement the experimental understanding we have described in this work.

We have only employed four videos in the experimental procedure. Previous work on content-based information retrieval from videos has emphasized the number of videos employed in similar experiments, because the respective algorithms treated the content of those videos. In this user-based work, we are not concerned with the content of the videos, but with the user activity on the videos. Nevertheless, it is worthwhile to explore the effect of more videos and interaction types. Therefore, the small number of videos used in the study is not an important limitation, but further research has to elaborate on different genres of video (e.g., news, sports, and comedy) and the semantic label of the interaction (e.g., answering who, what, and how).

Another significant open research issue is the number of thumbnails. We have already shown that Google YouTube (Figure 1) provides so many thumbnails that the user has to navigate through them by scrolling. This research issue has already concerned SmartSkip's developers [6]. They started out with ten thumbnails and after an early prototype test; they reduced the number of thumbnails to eight. According to the final user test, they suggested to reduce the number of thumbnails even further to five. Nevertheless, the number of scenes depends on several parameters, such as the type and length of the video. Therefore, it is unlikely that there are a fixed number of scenes that describe a particular video. If the required number of scenes is different for each video, then, besides the scene extraction technique, we need a method to select the most important of them.

6. Conclusion

In this research, we validated a method for scene detection in web videos. Our main goal is to understand the semantics of video content from users' interactions with the video player. In particular, we found that the aggregation of user *Replay* interactions with the video player stands for the most important segments of a video. The results of this type of study can be used to develop systems that understand important video's scenes, generate thumbnails, and create a video summary. We decided to explore a user-based approach, because previous works have already analyzed content-based methods and because of a growing number of web videos and the respective user interactions.

A direction for further research would be to perform data mining on a large-scale web-video database. Nevertheless, we found that the experimental approach is more flexible than data mining for the development phase of a new video retrieval system. In particular, the iterative and experimental approach is very suitable for user-centric information retrieval, because it is feasible to explore and associate user behavior with the respective data-logs. Moreover, 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. Finally, we suggest that user-based content analysis has the benefits of continuously adapting to evolving users' preferences, as well as providing additional opportunities for the personalization of content. For example, researchers might be able to apply several personalization techniques, such as collaborative filtering, to the user activity data. In this way, implicit video pragmatics is emerging as a new playing field for improving user experience on social multimedia on the web.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The work reported in this paper has been partly supported by Project CULT (http://cult.di.ionio.gr/). CULT (MC-ERG-2008-230894) is a Marie Curie project of the European Commission (EC) under the 7th Framework Program (FP7).

References

[1] M. Cha, H. Kwak, P. Rodriguez, Y. Ahnt, and S. Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," in *Proceedings of the 7th ACM SIGCOMM Internet Measurement Conference (IMC '07)*, pp. 1–14, ACM, San Diego, Calif, USA, October 2007.

[2] J. Yew and D. A. Shamma, "Know your data: understanding implicit usage versus explicit action in video content classification," in *5th Multimedia on Mobile Devices 2011; and Multimedia Content Access: Algorithms and Systems*, vol. 7881 of *Proceedings of SPIE*, San Francisco, Calif, USA, January 2011.

[3] E. G. Toms, C. Dufour, J. Lewis, and R. Baecker, "Assessing tools for use with webcasts," in *Proceedings of the 5th ACM/IEEE Joint Conference on Digital Libraries*, pp. 79–88, ACM Press, New York, NY, USA, June 2005.

[4] B. T. Truong and S. Venkatesh, "Video abstraction: a systematic review and classification," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 3, no. 1, article 3, 2007.

[5] 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*, pp. 820–828, ACM, Singapore, 2005.

[6] S. M. Drucker, A. Glatzer, S. de Mar, and C. Wong, "Smartskip: consumer level browsing and skipping of digital video content," in *Proceedings of the SIGCHI Conference on Human Factors in*

Computing Systems (CHI '02), pp. 219–226, Minneapolis, Minn, USA, April 2002.

[7] F. C. Li, A. Gupta, E. Sanocki, L. W. He, and Y. Rui, "Browsing digital video," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*, vol. 2, pp. 169–176, April 2000.

[8] L. Chen, G. Chen, C. Xu, J. March, and S. Benford, "EmoPlayer: a media player for video clips with affective annotations," *Interacting with Computers*, vol. 20, no. 1, pp. 17–28, 2008.

[9] 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*, vol. 64, no. 4, pp. 340–355, 2006.

[10] J. Kim, H. Kim, and K. Park, "Towards optimal navigation through video content on interactive TV," *Interacting with Computers*, vol. 18, no. 4, pp. 723–746, 2006.

[11] A. G. Money and H. Agius, "Analysing user physiological responses for affective video summarisation," *Displays*, vol. 30, no. 2, pp. 59–70, 2009.

[12] R. Hjelsvold, S. Vdaygiri, and Y. Léauté, "Web-based personalization and management of interactive video," in *Proceedings of the 10th International Conference on World Wide Web (WWW '01)*, pp. 129–139, 2001.

[13] R. Yan and A. G. Hauptmann, "A review of text and image retrieval approaches for broadcast news video," *Information Retrieval*, vol. 10, no. 4-5, pp. 445–484, 2007.

[14] C. G. M. Snoek and M. Worring, "Concept-based video retrieval," *Foundations and Trends in Information Retrieval*, vol. 2, no. 4, pp. 215–322, 2008.

[15] B. Yu, W. Y. Ma, K. Nahrstedt, and H. J. Zhang, "Video summarization based on user log enhanced link analysis," in *Proceedings of the 11th ACM International Conference on Multimedia (MULTIMEDIA '03)*, pp. 382–391, ACM Press, New York, NY, USA, November 2003.

[16] T. Syeda-Mahmood and D. Ponceleon, "Learning video browsing behavior and its application in the generation of video previews," in *Proceedings of the 9th ACM International Conference on Multimedia (MULTIMEDIA '01)*, pp. 119–128, ACM Press, New York, NY, USA, October 2001.

[17] R. Shaw and M. Davis, "Toward emergent representations for video," in *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*, pp. 431–434, ACM, New York, NY, USA, 2005.

[18] C. Gkonela and K. Chorianopoulos, "VideoSkip: event detection in social web videos with an implicit user heuristic," *Multimedia Tools and Applications*, 2012.

[19] K. Chorianopoulos, "Collective intelligence within web video," *Human-Centric Computing and Information Sciences*, vol. 3, article 10, 2013.

[20] I. Groma, F. F. Csikor, and M. Zaiser, "Spatial correlations and higher-order gradient terms in a continuum description of dislocation dynamics," *Acta Materialia*, vol. 51, no. 5, pp. 1271–1281, 2003.

[21] E. Vanmarcke, *Random Fields, Analysis and Synthesis*, MIT Press, Cambridge, Mass, USA, 1983.

[22] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Kogakusha, Tokyo, Japan, 9th edition, 1965.

[23] M. Zaiser, M. C. Miguel, and I. Groma, "Statistical dynamics of dislocation systems: the influence of dislocation-dislocation correlations," *Physical Review B*, vol. 64, no. 22, Article ID 224102, 9 pages, 2001.

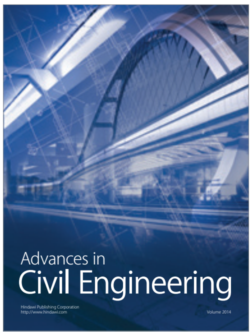
Journal of
Engineering

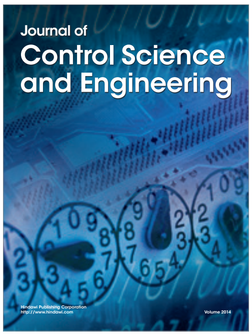
The Scientific World Journal

International Journal of
Rotating Machinery

Journal of
Sensors

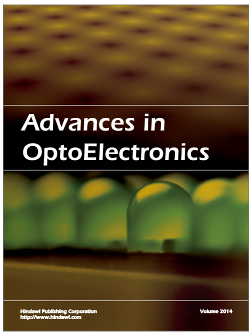
International Journal of
Distributed Sensor Networks


Advances in
Civil Engineering


Journal of
Control Science and Engineering

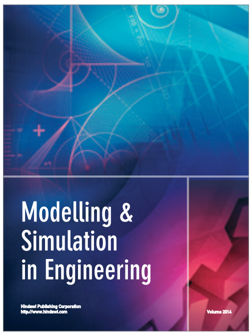
Journal of
Robotics

Journal of
Electrical and Computer Engineering

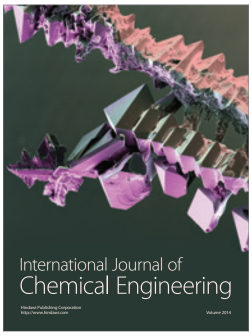

Advances in
OptoElectronics

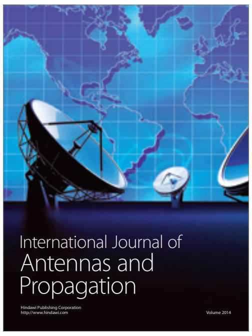
VLSI Design

International Journal of
Navigation and Observation

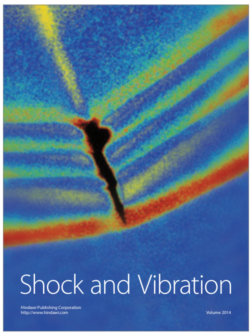

Modelling &
Simulation
in Engineering

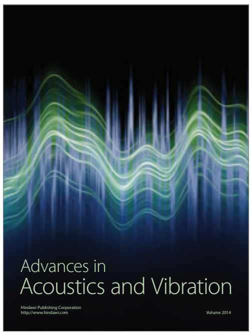
International Journal of
Aerospace Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration