The Vision of Crowds: Social Event Summarization Based on User-Generated Multimedia Content

Manfred del Fabro
Institute of Information Technology,
Klagenfurt University,
Universitätsstrasse 65-67
9020 Klagenfurt, Austria
manfred@itec.uni-klu.ac.at

Laszlo Böszörmenyi
Institute of Information Technology,
Klagenfurt University,
Universitätsstrasse 65-67
9020 Klagenfurt, Austria
laszlo@itec.uni-klu.ac.at

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
In this position paper we introduce the idea of generating a superior view of a large social event, based on user-generated – crowdsourced – content. Instead of just collecting and making them available in a raw form (as social platforms like YouTube), we automatically generate semantically coherent summarizations of the entire event. The individual consuming user gets thus a compact view generated by a large number of producing users. We call this idea the ”Vision of Crowds”.
A case study has been conduced at a social event where we used user-generated content to automatically generate live reports about that event. Furthermore, we have implemented a GUI that allows users to interactively compose personalized video summaries, based on the user-generated data collected at the case study.

Keywords
Vision of Crowds, Interactive Video Composition, Event Summarization, User-Generated Content

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User interfaces – Evaluation/ methodology

Introduction

At large social events, maybe beside one professional video production team, many visitors are producing and sharing photos and videos with their digital cameras or with camera-equipped mobile phones. This user-generated content can serve as additional source of information for the news coverage of social events. Reports may be enhanced by mixing professionally produced content with the contributions of visitors in order to present a comprehensive and multifaceted view of an event.

Every visitor has its own vision (view) of a social event. Different people may be interested in different activities or they may have different intentions when visiting an event. As the visitors are spread across the whole area where an event takes place, photos and videos are captured at different places. It is possible to identify the hot spots of an event, because we suppose that people tend to move to locations where something interesting is happening. If many people are in one place, it is more likely that also a lot of content will be produced there.

Composing the contributions of different visitors leads to presentations that express how a crowd of people is experiencing an event. All visitors are able to get an impression of the event by watching presentations not only through the eyes of a single director, but through the eyes of many other visitors: *The Vision of Crowds*.

The idea of the Vision of Crowds is a paraphrase of the well-known notion of the Wisdom of Crowds [3]. Decisions that are made by a group of people are often smarter than decisions that are only made by a single person. Many web applications rely on it to gather information, e. g. for image annotation. However, we are not interested in how a group of people decides, but how a group of people witnesses a social event. One way to get this information is to rely on crowdsourced data by looking at the photos and videos that individual visitors have captured. We do not only collect the individual impressions from visitors of an event, but with the help of the so-called ”Auto Director” component we also create meaningful and compact ”compositions” of them.

In video summarization [2, 4] it is common to try to find the most important key frames, shots or scenes in a single video. They are used to compose a shorter video or a compilation consisting of still images. Video summaries provide a short alternative to the original video. We go one step further. We do not summarize the content of single videos, but rather that of a whole social event. We try to identify photos and videos showing interesting situations in order to be able to compose semantically meaningful *event summaries*. Different situations and activities are shown that have happened during the event.

Case Study

In November 2010 an event called ”The Long Night of Research” took place at the Klagenfurt University. At 104 exhibition stands, spread across the whole campus, different research projects were presented to the public. More than 5000 people visited the university where they could attend presentations, watch demonstrations or participate in experiments. We were concerned with the question: *How to keep the visitors up-to-date at the Long Night of Research?* This question was answered by taking advantage of the idea of the Vision of Crowds. All visitors were encouraged to capture photos and videos and to upload them to our system. The collected content was used to inform all visitors about ongoing activities at the event. Figure 1 shows the architecture of the system.

Visitors with mobile phones could upload their photos and videos as attachment to an email. In the subject line they were able to annotate the content with tags. For people that used digital cameras a number of upload stations equipped

with card readers were available all over the campus area. All data uploaded to our system was stored on a file server.

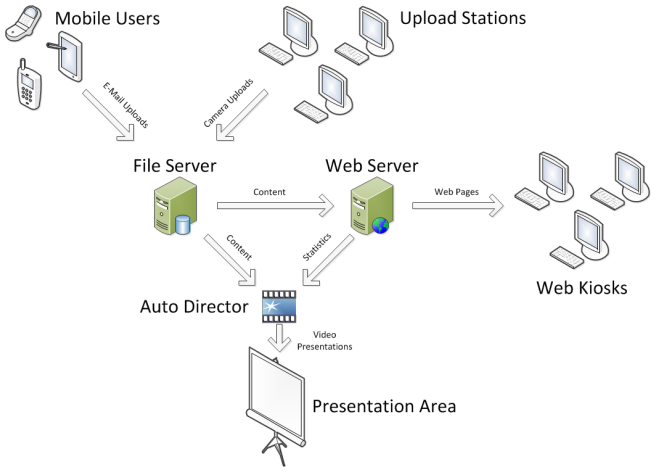

Figure 1: Architecture of the system used at the case study

The available content was presented to the visitors in two ways. Several web kiosks, also spread across the campus, could be used to browse the content using an interface similar to the popular photo sharing website Flickr. In addition, we installed big video screens in places where a lot of people were expected. We used our video browser [1] to show automatically generated summaries of the user-generated content on these screens. Our video browser can be used to show multiple videos in parallel. Therefore, we composed summaries that consisted of up to four parallel presentations. The decision which content to show at what time was made by a service called *Auto Director*. The Auto Director can be configured with predefined presentation profiles. Each profile consists of a combination of tags and the number of parallel presentations that should be shown. Based on several factors the Auto Director selects one profile and composes a presentation consisting of photos and videos that were annotated with the tags defined in the profile. The listing in Figure 2 shows the pseudo code of the composition algorithm. The decision which profile to use at a certain point in time is based on the popularity of the content, the most often searched key words, the user ratings (which can be entered in the web kiosk application), and the most recent content uploads. If a profile is selected, the Auto Director composes a presentation consisting of the most recent and the most popular photos and videos that matched that profile.

```
DO
        stats = getStatsFromWebKiosk()
        uploads = getRecentUploads()
        profile = selectProfile(stats,uploads)
        presentation = compose(profile,stats,uploads)
        showPresentation(presentation)
WHILE auto_director_alive
```

Figure 2: Main steps of the Auto Director

We made an empirical observation of the summaries shown by the Auto Director. The visitors behaved in a self-organizing way and reported mainly about places and situations, which are worth to be seen. As a result, the Auto Director composed presentations of these places, which attracted further people to go there. We got a novel, very interesting, semantically rich, multi faceted view. This view has never been seen in reality by any individual person - it is indeed a Vision of Crowds.

Interactive Event Summarization

In our current research activities we are facing the question how to inform people about a social event after it took place and which they might not have visited? What could be better than having a look through the eyes of other visitors? During our case study 1444 photos and videos were uploaded to

our system. Now we are developing a plug-in for our video browser [1] that enables users to interactively compose event summaries with that content. In Figure 3 a screenshot of the composition interface is shown.
By default, users can browse the whole content that has been uploaded. As different people are likely to have different interests, search filters can be used to narrow down the amount of available content. It is possible to define the time period when certain photos or videos were uploaded, the maximum duration and whether only videos or only photos should be shown. Furthermore, it is possible to indicate tags that the searched photos and videos must be annotated with.

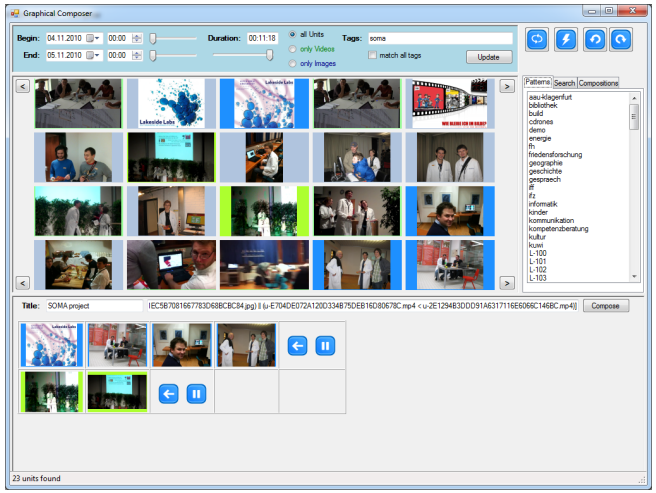

Figure 3: GUI for the composition of event summaries

At the center of the window, the search results are shown in a grid view. The background color indicates whether a thumbnail represents a photo (blue) or a video (green). At the bottom it is possible to manually compose event summaries by simply dragging and dropping thumbnails on the blue symbols. Dropping a thumbnail on the ← symbol adds the photo or video to the corresponding sequence, dropping it on the || symbol adds a new parallel stream to the presentation. The example shows a composition of two parallel streams. The first one consists of a sequence of four photos, while in parallel a sequence of two videos is going to be shown.
In addition to the manual composition it is possible to use the same profiles that are used by the Auto Director. In the right part of the window a list with all profiles is shown. A double click on the name of a profile results in an automatic composition of an event summary. Only photos and videos are included that meet the defined search constraints. Finally, by clicking on the compose button, the presentation is shown in the video browser.

Conclusion

In this position paper we introduced our idea of summarizing social events based on crowdsourced multimedia data. People can gather information about a social event by watching the content produced by other people with similar interests. It is like looking with the eyes of them. We call this principle the Vision of Crowds.
We conducted a case study and we implemented a first user interface prototype for event summarization where we take advantage of that. Empirical observations showed that people tend to capture only situations that are interesting for them. Other people with the same interests can be guided by that information.
We are currently working on some further issues. On the one hand we are going to perform simulations that allow us to replay our case study over and over again. With the help of this simulation we are able to adjust the configuration parameters of the Auto Director and to perform a detailed evaluation of our summarization strategies. On the other hand we are im-

proving the GUI for the interactive event summarization and preparing detailed user studies.

The summarization of social events is not the only possible field of application for our approach. We are going to investigate possibilities to integrate data from the well known photo and video sharing platforms on the Internet. Thus, we will be able to apply the Auto Director also on subsets of previously uploaded content. We could e.g. apply it on videos in YouTube tagged by a special tag for a given event. Furthermore, we intend to apply the approach to surveillance scenarios. We could e.g. generate a summary of the essential events that happened in a given part of a public motorway in a given time period.

References

[1] M. del Fabro, K. Schoeffmann, and L. Böszörmenyi. Instant video browsing: A tool for fast non-sequential hierarchical video browsing. In G. Leitner, M. Hitz, and A. Holzinger, editors, *HCI in Work and Learning, Life and Leisure*, volume 6389 of *Lecture Notes in Computer Science*, pages 443–446. Springer Berlin / Heidelberg, 2010.

[2] A. G. Money and H. Agius. Video summarisation: A conceptual framework and survey of the state of the art. *J. Vis. Comun. Image Represent.*, 19(2):121–143, February 2008.

[3] J. Surowiecki. *The Wisdom of Crowds*. Anchor, August 2005.

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