MULTIMEDIA TOOLS AND APPLICATIONS, vol. 70, no. 1, pp. 1-6, May 2014.

# Real-life Events in Multimedia: Detection, Representation, Retrieval, and Applications

Vasileios Mezaris, Ansgar Scherp, Ramesh Jain and Mohan S. Kankanhalli

The multimedia content that all of us frequently capture with our different multimedia enabled devices (e.g. cameras, smartphones, tablets) is typically the digital residue of a real-life event that unfolded before us, such as a graduation, a trip, a football game, or a natural disaster; it is this real-life event that we try to immortalize through the digital content creation process. Similarly, at an organizational (as opposed to personal) scale, multimedia content such as satellite images and radar signals are often captured in order to document natural or man-induced real-life events, such as weather phenomena or sea pollution incidents. Despite the central role that real-life events play in the generation and also the later interpretation of multimedia content, though, (since, for instance, what is the meaning and the value of a picture of a football field with some players on it, unless we can put it in event-context: which football game was that? And, what exactly happened in that game that makes it special?) the organization and retrieval of multimedia content are yet to fully embrace event-centric methodologies.

In response to this challenge, this special issue focuses on methods for the event-based processing and organization of multimedia content, with particular emphasis on the detection of real-live events in multimedia, the modeling of such real-life events, the sharing and the event-based retrieval of content, and the development of novel applications that jointly consider multimedia content and the real-life events that this content represents. The special issue includes nine papers that highlight different problems and solutions in these domains.

In the first paper, Scherp and Mezaris focus on the representation of events using suitable models. In particular, they conduct a survey of existing event models, and attempt to compare these models along a number of different dimensions that an event model should cover ("event aspects"). Further to this, they analyze how the different aspects of events relate to each other and how they can be applied together, and conclude with some references and thoughts on how the linking between the multimedia data and the events can be achieved, so as to provide the basis for future event-based multimedia applications.

The next three papers deal with event detection and event-based organization for still images. Dao, Dang-Nguyen and De Natale address the problem of associating personal image collections with events by analyzing the photo collection of an event as a whole, rather than looking at individual images. The objective is to detect event-types such as graduation, wedding, or different types of vacations and sports events, which describe the collection. In order to create a composite event signature for an image collection, they combine Saliency, Gist and Time information and introduce the notions of Gist-Saliency Signature Image Base (GS-SIB), which captures dominant colors and saliency information for all the images belonging to a photo collection, and Temporal SIB (T-SIB), which captures the temporal evolution of the images.

Ruocco and Ramampiaro look at the problem of event-based organization of images that are available in online photo-sharing applications such as Flickr. They propose a clustering approach, which takes into account textual annotations as well time and geo-location metadata of the images. To this end they extend the well-known Suffix Tree Clustering (STC) algorithm, originally developed for clustering text documents. They also investigate how the processing of the images at different time and space granularities affects their event-based organization.

Zigkolis, Papadopoulos, Filippou, Kompatsiaris and Vakali take the work on event-based organization of online images one step further, presenting a semi-automatic tool for the user-assisted generation of ground-truth image-event associations from online image collections. Specifically, they present CrEve, a collaborative event annotation framework which facilitates the annotation process and increases the coverage of the generated ground truth. Furthermore, the paper discusses the results of a user study that quantifies the contribution of different event dimensions in the event annotation process, confirming, for instance, the prevalence of spatio-temporal queries as the prime option of discovering event-related content in a large collection.

The next four papers deal with event detection and event-based organization of video content, with the latter content ranging from broad-domain user-generated videos to content captured under well-controlled conditions for supporting very specific applications. Cricri, Dabov, Curcio, Mate and Gabbouj exploit the readings of auxiliary sensors such as accelerometers and GPS receivers, which are typically included in camera-enabled devices, for detecting interesting events in user generated videos and for extracting high-level contextual information about the recording activity. In addition, they exploit multiple audio-visual recordings of a common event (e.g., music concerts), when available, to extract additional

V. Mezaris is with the Information Technologies Institute / Centre for Research and Technology Hellas, 6th Km Charilaou-Thermi Road, P.O.BOX 60361, Thermi 57001, Greece, bmezaris@iti.gr.

A. Scherp (corresponding author) is with the Institute of Computer Science and Business Informatics, University of Mannheim, 68131 Mannheim, Germany, ansgar@informatik.uni-mannheim.de

R. Jain is with the University of California, Irvine, CA 92697, USA.

M. S. Kankanhalli is with the Department of Computer Science, School of Computing, National University of Singapore, 117417 Singapore, Singapore.

The final publication is available at http://link.springer.com/article/10.1007/s11042-013-1426-8

event-related information such as regions of interest in the videos.

Poulisse, Patsis and Moens propose a method for identifying the semantic structure in long semi-structured video streams. They identify chains, i.e., local clusters of audio or visual features that are repeated in time, and use each chain as an indicator that the temporal interval it demarcates is part of a single semantic event. By layering all the chains over each other, dense regions emerge from the overlapping chains, revealing the semantic structure of the video.

The paper by Cheng, Liu, Zhao, Ye and Sun addresses the problem of detecting events in the daily activities of seniors within their home, for the purpose of monitoring the seniors' health. The authors introduce as part of their system a subspace Naive-Bayesian Mutual Information Maximization (sNBMIM) algorithm, which divides the feature space into a number of subspaces and allows the kernel and normalization parameters to vary between different subspaces. The presented senior home activity recognition system is evaluated for eight categories of everyday home events, such as sleep, eat, wash.

Spampinato, Palazzo, Boom et. al look at an environment-related application, and specifically present a system for understanding fish behavior during typhoon events by analyzing underwater video footage. The first step of the system that they propose involves the detection of "typhoon" events from the video recordings, which is based on video texture analysis and classification. Then, for understanding fish behavior during a typhoon event, they perform trajectory extraction and clustering from the video clip. They evaluate their approach on a set of underwater videos captured as part of the Fish4Knowledge EU project.

The special issue concludes with another paper on the detection of environment-related events. Talukder and Panangadan present an event-based multimedia processing framework for the detection, retrieval, and cross-media content assimilation of geo-spatiotemporal phenomena such as cyclones. They draw their primary content from various sources, such as remote satellites, in-situ visual sensors and weather bulletins, and use it for detecting the geo-spatiotemporal events of interest. Subsequently, they derive appropriate descriptions of the detected events, and use them for assimilating further event information across different media sources from the web.

In this special issue, we tried to include a rich set of works on the processing and organization of multimedia content according to real-life events. We hope that you will enjoy reading them. At this point, we would also like to take the opportunity to thank all the contributors of this special issue and all the reviewers who helped us in composing it.