Technical Report

The future of multimedia analysis and mining

Orgamizers: Nozha BOUJEMAA[1], Alexander G. HAUPTMANN[2], and Shin'ichi SATOH[3]
[1]INRIA, [2]Carnegie Mellon University, [3]National Institute of Information

ABSTRACT

Shonan meeting on "The Future of Multimedia Analysis and Mining," was organized from 3 to 6, November, 2012. This technical report summarizes the program of the meeting for record.

Overview of the meeting

We had a chance to organize a Shonan Meeting: Future of Multimedia Analysis and Mining. Luckily, the premier international conference ACM Multimedia was held in Nara, Japan, from October 29 to November 2, where top-class researchers gathered from all over the world. We decided to organize our Shonan Meeting right after the ACM Multimedia, namely, from November 3 to 6. The travel from Nara to Shonan was not trivial, and even some participants left Nara in the evening of November 2 after workshops of ACM Multimedia, and arrived at Shonan around midnight. Getting over the difficulty, we could attract many top-class multimedia researchers to the meeting, thanks to this collocation and succession. Related to the topic of our meeting: The Future of Multimedia Analysis and Mining, there are very successful precedents, ACM SIGMM retreat in 2003 [1] and the Dagstuhl seminar "Multimedia Research - where do we need to go tomorrow" [1)] in 2005. Although they were very comprehensive and well organized, our Shonan Meeting was intended to be more informal. However, we are very much confident that the discussion there was as exciting as the precedents, and of course more timely.

The meeting was composed of seven focused talks and five demos, mainly to share the ideas of cutting edge multimedia research and to stimulate each other towards the inspiration of the future multimedia research direction. The detail of the program can be found at: http://www.nii.ac.jp/shonan/seminar002/. The remaining time was mostly dedicated to free discussion. The discussion sessions were composed of series of plenary sessions and breakout sessions. The discussion was totally informal, and in the first plenary session, we discussed what to discuss during the meeting. The agreed and actually discussed topics are:

(1) Fundamental Multimedia Science
(2) Socially Motivated Multimedia Applications
(3) Education

After the first plenary session, we were divided into two to three groups and each group discussed one of these topics. We discussed on the topics in groups in breakout sessions and then reported and summarized the results in plenary sessions. We continue the series of sessions multiple times until we discussed on all topics (some topics were discussed by couple of different groups) and we summarized. We are thinking to report the summary of the discussion of the three topics in an appropriate forum. Finally we concluded the meeting by free discussion on possible research collaborations and sharable research resources among participants.

Interested readers may find detailed summary of the discussions made during the meeting in [2].

Participants

Nozha Boujemaa, INRIA (Organizer)
Alexander G. Hauptmann, CMU (Organizer)
Benoit Huet, Eurecom
Ichiro Ide, Nagoya University
Kevin Jing, University of Tokyo/Google
Kunio Kashino, NTT
Akisato Kimura, NTT
Rainer Lienhart, University of Augsburg
Tao Mei, MSRA
Frank Nack, University of Amsterdam

Received September 26, 2013; Revised December 2, 2013; Accepted December 5, 2013.
NII Shouan Meeting Report
The date of Meeting November 3–6, 2012, National Institute of Informatics

[1)] http://www.dagstuhl.de/de/programm/kalender/semhp/?semnr=05091

DOI: 10.2201/NiiPi.2014.11.6

Yuichi Nakamura, Kyoto University
Chong-Wah Ngo, City University of Hongkong
Vincent Oria, New Jersey Institute of Technology
Masanori Sano, NHK Science and Technology Research Labs
Shin'ichi Satoh, National Institute of Informatics (Organizer)
Alan Smeaton, Dublin City University
Hari Sundaram, Arizona State University
Marcel Worring, University of Amsterdam
Xiaomeng Wu, NII
Cai-Zhi Zhu, NII

Overview of the talks

Focused talks

Massive Change vs. Big-Data: The Role of Computing in Tackling Major Societal Problems
Hari Sundaram (Arizona State University)

Collective action problems, such as emission reductions to reduce risks of climate change or vaccination for infectious diseases, are prevalent in our society. The increasing globalization of social and ecological processes affects the scale of the collective action problems. The work of Nobel Laureate Elinor Ostrom and her colleagues has shown that small homogenous groups can overcome commons dilemmas as they can develop trust relationship in frequent interactions. However, in a globalizing world with increasing scale of collective action problems of a heterogeneous population the conditions for self-governance are more challenging. A common approach of addressing large scale collective action problems is a top down intervention from nation state governments. Unfortunately, a top-down approach is also not a panacea for collective action problems, illustrated by the lack of success on climate change policy during the last 20 years.

We need new approaches to understand, stimulate and sustain collective action in large heterogeneous populations. To engender cooperative behavior at large scales, we need to develop computational tools to facilitate the context for cooperation — homogeneity, effective communication — observed in smaller scale case studies and field experiments. I shall sketch out the some of the computational challenges that we need to address in order to scale up cooperation on collective action problems.

"Big-Data" is a key idea pervading much of contemporary computer science research. Big-Data research — including systems, machine learning algorithms and data-management amongst others — focuses on understanding massive datasets. Computing attempts to be value neutral — it provides services stitched together in support real-world problems. Information search engines are examples of computing algorithms used as a service; algorithms support queries independent of the use of the information resulting from the search.

Collective action problems are examples of big-data research, with an emphasis on purposeful social change. This talk shall explore challenges in pursuing research directed at positive social change.

Understanding Users from Large Scale Social Media Data
Tao Mei (Microsoft Research Asia)

People are contributing massive social media contents for knowledge sharing anytime and anywhere. Understanding users from social media therefore becomes important. There are four elements in the loop of understanding a user: user profile, context, multimodal input, and interactivity. This talk will introduce recent advances for understanding these key elements by leveraging machine intelligence and natural user interaction. Specifically, we will introduce: 1) how to predict user attributes from the heterogeneous forms of data by machine learning techniques, 2) how to leverage context information for personalized recommendation and easy task completion, 3) how to analyze multimodal user input by media content analysis, and 4) the design principles for natural user interaction. We will also showcase recently developed applications of this kind.

Any Places for Multimedia Analyses in Broadcasting Field?
Masanori SANO (NHK STRL, Science and Technology Research Laboratories)

One of the biggest changes which happened recently to broadcasting industry is a transition of work flow from analog to digital signal. This change has brought us the environment where we can feel it more natural to handle "TV program" as a digital content. In addition to this, with the movement toward fusion of broadcasting and telecommunication, TV program is no longer a just composition of video and audio data, but it has become the very multimedia content with various kinds of related information, such as EPG, closed caption and home page content.

Under this situation what the broadcasters need most is, needless to say, an effective management/handling technology of the digital content. Recently more importance has been put on the utilization of video archives in a value-for-money sense, and the methods by which user can retrieve what he/she desires easily is crucial. It has been said that the multimedia analyses can contribute to this part for a long time, but in reality not so many cases in which the multimedia analyses are employed on a massive and a full scale are reported so far. The main reason is an accuracy of the analyses results. But recently a new movement which can obscure the problem has occurred in the wake of the Great East Japan Earthquake.

In this talk, I will mention such the movements in broadcasting filed, and then I'd like to mention some activities related to introduction of multimedia analysis to TV production workflow in Europe. Finally I will introduce some efforts for this purpose in our research lab, i.e. NHK STRL.

Using Social Media to bring Context to Content for Event Detection and Recognition

Benoit Huet (Eurecom)

We have entered a ubiquitous world where technologies are enabling us tocapture, share and receive multimedia information regardless of our locationand activity. As the amount of multimedia data circulating on the Internetincreases at an overwhelming pace, the need for methods automaticallyextracting the semantic embedded within digital documents is gainingimportance. Without multimedia content understanding the Internet could soonbecome a large scale multimedia data graveyard... Events are a natural way for humans to structure and organize theiractivities. Whether it is a vacation, a meeting, a birthday or any otheractivity, events take place at a given time and place, and feature one ormore persons. Events can be categorized as public if open to all(presidential election ballot, a football game, a concert, etc...) orprivate if attended by invited people only (wedding, birthday, holiday, etc...). In this presentation, a novel framework is proposed to collect trainingsamples from online media data to model the visual appearance of socialevents automatically. The visual training samples are collected through theanalysis of the spatial and temporal context of media data and events. Whilecollecting positive samples can be achieved easily thanks to dedicated eventmachine-tags, finding the most representative negative samples from the vastamount of irrelevant multimedia documents is a more challenging task. Here,we argue and demonstrate that the most common negative samples, originatingfrom the same location as the event to be modeled, are best suited for thetask. A novel ranking approach is devised to automatically select a set ofnegative samples. Finally, the automatically collected samples are used to learn visual eventmodels using Support Vector Machine (SVM). The resulting event models areeffective to filter out irrelevant photos and perform with a high accuracyas demonstrated on various social events originating for various categoriesof events.

Multimedia Analytics: Towards a Research Agenda

Marcel Worring (University of Amsterdam)

Multimedia analytics, bringing together Multimedia Analysis and Visual Analytics, is an emerging field with applications in such diverse fields as medicine, security, forensics, and science. The current state of the field bears great similarities to the early years of content based image retrieval. CBIR required a truly multidisciplinary solution path joining the forces of database research and computer vision. But in the early years the datasets were considered far too small for the database community, while computer vision researchers considered the features used too trivial to be of scientific interest. Yet at that point in time those were the only feasible solutions and merely realizing the integration was needed to bring the field forward. Starting around 2000 the CBIR field matured and the problem has now been formalized to a point that benchmarks like PASCAL VOC for images and TRECVID for video have been established and are steering most of the current research. We are also reaching the point that the CVPR community has the lead in CBIR research. The field of multimedia analytics is new and exciting. Our multimedia community could play a leading role when we start valueing its potential and accepting the consequences such a new field brings to research. On the basis of that, we could start building a research agenda. In this talk I will try to define the field and show examples of solutions we and others have built in this direction. It hopefully will be the starting point for further interesting discussion sessions during the meeting.

Understanding Media Is Understanding Humans — Toward Human—Centric Media Understanding

Akisato Kimura and Kunio Kashino (NTT Communication Science Labs)

For understanding media content, such as images, videos and audio signals, we usually look into the media content itself. Various hand-crafted features, generative models and classifiers have been proposed for many specific tasks. Moreover, recent advances in machine learning techniques and availability of large amount of media data is encouraging us to design efficient features and better classifiers. However, when we think of "recognition" or "understanding" of a broader definition, it is also essential to take humans into account.

Toward "human-centric" media understanding, this talk will focus on the following two aspects:

(1) Human factors in the "sameness," and
(2) Human activity analysis in the media content generation.

The first aspect involves our hope to better model what humans recognize as a single "thing" or an "event," or the "same" things or events, which we call the "sameness." The sameness in the media content heavily depends on situations and circumstances. Therefore, only humans in a specific situation will be able to determine what should be recognized as the same things or the same events as something else.

Regarding the second aspect, we would like to analyze human behaviors in rich contexts and setups. We are now accessible to massive amount of information of various types associated with media contents or real-worlds events. They include (a) the ones intentionally authored by humans, such as geotags and microblogs, as well as (b) the ones automatically generated, such as Web browser operation history and travel records. Our assumption is that such data might be useful as an information source not only for human modeling but also for media content analysis itself.

For both aspects, we will introduce some of our early trials and discuss future directions.

Demos

Browse-to-Search: An Interactive Exploratory Visual Search System Realizing Your Online Shopping Experience
Tao Mei (Microsoft Research Asia)

Ultrahigh-Speed Commercial Detection and its Application to Knowledge Discovery
Xiaomeng Wu (National Institute of Informatics)

Fishing for a Needle in the Ocean — Instance Search from Large-scale Video Data Set
Cai-Zhi Zhu (National Institute of Informatics)

Snap-and-ask: An Example of Using Visual Instance Search for Question Answering
Chong-Wah Ngo (City University of Hong Kong)

mediaWalker: A Topic-based Browser for a News Video Archive
Ichiro Ide (Nagoya University)

References

[1] L. A. Rowe and R. Jain, "ACM SIGMM retreat report on future directions in multimedia research," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 1, no. 1, pp. 3–13, 2005.

[2] N. Boujemaa, A. G. Hauptmann, and S. Satoh, "The Future of Multimedia Analysis and Mining: Visions from the Shonan Meeting," *Media Impact Section, IEEE Multimedia*, vol. 20, no. 2, pp. 4–12, doi: 10.1109/MMUL.2013.30, April-June, 2013.

Nozha BOUJEMAA

Nozha BOUJEMAA is the director of the INRIA Saclay Ile-de-France research centre. Her research interests include multimedia content search, pattern recognition, and machine learning. Boujemaam has a PhD in computer science from the University of Paris V.

Alexander G. HAUPTMANN

Alexander G. HAUPTMANN is a principal systems scientist in the School of Computer Science at Carnegie Mellon University, with a joint appointment in the Language Technologies Institute. His research interests include multimedia analysis and indexing, speech recognition, interfaces to multimedia system, and language in general. Hauptmann has a PhD in computer science from Carnegie Mellon University.

Shin'ichi SATOH

Shin'ichi SATOH is a professor in the Digital Content and Media Sciences Research Division at the National Institute of Informatics (NII), Japan. His research interests include image and video analysis and database construction, management, image and video retrieval, and knowledge discovery based on image and video analysis. Satoh has a PhD in information engineering from the University of Tokyo.